Combined Quantum Mechanical / Molecular Mechanical Investigation of Enantioselective Reactions in Lipases

Inaugural-Dissertation
zur
Erlangung des Doktorgrades der
Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

vorgelegt von
Alexander-Nikolaj Otte
aus Marl

Juni 2006

Gedruckt mit Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

Referent: Univ.-Prof. Dr. Walter Thiel
Korreferentin: Univ.-Prof. Dr. Christel Marian
Tag der mündlichen Prüfung: 3. Juli 2006

Zusammenfassung

Die vorliegende Arbeit beschäftigt sich mit der theoretischen Untersuchung der enantioselektiven Esterhydrolyse, welche durch das Enzym Lipase A von *Bacillus subtilis* katalysiert wird. Ein Modell aus Enzym, Substrat und Solvens wird atomistisch mit einer Kombination von quantenmechanischen (QM) und molekülmechanischen (MM) Methoden beschrieben. Dabei werden für den QM Teil Dichtefunktionalmethoden und für den MM Teil das Charmm22-Kraftfeld verwendet.

Gemäss enzymkinetischen Untersuchungen an Lipasen und Serinhydrolasen erfolgt die Enantiodiskriminierung während der Acylierungsreaktion. Es ist allgemein akzeptiert, dass diese Reaktion über ein kurzlebiges tetraedrisches Intermediat verläuft, dessen Existenz allerdings experimentell für Lipasen noch nicht eindeutig nachgewiesen worden ist. Ein Teil dieser Arbeit befasst sich daher mit der theoretischen Charakterisierung dieses Intermediats. Hierzu werden Charmm22-Kraftfeldparameter für solche Intermediate durch Kalibrierung an *ab initio* Daten hergeleitet.

Die Enantioselektivität der Lipase-katalysierten Esterspaltung wird exemplarisch für das chirale Substrat 1-(2-Naphthyl)-ethyl-acetat untersucht. Verschiedene mögliche Bindungsmoden des Substrats im aktiven Zentrum werden zunächst durch MM-basierte Molekulardynamik(MD)-Simulationen identifiziert. Vom tetraedrischen Intermediat ausgehend werden dann die beiden kovalenten Teilschritte der Acylierungsreaktion (vom Michaelis-Komplex über das Intermediat zum Acylenzym) auf QM/MM- Niveau mit zwei verschiedenen Ansätzen behandelt. Zum einen werden Reaktionspfade und stationäre Punkte auf QM/MM-Potentialflächen durch Geometrieoptimierungen bestimmt, aus denen man die Aktivierungsenergien $\Delta E^{\ddagger}$ für die Acylierungsreaktion erhält. Zum anderen werden die entsprechenden freien Aktivierungsenthalpien $\Delta G^{\ddagger}$ aus MD-Simulationen durch Umbrella Sampling ermittelt. Die quantitative Auswertung der Enantioselektivität wird mit einem stochastischen kinetischen Modell durchgeführt, welches die Simulation der kinetischen Racematspaltung ermöglicht und die Konkurrenz zwischen den Enantiomeren richtig erfasst. Die auf QM/MM-Niveau berechneten Barrieren und Geschwindigkeitskonstanten dienen als Eingabeparameter für das kinetische Modell.

Die erhaltenen Ergebnisse spiegeln die Komplexität der untersuchten enzymatischen Reaktion wider. Es zeigt sich, dass das Substrat in verschiedenen Orientierungen binden kann und dass die Potentialkurven für die Acylierungsreaktionen je nach Bindungsmodus und gewählter Startgeometrie für die Optimierungen unterschiedlich sein können. Das tetraedrische Intermediat ist meist ein sehr flaches Minimum, in manchen Reaktionspfaden fehlt es jedoch, so dass technisch saubere Ergebnisse nur über ausgedehnte MD-Simulationen mit Umbrella Sampling zu erzielen sind. Auf diese Weise wird die experimentell gefundene R-Enantioselektivität für das untersuchte Substrat korrekt wiedergegeben, allerdings bleibt der berechnete **E**-Wert zu klein. Die QM/MM-Untersuchungen liefern somit detaillierte qualitative Einsichten in den Mechanismus, aber noch keine quantitativ zuverlässigen Vorhersagen der Enantioselektivität.

Abstract

The present work is concerned with the theoretical investigation of enantioselective ester hydrolysis as catalysed by the enzyme lipase A of *Bacillus subtilis*. A model consisting of enzyme, substrate, and solvent is described at the atomistic level with a combined quantum-mechanical (QM) and molecular-mechanical (MM) approach. Density-functional theory (DFT) is used for the QM part and the Charmm22 force field for the MM part of the model.

According to kinetic investigations of lipases and serine hydrolases, the discrimination of enantiomers occurs during the acylation steps of the reaction. It is commonly accepted that this step proceeds via an instable tetrahedral intermediate, whose existence has not yet been proven experimentally for lipases. One part of this work is therefore concerned with the theoretical characterisation of this intermediate. To this end, Charmm22 force field parameters were derived on the basis of *ab initio* data.

The enantioselectivity of the lipase-catalysed ester hydrolysis was analysed exemplarily for the chiral substrate 1-(2-naphthyl)-ethyl-acetate. Different binding modes of the substrate in the active site were identified initially using MM-based molecular dynamics (MD) simulations. Starting from the tetrahedral intermediate, both covalent steps of the acylation reaction (from the Michaelis complex via the intermediate to the acylenzyme) were treated at the QM/MM level using two different methodological approaches. On the one hand, reaction paths and stationary points on the QM/MM potential energy surfaces were determined by geometry optimisation, from which activation energies, $\Delta E^{\ddagger}$, of the acylation reaction were derived. On the other hand, the free energies of activation, $\Delta G^{\ddagger}$, were computed using the umbrella sampling method. The quantitative analysis of enantioselectivity was performed with a stochastic kinetic model of the kinetic resolution experiment, which accounts for the competition between the enantiomers. The barriers and rate constants computed at the QM/MM level served as input parameters for the kinetic model.

The results mirror the complexity of the enzymatic reaction. It is found that the substrate can bind in different orientations and that the potential-energy curves for the acylation reaction are strongly dependent on the binding modes and the starting geometries. The tetrahedral intermediate is often a shallow minimum, and is even nonexisting in some reaction paths, so that technically sound results can only be obtained using umbrella sampling simulations. In this manner the experimentally determined enantiopreference for the R-enantiomer is reproduced, however, with a too low **E** value. The QM/MM investigations are therefore able to provide qualitative insights into the mechanism, but not quantitative predictions of enantioselectivity.

Contents

Preface

Enzymatic enantioselectivity is both an absorbing and highly valuable phenomenon that can be exploited by organic synthesis for the production of drugs or precursors with high enantiomeric purity. In contrast to its widespread practical application are the sporadic attempts of a rationalisation by theoretical means. This disparity can most probably be traced back to the problem being a great challenge for any rational approach, due to the complexity of the models and the accuracy of the calculations that are required. Our ambition in this work is to probe the feasibility of an accurate and quantitative modelling of enzymatic enantioselectivity at the atomistic level. Studies in the literature up to the present point are limited to molecular mechanical approaches, where enantioselectivity can be inferred solely from empirical descriptors, and are thus restricted to a qualitative kind of argumentation. Quantitative approaches should focus on the evaluation of reaction barriers, which is not possible at a molecular mechanical level, but rather requires the use of quantum mechanical approaches to describe the breaking and forming of bonds. Since enantioselectivity is strongly believed to be caused by differential effects due to the environment of a reactive system, a detailed description of the former is desired. The use of combined quantum mechanical and molecular mechanical (QM/MM) potentials is an attractive choice that enables incorporation of the environment at moderate computational cost.

Organisation of the Document

In the following paragraphs we outline the organisation of this thesis. The volume consists of six chapters, plus appendices. While the main chapters are arranged in a way to facilitate comprehension of the subject matter when read through in the presented order, the appendices contain additional information separated from the main text to enhance readability. The appendices are referred to in the appropriate places, and may be required for a full comprehension of the work. In other cases they serve a different purpose. During the time spent on this project several theoretical approaches for the production of data and methods of analysis of this data were tested. Some of those proved to be more worthwhile than others for the present work. In cases where techniques were of less than central importance, but where the author thinks they may still prove worthwhile in the context of other projects, or where considerations of performance can help others to judge the utility of the method for their work, they were documented in the thesis in the form of appendices.

The Didactic Order

The ordering of the chapters corresponds to the best *a posteriori* way of reading through them. Chapter 1 contains a general analysis of the problem of enantioselectivity in the light of, but not exclusively, a computational approach, and therefore is a must-read. Chapter 2 focuses in more detail on the theoretical approaches that we use, and should be appreciated by those in the field of computational chemistry. Molecular modelling of the tetrahedral intermediate (TI) is described in chapter 3. All further work builds on the models generated by the approaches described here which are therefore of central importance to the project. Exploration of the potential energy surface of the acylation reaction starts from the TI, and proceeds from there to the Michaelis complex (MCC), or the acylenzyme, which we call the product complex (PDC), since the alcohol of the hydrolysed ester is still in the binding pocket. In chapters 4 and 5 we use two different approaches to calculate the barriers for these two reaction steps. In chapter 4 a static approach is used, where the potential energy surface is scanned by an adiabatic mapping procedure, at an effective temperature of zero Kelvin, which is often referred to as the method of the minimum energy reaction path. In chapter 5 the same reactions are treated by a molecular dynamics (MD) sampling approach to evaluate the reversible work along the reaction coordinate, referred to as the potential of mean force, from which we derive reaction free energy differences of the enantiomers. Although it might be appealing at this point to speak of a dynamic treatment of the reaction this does not reflect the fact that we are sampling from equilibrium distributions, which are stationary in the mean. Nevertheless the approach allows relaxation to occur on a larger scale than in the minimum energy reaction path method. At the end of chapter 5 the results of both methods are compared. To summarise, chapters 3, 4 and 5 contain the main results of this thesis and should be read in succession. In chapter 6 finally we present the parametrisation of the molecular force field of the TI for the Charmm22 force field. This chapter is aimed mostly at specialists and may be skipped by those not interested in parameter development.

The Chronological Order

The ordering of the subprojects into chapters is, as explained in the preceding, not in chronological order. Had the emphasis been on a chronological ordering we should have arranged the chapters (starting from 3) as follows. First a study of the gas phase model was conducted (shifted here to appendix C). This study showed the limits of such models in the serine hydrolase arena, but led to the definition of structural criteria that any subsequent modelling should meet. After this initial study a first attempt at a QM/MM treatment was undertaken, and it was observed that essentially all attempts at QM/MM geometry optimisation without proper description of the TI failed. It was reasoned that a force field to model the TI faithfully had to be parameterised in the context of Charmm22, which was our force field of choice for the MM part of the system. After the parametrisation new models for the TI were built for both enantiomers. By that time the enormous complexity of the configuration space accessible to the TI became evident, realisations of which are referred to as binding modes in the following. With the models thus constructed and further prepared by MD relaxation another attempt at geometry optimisation, equipped with a refined formulation for the computation of the reaction paths, was made. The energy barriers computed were rather different for all binding modes of the TI. This stressed the necessity of a complete picture of binding modes and associated barriers. The modelling of binding modes was thus reiterated and completed to the best of the author's knowledge. This part of the work can not be subjected to any 'mathematical' proof of completeness of the configuration space and can be seen as one of the bottlenecks towards an accurate evaluation of enantioselectivity. More advanced techniques than manual modelling, which may relieve the user of the painstaking work (which does not imply a warranty for completeness) and may be used in future studies are discussed in the main text. After the second round of modelling an attempt at free energy simulation using MD to sample configuration space was made, to get rid of artefacts encountered frequently with optimisations of minimum energy reaction paths. The density functional methods used in the minimum energy reaction path calculations were found to be computationally too expensive for any sensible computation and thus an alternative had to be found. The empirical valence bond (EVB, shifted to appendix F) method was considered a worthwhile complementary method to the calculations done so far. Although the formulation uses force fields to represent reactants and products it enables the simulation of the reaction with all bond-breaking and bond-forming events via a secular matrix in the spirit of valence bond calculations. Its use necessitated a prior implementation in the CHARMM[1] program. Some code supplied by Konrad Hinsen was reused in this context. Later that year, after the implementation and early tests of the EVB method, the self-consistent-charge-tight-binding DFT method became available to us. This formulation enabled genuine QM/MM approaches to be used and was assigned a higher priority than the EVB model. Using umbrella sampling with automatted histogram reweighting the final solution to the problem of enantioselectivity came closer. Free energy calculations have their own challenges nevertheless, and considerable time was spent on identification of artefacts

---

[1] CHARMM refers to the Charmm molecular modelling software package[30] , Charmm22 to the Charmm force field[108].

and enhancing the accuracy of predictions. The barriers thus calculated were used as input to kinetic models (see appendix D) which were then simulated to understand the origin of enantioselectivity in our system.

Comment on Experimental Reference Data

Computational approaches to complex systems require validation against experimental data. The information available at the beginning of the project was the crystal of the empty enzyme. A few months after initiation of the project the crystal structure of lipase A of *Bacillus subtilis* (BSLA), inhibited with chiral 1,2-O-isopropylydidene-sn-glycerol-phosphonate (IPGP), became available.

The chosen substrate was 1-(2-naphthyl)-ethyl-acetate (NAPHAC) which is cleaved by the enzyme with an **E**-value of 140 in favour of the R-enantiomer. No binding modes were known experimentally for this substrate, but it was discovered soon in our simulations that they were likely to be different from those of IPGP, and more numerous than expected. Kinetic data of BSLA with NAPHAC or any other substrate were not available so that the computed barriers could not be calibrated against experimental data. What was known experimentally, apart from the crystal structure without NAPHAC, was the **E** value of 140, which translates into a $\Delta\Delta G^{\ddagger}$ difference of 3 kcal/mol for the free energy barriers of the two enantiomers. This is a small energy difference that may be hard to capture in approximate calculations of complex systems with many degrees of freedom. It would clearly be valuable for computational studies such as ours to have reliable experimental data (e.g. on kinetics) for calibration and validation.

Chapter 1

Introduction

1.1 Chirality and Enantioselectivity

Lord Kelvin, in 1904[1], gave a definition of chirality, which has remained universally valid: *"I call any geometrical figure, or groups of points, chiral, and say it has chirality, if its image in a plane mirror, ideally realized, cannot be brought to coincide with itself."* The pair of nonsuperimposable entities are termed enantiomers. This definition in terms of geometry can be cast on molecules, which are required to have identical sum formulas and bonding patterns. Then, due to arrangement of the constituent groups in space the conditions for chirality can be met. It is furthermore necessary that this spatial arrangement discriminating the enantiomers be stable under predefined conditions, otherwise interconversion occurs and a racemic mixture (racemate) is formed as a 1:1 mixture of the enantiomers. In other words the enantiomers are required to be minima on the Born-Oppenheimer surface separated by a sufficiently high barrier to suppress interconversion[165]. Examples of chiral molecules are given in figure 1.1. Chirality in A) is due to the arrangement of the chlorine and hydrogen atoms. A rotation by $90°$ about the axis, in which all three carbon atoms lie, would make the molecule achiral, but the barrier for this rotation is too high for racemisation to occur (axial chirality). Molecules B) and C) show central chirality which is enforced by barriers of inversion about the center. In B), where only three substituents are attached to the central atom, this barrier may be quite low and species of this kind may thus have to be kept at low temperatures to avoid racemisation. The stereoconfiguration of enantiomers is characterised by labels R and S, according to the nomenclature of Cahn, Ingold and Prelog[2].Enantiomers cannot be distiguished from each other by most physical and chemical methods except when asym-

---

[1] in: *"Baltimore Lectures on Molecular Dynamics and the Wave Theory of Light"*

[2] Visit www.iupac.org or read R S Cahn, C K Ingold and V Prelog,*Angew. Chem. Internat. Ed.*,**5**, 385-415 (1966)

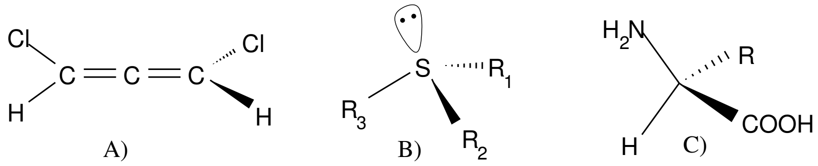


Figure 1.1: Chiral molecules.

metric probing is applied. Rotation of plane-polarised light is the physical method of choice for discrimination of enantiomers[3]. The magnitude of rotation is the same for both, but the sign is inverted. Chemical probes must be chiral themselves to distinguish between enantiomers. The interaction of probe and studied system then leads to a break in symmetry and allows a separation of the diastereomeric complexes by further chemical purification or physicochemical methods.

Chemical reactions with racemates in the presence of a chiral agent can occur at different rates for both enantiomers. The reaction is then called *enantioselective*. Enzymes, which are chiral and exist in enantiopure form fulfill those requirements in offering an asymmetric environment and catalytically active groups. From a macroscopic point of view the rate enhancement can be rationalised on the basis of the following kinetic consideration:

$$\begin{aligned} \mathrm{E}+\mathrm{R} &\overset{\mathrm{k_1}}{\rightarrow} \mathrm{E}+\mathrm{A} \\ \mathrm{E}+\mathrm{S} &\overset{\mathrm{k_2}}{\rightarrow} \mathrm{E}+\mathrm{B} \end{aligned}$$

In these equations the substrate and its stereoconfiguration is symbolised by R and S. E is the catalyst, which could be an enzyme. An experimental setup where R and S compete for the catalyst and react with different rates $\mathrm{k_1}$ and $\mathrm{k_2}$ to form products A and B is called *kinetic resolution*. This procedure can be exploited in synthetic chemistry to produce chiral molecules with high excess of one enantiomer. Enantiopurity of compounds is necessary in the syntheses of natural products and especially drugs, where disregard for this condition can have deleterious effects. The most infamous example is thalidomide, a sedative drug that was sold under the name Contergan. One enantiomer causes severely abnormal effects during embryogenesis, and victims are born with malformations of extremities[4]. The 2001 nobel prize in chemistry was awarded to Knowles, Noyori and Sharpless[5], honouring the complexity and importance of chiral catalysis.

---

[3] For racemates the rotation of plane polarised light is zero.

[4] In this special case preparing only one enantiomer does not solve the problem, since thalidomide is racemised in vivo.

[5] visit www.nobel.se

Returning to the kinetic resolution experiment indicated above, it should be stressed that k_1 and k_2 may be apparent rates for the formation of products resulting from a succession of several elementary steps. Assuming this simple model the enantioselectivity of an enzyme is quantified as $\mathbf{E} = \frac{\mathrm{k}_1}{\mathrm{k}_2}$. The dimensionless number **E** is termed the enantioselectivity factor; values larger than 1 indicate preferential reaction with one of the enantiomers, while a value of 1 means that there is no enantioselectivity. By convention, k_1 and k_2 are chosen such that $\mathrm{k}_1 > \mathrm{k}_2$.

Use of the Eyring equation $\mathrm{k} = \frac{\mathbf{k}\mathrm{T}}{\mathrm{h}}\kappa \cdot \exp\left(-\frac{\Delta\mathrm{G}^\ddagger}{\mathrm{RT}}\right)$, where $\Delta\mathrm{G}^\ddagger$ is the activation free energy, R the universal gas constant, T the absolute temperature, **k** Boltzmann's constant, h Planck's constant and κ the transmission coefficient leads to the following definition of **E** which highlights its link to thermodynamics:

$$\mathbf{E} = \frac{\mathrm{k}_1}{\mathrm{k}_2} = \frac{\kappa_1}{\kappa_2} \cdot \exp\left(\frac{\Delta\Delta\mathrm{G}^\ddagger}{\mathrm{RT}}\right). \tag{1.1}$$

Hence we have one contribution from the difference in activation free energies, and another one from the quotient of transmission coefficients. While the activation free energies have exponential weight in the rate expression, the pre-exponential factors make a linear contribution. The pre-exponential factor, or transmission coefficient is attributed to *dynamical* effects, while the activation free energies are equilibrium properties[212]. κ can take values in the range between 0 and 1. To have substantial influence on **E** the quotient should at least be on the order of 5-10, which requires $\kappa_2 < 0.2$. Theoretical studies summarised in a review article by J B Anderson[4] did not show any transmission coefficients below 0.2, and none above 0.9. The rotation of a tyrosyl ring on the surface of bovine pancreatic trypsin inhibitor[150] was on the low side, while the diffusion of oxygen out of myoglobin[125] had the highest transmission coefficient. This leads to the conclusion that in the present context free energy differences are more important than dynamical effects[212]. The simple kinetic model presented in this section is refined later to suit the enzymatic problem studied in this work.

1.2 Lipases

Lipases form a class of enzymes that primarily catalyses the breakdown of fats. An important class of substrates are triacylglycerols, which are esters of long chain carboxylic acids, but many lipases accept also other esters of lower molecular weight as substrates. The molecular weight of lipases range from 35 kDa, *Bacillus subtilis* lipase A (BSLA) is an exception at 19 kDa. Lipases are known to be catalytically active only at lipid-water interfaces occuring at micelles or membranes. This effect is called interfacial activation. BSLA is an exception also to this rule and catalyses ester cleavage in solution. This eliminates the need for lipid-water phase separations and makes the enzyme an interesting target for industrial applications.

By classification lipases belong to the family of serine hydrolases, that also contains members which cleave amide bonds and are indispensable in the metabolism of peptides and proteins and are thus termed proteases. Some members of this class of enzymes hydrolyse both amides and esters with good catalytic efficiency. A famous example is chymotrypsin, the first serine hydrolase for which an x-ray structure was available[139]. This enzyme was studied in detail by all disciplines of the life sciences. Another important serine hydrolase is acetylcholinesterase, which plays an important role in signal transduction. This enzyme has been under study for a long time, but it was only in 1991[199] that a crystal structure became available and enabled molecular modelling work. Due to the functional and structural similarities to lipases, concepts and insights gained on chymotrypsin and acetylcholinesterase are used whenever they are needed. In the following the term serine hydrolase is preferred over lipase when discussing general kinetic and mechanistic features.

Serine hydrolases can be highly selective for their substrates and enantioselectively cleave esters and amides. Chymotrypsin, for example accepts esters of L-aminoacids as substrates, but not D-stereoisomers of the same compound[64], which rather act as inhibitors. Although thermodynamic requirements are not met in vivo, serine hydrolases can also efficiently catalyse the back-reaction in vitro, esterification and peptide bond formation. These syntheses are performed in organic solvents with esters[119] or binary mixtures of solvents for peptide synthesis[27].

Lipases are favoured in organic synthesis because of their high availability and versatility, being able to perform enantioselective hydrolyses and transesterifications[181][201]. Recent applications include resolution of racemic thioesters by *Carica papaya* lipase[149] and preparation of hydrocinnamic esters[164]. A novel application is their use as highly selective filters in combination with liquid membranes[142]. This enables enantioselective separation of organic acids, such as ibuprofen. Lipases have been targeted with the techniques of molecular biology to enhance enantioselectivity for certain substrates. Such methods allow for *directed evolution*[170]. Recent successful applications include inverted and increased stereoselectivity of BSLA for a meso substrate[72]. Physicochemical properties can also be enhanced using directed evolution. Dramatically increased thermostability was reported for a triple mutant of BSLA[1].

1.3 Reaction Mechanism of Serine Hydrolases with Ester Substrates

The catalytic process that leads to the cleavage of the ester substrate (figure 1.2) can be split into two stages. The first is termed the *acylation* reaction and the second *deacylation*[97]. Acylation leads to the acylenzyme, which is an ester of the acid moiety of the substrate with the enzyme, which acts as an alcohol. This step is equivalent to reesterification. During deacylation the acylenzyme decomposes on reaction with water. Both nucleophilic steps, the attack of serine on the sp^2 carbon of the substrate ester as well as the reaction of the acylenzyme with water proceed via a *tetrahedral intermediate* (TI). Unstable tetrahedral species were shown to be on the reaction path of ester hydrolysis in alkaline aqueous solution by isotope marking experiments[15] and are also assumed to exist analogously in enzymatic hydrolysis. Complexes of enzymes with suicide inhibitors are often observed to adopt a tetrahedral configuration. The TI for good substrates is so unstable that it is considered a transition state by some researchers (see discussion in [97]).

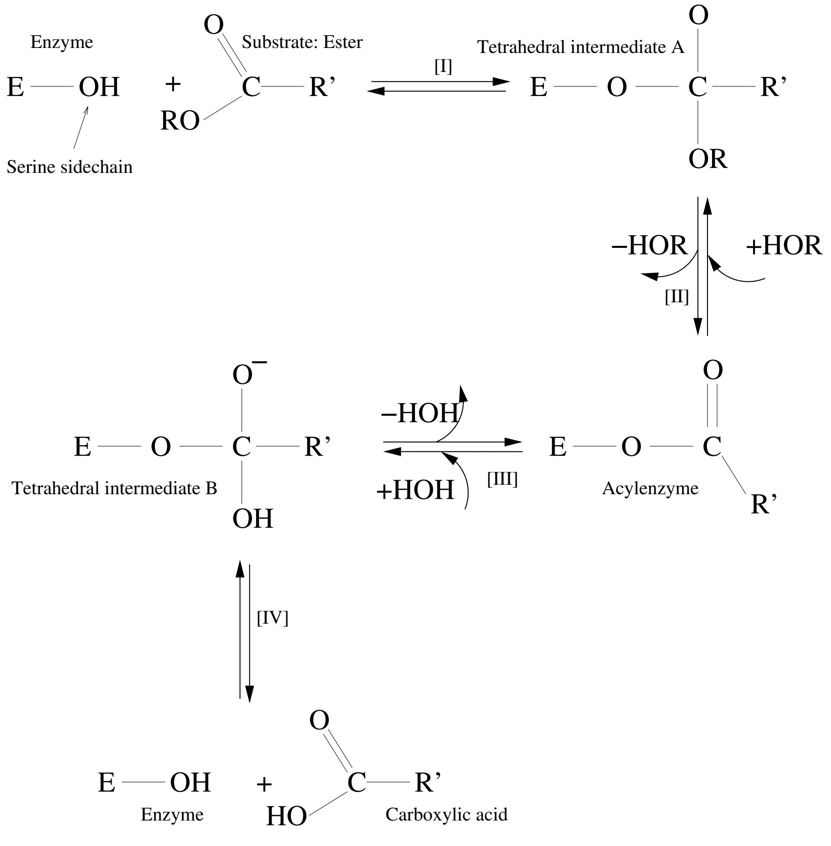


Figure 1.2: Schematic representation of the reaction mechanism of serine hydrolases. Steps [I] and [II] are the acylation reaction, steps [III] and [IV] form the deacylation reaction.

Components of the active site. An arrangement of three amino acids that is found in all serine hydrolases is called the *catalytic triad*[97]. It consists of a serine residue that is covalently modified as shown in figure 1.2, a histidine, and an aspartate (sometimes glutamate).

Serine was the first residue to be identified as taking part in the enzyme reaction. In experiments with potent inhibitors like diisopropyl-fluorophosphate (DFP) that bind irreversibly to alcohols, stoichiometric reaction with the enzyme was noted[103]. Schaffer, May and Summerson extracted[178] serine phosphoric acid from DFP-chymotrypsin and concluded one serine (out of 27) to be in a reactive state towards DFP. Using nitrophenyl ester as substrate at pH 5, the enzyme was found to be inactivated in an acetylated form[11], the reaction being stoichiometric as with the DFP. Recovery of the free and active form was possible by elevation of pH to 6-7. It was noted that esterification of ethanol could be performed by addition of the alcohol to acetyl-chymotrypsin and adjustment to pH 6. After extraction of short acetyl peptides from partially degraded acetyl-chymotrypsin, Oosterban and van Adrichem[152] concluded that the same serine as in the earlier experiments conducted with DFP was the reactive group.

Histidine exerts general acid/base catalysis during the reaction. In steps [I] and [III] the histidine acts as a base and enhances the nucleophilicity of the attacking group which is the serine in [I] and water in [III]. In the positively charged, doubly protonated form histidine acts as an acid donating the excess proton to the tetrahedral intermediates A and B. This ability of histidine to act both as an acid and a base is attributed to a pK value near neutrality in the enzyme as determined early by titration experiments[17] or by later NMR studies[8]. Direct evidence for the catalytic importance of histidine was gathered by mutation studies[33].

Aspartate was discovered in x-ray crystals of chymotrypsin[21]. This finding led to the proposal of different mechanisms for the nucleophilic activation of the serine residue. In the *charge-relay* system[21] the catalytic triad exists in two tautomeric forms (figure 1.3). The role of aspartate is that of a proton sink, which increases the basicity of the histidine, such that proton abstraction from the serine becomes likely. The alkoxide is the species attacking the substrate. This concept was modified by Hunkapiller[101] who argued that a concerted proton transfer on reaction ('proton shuttle') should be more likely than alkoxide formation.

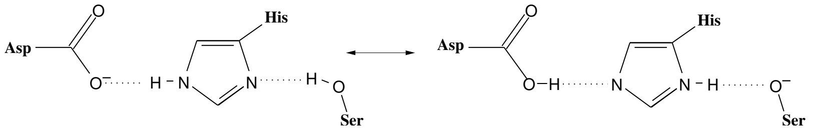


Figure 1.3: The charge-relay mechanism[21] considers tautomerism of the catalytic triad.

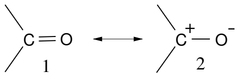


Figure 1.4: Resonance structures of the oxyanion moiety of the TI.

Experimental evidence against a protonated aspartate[136][124][8] ruled out both mechanisms. Modified theories were devised that do not require the proton to be located at the aspartate, but rather to share it with the histidine, thus creating an exceptionally strong hydrogen bond (*short-strong-hydrogen bond* (SSHB)[78] or *low-barrier-hydrogen-bond*(LBHB)[38][69], and were backed by interpretation of NMR data[69][68]. However, the prospects for enzyme catalysis were not explained satisfactorily and the concept was attacked in a later theoretical article[223]. From their point of view, Warshel, Papazyan and Kollman[223] argued that simple electrostatic stabilisation of the transition state by aspartate was sufficient to explain its role in catalysis.

The oxyanion hole is a structural feature composed of NH functions in the backbone of amino acids or hydrogen bond donating side chains (for example asparagine in subtilisin) in the vicinity of the catalytic serine. Its function is the stabilisation of the negative charge developing on the carbonyl oxygen of the substrate and the acylenzyme on formation of the TIs. It was discovered in enzyme crystals that had an inhibitor attached to the active site serine. Two or sometimes three groups serve as hydrogen bond donors, the oxyanion is the recipient. Using vibrational spectroscopy interactions between the carbonyl oxygen of the acylenzyme and the oxyanion hole were shown to exist[32]. Carey and Tonge also found a correlation between the C=O bond length of the acylenzyme and its deacylation rate. Acylenzymes with longer C=O bonds were more reactive than those with short ones. They estimated the increase in bond length for the most reactive acylenzyme to be 0.0025 Å compared to a standard C=O bond length of 1.22Å and explained this finding in terms of two valence bond structures shown in figure 1.4. The electronic interaction with the oxyanion hole is expected to admix significant amounts of configuration 2[32].

1.4 Kinetics of Ester Hydrolysis

According to Gutfreund[88] 'Kinetic investigations tend to expose open ended questions since one can continue to divide steps into ever increasing resolution'. Kinetic rates can be further resolved into more elementary steps, if necessary, but a qualitatively correct picture may be obtained from a coarse macroscopic description. In contrast to this, the application of the methods of computational chemistry requires that investigators face all microscopic details of the reaction, or part of the reaction, from the start. Therefore one should limit the study to an evaluation of the enantioselective steps.

The accepted representation[64] of the kinetics of ester hydrolysis is

$$\mathrm{E} + \mathrm{AcOR} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{E} \cdot \mathrm{AcOR} \overset{-\mathrm{HOR}}{\overset{k_2}{\rightarrow}} \mathrm{AcE} \overset{+\mathrm{H_2O}}{\overset{k_3}{\rightarrow}} \mathrm{E} + \mathrm{AcOH}. \tag{1.2}$$

Here the first step is the binding of the substrate ester AcOR to the protein E to form a *Michaelis complex* (MCC) denoted as $\mathrm{E} \cdot \mathrm{AcOR}$. This complex is assumed to occur in all enzyme reactions but usually lacks structural definition. The initial step is reversible except with inhibitors and reaches equilibrium quickly. The second step is formation of acylenzyme AcE and combines steps numbered [I] and [II] in figure 1.2. Rate k_2 is therefore a composite of two elementary rate constants: The ephemerality of the TI poses outstanding demands on experimental approaches such that it can normally be resolved neither kinetically nor structurally. To this end, and to keep the discussion simple, we consider the TI as a transition state in the following, thus contracting steps [I] and [II] into a single step. In chapters 4 and 5 we quantitatively consider both half-reactions, but this does not alter the qualitative picture needed for an understanding of enantioselectivity, which is developed here. Deacylation is irreversible in aqueous solution due to water competing with alcohol for nucleophilic attack on the acylenzyme. The concentration of water exceeds that of the alcohol by far and makes reesterification negligible. The third step, deacylation of the enzyme, recovers enzyme and completes ester hydrolysis by release of the carboxylic acid AcOH. This step is assumed to be rate-limiting for most ester substrates, although there are some exceptions (see chapter 7 in [64]). Deacylation was certainly rate limiting in the experiments conducted by Hartley and Kilbey[96]. In their kinetic analysis of the hydrolysis of nitrophenyl acetate by chymotrypsin, an initial burst of nitrophenol release was detected and attributed to a fast acylation of the enzyme followed by slow hydrolysis of the acylenzyme. Apart from mechanistic studies rate constants for individual steps on the reaction path were reported[16][20]. A complete reaction profile (according to equation 1.2) including free energies of Michaelis complex association, and of both acylation and deacylation for a number of substrates have been measured[138] using appropriate techniques[64].

Enantioselectivity can be measured by kinetic resolution of racemates. Assuming the same mechanism as above (equation 1.2) we may write

$$\begin{array}{lclcl}
\mathrm{E} + \mathrm{AcOR_R} & \underset{\mathrm{k_{-1}}}{\overset{\mathrm{k_1}}{\rightleftharpoons}} & \mathrm{E \cdot AcOR_R} & \overset{\mathrm{k_2}}{\rightarrow} & \mathrm{AcE} + \mathrm{HOR} \overset{\mathrm{k_3}}{\rightarrow} \; \mathrm{E} + \mathrm{AcOH} \\
 & & & \nearrow k_5 & \\
\mathrm{E} + \mathrm{AcOR_S} & \underset{\mathrm{k_{-4}}}{\overset{\mathrm{k_4}}{\rightleftharpoons}} & \mathrm{E \cdot AcOR_S} & &
\end{array} \tag{1.3}$$

Chiral esters are denoted as $\mathrm{AcOR_R}$ and $\mathrm{ACOR_S}$, the alcohol moiety carries the asymmetric information, whereas the constituent acids (acetic acid) are achiral. The alcohol leaves the enzyme environment upon formation of the acylenzyme. This leads to identical acylenzymes being hydrolysed, therefore this step cannot be enantiodiscriminative. The enantioselectivity factor **E** can thus be formulated as[193]

$$\mathrm{E} = \frac{\mathrm{k_{cat_R}/K_{m_R}}}{\mathrm{k_{cat_S}/K_{m_S}}}. \tag{1.4}$$

The Michaelis parameters for this kinetics are composite rate constants expressed as $\mathrm{k_{cat_R}} = \frac{\mathrm{k_2 k_3}}{\mathrm{k_2+k_3}}$ and $\mathrm{K_{m_R}} = \frac{\mathrm{k_{-1}}}{\mathrm{k_1}} \cdot \frac{\mathrm{k_3}}{\mathrm{k_2+k_3}}$. The relations for the S-enantiomer are analogous.

Enantioselection must occur at the binding steps and/or acylation, the relevant rates of which are $\mathrm{k_{-1}, k_2, k_{-4}}$ and $\mathrm{k_5}$. Here we assume that substrate insertion proceeds similarly fast and is essentially barrierless for both enantiomers. Rates $\mathrm{k_1}$ and $\mathrm{k_4}$ are thus not considered in the following. There are two limiting cases that recover the simple form of equation 1.1. Assuming different activation energies for acylation and identical rates of the preceding steps, one obtains $\mathbf{E} = \mathrm{k_2/k_5}$ (figure 1.5 A). On the other hand, if only the binding free energy differs, one arrives at $\mathbf{E} = \mathrm{k_{-4}/k_{-1}}$ (figure 1.5 B)). In real systems, the enantioselectivity will be determined by the free energy differences for both steps. In the hypothetical case C) depicted in figure 1.5, $\mathrm{k_{-1}, k_{-4}, k_2}$, and $\mathrm{k_5}$ are all different (note that in this example $\mathrm{k_{-4}/k_{-1}} = \mathrm{k_5/k_2}$). In examples A) and B) it would be sufficient to calculate the free energy difference of the TIs, and in C) that of the MCCs. In the absence of experimental data to support any of the simplified cases one must consider both the TI and the MCC as in D).

Figure 1.6 demonstrates the development of the product enantiomeric excess (e.e.) during the course of a kinetic resolution for different values of E. The reaction was simulated using Gillespie's algorithm (see appendix D). Model A) of figure 1.5 was applied by variation of $\mathrm{k_2}$ and $\mathrm{k_5}$. In table 1.1 the simulation parameters are listed.

In a typical experimental setting the degree of enantioselectivity of an enzymatic reaction is determined by measurement of the chiral product (or reactant) ratios in kinetic resolutions of racemates. The analysis of the composition of the reaction mixture can be done on chiral columns by gas chromatography after stopping the reaction. The experimentalist usually

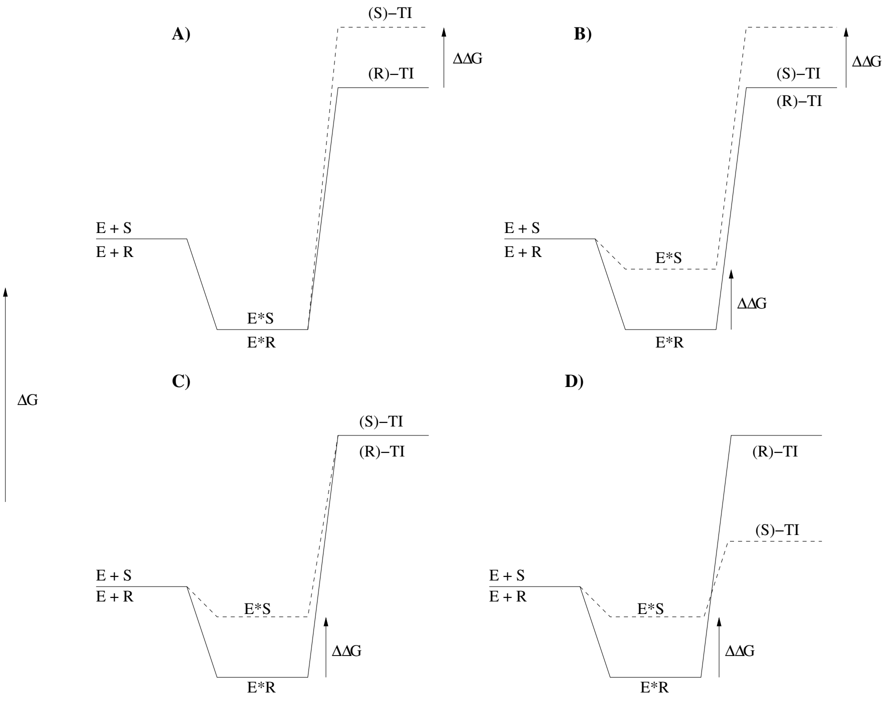


Figure 1.5: Scenarios leading to enantioselective ester hydrolysis in terms of free energy differences along the reaction coordinate. Initially ester and enzyme are well separated, providing a reference (enantiomers are degenerate). Formation of the Michaelis complex (E*R,E*S) and further reaction take place on the asymmetric surface of the enzyme and give rise to enantioselectivitiy. Here the TIs are treated as transition states, as is often done in the literature[97][40][22].

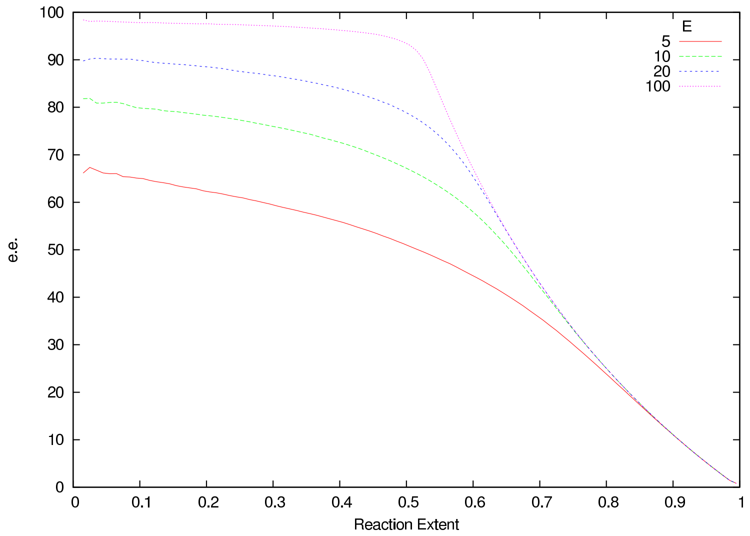


Figure 1.6: The enantiomeric excess depends on the extent of reaction in a kinetic resolution experiment. Here we record the enantiomeric excess of products for enzymes with different selectivities.

Rate parameters				Concentrations		
k_1	k_{-1}	k_2	k_3	E	R	H_2O
500	10	0.05	5000	5	5000	1000000
k_4	k_{-4}	k_5		S		
500	10	0.0005		5000		

Table 1.1: Parameters for stochastic simulation of the kinetic resolution of esters. R is the faster reacting enantiomer. The parameters above lead to E = 100; k_5 was varied to achieve additional E values. Figure 1.6 displays e.e. vs. total conversion for different E values.

determines the degree of total conversion of the reactants after time increments. Although it is logical that in kinetic resolutions at 100% conversion a racemic mixture of products is regained, one should have a minimum conversion of approximately 30% to minimise statistical error that lead to inaccurate estimation of E. For preparative purposes the desired e.e. is achieved by monitoring the degree of conversion and stopping the reaction in time. As seen in figure 1.6 a high yield of enantiopure R product is retrieved up to 50% conversion for high values of E. Vedejs and Jure wrote a review article on enzymatic kinetic resolution of racemates as well as for nonenzymatic catalysis recently[210]. A specialised review on the use of lipases in kinetic resolution is provided by Ghanem[80]. Sih and Wu[193] as well as Kagan and Fiaud[109] give general overviews including physicochemical considerations. Prochiral molecules are interesting precursors in asymmetric synthesis. While in kinetic resolution of racemates only 50% yield can be reached for any enantiomer, with prochiral molecules theoretically 100% of yield are possible. Cis-cyclopent-2-en-1,4-diacetate is such a valuable precursor[201]. Its chiral monoacetates that can be generated using lipases as catalysts are valuable building blocks in the synthesis of prostagladins, carbocyclic analogs of nucleosides and other natural products. BSLA is only slightly enantioselective towards this substrate in the wild type, but even single mutations enhance its selectivity considerably[60].

1.5 Microscopic Basis of Lipase Enantioselectivity

The thermodynamic principles that lead to enantioselective catalysis have been laid out in the preceding chapter. From a macroscopic point of view the basis is therefore solid. Different microscopic concepts have been devised, but there exists no consensus concerning the qualitative microscopic explanation of enantioselectivity. In principle all factors that govern ligand binding, hydration of the enzyme and ligand, or allosterism, can be important. Some concepts both by experimental and theoretical groups are briefly reviewed in the following. Analysis of the microscopic events leading to enantioselectivity relates free energy components. In terms of differential free energy, the fundamental relation is $\Delta\Delta G^\ddagger = \Delta\Delta H^\ddagger -$

$T\Delta\Delta S^{\ddagger}$. The free energies are often written as activation energies of a single virtual TS, although k_{cat}/K_m contains elements of both binding and catalytic steps[158]. Differential enthalpic contributions arise from charge-charge or general multipole-multipole interactions, for example, that may preferentially occur for only one enantiomer upon binding in the Michaelis complex or the TI. Enantioselective ester hydrolysis by a lipase was attributed to such interactions in a recent publication[22]. In this study the S enantiomer was the preferred substrate (E=51) in a mutant enzyme containing six amino acid exchanges. The TI of S was stabilised electrostatically by an additional hydrogen bond in the oxyanion hole, whereas the formation of this hydrogen bond in the TI of R was sterically hindered. In a theoretical study of enantioselectivity of subtilisin[41][40] the authors used different sets of charges for each enantiomer. These charges were derived using QM/MM methodology, placing both TIs in the active site, and were found to differ considerably between both enantiomers. The TI for S, which was the faster reacting enantiomer, had a higher concentration of negative charge on the oxygen atoms next to the tetrahedral carbon. This was interpreted to have a stabilising effect on the hydrogen bonding network of the TI. In a study of *Candida rugosa* lipase, chiral inhibitors (1R)-menthyl-hexyl-phosphonate and (1S)-menthyl-hexyl-phosphonate were reacted with the enzyme and crystallised[45]. The binding of the tetrahedral phosphonate moiety, the acid moiety and the cyclohexyl rings of the alcohol (menthol) were very similar with both enantiomers. By contrast, the position of the isopropyl chains of the menthol was largely different. This led to a displacement of the catalytic histidine in a way that distorted the hydrogen bond network in the active site for the slow reacting enantiomer. In some studies, the probability of hydrogen bond formation (with the oxygen of the TI belonging to the leaving alcohol moiety) was evaluated from MD simulations, and related to reactivity[23][184] (higher probability implying higher reactivity).

Similar to polar interactions, attractive Van-der-Waals (VdW) interactions with the enzyme can stabilise the TI. However, any binding event (of either nonpolar of polar origin) has an unfavourable component in terms of free energy since a loss in translational and rotational motion is associated with a decrease in entropy, and $T\Delta\Delta S^{\ddagger}$ then makes the process less spontaneous. In case of the mutant enzyme[22] in the preceding paragraph, the gain in terms of enthalpy will easily offset any loss of entropy, since the charge-dipole interaction renders $\Delta\Delta H^{\ddagger}$ the dominant contribution. In several studies[23][155][153] $\Delta\Delta H^{\ddagger}$ was found to be almost always negative, and to outweigh the counteracting $T\Delta\Delta S^{\ddagger}$ term ($\Delta\Delta S^{\ddagger}$ is normally negative). The apparent differential activation free energy will therefore tend to be too large when ignoring entropic terms, and E values thus calculated would be too high. There are exceptions, however; in one case reported by Ottosson[153], the differential activation entropy was positive, thereby increasing E. In kinetic resolution of fenoprofen thioester by *Carica papaya* lipase[149] $T\Delta\Delta S^{\ddagger}$ was the dominant contribution, and even reversed enantioselectivity.

Nonpolar interactions can also be thought to contribute in a different manner to enzyme enantioselectivity. As known from solvation thermodynamics, the *hydrophobic effect*[35] drives apolar compounds out of the aqueous phase, or at least minimises their surface with water. One possibility to do so is by clustering in the form of micelles. Another one is to absorb to the surface or a cavity inside a protein. The overall result is a loss of free energy, indicating a spontaneous process. If the enantiomers bind differently this can again contribute to $\Delta\Delta G^{\ddagger}$. To estimate this contribution one can compare the solvent accessible surfaces for the enantiomers for hydrophobic parts of the substrate[40][114][227]. The enantioselectivity with substrates governed by this effect should depend on the composition of the solvent. Hydrolase activity in nonaqueous media was studied and confirmed these notions quite some time ago[120]. The use of different solvents and binary mixtures allows enantioselectivity to be adjusted[154][119][156].

Philips[159] defined a racemic temperature $T_r = \Delta\Delta H^{\ddagger}/\Delta\Delta S^{\ddagger}$, where a temperature-dependent inversion of stereochemistry is predicted to occur. He employs a *two-site* model with large and small binding pockets where nonconvalent interactions of the enzyme substrate complex with these pockets determine the thermodynamic and kinetic properties of the complex and thus the stereochemical outcome. The basic assumption is a close balance of $\Delta\Delta H^{\ddagger}$ and $T\Delta\Delta S^{\ddagger}$, which is more likely to occur with substrates that have mainly nonpolar side chains. In terms of figure 1.7 this means enthalpic advantages exist for complex **a** due to favourable VdW contacts, but entropic considerations render **b** more likely. If $\Delta\Delta H^{\ddagger}$ dominates, the reaction should be carried out at the lowest practical temperature to obtain the highest e.e. possible. The reverse is true for systems dominated by $T\Delta\Delta S^{\ddagger}$.

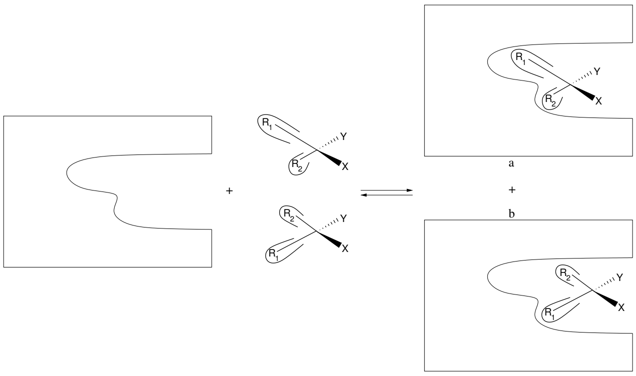


Figure 1.7: Two site model for stereoselectivity. Figure adapted from [159]

A detailed review article of lipase enantioselectivity is provided by Ema[57], including both a variety of experimental and theoretical examples. For studies of enantioselectivity in enzymes using QM/MM methodology the structure-reactivity concepts described above are useful for interpretation of the results. Our computational methods allow the bond making and breaking processes occuring in lipases to be studied and the free energy changes accompanying them to be calculated. Hence, enantioselectivity can in principle be predicted from theory.

1.6 Molecular Models

In computational enzymology the aim is to understand catalytic events at the atomistic level, and to include environmental effects originating from the active site, the remainder of the protein, and the aqueous phase surrounding it[137][218][75][219] [212][234]. Enzyme efficiency and stereoselectivity arise from the complex interplay of those partitions.
Models aimed at reproducing such complex and heterogeneous condensed phase systems at the atomistic level need to fulfill certain criteria:

1. Reproduction of intramolecular forces (e.g. bonding).
2. Proper description of bond breaking and forming.
3. Inclusion of environmental effects (e.g. solvent effects or interactions with macromolecular surroundings).

Quantum mechanical (QM) methods are in principle capable to satisfy all needs. The sole problem is the scaling behaviour of typical current methods from $\mathrm{O}(\mathrm{N}^3)$ up to $\mathrm{O}(\mathrm{N}^7)$ (N = number of electrons) which inhibits their use on large systems such as solvated proteins. With uncorrelated wave function methods or density functional theory (DFT) up to a thousand atoms can be treated, but if many energy and gradient evaluations are required to extensively study conformational states the maximum number is much smaller. Semiempirical methods[202] like AM1 or SCC-DFTB[56] allow larger atom counts, and their much higher speed also permits a massive increase in the number of conformational states that can be visited. The price of empirical parameters in semiempirical formulations is a limited accuracy in general, the restriction to atoms that have been parameterised, and artefacts for molecular properties not considered in the parametrisation. AM1 for example does not model hydrogen bonds at a useful level, and rotational barriers around double bonds are too small[143]. Correlated wave mechanics approaches perform well at reproducing molecular properties and geometries in general, and of weakly bound complexes in particular, but are limited to small systems of a few dozen atoms (MP2-level) or less (Coupled-Cluster theories)[105]. Such methods may sometimes still be used to treat model systems of the enzyme active site or substrates, to check the performance of computationally less intensive methods, or to parametrise molecular mechanical force fields. Average effects of isotropic solvents can be included in QM calculations using continuum solvation models[43][118].

Molecular mechanical (MM) methods are mainly applied to large macromolecular systems to sample configuration space and to study the interplay of solvent and solute. The formulation of force fields is empirical and their parametrisation applies to a limited class of molecules. The fitting of parameters is usually done against a limited set of properties and good performance with observables not within that set is not guaranteed. In the MM world the pattern of bonding cannot be established by the balance of forces due to the electronic distribution. In the simplest formulation, which is commonly used in biomolecular force fields, bonding is established by harmonic terms: $\mathbf{E} = \mathrm{k} \cdot (\mathrm{r}_{\mathrm{A-B}} - \mathrm{r}_0)^2$. Here k is a force constant, r_0 is the equilibrium distance, and $\mathrm{r}_{\mathrm{A-B}}$ the instantaneous bond length. Such a form does not allow for bond breaking to occur, and the formation of bonds is not modeled either. Rather, a *topology* is defined fixing the bonding pattern for all times, thereby violating criterion 2 in the list above. For this chosen topology all other force field terms in addition to bond terms are assigned, e.g., angle bends, Urey-Bradley terms, torsions, Coulomb interactions, and VdW terms[6]. Atoms are assigned unique types taking into account their chemical surrounding. By the time a simulation is started all parameters for bonded/nonbonded combinations of atom types encountered in the molecular system must be defined. The largest number of parameters are usually needed for torsions, where four atoms per term participate. The form of the Charmm22 force field and its fitting procedure are highlighted in a later chapter.

Hybrid quantum mechanical / molecular mechanical (QM/MM) methods are designed to unify the advantages of both the QM and the MM worlds. A quantum mechanical core region is typically embedded in a large molecular mechanical environment. Both partitions can interact with each other, and the nature and realisation of the coupling is variable. A taxonomy of coupling schemes was developed by Bakowies and Thiel[9]. The QM/MM methodology was introduced by Warshel and Levitt, who studied the reaction of lysozyme using a semiempirical method for the QM core[221]. The acceptance of QM/MM potentials was not immediate and it took several years until implementations in other molecular modeling packages became available. The Amber program was the first of the renowned suites to be equipped with QM/MM capability[195], followed by CHARMM[65].

---

[6] There can be many other terms, but these are the common ones in biomolecular force fields.

QM/MM is nowadays a widely used technique in studies of large molecular systems. A number of reviews that focus on biomolecules[70][144], and on general applications[133][191] have been published. Gao evaluated QM/MM as a method for treating solvation[73]. Recent studies from our group have provided mechanistic insight into enzyme reactions[186][183], and have also contributed to the technology of free energy calculations[187][59].

With special functional forms and parametrisations a treatment of covalent chemical reactions is possible even with MM force fields. Methods of reactive force fields were pioneered by Warshel, who introduced them as *empirical valence bond* (EVB) approach that was applied to a variety of chemical and biochemical problems[212][217]. In the spirit of QM valence bond approaches, a Hamiltonian is built from resonance structures of the reacting fragment, which are modeled by a force field. The mixing of resonance forms is achieved by off-diagonal elements that have a parametric dependence on atomic distances. The advantages of using this empirical formulation are speed of computation and parametrisability. However, to be able to parametrise, sound knowledge of the reaction mechanism is required. New mechanisms cannot be discovered. For this reason EVB is suited mainly to the evaluation of environmental effects on reaction barriers, occuring on mutation, or on asymmetric induction in enantioselective processes, for example.

Among the studies conducted with the EVB method are the first theoretical studies of the mechanism of serine proteases ([222] and references therein) taking into account the enzyme environment and solvation. Catalytic free energies in reactions of trypsin[224] and mutant enzymes were in accordance with experiment. More recent application in alkylation reactions highlighted the value of EVB in the study of substituent and environmental effects[211]. Reactive forecefield methods are used and developed in several other research groups. The *approximate valence bond* (AVB) model was applied to the reaction of phospholipase A_2 in quantum-classical MD[85][10]. Hammes-Schiffer and coworkers approached hydride transfer in liver alcohol dehydrogenase with EVB, focusing on contributions of enzyme motion to energy barriers and nuclear quantum effects[93]. Truhlar and coworkers developed *multiconfiguration molecular mechanics* (MCMM) including a sophisticated scheme for meticulous fits to *ab initio* energy surfaces and derivatives[132]. A conceptionally simpler connection of EVB to *ab initio* potential surfaces was elucidated by Chang and Miller[36], and Sierka and Sauer integrated EVB into a solid state chemistry package[192].

Chapter 2

Theoretical Background

2.1 Density Functional Theory

Density functional theory (DFT) rests on the theorems of Hohenberg and Kohn[99], the first of which states that the electronic energy of a system in its ground state is entirely determined by the electron density ρ. The mapping from density to energy is mediated by a functional, leading to the exact energy including all correlation effects if the exact mathematical form of the functional were known. The second contribution[99] is the proof of existence of a variational principle for the energy given the true density functional, which can be used to obtain the correct ground state energy. Since the exact functional is not known approximate forms have to be devised. Wave-function based methods depend on 3N variables (N=number of electrons). The electron density is determined by three coordinates independent of the number of electrons making it attractive for the treatment of large systems. The expression for the electronic energy within the DFT framework can be written as

$$\mathrm{E}_{\mathrm{elec}}[\rho] = \mathrm{T}[\rho] + \mathrm{E}_{\mathrm{ne}}[\rho] + \mathrm{J}[\rho] + \mathrm{E}_{\mathrm{ncl}}[\rho]. \tag{2.1}$$

Here $\mathrm{T}[\rho]$ is the kinetic energy, $\mathrm{E}_{\mathrm{ne}}[\rho]$ the nuclei-electron attraction, $\mathrm{J}[\rho]$ the classical Coulomb interaction, and $\mathrm{E}_{\mathrm{ncl}}[\rho]$ contains all non-classical contributions like exchange and correlation. Early attempts using the electron density[203][62] as the basic variable rather than wave functions were pursued out of intuition before the fundamental proofs of Hohenberg and Kohn[99], but performed poorly for molecules since bonding could not be described.

Modern DFT became workable due to the introduction of orbitals by Kohn and Sham[121]. Like in Hartree-Fock theory the aim is to make the variation of the energy stationary, under the constraint of orthonomality of the orbitals. The effective one-electron operator

$$\hat{\mathrm{H}}_{\mathrm{KS}} = -\frac{1}{2}\nabla^2 + \mathrm{V}_{\mathrm{ne}} + \int \frac{\rho(\mathbf{r})}{\mathbf{r}-\mathbf{r}'}\mathrm{d}\mathbf{r}' + \mathrm{V}_{\mathrm{ncl}}(\mathbf{r}) \tag{2.2}$$

is evaluated and the Kohn-Sham equations $\hat{\mathrm{H}}_{\mathrm{KS}}\phi_{\mathrm{i}} = \epsilon_{\mathrm{i}}\phi_{\mathrm{i}}$ are solved as a pseudo-eigenvalue problem. Since the energy depends on the total density, which is yet to be determined, the procedure is iterative, starting from an initial guess. The exchange-correlation functional $\mathrm{E}_{\mathrm{ncl}}$ can take different forms. It is customary to split it up into a pure exchange part E_{x} and a pure correlation part E_{c}. One possible formulation for the pure exchange, which is used mainly in this thesis, is due to Becke[14]. The correlation part can be expressed as proposed by Lee, Yang, and Parr[131][141]. Exchange-correlation functionals have a complicated appearance that does not allow simple interpretation of its components.

The semiempirical SCC-DFTB method[56], which is implemented in the CHARMM package[44], offers fast calculation of QM/MM energies and gradients, and it is thus applicable to MD simulations or geometry optimisations. The derivation from standard DFT proceeds by a second-order expansion of the energy with respect to a variation of the charge density[56][44]. All interaction terms that arise are transformed into atomic pairwise potentials. The energy expression reads

$$\mathrm{E}^{\mathrm{SCC-DFTB}} = \sum_{\mathrm{i}}^{\mathrm{occ}} \left\langle \phi_{\mathbf{i}} | \hat{\mathrm{H}}_{\mathrm{KS}} | \phi_{\mathbf{i}} \right\rangle + \frac{1}{2}\sum_{\mathrm{A,B}} \gamma_{\mathrm{AB}} \Delta \mathrm{q}^{\mathrm{A}} \Delta \mathrm{q}^{\mathrm{B}} + \mathrm{E}_{\mathrm{rep}}, \tag{2.3}$$

where $\hat{\mathrm{H}}_{\mathrm{KS}}$ is the effective Kohn-Sham Hamiltonian, and γ_{AB} a function modeling the second derivative of the total energy with respect to the charge density fluctuations. The sum in the second term of equation 2.3 runs over all pairs of atoms A,B, and the $\Delta\mathrm{q}^{\mathrm{A}}$ and $\Delta\mathrm{q}^{\mathrm{B}}$ terms approximate charge density fluctuations ($\delta\rho$) as the deviation in Mulliken population at the atomic centers from the number of valence electrons of the neutral atoms ($\Delta\mathrm{q}^{\mathrm{A}} = \mathrm{q}^{\mathrm{A}} - \mathrm{q}_0^{\mathrm{A}}$). $\mathrm{E}_{\mathrm{rep}}$ is a contracted expression to correct for double-counting terms from Coulomb and exchange-correlation contributions in matrix elements of $\hat{\mathrm{H}}_{\mathrm{KS}}$, and also contains the core-core repulsion energy. As in other semiempirical methods only the valence electrons are treated explicitly, and a minimal basis of Slater orbitals represents the Hamiltonian matrix elements $\left\langle \phi_{\mathrm{i}} | \hat{\mathrm{H}}_{\mathrm{KS}} | \phi_{\mathrm{i}} \right\rangle$. A two-center approximation is introduced and the resulting integrals are pretabulated with respect to interatomic distances. Likewise $\mathrm{E}_{\mathrm{rep}}$ is represented as a sum of pairwise atom-atom potentials which are parameterised against small molecular reference systems. The energy is minimised by iteration to self consistency with respect to Mulliken charges due to the dependence of equation 2.3 on the latter, and not to self consistency of

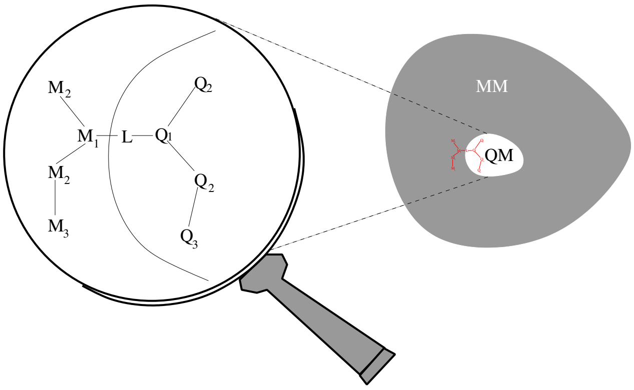


Figure 2.1: Atom labels in the vicinity of the QM/MM boundary (as in reference[191]). The Q atoms are the QM atoms and the M atoms belong to the MM partition. Q_1 and M_1 are the atoms involved in the frontier bond. L is a link atom.

the electron density, as in standard DFT.

2.2 The QM/MM Method

As discussed in the introduction, QM/MM methods provide a way to incorporate environmental effects at atomic resolution in systems where the use of quantum mechanics is mandatory in a region of limited size. Examples are reactions, electronic excitations, or spectroscopically accessible properties. If a system has two components as in figure 2.1, a QM and an MM partition, the energy of the whole system can be expressed as a sum

$$\mathrm{E} = \mathrm{E}_{\mathrm{QM}} + \mathrm{E}_{\mathrm{MM}} + \mathrm{E}_{\mathrm{QM/MM}} \ . \tag{2.4}$$

Here E_{QM} and E_{MM} are the energies of the isolated subsystems, and $\mathrm{E}_{\mathrm{QM/MM}}$ results from the interactions of both regions. Any QM method may be chosen in principle, but DFT and semiempirical theories seem to dominate in contemporary literature. The most flexible aspect are interactions between the QM and MM subsystems captured by $\mathrm{E}_{\mathrm{QM/MM}}$. First we consider two isolated fragments (not connected by covalent bonds as in figure 2.1). A simple type of coupling is provided by the VdW terms between QM and MM atoms using the expression of the classical force field. Compared to a vacuum, this coupling implies that the QM region is now subject to steric effects. Similarly, the electrostatic terms may be included via the force field, but this approach is unsatisfactory since the wavefunction or electron density does not ’feel’ the surrounding charges directly. When the interactions between QM and MM fragments are handled at the classical level the QM/MM scheme is termed *mechanical embedding*. A more realistic model is achieved by inclusion of point charges of the force field in the QM Hamiltonian, thus being able to account for polarisation effects in the QM-region. This is the *electrostatic embedding* scheme[9], which is used throughout this work. Two additional coupling models were defined allowing for the treatment of polarisation of the MM region (*polarised embedding*), either non-self-consistently[9], or self-consistently[5].

In cases where one or several bonds exist between QM and MM regions one must augment the QM/MM approach with an adequate treatment of the boundary. This aspect generates the largest differences in present QM/MM implementations, and makes a thorough comparison of results between different programs difficult. The goal is to retain the bond(s) and intramolecular forces across the boundary, while satisfactorily terminating both the QM and the MM region. Here the most difficult problem arises with the QM partition, where the cutting of covalent bonds usually generates open valencies, that in turn would generate unrealistic and unacceptable electron distributions. A useful remedy is provided by the introduction of *link atoms*, typically hydrogen[195][65][171], which serve to saturate the QM region (see figure 2.1). In another approach pseudobonds are constructed to terminate C-C single bonds[237][238]. The QM/MM boundary is not across the $\mathrm{Q_1}$-$\mathrm{M_1}$ bond in this case (figure 2.1), but coincides with $\mathrm{M_1}$. $\mathrm{M_1}$ is included in the QM Hamiltonian with one free valence, and assigned a fluorine basis set. An effective core potential is introduced and parametrised to mimic realistically the bond length and forces. In the same spirit *connection atoms* were introduced with semiempirical HF methods[6]. A third type of coupling is provided via implementation of hybrid orbitals as in the groundbreaking study of Warshel and Levitt[221][1].

The link atom approach is used exclusively throughout this work since it is readily available in the programs Chemshell and CHARMM. Some further amendments have to be added to the QM/MM boundary, though, since all that was achieved up to this point is a saturation of the QM region. To actually bind both fragments together an MM bond stretch potential is applied to $\mathrm{Q_1}$ and $\mathrm{M_1}$. Introduction of the link atom introduces an artificial degree of freedom, which is undesirable in optimisations or MD. This could be corrected using a constraint to force the atom on the bond vector $\mathrm{Q_1M_1}$, which would require a coordinate transformation[83]. A simple solution in Cartesian coordinates is to express the link atom forces in terms of forces on the frontier atoms ($\mathrm{Q_1}$,$\mathrm{M_1}$) by application of the chain rule

$$\frac{\partial \mathrm{E}}{\partial \mathbf{r}_{\mathrm{M_1}}} = \frac{\partial \mathrm{E}}{\partial \mathbf{r}_{\mathrm{M_1}}} + \frac{\partial \mathrm{E}}{\partial \mathbf{r}_{\mathrm{L}}} \frac{\partial \mathbf{r}_{\mathrm{L}}}{\partial \mathbf{r}_{\mathrm{M_1}}} \, . \tag{2.5}$$

To couple the QM fragment even closer to the MM part, other intramolecular potentials

---

[1] For a more complete overview of coupling methods please refer to one of the review/methodological articles[70][133][190][171][73][61]

may be looped through the frontier bond. In the Charmm22 force field those would be angle bends, Urey-Bradley terms, and torsion potentials. A double-counting of interactions must be avoided here, since bending the M_1-Q_1-Q_2 angle results in a force acting on M_1 via L_1 and the chain rule. The force field terms for M_1-Q_1-Q_2 are obviously not included, and neither are Urey-Bradley terms of the type M_1-Q_2. Finally we exclude M_1-Q_1-Q_2-Q_3 from computation.

The proximity of MM charges at the QM/MM interface can lead to problems when a covalent bond between both regions exists, as the distance between the fragments is small, particularly for link atoms. A possible solution is the deletion of all charges of the first charge group on the MM side of the boundary[2][61]. If this is not acceptable, a *charge shift*[191] scheme may be used, where only the charge on M_1 is set to zero. The charges of the M_2 atoms are then adjusted to maintain integral charge, and pairs of equal and opposite point charges, placed near the M_2, and along the M_1M_2 bond vector reproduce the dipole moment of the original group. The charge shifting scheme is the preferred electrostatic boundary correction scheme in Chemshell.

2.3 Chemical Reactions: The Computational Approach

The definition of molecular models (*vide supra*) contains no *a priori* prescription of how to actually use them. To establish a connection with experiment, algorithms are needed that *employ* those models. In the following an outline of the approaches used in this work is given, covering potential energy surface (PES) based methods, and free energy techniques. The section is closed with an overview of novel methods that may prove useful in future studies.

The course of a chemical reaction can often be characterised in terms of progress along a reaction coordinate, as shown in figure 2.2. A system that is initially in a reactant state climbs a barrier, passes over an energy maximum called a transition state (TS), and then descends towards the products for high values of the reaction coordinate (RC). The RC, sometimes also called a progress variable, may be chosen to be a function of the molecular coordinates (*vide infra*), although more abstract definitions in terms of nonspatial properties are possible. Definition of a low-dimensional RC provides a convenient way to reduce the dimensionality[3] of the problem. Thus we separate out a (small) set of atoms that participate directly in

---

[2] Charmm22 sorts atoms into groups, which carry integral, but not necessarily zero charge. A C=OC(R)NH amino acid group excluding R *is* uncharged. Assigning the side chain group R as QM region is a typical case.

[3] The PES has 3N dimensions in Cartesian space, N being the number of atoms in the system.

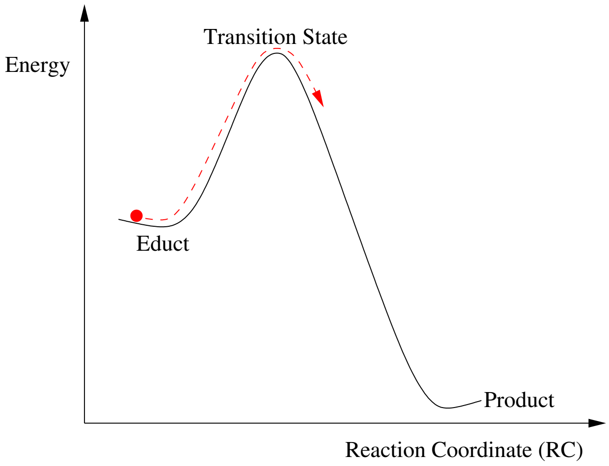


Figure 2.2: Reacton profile

the reaction. A RC $\xi(\mathbf{r_1}, \mathbf{r_2}, ..., \mathbf{r_n})$[4] is then defined to be a function of the atomic positions and relates to the chemical transformation under study. ξ should lead from the reactant to the TS and from the TS to the products along a minimum energy path in the $\mathbf{r_1}, \mathbf{r_2}, ..., \mathbf{r_n}$ subspace. The TS occurs at $\xi^{\ddagger}$, the value on the RC which separates reactant and product valleys and appears as a maximum in the potential energy (or free energy) plot. Segregation of a set of coordinates $(\mathbf{r_1}, \mathbf{r_2}, ..., \mathbf{r_n})$, which are considered essential, from the rest of the system $(\mathbf{r_{n+1}}, \mathbf{r_{n+2}}, ..., \mathbf{r_N})$, that is thought to contribute as 'bystanders', has an implication. The TS is not a single structure, but rather a collection of structures determined by variations of the less important coordinates. For statistics one collects data from a small intervall defined as $|\xi - \xi^{\ddagger}| < \delta$.

2.4 Static Treatment of a Reaction

It was practiced, but not strictly necessary to employ an approximate RC as above to locate TSs. A TS is a feature of the underlying PES (the graph of the function), and is characterised as such by *one* negative eigenvalue of the systems' Hesse matrix, and termed first-order saddle point. Normal mode analysis[83], performed for the saddle point configuration, returns (in harmonic approximation) an all-atom RC[5]. This complicated RC may be used to assess the quality of approximate RCs, that have a simple functional dependence on a few atoms only. From the opposite point of view, an approximate saddle point found along a simple RC may be checked for its being a true TS using normal mode analysis. Product and reactant geometries may be reached from a transition state structure by adding/subtracting

---

[4] We assume ξ to be one-dimensional in the discussion, but the RC can generally be multidimensional.

[5] Which is strictly valid only at the TS.

increments of the normal mode eigenvector corresponding to the imaginary frequency, and subsequent geometry optimisation [6].

Up to this point the RC was a diagnostic tool to describe the path from reactants to products. Now it will be used drive a chemical reaction. Typical RCs are distances between atoms, or linear combinations thereof. The corresponding harmonic potentials can be added to the total energy of the system as

$$\mathrm{f}_{\mathrm{restraint}} = \mathrm{k}(\mathrm{r}_{\mathrm{A-B}} - \mathrm{S})^2, \tag{2.6}$$

$$\mathrm{f}_{\mathrm{restraint}} = \mathrm{k}(\mathrm{r}_{\mathrm{A-B}} - \mathrm{r}_{\mathrm{C-D}} - \mathrm{S})^2, \tag{2.7}$$

$$\mathrm{f}_{\mathrm{restraint}} = \mathrm{k}(\mathrm{r}_{\mathrm{A-B}} - \mathrm{r}_{\mathrm{C-D}} - \mathrm{r}_{\mathrm{E-F}} - \mathrm{S})^2. \tag{2.8}$$

Here k is a force constant, r denotes a distance, and S is a variable that defines the equilibrium position. Potential 2.6 is identical to the bond term that is used in many force fields. Other possible functional dependencies between atoms may be expressed using bond angles or dihedral angles. When a geometry optimisation (minimisation) is conducted in the presence of such potentials the system is forced from its current position to a new stationary point in an attempt to satisfy the equilibrium position S of the additional term. To this end rather large force constants (k) are chosen (1-3 a.u. are useful, see for example [238]). Iterative optimisation and adjustment of S lead to a cut through the PES in the sense of figure 2.2. This procedure is sometimes referred to as *adiabatic mapping*[220] or the *distinguished-coordinate method*[228]. In our study we are concerned mainly with reaction pathways of chemical transformations, therefore the term *reaction path method* is used preferentially.

For optimisation purposes a gradient expression is needed. The gradient of the restraint term is given below for the potential term in equation 2.6, the functional forms for 2.7 and 2.8 are similar:

$$\begin{aligned} \frac{\mathrm{df}}{\mathrm{d}\mathbf{r}_{\mathrm{A}}} &= 2\mathrm{k}(\mathrm{r}_{\mathrm{A-B}} - \mathrm{S}) \left(+ \frac{1}{\mathrm{r}_{\mathrm{A-B}}} \mathbf{r}_{\mathrm{A-B}} \right) \\ \frac{\mathrm{df}}{\mathrm{d}\mathbf{r}_{\mathrm{B}}} &= 2\mathrm{k}(\mathrm{r}_{\mathrm{A-B}} - \mathrm{S}) \left(- \frac{1}{\mathrm{r}_{\mathrm{A-B}}} \mathbf{r}_{\mathrm{A-B}} \right) \end{aligned} \tag{2.9}$$

For large and unknown systems it is difficult to guess structures that are close enough to a TS for optimisers to converge towards it. The simple RC method is a good option to gain experience on the system, and to get close to a TS. The latter should then be asserted

---

[6] Minimisation of the energy with respect to variation of the coordinates using numerical schemes, such as steepest descent, conjugate gradients or BFGS procedures[163].

(*vide supra*) or be further refined using a different method, for example the microiterative optimisation offered in HDLCopt[19][7]. The RC can then be corrected if necessary, and used on different structures of the same system that were generated by MD, for example. The aim is to use adiabatic mapping unsupervised, since it is a straightforward method.

2.5 Free Energy Methods

Thermodynamic potentials play an important role in quantitatively characterising chemical reactions. In this context, the free energy is particularly useful. From the macroscopic observations made by experimentalists, for example in kinetic studies or measurements of E, all of which can be related to free energy differences (*vide supra*), a connection must be made to the microscopic simulations which we can perform.

In the following the system is not treated statically, as in the preceding paragraph, where the notion of temperature did not exist. In this section MD simulations at finite temperature are run, generating trajectories via integration of Newton's equations of motion[83]. Thus we gain access to thermal quantities like free energy or entropy. With thermal quantities there is usually a problem: They cannot be measured as absolute values, but only as differences to a reference, or comparison state. This is true both for experiment and molecular simulation. To obtain reliable ΔG's specialised techniques are clearly needed. Here we focus on *free energy perturbation* (FEP) and *umbrella sampling*. Other approaches not discussed in detail are *thermodynamic integration*[18], *slow-growth*[198], or novel methods based on *Jarzynski's* inequality[104]. The book 'Molecular Modelling' by A Leach[130] serves a good primer for free energy simulation techniques.

Free energy perturbation

FEP is often attributed to Zwanzig who derived the basic equations in his famous 1954 paper[239]. The aim is to transform G, which is not directly accessible in the simulation, into ΔG's, that are quantities *measurable* as ensemble averages[8]. In the condensed phase $\Delta \mathrm{G} \approx \Delta \mathrm{A}$, and we use the Helmholtz free energy A for the sake of a more compact notation in the following derivation. The basic relation for a free energy difference between two

---

[7] Using P-RFO steps for a central, small core to search uphill, and BFGS minimisation for the rest of the system.

[8] The quasiergodic hypothesis[84] asserts that time averages may be equated to ensemble averages. Time averages are directly accessible from trajectories.

systems in terms of the underlying partition functions is

$$\begin{aligned}\Delta \mathrm{A}_{\mathrm{A}\to\mathrm{B}} &= \mathrm{A}_{\mathrm{B}} - \mathrm{A}_{\mathrm{A}} \\ &= -\tfrac{1}{\beta}\ \ln \tfrac{\mathrm{Q}_{\mathrm{B}}}{\mathrm{Q}_{\mathrm{A}}} \\ &= -\tfrac{1}{\beta}\ \ln \left\{ \frac{\int \mathrm{d}\mathbf{p} \int \mathrm{d}\mathbf{r}\ \exp[-\beta \mathrm{H}_{\mathrm{B}}(\mathbf{p},\mathbf{r})]}{\int \mathrm{d}\mathbf{p} \int \mathrm{d}\mathbf{r}\ \exp[-\beta \mathrm{H}_{\mathrm{A}}(\mathbf{p},\mathbf{r})]} \right\}\ .\end{aligned} \tag{2.10}$$

Here, Q_{A} and Q_{B} are the canonical partition functions of systems A and B, and β is defined as $\beta = \frac{1}{\mathrm{kT}}$. The partition functions are expressed as integrals over phase space in the second step. H_{A} and H_{B} are Hamiltonians of both systems. The trick is now to formulate an expression for the difference of both Hamiltonians, and to resolve it cleverly.

$$\begin{aligned} &\Delta \mathrm{H}_{\mathrm{AB}}(\mathbf{p},\mathbf{r}) &&= \mathrm{H}_{\mathrm{B}}(\mathbf{p},\mathbf{r}) - \mathrm{H}_{\mathrm{A}}(\mathbf{p},\mathbf{r}) \\ \Leftrightarrow\ &\mathrm{H}_{\mathrm{B}}(\mathbf{p},\mathbf{r}) &&= \Delta \mathrm{H}_{\mathrm{AB}}(\mathbf{p},\mathbf{r}) + \mathrm{H}_{\mathrm{A}}(\mathbf{p},\mathbf{r})\ .\end{aligned} \tag{2.11}$$

This expression can be substituted into equation 2.10

$$\begin{aligned}\Delta \mathrm{A}_{\mathrm{A}\to\mathrm{B}} &= -\tfrac{1}{\beta}\ \ln \left\{ \frac{\int \mathrm{d}\mathbf{p} \int \mathrm{d}\mathbf{r}\ \exp[-\beta \mathrm{H}_{\mathrm{A}}(\mathbf{p},\mathbf{r}) \exp[-\beta \Delta \mathrm{H}_{\mathrm{AB}}(\mathbf{p},\mathbf{r})]]}{\int \mathrm{d}\mathbf{p} \int \mathrm{d}\mathbf{r}\ \exp[-\beta \mathrm{H}_{\mathrm{A}}(\mathbf{p},\mathbf{r})]} \right\} \\ &= -\tfrac{1}{\beta}\ \ln \left\langle \exp[-\beta \Delta \mathrm{H}_{\mathrm{AB}}(\mathbf{p},\mathbf{r})] \right\rangle_{\mathrm{A}}\ .\end{aligned} \tag{2.12}$$

This final equation now expresses $\Delta \mathrm{A}$ in form of an ensemble average. This can be turned into a practical recipe. The prescription is to run a MD simulation on potential A. For each configuration thus generated one calculates also an energy using potential B. Then one computes the time average of the exponential $\langle \exp[-\beta \Delta \mathrm{H}_{\mathrm{AB}}(\mathbf{p},\mathbf{r}) \rangle_{\mathrm{A}}$ which is $\Delta \mathrm{A}$. In cases where the free energy difference to be calculated differs by more than $\sim\ 2\mathrm{kT}$ convergence of the free energy to be calculated is slow or not reachable. The computation between two states then has to be split into several intermediate steps. This *modus operandi* is utilised in *alchemical transformation* studies. Here $\Delta \mathrm{A}$ is calculated between two molecular species by changing gradually from Hamiltonian A to B in dependence of a coupling parameter λ, which is varied stepwise from 1 to 0

$$\mathrm{H}_{\mathrm{mix,i}} = \lambda_{\mathrm{i}} \cdot \mathrm{H}_{\mathrm{A}} + (1 - \lambda_{\mathrm{i}}) \cdot \mathrm{H}_{\mathrm{B}}\ . \tag{2.13}$$

At step i in the computation $\mathrm{H}_{\mathrm{mix,i}}$ is the 'driver' Hamiltonian, and $\mathrm{H}_{\mathrm{mix,i-1}}$ is the comparison Hamiltonian in the sense of equation 2.12. Alchemical transformation has been used with MM force fields in studies of relative binding free energies of ligands in the active sites of proteins, and of differential solvation free energies in aqueous or organic solvents[123][122][67]. In a similar spirit the $\Delta \mathrm{A}$ value between diastereomeric TIs of subtilisin was computed[41]. Note that in alchemical transformation the RC is λ, but since the reaction path is artificial intermediate steps should not be interpreted as physical entities, e.g transition states. They are just a means to bridge the gap between both molecular species to accelerate convergence of $\Delta\mathrm{A}$.

FEP may also be used to calculate free energies along a physical reaction path, for example in the EVB method, where a chemical reaction takes place for $\mathrm{A} \rightarrow \mathrm{B}$ (A reactants, B products), and a TS is passed. Zhang, Liu and Yang made FEP on reaction paths viable with QM/MM potentials. They assumed the largest contribution to $\Delta\mathrm{A}$ to arise from fluctuations in the surrounding of the active site, and not the reacting fragment (the QM part) itself. A chemical reaction is thus forced to happen by adiabatic mapping, and the structures from each mapping point are saved. During FEP simulation the QM fragment is always fixed[9], the perturbations are done between successive sets of frozen coordinates. This yields the approximate $\Delta\mathrm{A}^{\mathrm{QM/MM}}$, where the QM-QM contribution to $\Delta\mathrm{A}$ is approximated by $\Delta\mathrm{E}^{\mathrm{QM}}$:

$$\Delta\mathrm{A}_{\mathrm{A}\rightarrow\mathrm{B}} = \Delta\mathrm{E}^{\mathrm{QM}}_{\mathrm{A}\rightarrow\mathrm{B}} - \frac{1}{\beta}\ \ln\left\langle \exp[-\beta(\mathrm{E}^{\mathrm{QM/MM}}_{\mathrm{B}} - \mathrm{E}^{\mathrm{QM/MM}}_{\mathrm{A}})]\right\rangle^{\mathrm{MM}}_{\mathrm{A}} . \tag{2.14}$$

Umbrella sampling and WHAM

An important contribution to free energy simulation was made by Torrie and Valleau in 1977 with the introduction of umbrella sampling[207], which they used to compute free energy differences between two different model liquids. One of the early users of this method for biomolecular systems were Karplus and coworkers[150], who calculated the *potential of mean force* (PMF)[10] for the rotation of a tyrosyl ring on the surface of a protein. The relation of the PMF to the free energy of activation, and thus transition state theory was recently analysed by Schenter, Garrett and Truhlar, who showed that the PMF can be converted into a free energy of activation by multiplication with a correction factor[179]. Since we study here differential PMFs this correction term is assumed to cancel out. For this reason and a better appreciation of the presented material the term free energy is preferred over PMF in the later chapters.

Umbrella sampling conceptually works with the distribution function $\rho[\xi(\mathbf{r_1}, \mathbf{r_2}, ..., \mathbf{r_n})]$ of some RC $\xi(\mathbf{r_1}, \mathbf{r_2}, ..., \mathbf{r_n})$. When free energy barriers are present along the RC, normal sampling with unbiased MD will not generate a useful distribution function. The simulation then spends most of its time in the regions of low free energy of the RC, rarely visiting the barrier regions where transition states are located. To this end a biasing potential $\mathrm{W}(\ \xi(\mathbf{r_1}, \mathbf{r_2}, ..., \mathbf{r_n})\ )]$

---

[9] This means the ensemble average $< ... >$ is for the MM part.

[10] The PMF was introduced by Kirkwood for theoretical studies of monoatomic liquids, to designate the reversible work for approach of two tagged particles[117][34].

is introduced, which forces the MD to visit the barrier regions and thus produce an even distribution. Such potentials are often of the harmonic type presented above in equations 2.6-2.8, but a much smaller force constant is used than in geometry optimisation. The distribution corresponding to the original, unbiased system ρ can then be retrieved from the biased one ρ^* as

$$\rho[\xi(\mathbf{r_1},\mathbf{r_2},...,\mathbf{r_n})] = \frac{\rho^*[\xi(\mathbf{r_1},\mathbf{r_2},...,\mathbf{r_n})] \cdot \exp[\ +\beta \mathrm{W}(\ \xi(\mathbf{r_1},\mathbf{r_2},...,\mathbf{r_n})\ )]}{\langle \exp[\ +\beta \mathrm{W}(\ \xi(\mathbf{r_1},\mathbf{r_2},...,\mathbf{r_n})\ )]\rangle}\ . \tag{2.15}$$

The connection to the PMF is then made by noting that

$$\mathrm{PMF}(\xi(\mathbf{r_1},\mathbf{r_2},...,\mathbf{r_n})\ ) = -\frac{1}{\beta}\ln \rho[\xi(\mathbf{r_1},\mathbf{r_2},...,\mathbf{r_n})]\ . \tag{2.16}$$

For most practical problems, one biased simulation is not enough to cover the entire RC. Rather, one performs several simulations with varying potentials $\mathrm{W}(\ \xi(\mathbf{r_1},\mathbf{r_2},...,\mathbf{r_n})\ )$, and combines the unbiased data. The PMF in each such *window* is then given by

$$\mathrm{PMF_i}(\xi(\mathbf{r_1},\mathbf{r_2},...,\mathbf{r_n})\ ) = -\frac{1}{\beta}\ln(\rho^*_\mathrm{i}[\xi(\mathbf{r_1},\mathbf{r_2},...,\mathbf{r_n})]) - \mathrm{W_i}(\ \xi(\mathbf{r_1},\mathbf{r_2},...,\mathbf{r_n})\ ) + \mathrm{F_i}, \tag{2.17}$$

where the $\mathrm{F_i}$'s are undetermined free energy constants $\frac{1}{\beta}\ln \langle \exp[\ +\beta \mathrm{W}(\ \xi(\mathbf{r_1},\mathbf{r_2},...,\mathbf{r_n})\ )]\rangle$ arising from the introduction of the bias potential. To obtain a smooth free energy curve the constants $\mathrm{F_i}$ must be chosen such as to make adjacent curve fragments from the simulation windows match. This can be done by hand or least squares fitting[140]. Another elegant approach to unbiasing and combining simulation data is the *weighted histogram analysis method* (WHAM)[126][127] which is based on techniques described in Alan Ferrenberg's dissertation[63]. WHAM was simplified for use in constant temperature simulations, extended to multidimensional free energy surfaces[24][174], and further refined for application in combined FEP and umbrella sampling studies[197]. We consider once more expression 2.15, which can be written in the form

$$\rho[\xi(\mathbf{r_1},\mathbf{r_2},...,\mathbf{r_n})] = \rho^*[\xi(\mathbf{r_1},\mathbf{r_2},...,\mathbf{r_n})] \cdot \exp[\ +\beta(\mathrm{W}(\ \xi(\mathbf{r_1},\mathbf{r_2},...,\mathbf{r_n})\ ) - \mathrm{F}\ )]. \tag{2.18}$$

If the total distribution is spliced together from several slices, it may be recast as a weighted sum of individually unbiased distributions

$$\rho[\xi(\mathbf{r_1},\mathbf{r_2},...,\mathbf{r_n})] = \mathrm{C}\sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{p_i} \cdot \rho[\xi(\mathbf{r_1},\mathbf{r_2},...,\mathbf{r_n})], \tag{2.19}$$

where C is a normalisation constant, $\mathrm{p_i}$ are weighting functions subject to the condition $\sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{p_i} = 1$, and N is the number of individual simulations. The weights are chosen in a way that minimises the error on the total distribution ρ (derivation in [197]), leading to

$$\rho[\xi(\mathbf{r_1}, \mathbf{r_2}, ..., \mathbf{r_n})] = \mathrm{C} \sum_{\mathrm{i}=1}^{\mathrm{N}} \frac{\mathrm{n_i} \cdot \rho_{\mathrm{i}}^{*}[\xi(\mathbf{r_1}, \mathbf{r_2}, ..., \mathbf{r_n})]}{\sum_{\mathrm{j}=1}^{\mathrm{N}} \mathrm{n_j} \exp[-\beta(\mathrm{W_j}(\ \xi(\mathbf{r_1}, \mathbf{r_2}, ..., \mathbf{r_n})\ ) - \mathrm{F_j}\ )]}\ . \tag{2.20}$$

Here, $\mathrm{n_i}$ and $\mathrm{n_j}$ are the number of data points, that were used to estimate the distributions in windows i and j, respectively. The free energy constants $\mathrm{F_i}$ occuring in 2.20 are computed from the optimal distribution function,

$$\exp(-\beta\mathrm{F_i}) = \langle \exp[\ -\beta\mathrm{W_i}(\ \xi(\mathbf{r_1}, \mathbf{r_2}, ..., \mathbf{r_n})\ )] \rangle \tag{2.21}$$

$$= \int \exp[\ -\beta\mathrm{W_i}(\ \xi(\mathbf{r_1}, \mathbf{r_2}, ..., \mathbf{r_n})\ )] \rho[\xi(\mathbf{r_1}, \mathbf{r_2}, ..., \mathbf{r_n})] \mathrm{d}\xi \tag{2.22}$$

Because $\rho[\xi(\mathbf{r_1}, \mathbf{r_2}, ..., \mathbf{r_n})$ depends on the constants $\mathrm{F_i}$, which are not known at the beginning, an iterative procedure is adopted to solve 2.20 and 2.22 self-consistently.

2.6 New Techniques in Studies of Activated Processes

The preceding chapter dealt with free energy calculations of chemical reactions. Here it proved useful to introduce a reaction coordinate to drive the system out of its equilibrium wells to make the study efficient and statistically sound. Sometimes it is not possible to devise a simple RC, or this may introduce artefacts into the simulation, for example by spatial distortion of the molecular model. In such cases a definition of order parameters to measure the reaction extent may still be feasible. Furthermore, order parameters can be rather abstract measures, that are not easily translated into the form of restraints. The *transition path sampling* (TPS) technique developed in the group of Chandler[26][48] depends solely on the ability to define a reactant and a product basin in terms of arbitrary order parameters. Starting from an initial reactive trajectory an ensemble of paths connecting the reactant and product basins is generated, from which a reaction rate constant can be calculated[47][208]. TPS was shown to be useful in the treatment of large systems, for example folding and conformation isomerisation pathways of biomolecules[25][167], study of reaction mechanisms within proteins (using QM/MM methods)[13], and ligand exchange in organometallic complexes[196]. Inspired by TPS, Radhakrishnan and Schlick recently designed a path sampling scheme named BOLAS [166], which allows the computation of the PMF essentially by umbrella sampling without geometric restraints. An application has been reported by the same authors[168].

An approach to *discovering* structural transition and reaction mechanisms is *conformational* or *chemical* flooding[87][146]. This method uses information on the deep local energy wells that trap the system. Unimolecular chemical reactions and structural isomerisations are suited for analysis by chemical/conformational flooding. MD is used to perform a quasiharmonic analysis of the nuclear dynamics[74], the choice of atoms included here is a control parameter that can be used to test for the existence of alternative mechanisms. A potential is then constructed to fill the potential well and drive the system away from its current position in configuration space. The mechanism of ring opening of bicyclopropylidene was studied in this manner[146].

Other methods were developed recently that modify the PES and are also aimed at faster exploration of configuration space, namely *accelerated molecular dynamics*[91][92], which is based on Voter's *hyperdynamics*[214], and *metadynamics*[129][229]. In accelerated MD a *boost* potential is applied to enhance the escape probability from local minima. For a test system (hepta-alanine) with high torsional barriers a boost potential was constructed from the sum of dihedral and 1-4 nonbonded interactions. After reweighting a two-dimensional PMF of backbone torsional angles was obtained that exhibited much better sampling than normal MD away from the local minima [90]. Metadynamics is a restrained MD which acts on a predefined set of collective coordinates, for example selected dihedral or bond angles. The restraints are used to bias the systems out of known regions in the space of collective coordinates, and towards new configurations, by means of a Gaussian repulsive potential. This potential is a sum of Gaussian kernels that is constructed and updated during simulation. Prospective targets for this method are docking studies, where one usually encounters rough energy landscapes with steep potential wells, especially for rotation of ligands or parts thereof, in the binding pocket[29][79].

Chapter 3

Modelling the Tetrahedral Intermediate

The first step towards simulation of the acylation reaction was the generation of a model system consisting of enzyme, substrate, and water. More specifically we used a crystal structure of the enzyme BSLA, which was augmented with the substrate in its TI form. After solvation of the enzyme the system was subjected to an iterative procedure of minimisation, MD, and resolvation. For this purpose (and in this chapter in general) only MM methods were employed. In a final long MD simulation we sampled the model system at equilibrium. The trajectories thus generated were raw material for forthcoming chapters in that configurations were harvested from them as snapshots, which in turn were used as input to QM/MM calculations in chapters 4 and 5. The trajectories were analysed considering aspects like conformational flexibility and stability, and hydrogen bonding patterns important for the catalytic activity of the enzyme. The considerable number of possible binding modes found for the NAPHAC substrate required a thorough structuring of the data. As another result of this conformational complexity multiple models for each enantiomer were considered in later chapters.

3.1 The Crystal Structure

The protein coordinates for our enzyme model are derived from x-ray diffraction data. The first structure of *Bacillus subtilis* lipase was reported in 2001[162], determined at a resolution of 1.5 Å (code 1I6W). Presently six structures are deployed in the protein data bank, communicated by three different research groups. The original structure[162] was pure lipase, without inhibitor attached to the active site serine. This is worth noting, since often serine hydrolases are crystallised in the presence of an inhibitor in order to help understand substrate binding. In addition, this ensures that the enzyme is in its active form, where the hydrogen bond network and the catalytic triad are in their active geometries. Examples exist where the reactive network is created only in the presence of a ligand[115], but in BSLA it exists in the native state[162]. BSLA with a chiral inhibitor, 1,2-O-isopropylydidene-sn-glycerol-phosphonate (IPG-phosphonate), were submitted to the Brookhaven protein data bank in 2004 (codes 1R50 and 1R4Z), but preliminary structures were made available to us in late 2001[161] and used in our model building process. A structure with slightly higher resolution than in the original study of BSLA[162] was solved by a Japanese group[113] (code 1ISP). In addition to higher resolution they proposed subtle conformational variants of the active site serine. In another interesting development directed evolution methods were used to create thermostable variants of BSLA[1], and several mutants were discussed and crystallised (codes 1T2N and 1T4M).

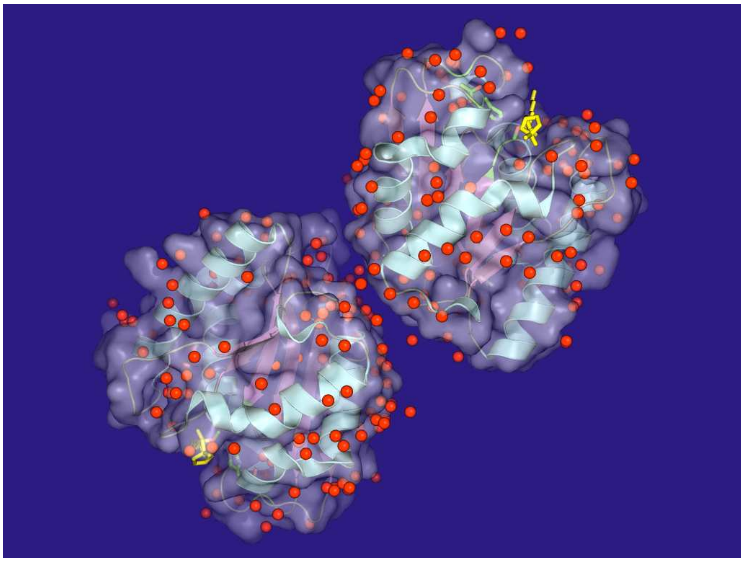


Figure 3.1: The asymmetric unit of the 1R4Z crystal contains two molecules BSLA. The (S)-IPG-phosphonate inhibitor is bound as TI and drawn yellow. Oxygen atoms of the crystal water are coloured red.

The enzyme BSLA has a minimal α/β hydrolase fold[162]. The structure is composed of a central six-stranded parallel β-sheet which is flanked by five α-helices. The catalytic triad[64] is formed by residues Ser77, Asp133 and His156, the oxyanion hole by the backbone amide groups of residues Ile12 and Met78. At 19 kDa (181 residues) BSLA is a very small lipase, comparable in size to a lipase of *Bacillus pumilis*. It has a globular structure of 35 Å x 36 Å x 42 Å. We suppose BSLA to be an enzyme of lesser dynamics, on account of its globularity and its central β-sheets, which are sensitive to water contact[213] that would likely cause disintegration of the enzyme. A hydrogen bond between His156 and Ser77 is assigned by the crystallographers[113]. However, it should be weak in the free enzyme, since both published structures[162] [113] 1I6W and 1ISP reveal an angle between serine oxygen and the imidazole which is significantly nonlinear. In the inhibited enzymes 1R4Z and 1R50 a hydrogen bond between the presumably doubly protonated His156 and the TI is nonexistent. In 1I6W and 1ISP a water molecule occupies the oxyanion hole. In 1R4Z and 1R50 this position is occupied by an oxygen atom of the TI. A different water molecule, which is assumed to form a hydrogen bond to the backbone of Gly11 (see figure 3.2) is resolved in all crystal structures. Its structural stability can be further enhanced by hydrogen bonding to His76. The atom assignment at the His76 residue in the crystal structures 1R4Z and 1R50 does not seem sensible, as it obstructs this hydrogen bonding. Since at the experimental resolution nitrogen and carbon cannot be differentiated, one may apply a histidine *flip*, where atoms in the imidazole ring are mirrored in a plane that runs through atom CG and whose normal is the CE1-NE2 bond vector[1]. Doing so the hydrogen bond between the crystal water molecule and His76 is enabled, as well as another hydrogen bond of His76 to the backbone of His156. The importance of residue His76 was assessed independently by experiment and theory[72], His76 is found to be a relay position, where enantioselectivity changes drastically on mutation, or even reverses. The origin of this effect is not well understood, but it is likely that subtle changes in the local hydrogen bond network are an essential ingredient. An unoffical preliminary structure, which corresponds to 1R4Z, was

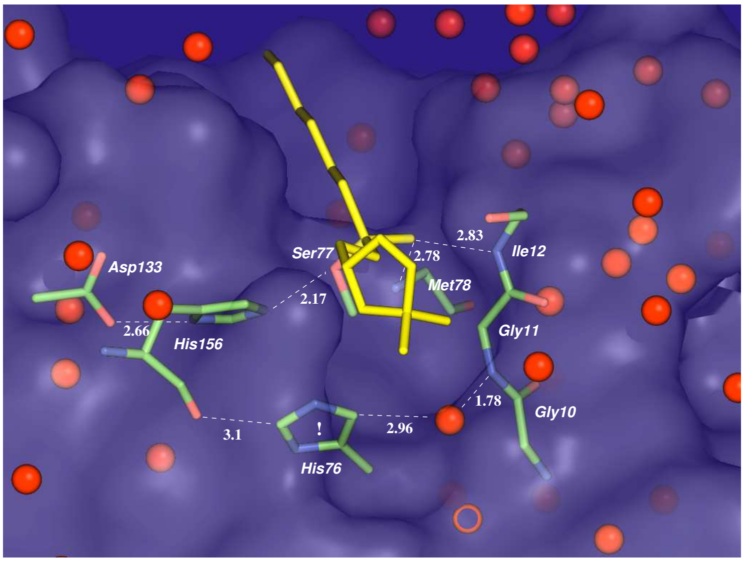


Figure 3.2: Local view of the active site of chain a in the asymmetric unit of the 1R4Z crystal. The distances (in Å) are drawn for positions where hydrogen bonds exist, or are assumed to exist. In the case of His76, which is marked by an exclamation mark, it seems justified to apply a flip, in order to enable hydrogen bonding with the backbone of His156 and a crystal water.

---

[1] Note that the structure used in the validation of the QM/MM method in the next chapter, ANRU453, adopts the His76 conformation of the crystal structure. All other production geometries discussed in this work contain the flipped histidine.

used in building the model. This structure contains the S enantiomer[2] of the IPGP inhibitor. 1R4Z has two enzyme molecules in the asymmetric unit, just like 1I6W. Subunit B has crystal contacts at the active site and is structurally slightly different from subunit A. The latter faces bulk solvent and is thought to assume a more solution-like structure. The same is true for our preliminary version of the crystal structure[3]. Therefore this subunit is the preferred basis for all subsequent model building. The unoffical 1R4Z release is almost identical to the official release, but Arg33 and Asn181 were assigned different side chain conformations. The large distance to the active site renders these slight changes irrelevant.
The raw enzyme structure was checked by WHATIF[215][4], to generate proposals for protonation states of amino acid side chains, and identify flipped positions in asparagine, histidine or glutamine residues. Based on the output of WHATIF we decided on the following protonation scheme for histidine: His3 (HSE), His10 (HSD), His152 (HSE), His156 (HSP), His152 (HSE), His76 (HSE). All other polar and potentially charged amino acids were used in their ionised form. Flips were applied to His152 and His76, as well as Asn82 and Asn138.

Some enzymes have large hydrophobic internal cavities in their X-ray structures which may contain water molecules that are not resolved because of their mobility. Cytochrome P450[188] is an example in this respect[5]. In BSLA there are no such buried cavities that may contain unresolved water molecules[162].

3.2 Binding Modes of the Substrate

An important component of the structural basis of our model, the enzyme itself, was discussed above. A more complicated problem is posed by the unknown conformations of the NAPHAC enantiomers on the surface of the enzyme. Successful substrate binding culminates in the TI, which is viewed as a portrait of the prior Michaelis complex. In this line of thought we must search through conformation space of the TI as realisations of different binding modes of the substrate. For IPGP such geometries are known from crystals structures 1R50 and 1R4Z (figure 3.2). Our experimental partners[6] have repeatedly tried to synthesise the inhibitor of NAPHAC and to grow crystals of BSLA in the presence of this inhibitor, but these attempts failed, probably due to solubility problems. Molecular modeling

---

[2] The inhibitor is erroneously classified R in the header of 1R4Z. In 1R50 it is R instead of S.

[3] Note that the chain assignment is reversed in the unofficial structure, e.g. A=B, B=A.

[4] An online service was used at *biotech.ebi.ac.uk*, where PDB structures can be uploaded and processed by WHATIF[215].

[5] Solutions to the problem of occupation number of cavities have recently been discovered[39][102].

[6] Our partners are Prof. Reetz, Mülheim, for organic synthesis, and Prof. Quax, Groningen, for crystallography.

is therefore required to fill the gap and provide a structural basis for further work. One of our main assumptions is that the binding mode of the TI in the active site affects the reaction barriers significantly. A complete picture of possible binding modes is needed for a fair comparison of rates at the most efficient catalytic geometries of both enantiomers. Thus we do not simply face a docking problem, since the reaction barrier has to be accounted for[7]. For any distinct binding mode discovered at the modelling stage, extensive barrier calculations were performed later (Chapters 4 and 5). We also test some nonproductive binding modes for a more complete spectrum of reaction barriers.

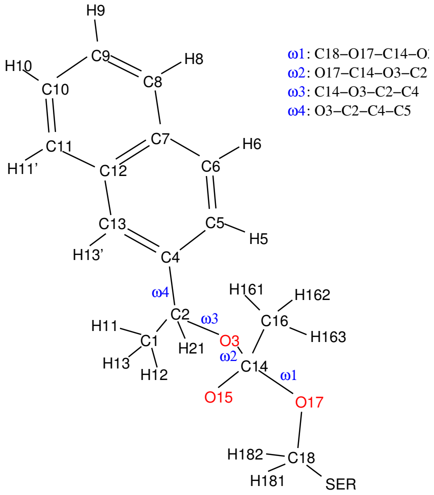


Figure 3.3: Naming scheme for atoms of the TI of the NAPHAC substrate, and for the dihedral angles. O15 is the oxyanion, O3 and O17 are the sites of cleavage of the TI. Hydrogen transfer to O3 leads to the acylenzyme, and hydrogen transfer to O17 to the Michaelis complex. Greek letters indicate the torsion about the respective bonds. ω_3 and ω_4 offer the largest potential for conformational diversity, ω_1 and ω_2 are confined to smaller intervals due to restrictions on productive binding.

The NAPHAC substrate contains a sterically demanding naphthyl ring which may limit the conformational degrees of freedom for binding of this substrate. However, there are several torsional degrees of freedom in the TI that complicate the situation (figure 3.3). In the next section technical details of the model building procedure are described.

3.3 Model Setup and Dynamics

Building TIs was performed in MOLOC[77], which offers convenient and comparatively fast model building. One of the components that make it fast is that it has its own force field MAB[77], which supports automatic assignment of parameters and charges[76] for almost

---

[7] See analysis of the kinetics in chapter 1.

any kind of molecule. This puts the user in the position to perform geometry optimisations and MD calculations. The former are used to remove steric contacts from the manually built structures, but also to probe thoroughly other modes of insertion.

First, we cleaved the IPGP from the structure and replaced it by NAPHAC. A range of conformationally distinct starting positions was created by rotation about torsion angles. This was in most cases followed by structural relaxation of the TI with the MAB force field, keeping the crystal coordinates at their initial positions. In so doing the TI is made to fit into the enzyme, and the modeller gets a first opportunity to judge whether the proposed binding mode is stable.

The whole structure is then saved and tranferred to the CHARMM[31] program. For any molecule imported into the CHARMM program, a residue topology and force field parameters are needed. As far as proteins alone are simulated this is not a problem, since the Charmm22 force field was created for proteins and amino acids. The component amino acids and nucleobases are predefined in the *residue topology file*, but for the majority of ligands or cofactors such a definition does not exist and must be created by the user. We underwent this process, and created a residue definition and parameters. This procedure and its results are discussed in chapter 6, in this chapter and the following two they are assumed to exist. This puts us in the position to perform geometry optimisations and MD calculations with our enzyme model. The parameters have been been used in an MM study of lipase enantioselectivity[22] and perform well.

We now describe the CHARMM-based setup procedure. First, hydrogen atoms, which are not resolved in the x-ray structure, are added. Then the TI is energy minimised (100 steps ABNR) and relaxed in a dynamics run (1 ps at 600K). Thereupon a hydration shell is added to the enzyme, which consists of a spherical volume of pre-equilibrated TIP3[107] water molecules of radius 25Å. A typical cutoff value of 2.8Å around heavy atoms for the insertion of water molecules is applied. The solvent does not cover the entire enzyme, only that part which contains the active site, and some of the surrounding. Some part of the model thus faces the vacuum and must be fixed in the modelling process to prevent formation of artificial geometries. The set of fixed atoms of the protein consists of all atoms beyond a spherical cutoff of 12.5 Å, counted from the origin, which was set to the initial position of the hydrogen atom HE2 of His156. Water molecules were always restrained by a quartic potential to keep the shape of the solvent spherical and to prevent evaporation of water in dynamics simulations. The 'GEO SPHERE QUARTIC' command in the MMFP module of the CHARMM program was used here. Thereafter we minimise (500 steps SD and 500 steps ABNR) the entire system, and run a heating dynamics simulation (50000 steps) starting at

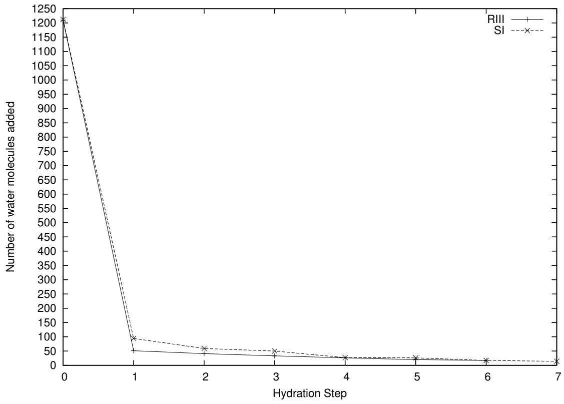


Figure 3.4: Number of water molecules added to the model. Clearly most water enters in the first hydration step. The next rehydration still adds on the order of 100 molecules, which is a non-negligible number. Thereafter less water is introduced, in each step. After approximately seven cycles the procedure was terminated with less than 20 molecules added in the last step.

50 K and ending at 300 K. The Verlet algorithm is used with a timestep of 1 fs, and all bonds to hydrogen are constrained by SHAKE. We then rehydrate the system and repeat the minimisation and dynamics steps seven times (see figure 3.4). Rehydration is necessary since volume contraction of the solvent results from enhanced interaction and relaxation of water on the enzyme surface. Water does not enter the protein interior on resolvation. An equilibrium dynamics of 1.1 ns is finally run and completes the preparation of the model systems. In most of these trajectories no serious structural reorganisation was noticed over the entire 1.1 ns. The temperature and total energies are equilibrated typically at $\approx$ 150 ps after the initial heating period. Nonetheless we regarded the first 500 ps as equilibration time by default. Any analysis starts at 500 ps, which is also the point in time where snapshots are extracted in later chapters.

3.4 Characterisation of Binding Modes

In light of the diversity of binding modes it is helpful to structure the results, and impart a spatial impression to the reader. The topology of the active site is visualised in figure 3.5, which shows an empty binding cleft. It reveals a tangible analogy, a clawed hand, which can be considered to keep hold of the TI during the catalytic cycle. This is a front view of the active site, which is also used in the graphical presentation of the actual binding modes in figures 3.7, 3.8 and 3.9. In these illustrations the front view corresponds to the left hand picture, the top view to the central image, and a view from the back to the right hand representation. Binding modes, which equate to individual trajectories, are indicated in yellow font. The naming scheme accounts for the stereoconfiguration as R and S, and roman numbers indicate the number of the model in the chronological order in which they were created. RIII, for example, is henceforth used to indicate the third model built for (R)-NAPHAC. In figure 3.6 lettered planes symbolise the orientation of the naphthyl ring as a representative of the stable conformations which the TI can assume in BSLA. The top view shows a broken line to further subclassify the binding modes. Geometrical membership of the binding modes will be specified by the capital letter of the respective plane in calligraphic font, and a lower case letter indicating whether the orientation is a(nterior) or p(osterior).

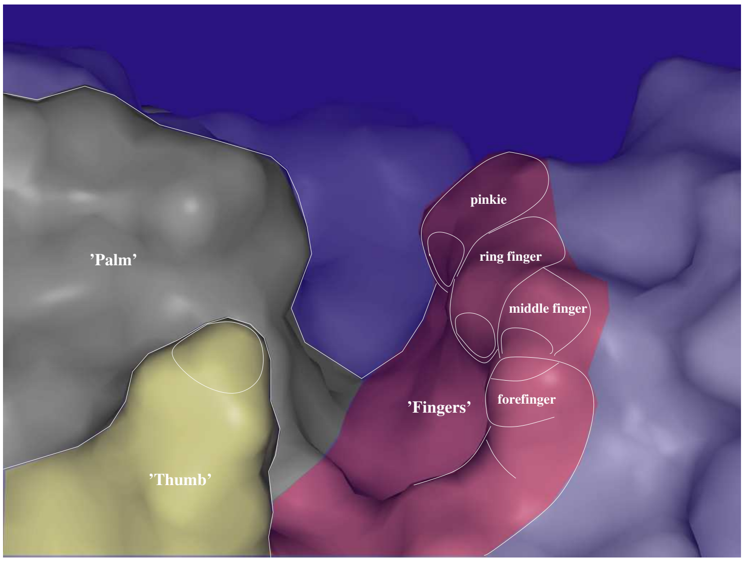


Figure 3.5: The active site of BSLA resembles a clawed hand. The 'palm' domain hosts His156 and Asp133, which are members of the catalytic triad. Ser77 is positioned at the boundary of 'fingers' and 'palm'. The 'fingers' cover the oxyanion hole, which is located on the right of the binding cleft. The 'thumb' is formed by the side chain of Ile157.

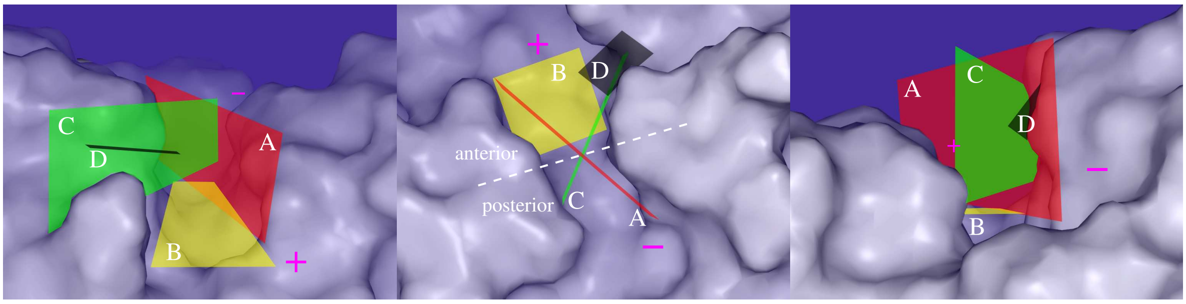


Figure 3.6: Front (left), top (center), and back (right) view of the active site of BSLA. The symbols in magenta serve as orientational aid. The broken white line in the top view indicates that we conceptually separate an *anterior* and a *posterior* region. The substrate is considered to enter via the anterior region, which is wider than the posterior region. The coloured planes denote preferential orientations of the naphthyl ring of the NAPHAC substrate. It is coplanar to these planes.

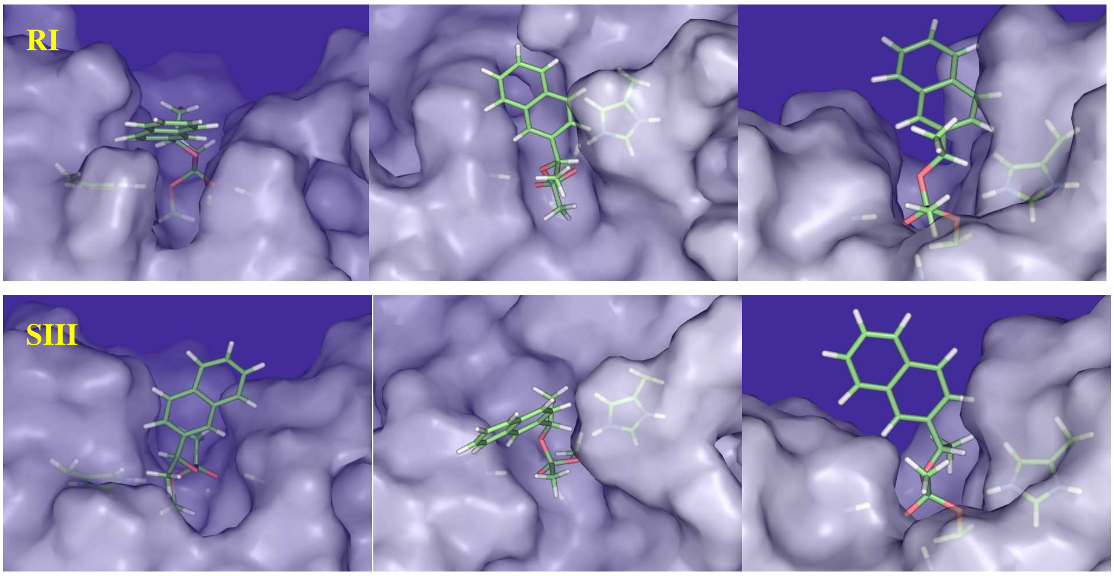


Figure 3.7: Nonproductive conformations of the TI.

RII	RIII	RIV	RV	RVI	RVII	SI	SII	SIV	SV	SVI
$\mathcal{AB}a$	$\mathcal{A}p$	$\mathcal{C}a$	$\mathcal{C}a$	$\mathcal{C}a$	$\mathcal{A}p$	$\mathcal{B}a$	$\mathcal{C}a$	$\mathcal{B}a$	$\mathcal{D}a$	$\mathcal{A}p$

Table 3.1: Geometrical classification of binding modes.

For RIII a look at figures 3.8 and 3.6 confirms the classification as $\mathcal{A}p$. Binding modes from all other trajectories are given in table 3.1. RI and SIII (figure 3.7) are excluded, because they are nonproductive binding modes, by virtue of the orientation of the tetrahedral moiety (*vide infra*). They were created as representatives of the putatively highest possible reaction barriers.

Table 3.1 shows only the most dominant orientations of the TI during simulations. In some cases transitions are made from one binding mode to another. This is the case in RV, where after 1.05 ns a $\mathcal{AB}a$ geometry is adopted. SV assumes its final position, which is displayed in 3.9 after 500 ps, and SVI makes a transit to $\mathcal{C}a$ after 1 ns of simulation time.

Distribution of Dihedral Angles

The large variety of conformational orientations of the TI is ascribed to four torsion angles, ω_1-ω_4 (figure 3.3). These torsions were positioned manually during model building, and then left free to evolve in subsequent MD simulations. Their assumed influence on reaction barrier heights makes an analysis of their characteristic behaviour mandatory. Specifically one would like to know whether stable states in dihedral space exist, and if so, how many there are. This is necessary for the selection of representative snapshot configurations from the equilibrium trajectories which are used in chapters 4 and 5 to assess reaction barriers.

Histograms are a natural and unbiased representation of the torsion angle distribution.

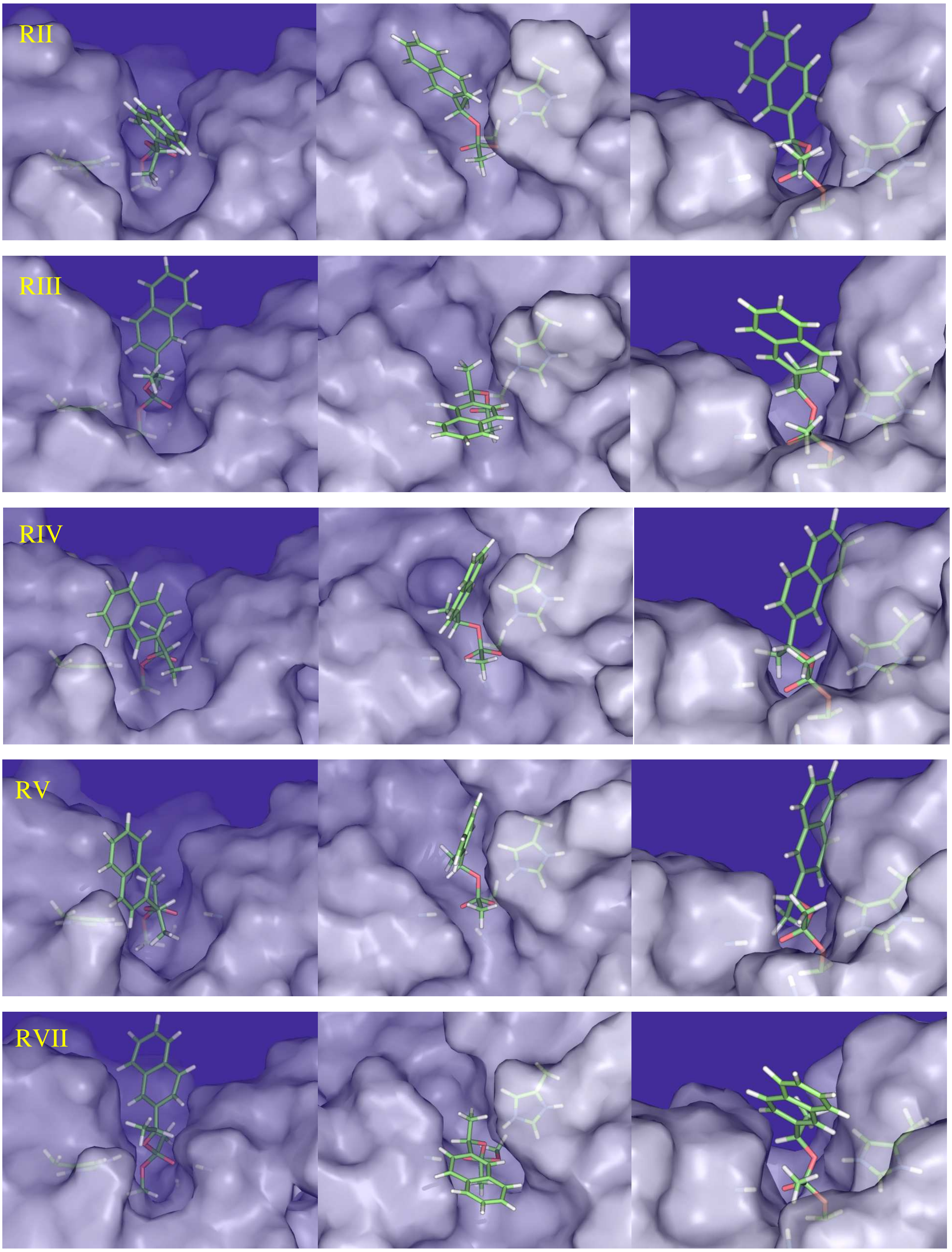


Figure 3.8: Orientations of the TI of **R-NAPHAC** as obtained from equilibrium MD with a classical force field.

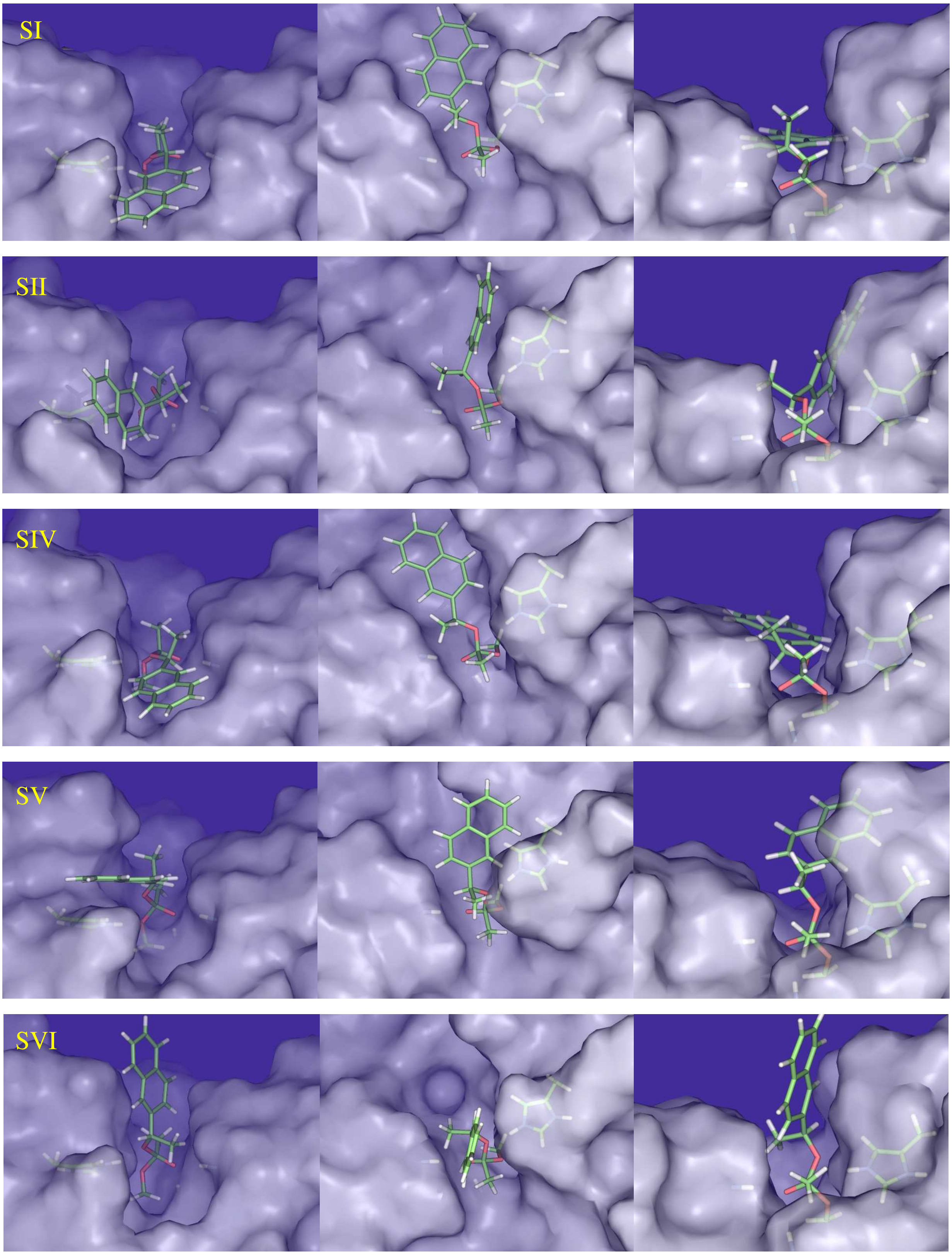


Figure 3.9: Orientations of the TI of **S-NAPHAC** as obtained from equilibrium MD with a classical force field.

These histograms along with the MM PES from the parametrisation of the TI are shown in figures 3.10 and 3.11. For ω_3 and ω_4 the PES was mirrored, since C2 is the enantiodiscriminating atom of the TI. A quick overview is given in table 3.2, and can help to identify similar configurations. RV and RVI are almost identical by visual inspection of the trajectory, the values shown here confirm this impression. RIII and RVII differ by an inversion of the orientation of the naphthyl ring. This is realised by a rotation of ω_4 about $180°$, a condition which is met here almost perfectly. The same symmetry relationship holds for RIV and RV. Torsions ω_1 and ω_3 assume identical values, ω_2 is very close as well, which is best appreciated in figure 3.10. Another pair of geometries with an analogous relationship is SI and SIV, where ω_4 differs by $180°$. The values of torsions ω_1 and ω_3 are quite similar, but those of ω_2 seem to be significantly different according to table 3.2. An inspection of figure 3.10 clarifies this impression. Torsion ω_2 populates the region around the discontinuity at $(180°; -180°)$, which distorts the simple arithmetic averages in table 3.2.

	RI	RII	RIII	RIV	RV	RVI	RVII	SI	SII	SIII	SIV	SV	SVI
ω_1	-100	-86	-93	-69	-70	-69	-93	-67	-91	-64	-65	-99	-99
ω_2	-72	109	155	150	151	154	142	107	76	-65	146	138	-19
ω_3	149	152	86	126	123	126	86	-139	-155	-100	-146	-151	68
ω_4	-141	24	19	-99	73	73	-153	147	145	127	-24	128	34

Table 3.2: Average values of dihedral angles. The torsions are specified in figure 3.3, page 39. This table serves only as a quick overview. The histograms in figures 3.10 and 3.11 give and unsophisticated impression of the distributions.

The comparison with the potential curves for these torsions, which are calculated in vacuum[8], and the distributions as they occur in the actual system gives an impression of the conformational strain of the TI. Torsions ω_1 and ω_2 have a narrow range of accessible values for catalysis to occur effectively, ω_3 and ω_4 can be adjusted more freely, but are subject to steric requirements.

The values of ω_1 are restricted to the vicinity of the local minimum which occurs at $-65°$ (figure 3.10). Two groups are identified, one which is in direct accordance with the local minimum, the other being shifted towards the trans configuration by $30°$. The local minimum at $75°$, and the global minimum at $180°$ are not populated in any trajectory. This is expected by considerations of steric requirements alone. At $75°$ the naphthyl ring would collide with the 'fingers' of the active site, at $180°$ the methyl group around carbon C16 cannot be accomodated. Torsion ω_2 primarily assumes a trans configuration. The minimum at $-75°$ is occupied solely by RI and SIII, both of which were prepared to study nonproductive binding (figure 3.7). This alleged nonproductivity results from O3 being in an awkward po-

---

[8] Performed as a part of the parameterisation of the TI, see chapter 6.

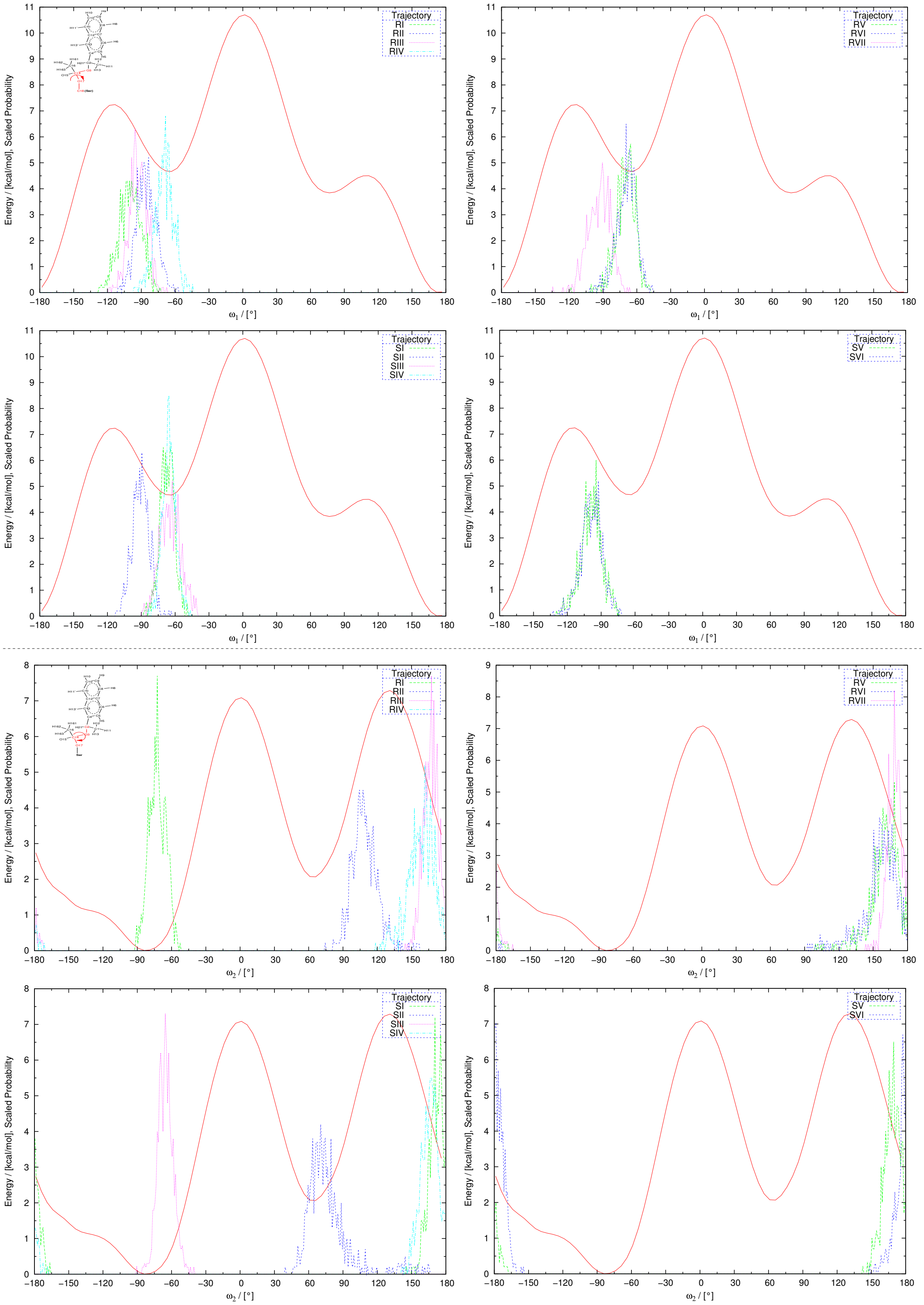


Figure 3.10: Torsions ω_1 (top) and ω_2 (bottom) of the TI.

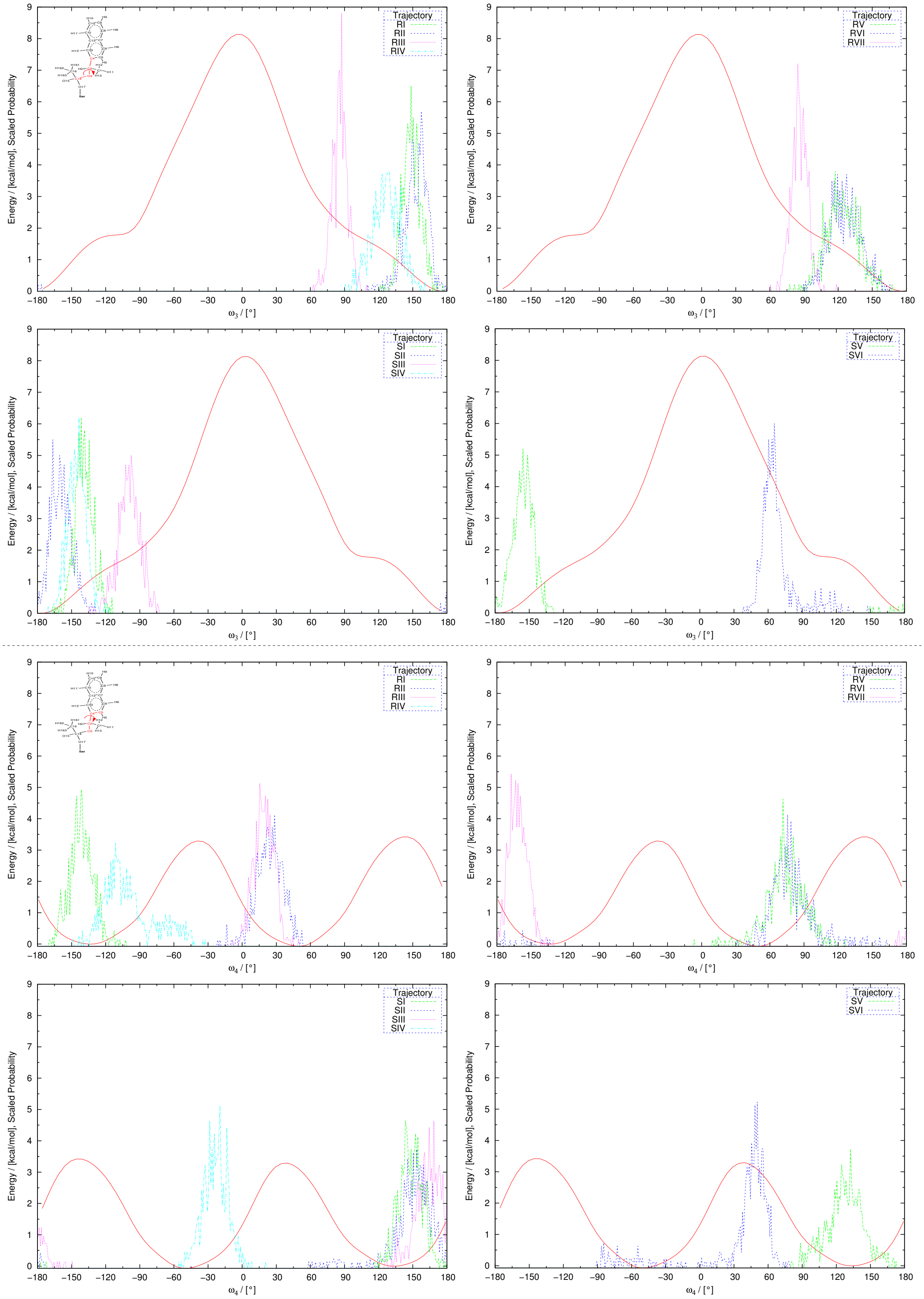


Figure 3.11: Torsions ω_3 (top) and ω_4 (bottom) of the TI. The torsion potential curves are mirrored for the S enantiomers.

sition for proton transfer. The acylation reaction is thus thought to be arrested at the TI. To a lesser degree such conduct is expected from RII and SII, which are centered about $100°$ and $75°$, respectively. The accessible values of ω_3 extend from the trans configurations down to $60°$ with R, and up to $-75°$ with the S enantiomer. For positive values the methyl group of carbon C1 points downwards, into the enzyme, and for negative values into the bulk solvent. This is best appreciated in figures 3.8 and 3.9. The potential curves in figure 3.11 are mirrored because of ω_3 passing through the asymmetric atom C2. SVI is the only binding mode of the S enantiomer which assumes a positive value of ω_3. On the contrary we could not prepare a single configuration of R with ω_3 in the negative range. Such geometries were found to be instable, and converted back to positive values of ω_3 readily. Torsion ω_4 has two minima and is almost C_2 symmetric about its axis of rotation. The symmetry is not perfect, nevertheless we shall assess the effect on reaction barriers in chapter 5 on RIII and RVII.

The TI is subject to remarkable conformational strain on the surface of the enzyme. This strain results primarily from torsions ω_1 and ω_2, the values of which are confined to a limited range of values for productive geometries; the local minima at $180°$ and $-75°$ are unaccessible (ω_1), or unproductive (ω_2). Torsions ω_3 and ω_4 can assume a larger range of values without compromises in functionality. Most binding modes are close to the minima of the PES, or accept some minor strain ($\approx$2 kcal in ω_3). Population of the maxima of the PES leads to unstable geometries, such as SVI, which releases the strain arising from ω_3 and ω_4, and makes a transition to a more stable conformation. This is seen in the diagrams in figure 3.11. The tail in these histograms is a consequence of this behaviour.

3.5 Active Site Fluctuations

In later chapters (4 and 5) starting geometries are selected from the MD trajectories, and used in minimum energy path calculations or free energy calculations. Large scale motions occuring within and around the active site presumably affect the barriers of the acylation reaction calculated from these geometries. Therefore it is advisable to identify major sources of structural diversity, and to account for a variety of stable conformational states, if necessary. Important positional fluctuations can be identified by computing the root-mean-square-deviation (RMSD) of all protein atoms relative to the initial structure. Residues that contain atoms which moved by more than 1.5 Å are highlighted (table 3.3). For a better comprehension of the spatial relationship to the active site and the TI these residues are highlighted in figure 3.12. The following discussion proceeds in the order of table 3.3.

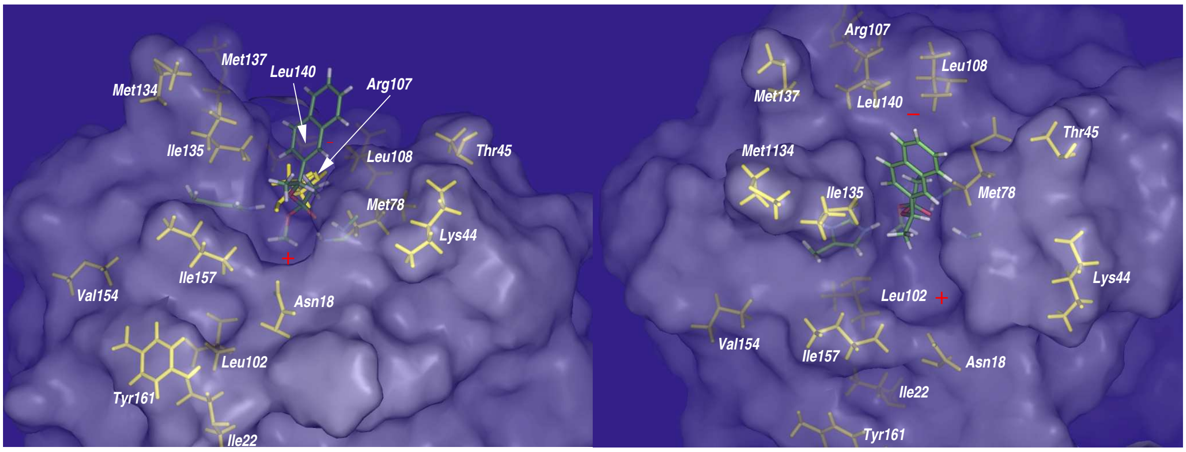


Figure 3.12: Front and top view of BSLA with residues causing notable RMSD highlighted. Orientational aids are coloured red.

	Location	RII	RIII	RIV	RV	RVII	SI	SII	SIV	SV	SVI
ASN18	surface	X	X	X	X	X	X	X		X	X
ILE22	interior						X				
LYS44	surface			X							
THR45	surface	X	X			X	X		X	X	X
TI	surface				X	X				X	X
MET78	surface	X	X			X			X		X
LEU102	interior			X				X			
ARG107	surface	X	X	X		X			X		X
LEU108	surface				X					X	
MET134	surface	X	X	X	X	X	X	X	X	X	X
ILE135	surface									X	
MET137	surface	X	X	X	X	X	X	X	X	X	X
LEU140	surface	X	X				X	X			X
VAL154	surface					X					
ILE157	surface	X	X	X	X	X		X			
TYR161	surface			X					X	X	X

Table 3.3: RMSD in the trajectories of BSLA with the TI of NAPHAC. Marked are residues which contain at least one atom which moves more than 1.5 Å (black X), or more than 2 Å (red X). The location indicates the position of the amino acid in the enzyme.

Asn18 exhibits notable mobility in all simulations, except that of SIV. The motion can be captured by monitoring the torsion[9] $\chi_{\mathrm{i}}^{3,2}$, which occasionally flips. Position 18 is known to be important for the enantioselectivity of BSLA towards meso-1,4-diacetoxycyclopentene (DIAC). Insertion of serine at this position inverts enantioselectivity[60]. **Ile22** is a rather settled residue, but experiences a transient shift of 90° about $\chi_{\mathrm{i}}^{3,1}$ in trajectory SI. RIV is special due to **Lys44** showing significant activity rooted in torsions χ_{i}^{2} and χ_{i}^{4}, and because of Tyr161, where a rotation about χ_{i}^{6} leads to a change of the hydrogen bond network. Lys44 is located on the surface of BSLA and not engaged in salt bridges with the enzyme. It car-

---

[9] The IUPAC definition of symbols and nomenclature for the description of amino acid and peptide conformations is documented in [151].

ries a positive charge which has an effect on the reaction barriers, as will be discussed in chapter 4. **Thr45** has a tendency to revolve on χ_1 in most simulations, and as direct neighbour of Lys44 it is a surface residue. The structural impact is therefore considered negligible. Four trajectories display a noteworthy RMSD of the **TI**. SV represented a $\mathcal{A}p$ structure, was found unstable, and collapsed to $\mathcal{D}a$ after 500 ps of simulation. SVI was prepared as $\mathcal{A}p$, but transformed to $\mathcal{C}a$ after one nanosecond of simulation time. The stability of posterior orientations SV and SVI was highly dependent on the initial position set at the modelling stage; these orientations do not exhibit long-term stability and seem to be disfavoured binding modes of the S enantiomer. A conformational transition is also noted for RV, which starts as $\mathcal{C}a$ and makes a transition to $\mathcal{AB}a$, which closely resembles RII, after 1.05 ns. The table indicates some mobility for the TI in RVII, but, the system remains $\mathcal{A}p$ throughout the entire MD simulation (1.1 ns). **Met78** has some space for flips of torsion χ_i^3, which is realised in five trajectories. The backbone is unaffected, and its function as part of the oxyanion hole is conserved, as inspection of table 3.5 reveals. **Ile102** exhibits a transient shift of χ_i^1 in both RIV and SII, no reorganisation of hydrogen bonds in the interior of the protein is involved. **Arg107** is located on the surface of BSLA, and is involved in a salt bridge to Asp144. The binding to the tail group is quite firm, and consequently the motion of Arg107 is confined to a concerted rotation of χ_i^1 and χ_i^2, which leaves the charged group in its initial position, but leads to transient twists of CG. The side chain of **Leu108** has freedom to rotate on the surface of the protein. Significant motion about χ_i^1 and χ_i^2 is registered in two simulations. **Met134** and **Met137** protrude from the enzyme surface, their motion is largely chaotic and has a large amplitude in all simulations. They are not in steric contact with the TI, and do not carry a charge to act on the barriers. Still the role of Met134 and Met137 may be significant in binding of triglycerolesters, where one of the fatty acid chains of the substrate supposedly threads through them[162]. Geometry $\mathcal{D}a$ of SV requires the entire **Ile135** to move outward to accomodate the naphthyl ring of the TI. The backbone of Met134 is also displaced, with grave consequences: water gains access to Asp133 and His156 of the catalytic triad. The naphthyl ring pictorially pries open a forbidden gate, and obliterates the activity of the enzyme. Geometries $\mathcal{D}a$ are consequently deemed nonsignificant, as they open a predetermined breaking point of the enzyme structure. This demonstrates the need for critical evaluation of modelled enzyme substrate complexes in terms of structural integrity. **Leu140** is spatially close to Leu108, and also located at the surface. The side chain is rotatable without much hindrance. The same is true for **Val154**. More significance is attributed to **Ile157**, which is potentially in steric contact with the naphthyl moiety of the TI. Ile157 is the 'thumb' in the surface representation in figure 3.5, its degrees of freedom are χ_i^1

and $\chi_{\mathrm{i}}^{2,1}$. Configurations with the substrate being in touch with Ile157 are SI, SII, SIV, and SVI. RIV is the sole representative of the R enantiomers. SI and SIV have the most intense contact to Ile157, here the substrate can be imagined as being grabbed by the binding pocket like a medium sized pebble by a human hand. Due to its distinct contacts with the TI, Ile157 is expected to be a key residue for the enantioselectivity of BSLA. This importance may not be recognised experimentally by site directed mutagenesis, since position 157 is so central to the active site that replacements may be lethal with regard to the activity of the enzyme. No steric contact or hydrogen bonding ability to active site residues is available for **Tyr161**, which is located on the surface in the anterior region of BSLA. Its degrees of freedom are χ_{i}^{2} and χ_{i}^{6}. This residue seems less relevant for enantioselectivity in catalytic cleavage not indicated.

Conclusion. All mobile amino acids except Ile22 and Leu102 are located on the surface of BSLA, and are scattered over the enzyme, as seen in figure 3.12. In all cases fluctuations are caused by side chain torsions, while the backbone position is conserved. Two thirds of the amino acids have nonpolar side chains, and consequently are not engaged in hydrogen bonding . No concerted motion or drift of domains has been encountered apart from SV. In all other cases the structural integrity of the model is assured. Overall the ability of BSLA to adapt to various substrates seems rather limited. Examination of the active site geometry in the top view of the enzyme in figure 3.6 shows why. Spatial limitations are primarily due to the 'fingers', and the 'palm' region of the active site. These domains host catalytically indispensable groups, the oxyanion hole and parts of the catalytic triad. The spatial position of both fragments must be maintained and shielded from bulk solvent to retain the activity of the enzyme. Since the oxyanion hole formed by His156 and Asp133 is covered by a molecular layer that is only one side chain thick, BSLA lacks the room for adaption to substrates.

3.6 Hydrogen Bonding in the Tetrahedral Intermediate

The function of serine hydrolases depends on the local hydrogen bond network of the active site. Figure 3.13 shows a closeup view of the active site. Three different components of this network are discussed in the following. Firstly we focus on the hydrogen bond between His156 and the TI. Then we turn to the stabilisation of the oxyanion in the oxyanion hole. A third structural characteristic of *some* binding modes is the presence of an additional water molecule below the TI in the anterior region, which may improve the catalytic proficiency of BSLA.

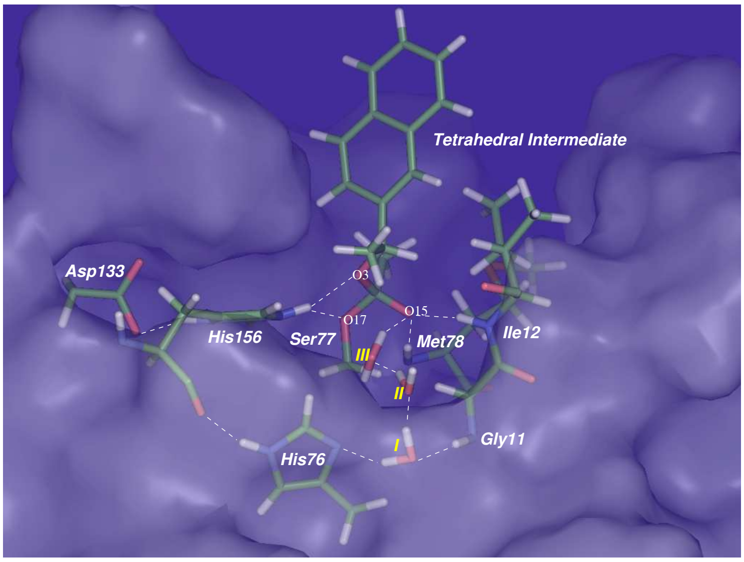


Figure 3.13: Active site hydrogen bond network (this example: RIII).

The Catalytic Histidine

Substrate preferences have been inferred from the distribution of hydrogen bonds between His156 and the TI obtained from equilibrium MD at the MM level [23]. For an efficient transport of hydrogen away from the catalytic histidine (His156), short hydrogen bonds are presumed to exist between the histidine atom NE2 and the reactive oxygen atoms O3 and O17 (figure 3.3, page 39). At the TI both reaction channels, one leading to the acylenzyme (cleavage at O3) and the other one to the Michaelis complex (cleavage at O17), are basically open. A preference for one of the reaction channels seems plausible if the mean hydrogen bond distances between NE2-O3 and NE2-O17 differ strongly . Furthermore, a weak (elongated), or nonexistent hydrogen bond at equilibrium indicates a barrier that must be overcome to enter this specific reaction channel, and would be thus considered a penalty. Short hydrogen bonds in the acylenzyme channel have been found preferentially for the fast-reacting enantiomer[23].

Averages of the hydrogen bond distances (the data conform approximately to a normal distribution) are displayed in table 3.4. For RIII, RIV, RV, RVI, and RVII as realisations of the TI with (R)-NAPHAC we find a slight asymmetry of the hydrogen bond distances in favour of the forward reaction channel, which leads to the acylenzyme. The same is found for SI, SIV, and SVI. In SV the distribution is symmetric. All systems mentioned so far correspond to

	RI	RII	RIII	RIV	RV	RVI	RVII	SI	SII	SIII	SIV	SV	SVI
NE2-O17	2.9	3.0	3.2	3.2	3.1	3.1	3.2	3.1	3.0	2.9	3.1	3.1	3.2
NE2-O3	4.1	3.2	3.0	3.0	3.0	2.9	2.9	3.0	3.4	4.2	3.0	3.1	2.9

Table 3.4: Hydrogen bond distances (Å) between NE2 of His156, and O3 or O17.

geometries which are assumed to be successful candidates, which means they should have the lowest possible barriers for acylation. The other geometries are strained to some extent, like RII or SII, or even expected to be non-reactive, like RI and SIII. Here a slight (RII and SII), or even pronounced asymmetry (RI and SIII) in disfavour of the forward reaction is noted.

At the TI hydrogen bonds are realised to both O3 and O17 for all supposedly catalytic configurations. This suggests that both reaction channels accessible. These observations tell nothing about the reaction barrier height, which is the target quantity of later chapters (chapters 4 and 5), but rather they provide some indirect evidence about expected relative barriers for decomposition of the TI. While the analysis can identify *very* poor candidates, such as RI and SIII, that were created here on purpose, it gives similar results for all the other geometries, and hence seems a rather insensitive. From the facile positional reorientation of the histidine towards both O3 and O17 we may deduce that a 'flip' mechanism such as proposed by Bachovchin[89] is not necessary for the catalytic activity of BSLA.

The Oxyanion Hole

	RI	RII	RIII	RIV	RV	RVI	RVII	SI	SII	SIII	SIV	SV	SVI
N_{Ile12}-O15	2.8	2.8	2.9	2.8	2.8	2.8	2.8	2.8	2.9	3.1	2.8	2.9	2.9
N_{Met78}-O15	2.7	3.1	3.2	3.1	3.1	3.1	3.2	3.1	3.0	2.8	3.1	3.2	3.4

Table 3.5: Average hydrogen bond distances (Å) between backbone amide of Ile12, Met78 and the oxyanion O15.

Optimum stabilisation of the TI is expected to occur only when hydrogen bonds are donated to the oxyanion O15 by the backbone amide groups of Ile12 and Met78. Inspection of the data in table 3.5 reveals the presence of such hydrogen bond in all trajectories. The hydrogen bond to Ile12 is generally shorter and therefore stronger than that to Met78, with the exception of RI and SIII, where the situation is reversed.

The distributions are quite uniform for all conformations in table 3.5, which precludes these two hydrogen bonds in the oxyanion hole from being the source of enantioselectivity.

The Water Network

Binding modes RIII, RVII, SV, and SVI feature an additional water molecule that is coordinated to the oxyanion (atom O15) from underneath the TI. This water molecule (labeled III in figure 3.13) is further held in position by crystal water I, and water II. During MD simulation waters I and III remain localised, water II exchanges occasionally. The water chain provides additional electrostatic stabilisation, mediated by water III, which acts as an extension of the oxyanion hole to three hydrogen bond donors. This effect may prove as an advantage for these binding modes.

3.7 Subset Energetics

In a force field calculation one can trivially separate and study interactions between different parts of the molecular system. This is frequently done in the analysis of MD simulations, which generate a trajectory file with coordinates written at regular intervals. One then iterates over this trajectory and computes the potential energy including only the contributions[10] and subsets of atoms which are of principal interest. Energy-based subsets have provided a qualitative explanation for the enantioselectivity in mutants of *Pseudomonas aeruginosa* lipase[22]. The subsets were chosen to understand the stabilisation of the TI[11], in terms of the interaction energies of the TI and the protein. All investigated mutants were more enantioselective than the wildtype and this was reflected in the preferred enantiomer having more favourable interactions with the surrounding system than the slow enantiomer. An implicit assumption in this investigation was that the rate determining step is the formation of the TI (step II in figure 1.2 on page 9). This supposition can fail, especially for nonproductive binding modes, when the rate determining step is formation of the acylenzyme (step III in figure 1.2).

In addition to analysing the interaction energy with the entire system we include the strain energy, measured as the interaction of the TI with itself. The strain is defined as the energy needed to distort the substrate as it enters the enzyme and forms the TI. This may easily modulate the reaction barriers and is one likely cause for enantiodiscrimination. Water is assumed to play an important role as one additional molecule is present in RIII, RVII, and SVI. We separated bulk solvent from the contribution of a hydration shell. This hydration water subset was defined to contain water molecules within 5 Å of the side chain of His76, or atoms C14 and O15 of the TI. His76 is an important amino acid for enantioselectivity of BSLA[72], and is located directly below the TI, in the anterior region of BSLA. The subset thus created is a measure for water in the site of entry. The energies in table 3.6 represent averages of the simulation starting at 500 ps. The effect of the conformational drifts of SVI and RV on the interaction energies is marginal and not considered here.

---

[10] Refer to chapter 6 for the energy expression of the Charmm22 force field.

[11] The study also included Michaelis complexes, but discrimination of enantiomers was not evident for these complexes. This is in agreement with Ema's observation that enantioselection occurs not upon binding of the substrate, but during the covalent steps of the reaction[57].

	Total	TI↔ ¬TI	TI↔Protein	TI↔Water	TI↔Hydr. Water	TI↔Bulk Water	TI↔TI
RI	-135.7	-178.4	-116.1	-62.3	-28.0 (7.1)	-34.3	42.6
RII	-132.1	-168.2	-119.3	-48.9	-13.0 (7.5)	-35.9	36.1
RIII	-137.9	-174.9	-113.9	-61.0	-24.4 (7.5)	-36.7	37.0
RIV	-125.6	-161.0	-121.3	-39.7	-2.7 (6.1)	-37.0	35.3
RV	-126.8	-161.3	-120.7	-40.5	-2.3 (6.2)	-38.2	34.5
RVII	-139.4	-176.3	-113.0	-63.3	-25.9 (7.1)	-37.4	36.9
SI	-132.8	-168.1	-131.8	-36.2	-8.4 (4.7)	-27.9	35.2
SII	-137.2	-171.6	-121.7	-49.9	-13.3 (5.9)	-36.7	34.4
SIII	-134.8	-171.2	-109.3	-61.9	-19.5 (8.0)	-42.4	36.4
SIV	-133.3	-168.7	-132.0	-36.7	-8.1 (4.1)	-28.6	35.4
SVI	-128.4	-174.8	-109.0	-65.7	-27.0 (8.1)	-38.7	46.3

Table 3.6: Subset interaction energies in kcal/mol. The second column ('Total') presents the sum of interaction and intramolecular strain energy (columns three and eight). The subset '¬TI' is the union of 'Protein' and 'Water', excluding the 'TI'. Column three therefore is the sum of column four and five. The water subset is the union of 'Hydration' and 'Bulk' water. Values in brackets specify the average number of water molecules in the hydration shell. The subset named 'TI' contains all atoms of Ser77 and the covalently attached NAPHAC except the amino acid backbone. See text for further details.

The best performing binding mode in terms of interaction energy with the surrounding (column 3 in table 3.6) is RI with a value of -178.4 kcal/mol. Models RVII (-176.3), RIII (-174.9) and SVI (-174.8) follow. Medium values are adopted by SII (-171.6), SIII (-171.2), SIV (-168.7), RII (-168.2), and SI (-168.1). At the lower end we find RIV (-161.0) and RV (-161.3). Considering both the interaction energy and intramolecular strain (column 2 in table 3.6) the ordering is reshuffled. RVII (-139.4), RIII (-137.9) and SII (-137.2) become the leading binding modes, RI (-135.7) falls back to the fourth place. The order of the other configurations is SIII (134.8), SIV (-133.3), SI (-132.8), RII (-132.1), SVI (-128.4), RV (-126.8), and finally RIV (-125.6).

The interaction energy with water (column 5 in table3.6) is most stabilising for those configurations which have an additional water in the anterior region of the TI, i.e., SVI (-65.7), RVII (-63.3), and RIII (-61.0). The nonproductive binding modes RI (-62.3) and SIII (-61.9) perform surprisingly well. An inspection of the trajectories reveals that water coordinates to atom O3, which is accessible to solvent. This is thought to induce subtle changes in the local water network. The other configurations follow at a distance. SII (-49.9) and RII (-48.9) have O3 tilted upwards, facing the solvent, which opens a possibility for hydrogen bonding. All remaining configurations are clustered around -40 kcal/mol or higher, such as RV (-40.5), RIV (-39.7), SIV (-36.7), and SI (-36.2). The number of water molecules in the hydration shell of the TI was allowed to fluctuate, since the subset definition was refreshed at each step of the trajectory analysis. Table 3.6 reveals that SVI has most water in the hydration shell, and SIV the least. The number of water molecules in the hydration shell and energy interaction with water are roughly correlated.

This analysis shows that interaction energies between subsets of the model must be used with care, and may easily be misleading when one tries to predict properties like relative reactivities. The results are strongly dependent on the partitioning of the model system and the energy terms which are used with the different partitions.

3.8 Conclusions

The generation of suitable input geometries for QM/MM calculations is a laborious process and needs to be executed with care. All results obtained at a later stage depend on the quality of the data produced at this stage. When experimental information about binding modes of the studied substrate or intermediate is absent, the accessible conformational space of the complexed or bound species must be explored. Only an exhaustive search can warrant coverage of all possible binding modes. A realistic modelling should furthermore account for relaxation of the enzyme to accomodate substrate or intermediate for any trial configuration. This task was solved here by manual molecular modelling at the docking stage, and the response of the surrounding medium was captured by MD simulations performed subsequently.

An interesting alternative to manual docking and relaxation is the use of automatted procedures like metadynamics. This method was used successfully in such a setting[79]. Success of a metadynamics simulation is dependent on a good choice of *metacoordinates*, which are the conformational degrees of freedom that are sampled or aid in the exploration of others. In the current system, one would obviously try to use torsion angles ω_1-ω_4 or a subset thereof as metacoordinates, and probably also the bond distance C14-O17 to facilitate conformational transitions and to study the dependence of the roughness of the binding surface on the distance of the ligand.

No matter what method is used, a complete coverage of the relevant conformational space cannot be guaranteed, which is unsatisfactory, since important binding modes may be overlooked. In the present work, we took care to include all sensible and accessible binding modes, and also checked two apparently nonproductive conformations.

As indicated by the experience with binding mode SV, it is very important to allow relaxation of the enzyme and to define criteria which separate useful from irrelevant geometries. To this end geometric criteria were found to be indispensable. We used the crystal structure(s) as reference, and scanned for systematic deviations like the ones reported for SV. Examination of the RMSD of the component residues of a macromolecule s a suitable measure to screen for geometric abnormalities.

Chapter 4

The Acylation Reaction: Potential Energy Surface Exploration

In the study of enantioselective ester hydrolysis of the NAPHAC substrate the acylation reaction is the phase where discrimination of the enantiomers takes place. Consequently we strive for a better understanding of the potential surface of this event. This chapter is organised as follows. At first we present a validation study to determine an appropriate level of theory and the details of the model for further use. Then some vital aspects of the reaction mechanism are discussed as an extension to the presentation in chapter 1, concerning features of the PES in the vicinity of the TI. Subsequently we evaluate the PES of the acylation reaction, computing both steps [I] and [II] (see figure 1.2 in chapter 1) for a large number of snapshot geometries, which were extracted from the trajectories of the TI (chapter 3). Finally light is shed on the electrostatic influence of individual amino acid side chains on the reaction barrier by a perturbation approach.

4.1 QM/MM Validation Study

The first choices to be made in order to be able to perform sensible QM/MM calculations of the acylation reaction are the theoretical level and the size of the QM-region within the enzyme model. Furthermore we need a meaningful reaction coordinate that allows a straightforward approach to the PES of a large number of structures, which are picked from the set of structures generated by the procedures described in the preceding chapter. For validation purposes we choose a test system which is henceforth referred to as ANRU453[1]. The vali-

---

[1] This model is not from the set of structures of the preceding chapter, but corresponds to an earlier setup, where the conformation of His76 was adopted from the crystal structure. A switch in the side chain conformation is belived to improve the local enzyme structure considerably (see preeding chapter for details), and therefore this earlier line of models was abandoned in production runs. This difference does not affect the usefulness of ANRU453 for evaluation purposes.

dation is done against the surface of step [II]. Due to the chemical similarity of both steps, there is no need for additional validation studies on step [I].

The level of theory was chosen in accordance with the general setup of the project. In the absence of experimental data derived from enzyme kinetics of BSLA with enantiopure substrates, a calibration against experimental barriers was not possible. All information that was available were E-values for the transformation of racemic NAPHAC by BSLA, i.e., information on relative reaction rates under the conditions of kinetic resolution. Consequently we concentrated on the relative values of activation energies predicted at different levels of theory, accepting the benefits from error cancellation in the comparative studies that we conducted. From earlier work done on a gas-phase model system of lipases, which we treated at the QM-level (see appendix C), we learned that DFT methods were in principle accurate enough to model hydrogen-bonded geometries and the PES of the acylation reaction.

Technical Considerations

The program Turbomole[2] was accessed via the Chemshell[191] interface and used for all calculations which employ the BLYP[131][141] and the B3LYP[14] functionals. The B3LYP functional is known to yield reasonable results for geometries and reaction energies of organic molecules[105]. Its computational cost limits its usefulness in repetitive studies, and it is more desirable to use the BLYP functional instead, where the RI approximation is applicable, speeding up optimisations by a factor of five for our system. For optimisation of reaction paths the BLYP density functional was thus used. However, we optimised a TS also at B3LYP/6-31+G* level, using a TS structure obtained at BLYP/6-31+G* level as initial guess.

Our preferred basis set was 6-31G[51] augmented with polarisation[95] and diffuse[37] functions for the heavy atoms for the TI as an anionic species. This basis set has proven useful also in other studies of serine hydrolases[100][236]. Additionally, we used the SVP[176] and TZVP[177] basis sets to check possible aberrations caused by the basis set. In all cases where the BLYP functional was used we also employed the RI approximation[55] using TZVP as auxiliary basis.

The semiempirical SCC-DFTB[56] method as implemented in the CHARMM program package starting from version 31b1 was also tested for its performance in the acylation reaction. SCC-DFTB became available to us only when the project was advanced beyond the reaction path calculations, which are described in the present chapter, and proved its utility in free energy calculations which are presented in chapter 5. However, it is included in the validation study to provide a complete picture. Geometry optimisations in CHARMM were

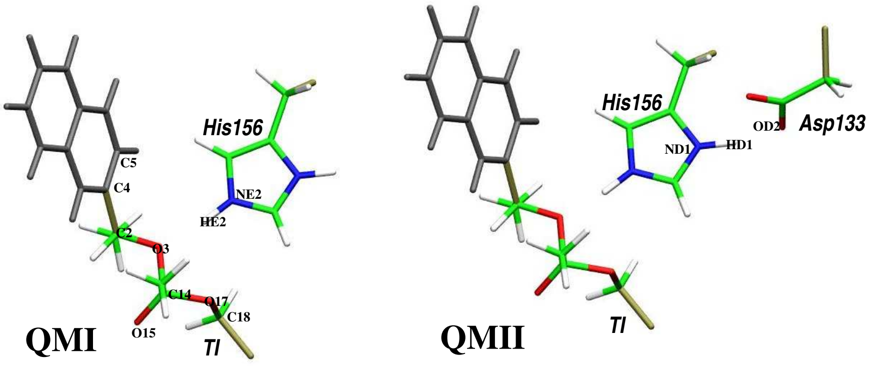


Figure 4.1: QM regions that were used in QM/MM calculations. In QMI 32 atoms (including link atoms) are treated at QM level, and 39 atoms in QMII. Atoms treated by molecular mechanics are coloured grey, whereas quantum mechanical atoms have unique colors (oxygen=red, carbon=green, hydrogen=white, nitrogen=blue). Frontier bonds between quantum mechanical and molecular mechanical regions are highlighted with tan colour

performed with the ABNR optimiser under conditions which fit closely those of the DFT calculations in Chemshell.

Choice Of QM Region

Two sizes of the QM region were tested, named QMI and QMII (see figure 4.1). In the spirit of a comparative study we tried to find a size that was both representative and small enough to ensure computational efficiency. As another boundary condition we want to use the same QM-region for both steps [I] and [II] to obtain comparable energy barriers. The chemical nature of the tetrahedral moiety helps in this regard and permits use of a QM-region which is chemically identical for [I] and [II]. The chemistry of acylation requires bonds NE2-HE2, C14-O17 and C14-O3 to be scissile. This minimal set is the core of QMI. Optionally one may include Asp133 to enable proton transfer from His156 to Asp133, which leads to QM-region QMII. The rationale behind QMII is to test the role of Asp133, which might participate in a low-barrier hydrogen bond or a related chemical construct (see chapter 1). Results obtained with a model system (*vide infra*), derived from the crystal structure of BSLA, indicate that in the gas-phase the proton may even shuttle between both residues. On the contrary, a QM/MM study of acetylcholinesterase[235] showed that proton tranfer was energetically unfavourable in this enzyme, and that the role of Asp133 could be rationalised by electrostatics. Therefore it was necessary to include Asp133 in validation calculations to find the minimal QM region consistent with the requirements of BSLA.

Generally it should be avoided to cut through polar bonds, furthermore one must include integral charge groups[2]when using ChemShell. Thus we included amino acids His156 and Asp133 with their complete side chains, which is usually the simplest choice for all amino acids. The TI with the complete NAPHAC-substrate contains too many atoms to be tractable in repeated calculations for a large number of snapshots. Therefore, the naphthyl ring, which is expected to contribute only sterically, was put in the MM region. Subsystem QMI thus contained three link atoms, while in QMII four link atoms had to be added.

Reaction Coordinate

The aim is to drive the system from the TI to either MCC or PDC, through reaction channels [I] or [II], respectively. This is possible via the use of contraints or restraints (see page 27), which are employed in PES scans (*vide supra*). In early QM/MM studies, RCs using one distance (equation 2.6) or a difference of two distances (equation 2.7) produced unreliable results. The resulting potential curves did not lead into the vicinity of the relevant transition states, and thus did not provide geometries that could be fed into geometry optimisers. Another RC, which is composed of three distances was successful in studies of acetylcholinesterase conducted in McCammon's group[235]. This RC allows treatment of either acylation step [I] or [II] in any single calculation and the restraint acts on a different set of atoms in each case:

$$\begin{aligned} \mathrm{Step[I]:}\quad & \mathrm{f}_{\mathrm{restraint}} = \mathrm{k}(\mathrm{r}_{\mathrm{O17-HE2}} - \mathrm{r}_{\mathrm{NE2-HE2}} - \mathrm{r}_{\mathrm{C14-O17}} - \mathrm{S})^2 \\ \mathrm{Step[II]:}\quad & \mathrm{f}_{\mathrm{restraint}} = \mathrm{k}(\mathrm{r}_{\mathrm{O3-HE2}} - \mathrm{r}_{\mathrm{NE2-HE2}} - \mathrm{r}_{\mathrm{C14-O3}} - \mathrm{S})^2. \end{aligned} \tag{4.1}$$

This RC (compare to equation 2.8) was adopted for our studies and provided better results with the acylation reaction than the other, simpler restraints (equations 2.6 and 2.7). We chose a force constant of 3.0 a.u., which is of the same order of magnitude as those found in the literature[238][235]. The initial value of S was set to the value obtained at the structure optimized without restraint. In analogy to quantum chemical computations on small systems this structure may be regarded the equilibrium structure of the TI. Stepsizes between 0.01 Å and 0.125 Å were used, although the former is too fine to be practical for typical production runs. Such a small stepsize should only be used when a true TS is to be refined from a reaction path structure using uphill search methods[19], which require the initial structure to resemble the true TS rather closely.

---

[2] Based on the topology of the Charmm22 force field.

Geometry Optimisations

The procedure described in the following is very similar to that used in the later production runs. In our approach the TI is the basis for all further studies. Steps [I] and [II] can be accessed from the TI in parallel, under identical starting conditions.

The initial structure for geometry optimisation was always a randomly picked snapshot from a MD trajectory of the TI. At the MM level, using the CHARMM program, the system was energy minimised (500 steps SD, then 10000 steps ABNR), including all residues which had at least one atom within a sphere of 12.0 Å around the origin. This type of selection, where entire residues are selected, can be enforced by submitting the 'BYRES' keyword to the CHARMM interpreter. This initial MM refinement was necessary to keep the computational costs of the QM/MM level at bay. The utility of a force field representation including reasonable parameters for the non-proteinogenic residues cannot be overemphasised at this point (for parametrisation see chapter 6). If parameters for non-protein residues are missing, one might consider to keep such fragments fixed in the MM refinement, or to use unminimised snapshot geometries. Both these options were tested and found to be impractical in the current study: any geometry optimisation at the DFT/MM level attempted in this fashion failed.

The structure minimised at the MM level was then subjected to optimisation at the QM/MM level using the same optimised region as in the MM calculations. An initial QM/MM calculation was done to locate the TI, which was assumed to be a minimum on the PES[3]. The HDLC optimiser was used to perform the energy minimisation. Thereafter, the energy optimised structure was submitted to a reaction path calculation, through an automatted procedure executed in ChemShell. In this procedure, subsequent minimisations were done in a loop where at each step the constant S of the restraint was adjusted. The step size was set to 0.05 Å for the BLYP/6-31+G* scan and to 0.125 Å for the scans with SVP and TZVP bases. The convergence criteria of the HDLC optimiser are listed in table 4.1. Optimisation of the reaction path at SCC-DFTB level was performed in the CHARMM program. The ABNR optimiser was used with convergence criteria for both average step and gradient tolerance set to 0.0005 Å and 0.0005 $\mathrm{kcal/Å}$, respectively.

As a check of the RC we located true TSs for the QMII/BLYP/6-31+G* and QMII/B3LYP/6-31+G* cases. A well behaved RC should lead from the energy minimised reactant structure via the TS to the products. In order to get closer to the TS a reaction path scan with a resolution of 0.01 Å was executed in the vicinity of the barrier. A TS search was then performed with the HDLC optimiser using the structure highest in energy as an initial guess. The TS

---

[3] This assumption was relaxed during the course of the project.

	Energy minimisation	**TS optimisation**	
		Reaction core	Surrounding
maximum step component	0.0054	0.0054	1.0
RMS of step vector	0.0036	0.0036	1.0
maximum gradient component	0.00135	0.00135	0.00045
RMS of the gradient	0.0009	0.0009	0.0003

Table 4.1: Convergence criteria of the QM/MM geometry optimisations expressed in atomic units (angles in rad).

thus found was confirmed by normal mode analysis. The TS search performed best when mode switching was enabled during the optimisation. The core region in the HDLC optimisation contained four atoms: NE2, HE2, O3 and C14 (see figure 4.1). This seemed the most natural choice since these atoms participate directly in bond breaking and formating. The convergence criteria for these optimisations are given in table 4.1.

Results of the Validation Study

Figure 4.3 shows the energy profile of the reaction path derived for ANRU453 at BLYP/6-31+G*/QMII level. The minimum on the right hand side corresponds to the TI. The energy rises in going from the TI to a maximum, marking a TS, in direction of more negative values of the RC. Following the graph onwards leads to another minimum on the left hand side of the graph, which is identified as PDC. In the same diagram another curve shows results obtained at B3LYP/6-31+G* level of theory. These energies are single point energies at the BLYP/6-31+G* level. The curves have identical shapes and are both smooth, but the B3LYP/6-31+G* barriers are higher, which is a common observation[58]. The coherence of both graphs lets one suspect already at this point that the geometrical preferences of both functionals are quite similar in the present study. Also shown in figure 4.3 are BLYP/6-31+G* single point energies at the stationary points of the reaction path, but calculated without the restraint term (equation 4.1). Both restrained and unrestrained energies are essentially identical. This is consistent with the general observation that the restrained distances relax quickly to their new equilibrium position by virtue of the large force applied to them, and then contribute no further to the total energy.

Tables 4.2, 4.3, and 4.4 present geometric parameters of model structures corresponding to TI, TS, and PDC. Those structures were derived in different ways. Geometries at the QMII/BLYP/6-31+G* and QMII/B3LYP/6-31+G* level of theory were refined from the restrained reaction path BLYP structures by unrestrained optimisation. To check the capability of our RC to get close to a TS, we compare the TS structure to the one from the top of the barrier ('SNAP20') of the QMII/BLYP/6-31+G* reaction path scan (table 4.3). All other structures were extracted directly from the corresponding reaction path. The geometries of the TI are very similar with all methods. Minor deviations exist between QMII/BLYP/6-31+G* and QMI/BLYP/6-31+G* for the distance OD2-HD1, which is marginally smaller in the latter case. Major effects could have been expected to occur between analogous QMI and QM II calculations due to the interaction with Asp133 being treated differently, but the difference remain small. The largest deviations in bond angles occur with QMI/SCC-DFTB which overestimates the angle NE2-O3-HE2 consistently by $3-6^{\circ}$. This angle can be seen as a measure of the hydrogen bond strength between His156 and the TI. QMI/SCC-DFTB is known to underestimate the strength of hydrogen bonds[44], which may cause this deviation. In addition, too short bond lengths for C14-O3 are found in TS and PDC optimised at the QMI/SCC-DFTB level, the acyl group is not quite planar and there are significant deviations of the torsions ω_1-ω_4 compared with the BLYP and B3LYP results.

The data in table 4.3 confirm that our choice of RC is justified since there is a close match of the geometry of genuine TS at the QMII/BLYP/6-31+G* level of theory and the structure of an approximate TS ('SNAP20'), extracted from the reaction path. The frequency analysis (see figure 4.4) demonstrates that we have located a true TS for the reaction under study. A transfer of hydrogen HE2 from NE2 to O3 makes the largest contribution to the TS mode, but there are also contributions by carbon C14 and oxygen O3 which indicate motion in opposite directions and therefore a breaking of the C14-O3 bond. The difference in ΔE between the genuine TS and reactants on one hand, and the barrier height read from the reaction path diagram on the other hand, is negligible.

The energy barriers calculated with the various methods are displayed in table 4.5. Comparison of the data obtained with QMI/BLYP/6-31+G* and QMII/BLYP/6-31+G* shows that the inclusion of Asp133 in the QM region does not change the picture significantly, both barriers are of comparable height. Larger differences are found for the stabilisation of the TI ($\mathrm{TI} \rightarrow \mathrm{TS}$). They are largest in QMII/B3LYP/6-31+G* and smallest with QMI/SCC-DFTB. The QMII/BLYP/TZVP barrier is the smallest of the *ab initio* barriers, missing diffuse functions may explain this phenomenon over the results with QMII/BLYP/SVP+. Although TZVP (triple ζ) is generally expected to be a better basis, the diffuse functions used with the SVP+ (double ζ) lead to a stabilisation of the anion, and thus a higher activation barrier for dissociation of the TI. The barrier for nucleophilic attack on the carbonyl group of the acylenzyme ($\mathrm{PDC} \rightarrow \mathrm{TS}$) is largest in QMII/B3LYP/6-31+G* and QMI/SCC-DFTB (7.7 kcal/mol). The other DFT results with the 6-31+G* basis are slightly smaller (5.8-6.7 kcal/mol) while QMII/SVP+ again seems to underestimate the barrier (4.4 kcal/mol).

To check for a low-barrier hydrogen bond (see chapter 1) between His156 and Asp133, or even a proton shift (as in a gas-phase model of lipases, see appendix C), we performed a scan of the proton HD1 position using reaction coordinate 2.7 (r_{A-B}=ND1-HD1, r_{C-D}=HD1-OD2). The resulting energy curve is presented in figure 4.2. Graph A shows the results of the scan obtained with the standard set of charges for the MM region. In reaction B, charges on residues adjacent to Asp133 were deleted. Those residues were two crystal water molecules, and the backbone of amino acids Met134, Ile135, and Val136, which are partially hydrogen bonded to Asp133. The immediate environment was therefore less polar, and can be considered as a first step towards the 'gas-phase' model. The initial position of the scan is on the left hand side and it proceeds towards more positive values on the right hand side. For negative values of the RC the proton is found closer to His156, and vice versa. At a value of 0.0 Å the proton is shared between both residues. We see that in scan B the initially minimised position of the proton is a little further away from His156 than in A (B: 1.11 Å, A: 1.08 Å). Oxygen OD2 of Asp133 is closer to the His156 in the initial structure in B (B: 1.50 Å, A: 1.60 Å), possibly as a result of structural destabilisation due to the deletion of charges. The modified amino acids were already observed to play and integral role in the stabilisation of the catalytic triad in chapter 3. In neither A nor B another minimum is seen close to Asp133, but the increase in energy is steeper in A. In the case (B) of an apolar surrounding, the barrier for proton abstraction from the histidine is lowered by a significant amount, a circumstance which may enable low-barrier hydrogen bonds. In the present case (A), the surrounding must be polar to stabilise the catalytic triad. Hence, the proton is tightly located at His156, and the hydrogen bond is expected to be mainly electrostatic in character. The similarity of barriers for step [II] of the acylation reaction computed with QMI and QMII at BLYP/6-31+G* level support this assumption.

We finally address the C14-O15 bond length in PDC, which has been determined by Tonge and Carey[205][204] (see chapter 1). In PDC the acylenzyme has a longer C14-O15 bond as compared to that of a 'standard' ester. Tonge and Carey cited a C=O bond length of 1.217 Å [204] for methyl-acetate as a reference for a free ester [4]. In the acylenzyme they observed an elongation of this C=O bond of about 0.015 Å, and estimated that changes up to 0.03 Å compared to a 'standard' ester could be expected. From table 4.4 a value of 1.24 Å (QMI/BLYP/6-31+G*, QMII/BLYP6-31+G*), or 1.23 Å (QMII/B3LYP/6-31+G*) is observed for the C14-O15 bond. This corresponds to C=O bond elongation by 0.023 Å or 0.013 Å, which is within the range of 0.015 Å-0.03 Å that Tonge and Carey predicted.

---

[4] B3LYP/6-31G* geometry optimisation using Gaussian03[71] yields 1.212 Å.

Conclusions

Geometric properties, especially of hydrogen bonds in the active site, are almost identical with BLYP and B3LYP. Energy barriers are higher with B3LYP than with BLYP, but both functionals would seem suitable for relative reaction barriers of enantiomers. Therefore, we decided to use the BLYP functional in the following study. The 6-31+G* basis has already proven its utility in studies of serine hydrolases not conducted by us[100][236], and the additionally tested SVP+ and TZVP bases did not produce any contradictory data, so that the decision was in favour of 6-31+G*. Subtle changes such as C14-O15 bond elongation observed experimentally by Tonge and Carey could be mapped by our model at QMI/BLYP/6-31+G* level, which enhances our confidence into the chosen approach. In QM/MM free energy calculations, where an excessive number of energy and gradient evaluations is necessary to obtain statistically sound results, we shall use the SCC-DFTB method for quantum chemical representation of the QM region. The differences between QM regions QMI and QMII are quite small, and Asp133 does no seem to take part in any low-barrier hydrogen bond or a similar construct. Therefore QMI was chosen to represent our QM region in all further studies.

The reaction coordinate composed of three distances (equation 4.1) produces a clean and continuous energy profile, leading from TI to PDC, and passing by a TS. The precise location of the TS was confirmed by a TS optimisation using the corresponding structure from the reaction path as input, and found to be very close to the latter regarding both energetic and geometric criteria. We thus conclude that regular TS searches can be avoided in the production phase, and be replaced by reaction paths, the handling of which is more straightforward, and does not require restarts and manual inspection to the same extent as TS searches do.

4.2 Reaction Path Study

The aim after the initial validation was to evaluate the reaction thermodynamics of ester cleavage during the acylation reaction. This process has been discussed in the light of enantioselectivity in chapter 1. Since enantioselectivity is governed by small free energy differences of typically less than 3 kcal/mol for the reaction pathways of the two enantiomers, special heed has to be paid to an accurate calculation of the reaction barriers. Therefore, it is not sufficient to rely on single reaction path calculations for each enantiomer, but it is necessary to perform multiple calculations on several snapshots from the trajectories discussed

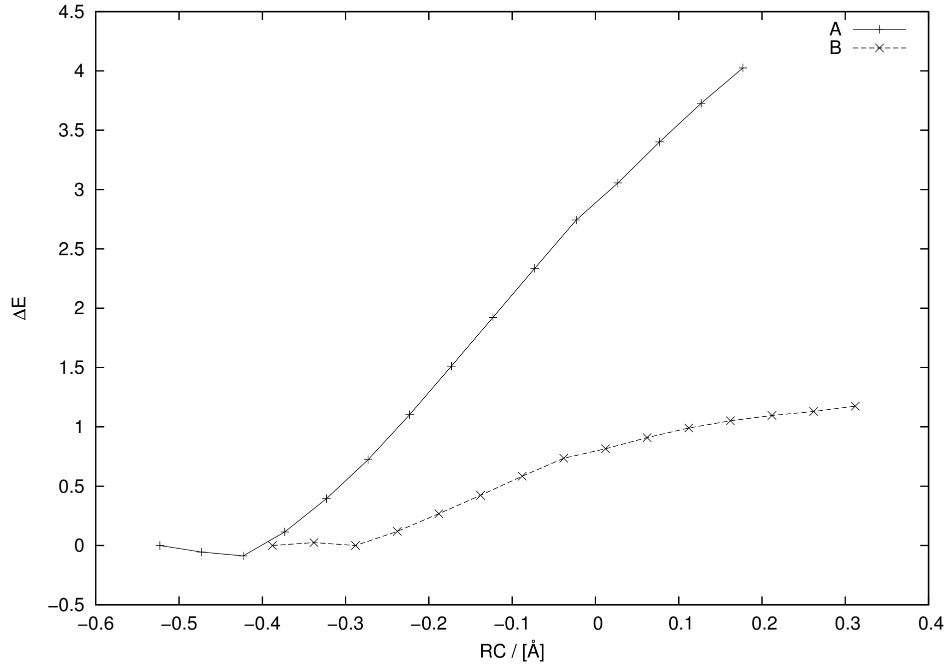


Figure 4.2: PES scan of proton transfer from His156 to Asp133 in ANRU453. The original charges were used in the MM region in scan A. The charges of residues adjacent to Asp133 were deleted in scan B.

Geometrical feature	Level of theory					
	QMI/ BLYP/ 6-31+G*	QMII/ BLYP/ 6-31+G*	QMII/ BLYP/ SVP+	QMII/ BLYP/ TZVP	QMII/ B3LYP/ 6-31+G*	QMI/ SCC-DFTB
Bond-length						
O3-HE2	1.65	1.64	1.57	1.58	1.66	1.59
NE2-HE2	1.07	1.07	1.09	1.08	1.05	1.09
C14-O3	1.59	1.58	1.60	1.63	1.53	1.66
C14-O15	1.30	1.30	1.29	1.28	1.30	1.30
C14-O17	1.50	1.51	1.49	1.50	1.47	1.52
ND1-HD1	1.07	1.08	1.10	1.09	1.07	1.05
OD2-HD1	1.53	1.60	1.56	1.57	1.61	1.60
ILE12-O15	1.82	1.81	1.81	1.81	1.81	1.77
MET78-O15	1.86	1.86	1.88	1.88	1.84	1.85
Angle						
NE2-O3-HE2	4.45	3.33	2.78	2.99	3.39	7.55
OD2-ND1-HD1	5.83	6.68	6.56	6.49	6.98	6.29
Torsion						
ω_4	87.3	87.4	88.0	87.6	87.3	87.1
ω_3	124.5	125.3	124.3	125.1	125.5	129.6
ω_2	155.4	155.2	155.0	154.6	155.1	150.5
ω_1	−79.5	−80.1	−79.6	−80.6	−80.3	−81.1

Table 4.2: Key geometric descriptors of the TI of ANRU453. See figure 4.1 for atom nomenclature. The quantum mechanical region is indicated above the basis set specification. Data corresponding to QMII/6-31+G* was obtained from optimisations without restraint term. The other structures correspond to minima on the reaction path.

Geometrical feature	**Level of theory**						
	QMII/ BLYP/ 6-31+G*	QMII/ B3LYP/ 6-31+G*	QMII/ BLYP/ SNAP20	QMI/ BLYP/ 6-31+G*	QMII/ BLYP/ SVP+	QMII/ BLYP/ TZVP	QMI/ SCC-DFTB
Bond-length							
O3-HE2	1.25	1.25	1.26	1.27	1.27	1.33	1.41
NE2-HE2	1.26	1.25	1.26	1.25	1.25	1.19	1.15
C14-O3	1.91	1.88	1.90	1.89	1.88	1.89	1.78
C14-O15	1.27	1.26	1.27	1.27	1.26	1.26	1.28
C14-O17	1.42	1.40	1.43	1.43	1.42	1.43	1.49
ND1-HD1	1.07	1.05	1.07	1.06	1.08	1.08	1.05
OD2-HD1	1.68	1.69	1.68	1.56	1.62	1.63	1.62
ILE12-O15	1.85	1.84	1.85	1.85	1.84	1.84	1.79
MET78-O15	1.90	1.88	1.89	1.89	1.90	1.90	1.88
Angle							
NE2-O3-HE2	1.96	1.51	1.92	2.65	1.33	1.55	5.60
OD2-ND1-HD1	7.06	7.60	7.14	6.22	6.95	6.87	6.54
Torsion							
ω_4	89.3	88.8	88.8	89.3	88.6	88.2	88.0
ω_3	125.1	123.5	124.6	123.9	123.1	123.9	128.6
ω_2	151.1	152.9	151.8	151.6	152.7	152.8	149.2
ω_1	−79.8	−78.4	−79.4	−79.0	−78.5	−79.3	-80.3

Table 4.3: Key geometric descriptors of the genuine TSs obtained using QMII/BLYP/6-31+G* and QMII/B3LYP/6-31+G* (left hand side, first 2 colums) versus structures picked from the top of the barrier of the reaction paths (right hand side, last 5 columns). See figure 4.1 for atom nomenclature. The quantum mechanical region is indicated above the basis set specification. SNAP20 corresponds to the structure found at the hilltop of the QMII/BLYP/6-31+G* scan.

Geometrical feature	**Level of theory**					
	QMI/ BLYP/ 6-31+G*	QMII/ BLYP/ 6-31+G*	QMII/ BLYP/ SVP+	QMII/ BLYP/ TZVP	QMII/ B3LYP 6-31+G*	QMI/ SCC-DFTB
Bond-length						
O3-HE2	1.01	1.02	1.02	1.01	1.00	1.00
NE2-HE2	1.80	1.72	1.69	1.69	1.73	1.91
C14-O3	2.52	2.50	2.44	2.45	2.49	2.16
C14-O15	1.24	1.24	1.24	1.23	1.23	1.25
C14-O17	1.37	1.37	1.36	1.37	1.35	1.39
ND1-HD1	1.06	1.06	1.07	1.06	1.04	1.05
OD2-HD1	1.54	1.71	1.66	1.67	1.72	1.61
ILE12-O15	1.95	1.95	1.94	1.93	1.94	1.85
MET78-O15	1.88	1.86	1.88	1.88	1.85	1.88
Angle						
NE2-O3-HE2	6.50	5.11	4.80	4.85	5.20	10.95
OD2-ND1-HD1	4.97	7.06	6.84	6.76	7.28	5.71
Torsion						
ω_4	91.2	90.1	88.6	89.3	90.0	86.7
ω_3	115.9	110.8	116.9	115.8	110.5	135.5
ω_2	154.0	160.6	155.4	156.5	160.5	141.5
ω_1	−73.5	−72.1	−73.2	−77.3	−72.6	−83.7

Table 4.4: Key geometric descriptors of the PDC of ANRU453 (acylenzyme + alcohol). Data corresponding to QMII/6-31+G* was obtained from optimizations without restraint term. The other structures correspond to minima on the restrained path.

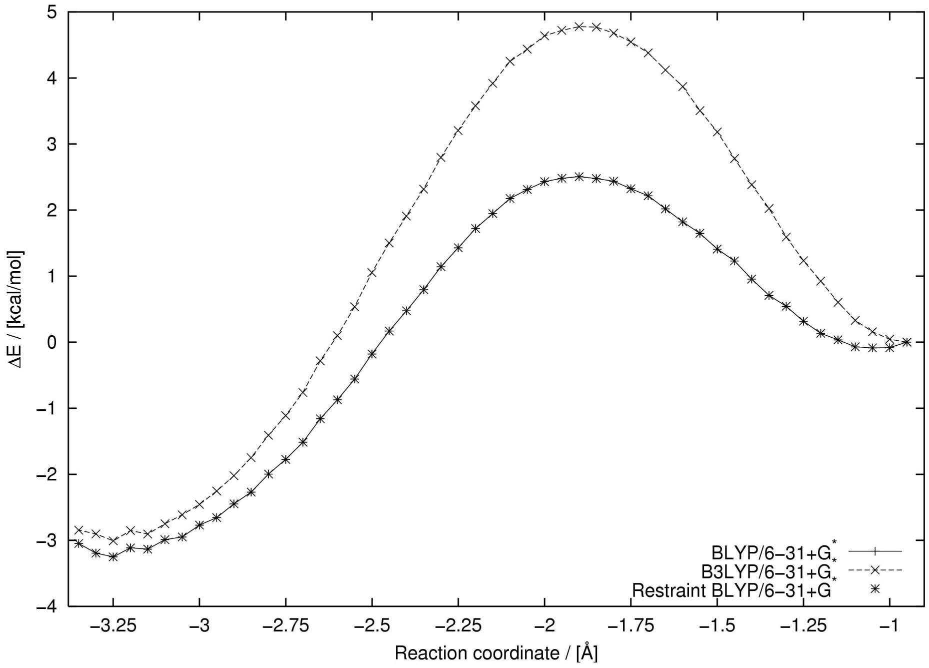


Figure 4.3: QMII/BLYP/6-31+G* optimized path and single point energies at QMII/BLYP/6-31+G* and QMII/B3LYP/6-31+G* level, both without the restraint term. At the stationary points the reaction coordinate term makes no significant contribution to the energy.

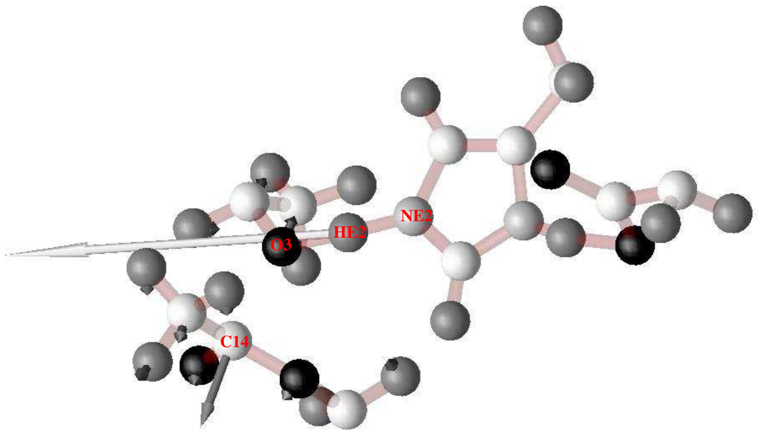


Figure 4.4: View of the reaction center of the lipase in the QMII/BLYP/6-31+G* transition state geometry with the normal mode vector corresponding to the imaginary frequency plotted onto the atomic positions. All atoms shown were unconstrained in the calculation of the Hessian using the whole QM/MM system. The labeled atoms were used as core atoms in HDLCopt during optimization of the transition state.

in the preceding chapter, in order to average out noise due to fluctuations in the protein environment.

One of the problems is the large number of possible binding modes of the TI, which we

QMI/6-31+G*	QMII/6-31+G*	QMI/SCC-DFTB	
2.3	2.5	0.2	TI → TS
6.7	5.8	7.7	PDC → TS
QMII/SVP+	QMII/TZVP	QMII/B3LYP/6-31+G*	
1.6	0.8	4.9	TI → TS
4.4	6.3	7.7	PDC → TS

Table 4.5: Barrier heights as calculated from energies corresponding to the geometries in figures 4.2, 4.3, 4.4. The BLYP functional was used in DFT calculations unless specified otherwise. All energy values are given in kcal/mol.

found by the model building and trajectory analysis described in chapter 3. By the time the reaction path study was conducted eight binding modes of the TI had been discovered. Those were RI, RII, RIII, RIV, and RV for the R-enantiomer, and SI, SII, SIII, and SIV for the S-enantiomer. Due to stability problems, binding mode SVI was discovered later during the work and could not be considered in the reaction path calculations. Also RVII was built late in the work and was considered only in free energy calculations. Due to structural affinity with RIII, the main difference being a rotation of the naphthyl ring by 180°, the catalytic proficiency of RIII and RVII is expected to be similar.

In the study of acylation reaction, we had to decide first which criterium should be applied for a representative selection of snapshots from available MD trajectories. Due to a lack of experience with the system and in view of the potential influence of *any* geometric parameter on the calculated reaction rates, the simplest possible ansatz was adopted, i.e., we picked geometries at fixed intervals from the trajectories. Such a strategy had also been employed by Zhang et al.[236], and was found useful for determining the fluctuations in the calculated QM/MM reaction barriers. Generally speaking, if one would know that only a certain range of values for one or more geometric parameters was acceptable, then snapshots could be selected according to a statistics over these parameters. Useful techniques for this purpose include cluster or principal component analysis[200][94]. An example for the use of cluster analysis, and an implementation, is provided in appendix G.

The reaction path optimisations were performed in two batches. The first batch comprised the snapshots taken from the trajectories at 500, 700, and 1000 ps, and the second one those at 600 ps. Optimisations were done as described in the preceding validation study, with minor alterations. The optimised region was expanded around the origin by 0.5 Å to a value of 12.5 Å. Furthermore the optimisation included the water molecules within 9.0 Å around the TI. The rationale was that for a quantitative evaluation of the barriers, the dissociating alcohol or ester should not experience an artificial restoring force stemming from an immobile solvent. The number of optimised atoms was then around 1500 out of a total of approximately 7500. Reaction path studies were always preceded by an energy minimisation of the TI, using the same HDLC options as before. At the stable intermediates, thus optimised, the corresponding value of the RC was determined. The first reaction path optimisation then used this value of the RC plus the standard increment of 0.125 Å as the starting point. The end value was arbitrarily set to an RC value of -3.5 Å, and later expanded to -4.5 Å for several paths to prove that a minimum had actually been found. In many cases a scan in the back direction starting from the end points of the reaction path was performed. This provides an intrinsic quality control, as both forward and backward reaction path should match structurally and energetically[238].

Together with the second batch another strategy aimed at acceleration of the barrier calculations was tested. Knowledge of the approximate locations of TSs from the first batch of optimisations was employed to attempt direct optimisation of TSs from the trajectory snapshots, omitting the laborious reaction path scans. From any TS thus located one can then add and substract increments of the normal mode corresponding to the imaginary frequency, and optimise for the minima, i.e., the TI and MCC, or PDC, respectively. The difference in energy between TS and the minima then gives the desired barriers. This strategy was put into practice in the following way. First the snapshots were energy minimised at the MM level, employing the same technique as before. Then the RC (defined in equation 4.1) was applied with a value of -1.8 Å, which was the position where most TSs were found in the first batch (see tables 4.7 and 4.8), and then the system was forced towards this value in a geometry optimisation, in the same manner as above. The optimisation of minima was performed after addition of the normal mode increment, without any restraint, using the same cutoff criteria with the HDLC optimiser as before (table 4.1). All snapshots taken at 800 ps and 900 ps were subject to this alternative procedure.

Consideration Of Computer Time

The large number of trajectories just mentioned, and the snapshots needed for representative sampling pose large demands on the local computer infrastructure. All jobs were run serially, which was the only sensible option regarding the large number of reaction paths needed to complete the study. From the experience gained we conclude that the computer time needed with a 2.6 GHz CPU of the Intel Xeon generation is on the order of four to six weeks for a reaction with 32 QM atoms treated at the BLYP/6-31+G* level (use of QMI), a region of 1500 optimised atoms out of 7500, and approximately 25 optimised geometries. Large fluctuations in the required computation time were observed and found to be caused by the load of the Xeon dual-CPU workstations whose memory adapter turned out to be a bottleneck when jobs were processed by the second CPU. With AMD Opteron workstations at 2.0 GHz the same job consumes approximately three weeks of CPU time, without dependence on other jobs.

Definition Of A Naming Scheme

The large number of snapshots and trajectories necessitates an efficient notation to simplify reference to the results. Similar to chapter 3 the corresponding binding modes are referred to as for example RI, RII, RIII. Additional pieces of information must specify the trajectory time of extracted snapshots, and the *direction* of the scan. As *forward* (F) direction we will always classify a scan starting from a TI, proceeding either to PDC or MCC. The term *backward* (B) direction will be applied when the first structure on the path is either a PDC or MCC, and the end point is the TI. To designate the forward scan performed on the snapshot extracted from the equilibrium trajectory of binding mode III of the R enantiomer after 500 ps example we shall write: **RIII500F**. The backward scan would be **RIII500B**, respectively.

4.2.1 The Tetrahedral Intermediate

The existence of a TI on the acylation pathway is a consensus belief of those who are active in the field of serine hydrolases. To the present day, however, its existence has never been proven experimentally. The difficulties in obtaining information about the TI are rooted in its short lifetime which would require ultrafast time-resolved spectroscopic methods for proper characterisation. Theoretical predictions can fill the gap for the most important geometrical properties which are characteristic of the TI.

The TI is a species which is often intuitively characterised as being close to a TS[97]. For an ester substrate being attacked nucleophilically at the carbonyl group this should imply a lengthening of the C=O (C14-O15) bond, which attains a significant amound of single bond character as an anion forms at O15. In the TI as representative of the anionic intermediate we find an average C=O bondl-ength of 1.29 Å, which is significantly larger than that of a standard ester, for example methyl-acetate[5], at 1.21 Å. Oxygen O15 is positioned such that it can form strong hydrogen bonds with the residues forming the oxyanion hole. Table 4.6 shows the averaged distance between O15 and the hydrogen atoms of the backbone of the indicated residues.

The C-O (C14-O3 and C14-O17) ester bonds are elongated with each alcohol moiety being a potential leaving group. Comparing to a standard ester[6] we find that the bonds C14-O17

---

[5] Geometry optimised at B3LYP/6-31G* level in Gaussian03[71].

[6] Methyl-acetate optimised at B3LYP/6-31G* level in Gaussian03.

and C14-O3 are significantly elongated (TI: 1.56 Å; Methyl-acetate: 1.35 Å). At the TI both attacking (C14-O17) and leaving (C14-O3) alcohols are equivalent in an average sense as judged by the data in table 4.6. The force field used to simulate the TI was trimmed for this equivalence. At the QM level this view can be differentiated. To see this point the reader is referred to fable 4.6. Here the C14-O17 or C14-O3 bond lengths are drawn on the abscissa, and the distance of both oxygens to HE2 at His76 is plotted on the ordinate. Focusing first on C14-O17 (left part of figure 4.6) three clusters are observed, one at 1.50 Å, another at 1.60 Å, and a minor one centered about 1.55Å. Some outliers are noted at values above 1.65Å, which are produced by the snapshots of SII and RIV600B. The distribution is non-normal and best characterised by two accumulations at 1.50Å and 1.60Å. The C14-O17 bond seems correlated with the parameter on the ordinate, one may even be tempted to interpret this as a linear relationship: at small distances O17-HE2 an elongation of the bond C14-O17 is observed, and vice versa. The distance O3-HE2 is short when O17-HE2 is large, and long when O3-HE2 is short. This points to O3 and O17 being sole basins of attraction for the hydrogen. Some white space caused by an absence of values on the ordinate is observed around 2.0 Å. This oxygen-hydrogen distance is not found in the QM/MM minimised model systems. Similar observations are made for the C14-O3 bond displayed in the graph on the right hand side in figure 4.6 which shows less structure and have a more scattered distribution than that for C14-O17. At the lower end of the scale of C14-O3, close to 1.45 Å, one find the models which have the largest C14-O17 distances indicating a correlation between these bond-lengths.

The overall behaviour suggests hydrogen bonding to be the cause for elongation/shortening of the C14-O17/C14-O3 bonds. In the studied systems either O3 or O17 *attracts* HE2, as is readily inferred from the plots. Intermediate values of O17-HE2 and O3-HE2 are not found due to an expected small barrier around 2.0 Å, separating the minima which exist due to hydrogen bonding. We can further corroborate this hypothesis by a plot of C14-O17 and C14-O3 versus the angles NE2-O17-HE2 and NE2-O3-HE2, as depicted in figure 4.7. The trends observed here are in accordance with the previous findings, and confirm the structural effects arising from a hydrogen bond at the TI. An illustration of this conformational feature is given in figure 4.5. A small barrier was also found in a gas-phase model of serine hydrolases[100]. The present results indicate that this barrier is small and uniform for the whole range of binding modes. Small rotations of the histidine during the lifetime of the TI serve to switch access to reaction channels [I] and [II]. This forced sequential behaviour in the protected enzyme environment is different from ester hydrolysis in solution, where catalytic groups may be present at both oxygen sites and effect steps [I] and [II] to become

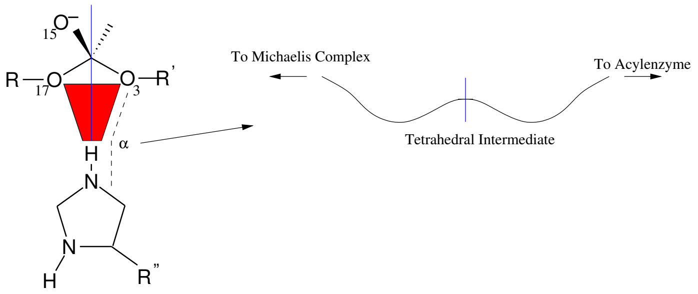


Figure 4.5: As the angle α is swept through, hydrogen bonds from histidine to the TI are broken and formed. This is thought to lead to a local double well structure at the TI shown on the left hand side. The barrier is located at the intermediate position, where no hydrogen bonding to either of the oxygens is possible.

concerted processes[175]. Our study so far suggests that a complicated mechanism of histidine relocation like the 180° flip of the catalytic histidine about the CB-CG bond proposed by Bachovchin[89] is not necessary; simple rotation is sufficient for a proper functioning of serine hydrolases.

A standard biomolecular force field like Charmm22 is not capable of reproducing the subtle effects like bond elongation on hydrogen bonding which are discussed in the preceding paragraphs. These require a quantum mechanical representation of the fragments, or a force field which assumes a special functional form to accomodate such effects. The EVB model described in the appendix F, for example, can be designed to model the binding in the active site more accurately.

C14-O17	1.56	NE2-HE2	1.06
C14-O3	1.56	ILE12-O15	1.87
C14-O15	1.29	MET78-O15	1.95

Table 4.6: Some average geometric parameters of the TI at QM/MM energy minimised structures.

4.2.2 The Acylation Reaction Revisited

Reaction Profiles

In chapter 1 the kinetics of the acylation reaction were treated in a simplified fashion. Steps [I] and [II] of the reaction (figure 1.2) were contracted into a single step, and the TI was assumed a TS. In some studies[235][236], researchers focus on step [I] and equate its barrier to that of the acylation or even the overall hydrolysis. While this may be the valid for amide substrates, the symmetry of the TI and the small differences that are expected to be decisive for enantiodiscrimination make such an approach prone to failure in the case

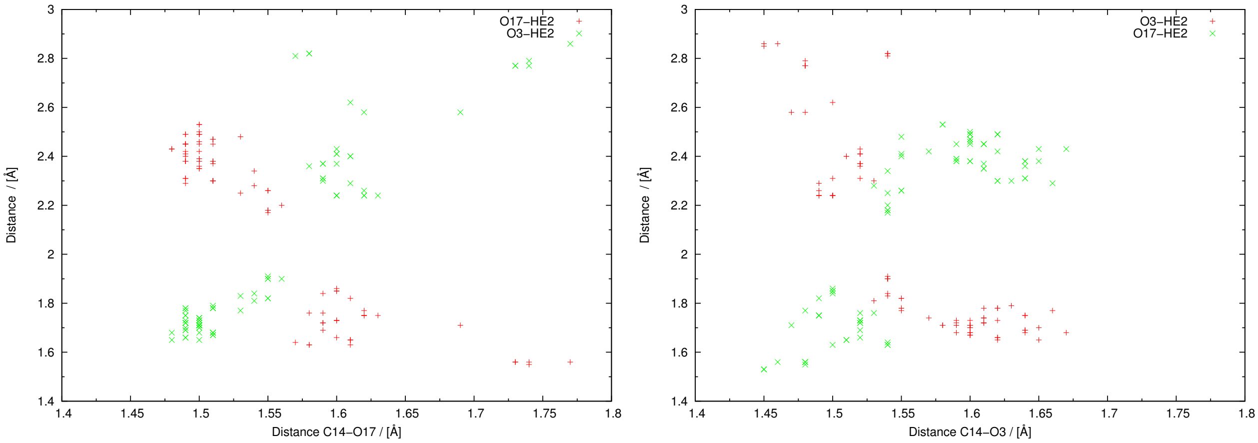


Figure 4.6: Correlation in the distributions of length of the scissile C14-O17 and C14-O3 bonds with the distance between the transferable hydrogen atom HE2 and the potential acceptor oxygen atoms O17 and O3.

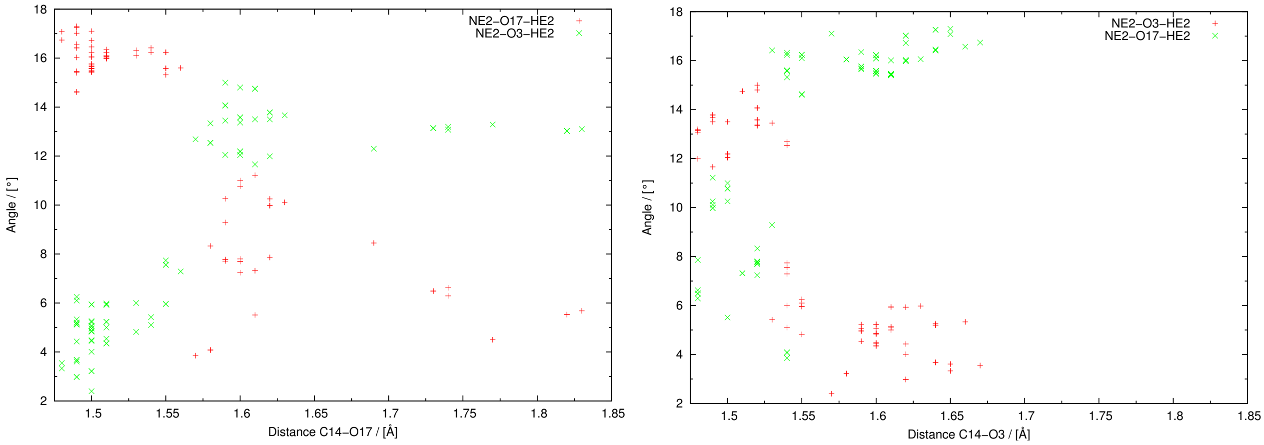


Figure 4.7: Correlation in the distributions of the length of the scissile C14-O17 and C14-O3 bonds with the bond angle formed by nitrogen NE2 of His156, O3 or O17, and hydrogen HE2.

of ester substrates. In a quantitative treatment of enantioselective ester hydrolysis reaction steps [I] and [II] must be considered[7], and one needs to explore the qualitatively different combinations of barriers of steps [I] and [II], shown in figure 4.8. Step [I] is assumed to be fully reversible, while [II] is treated as being irreversible, and therefore only the forward direction leading from TI to PDC is accounted for. In A), B) and D) the highest barrier is encountered at step [I] and the effective barrier is simply $\Delta E^{\ddagger}_{\text{Acylation}} = \Delta E^{\ddagger}_{\text{MCC}\rightarrow\text{TS}}$.Following Yagisawa[230], the rate determining barrier in C) is found at [II], and the effective acylation barrier is $\Delta E^{\ddagger}_{\text{Acylation=}}\Delta E^{\ddagger}_{\text{MCC}\rightarrow\text{TS[I]}} + \Delta E^{\ddagger}_{\text{TI}\rightarrow\text{TS[II]}} - \Delta E^{\ddagger}_{\text{TI}\rightarrow\text{TS[I]}}$. MM models assume that the TS can be approximated by the TI, which renders them incapable of reproducing the details of the profile.

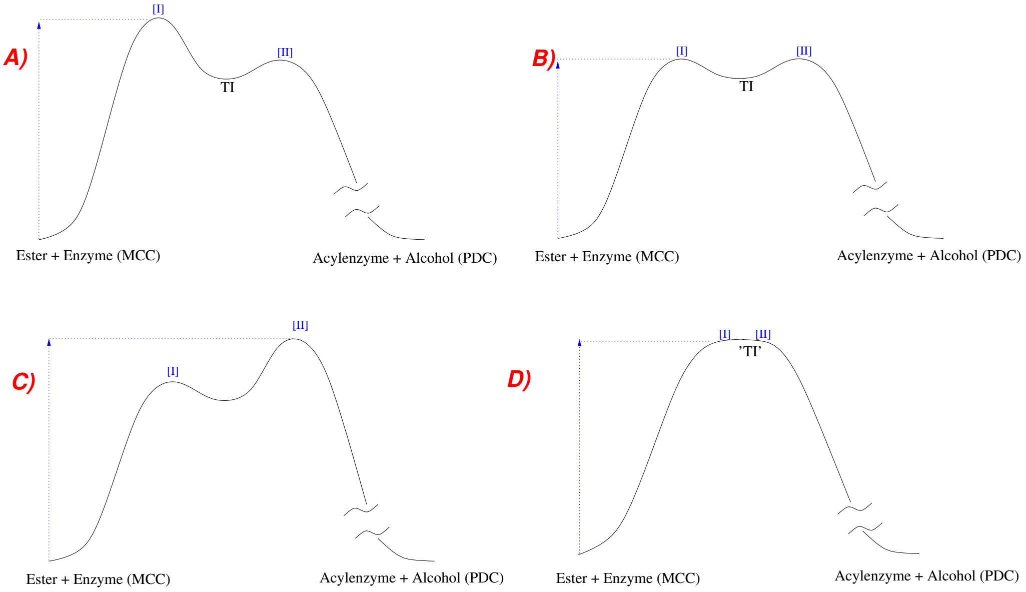


Figure 4.8: Possible shapes of potential surfaces for the enantiodiscriminative steps of ester cleavage in lipases. The enantiomeric fragment of the substrate is the alcohol, so that the enantioselectivity is governed by the acylation reaction.

Quantitative Kinetics

In discussions of product distributions of complex reactions qualitative arguments using effective (rate-determining) barriers are often applied. In this spirit the enantioselectivity of BSLA with the NAPHAC substrate can be evaluated as the quotient of reaction rates, which are computed from the rate determining barriers of the acylation reaction with each enantiomer using the Eyring equation. This amounts to calculating $\mathbf{E} = \frac{k_1}{k_2}$ (see also chapter 1, and the appendices D and E on simulation of chemical kinetics and kinetic resolution).

---

[7] When the chiral moiety of the ester is the *acid*, both steps of the deacylation reaction must be considered in addition.

Relying on a qualitative picture of the kinetics can be dangerous. Inaccuracies arise for example due to omission of the back-reactions. For a quantitatve treatment of the kinetics it is mandatory to include all steps and reaction barriers in a mathematical model of the chemical system. This includes faithful reproduction of the physicochemical conditions of the experiment in order to enhance comparability of theoretical and experimental data. This implies that a kinetic resolution experiment should be modeled. The mathematical model which was used for data analysis in this thesis is described in the appendix D on simulation of chemical kinetics. Where appropriate, the enantioselectivity computed from this model is compared to that obtained by the qualitative approach.

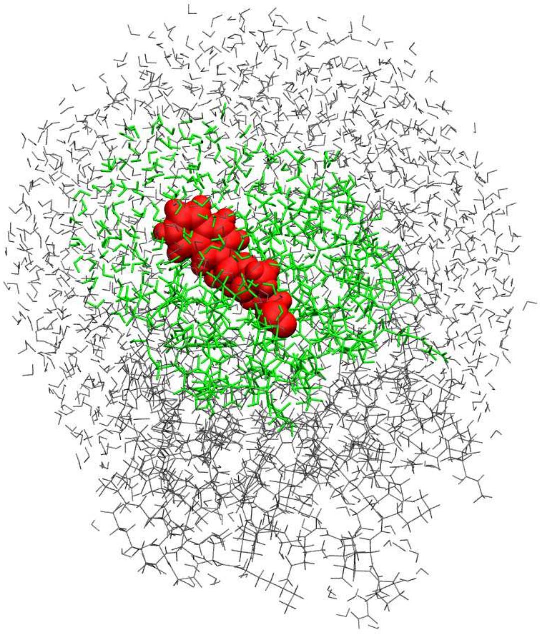


Figure 4.9: Size of the optimised region in the QM/MM reaction path studies. All atoms that were mobile are coloured, the TI is coloured red. Note that the QM region is aspherical, since in addition to a 12.5 Å spherical selection around the origin we selected all residues containing atoms within 9 Å of any atom of the red set to be active.

One Cycle of the Acylation Reaction

We now describe the geometrical changes that accompany steps [I] and [II]. The example is taken from RIII800‡, and is a prototypical case study.

The full cycle is displayed in figure 4.10. We show only those atoms that belong to the reactive core or are indispensable for the reaction. The reaction starts at A in the upper left corner of the figure where the MCC of NAPHAC with BSLA is shown. From this position the nucleophilic attack by serine oxygen O17 (atom labels displayed in E) on carbonyl carbon C14 is initiated. The carbonyl group around C14 is planar and the distance between O17 and C14 amounts to 2.70 Å. NAPHAC is held in position by a hydrogen bond between O15 and the backbone amide group of Ile12. Hydrogen HE2 is bound to O17, the distance between HE2 and NE2 suggests a very strong hydrogen bond, and the small angle NE2-O17-HE2 of 7.5° supports this interpretation. At the TS of step [I] (B) the C14-O15 distance is 2.05 Å, and thus noticeably reduced compared to MCC. HE2 is shared between NE2 and O17, and is in our example somewhat closer to NE2; in other cases HE2 is found to be shared evenly between both atoms. Although this is not obvious in figure4.10B the carbonyl group of C14 is not planar anymore, but on its way to sp^3 hybridisation. As the ester is being pulled into the active site the hydrogen bond to Met78 shortens and gets stronger, while that to Ile12 remains approximately equidistant. The TI is depicted in figure4.10C, and can be seen as a continuation of what has become apparent in the TS. A covalent bond is clearly established between C14 and O17. HE2 is now bound to NE2, but remains in hydrogen-bonding distance to O17, which is why the C14-O17 bond is elongated compared to C14-O3 (*vide supra*). Our MD studies suggest that His156 is mobile and shuttles between O3 and O17, so that the TI, which is depicted here, must be seen as only a single realisation of a class of structures. The TS of step [II] (figure4.10D) is quite similar to that of step [I], after substitution of O3 for O17, as a comparison of B and D confirms. The PDC (E) shows an esterified enzyme, named acylenzyme, which is in complex with the leaving naphthylethanol of the cleaved NAPHAC molecule. The hydrogen bonds of O15 withIle12 and Met78 in the oxyanion hole are of equal strength. The C14-O3 distance at 2.69Å is comparable to the C14-O17 distance in A. A large angle NE2-O3-HE2 (15.4) suggests a weak hydrogen bond between NE2 and the leaving alcohol. This makes sense as the enzyme should be trimmed for high throughput of substrate and not be inhibited by product.

As the visualisation of figure 4.10 already suggests, there are no large conformational changes in the reacting system which would involve displacements of the torsion angles ω_1-ω_4. This observation holds for the entire range of optimisations. The differences in the torsion angles of TI, TS, and MCC or PDC (read from tables 4.11 and 4.12) reflect, to a large degree, the local changes in hybridisation occuring at C14.

Adaption of the surrounding to the reactive core is driven largely by the need to fulfill steric requirements arising from the progress of the reaction. Reorganisations of the local hydrogen bond network were not observed, but in some cases discontinuities (*vide infra*) of the PES were found that could be attributed to jumps of single water molecules. This happens only rarely, however, as the optimisations tend to drive the system to the closest local minimum without large-scale rearrangements. Simulation approaches which are described in the next chapter allow such rearrangements and improve relaxation.

Determination of Reaction Barriers

The PES computed during the validation study is very smooth (see figure 4.3), and it was hoped that all reaction paths would exhibit equally benign behaviour. Unfortunately this was not the case, and a larger proportion of systems exhibit at least some roughness, if not jumps, in the calculated PESs. The source of such noise is investigated in more detail in a later section. We observed cases where a TS could not be localised either because (a) the acylation profile was of type D (figure 4.8) and thus the TI is not stabilised at all, or (b) the surface was rippled which interfered with an unambiguous determination of a TS. Therefore, a policy was needed to cope these special cases.

When a TS could not be located as a result of the roughness of the PES, but the TI still appeared as a local minimum, we used the highest point on the surface as a replacement for the energy of the TS. When this replacement TS was found in the RC interval between -1.5 Å and -2.5 Å the geometric parameters specified in the tables correspond to this geometry. Otherwise the parameters are from the structure closest to -1.8 Å on the RC, which is the location where genuine TSs are usually found in our system (for example in RVS900, SI800, SI900).

When the energy was monotonically decreasing during the scan, i.e., the TI was not a local minimum, we specify the geometrical parameter obtained at an RC value of -1.8 Å. The barrier $\mathrm{MCC} \rightarrow \mathrm{TS}$ was then taken to be the energy difference between MCC and TI, and the value for $\mathrm{TI} \rightarrow \mathrm{TS}$ was consequently set to zero (marked as '$-$' in the tables).

4.2.3 Acylation: General Observations

The gemetrical results presented in this section and the remainder of the chapter refer to values tabulated in tables 4.9 and 4.10 for bond-lengths and angles, and tables 4.11 and 4.12 for the torsions ω_1-ω_4 of the TI. Potential energy barriers are listed in tables 4.7 and 4.8. In the following we first report on characteristic features of reaction paths obtained for each of the different binding modes before we enter a general discussion. Note that the geometric parameters of the TI for 'F' scans must be the same in the tables for steps [I] and [II] with identical binding modes and snapshots, since the same starting geometry was used. An exception are structures derived from TS searches, and all TIs derived in 'B' scans.

We first discuss the computed barriers. The QM contribution stabilises the TI and thus

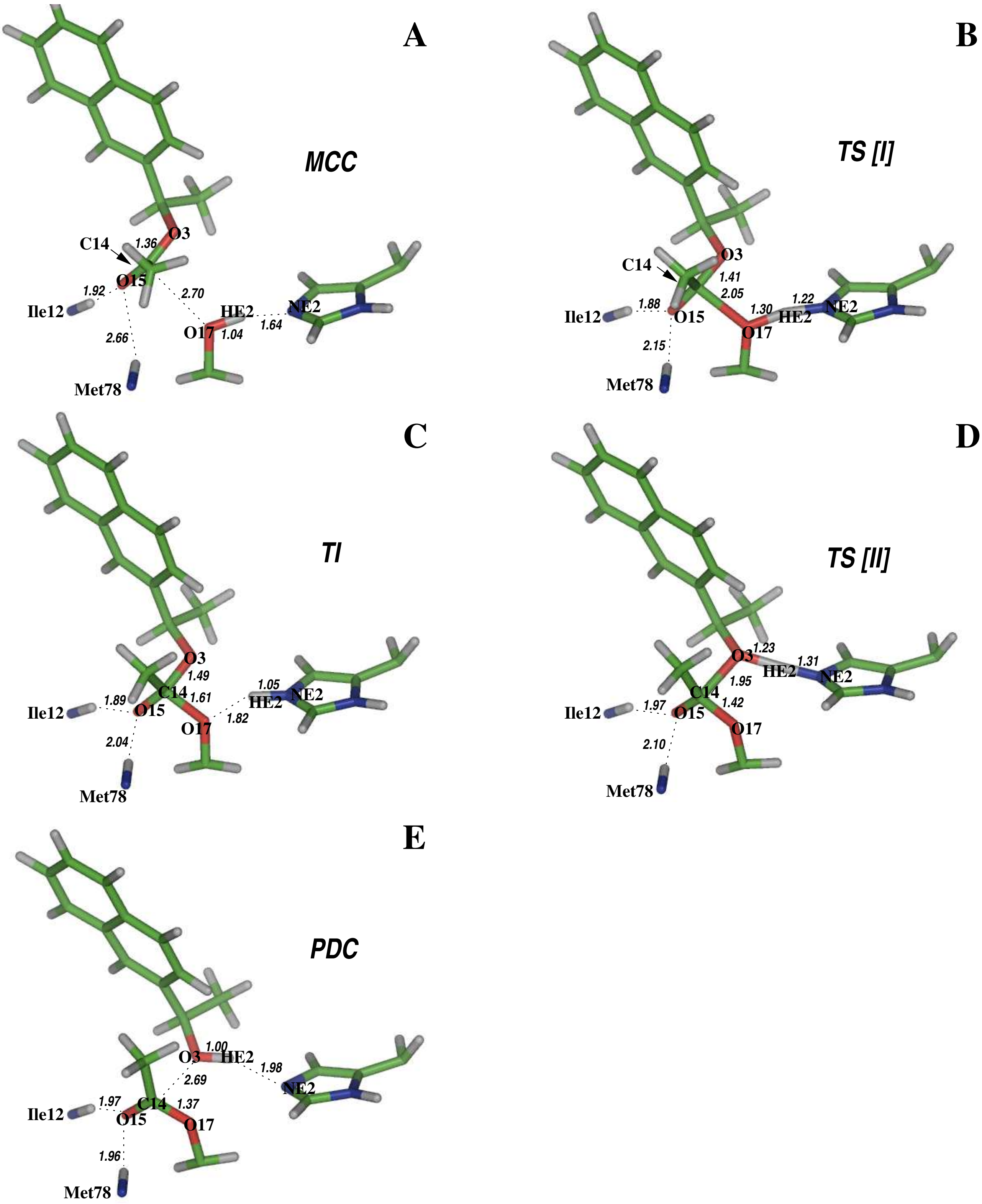


Figure 4.10: One cycle of the acylation reaction of a lipase comprising steps [I] and [II]. The MCC is shown in A. Nucleophilic attack of O17 on C14 leads via TS I (displayed in B) to the TI (figure C). Transfer of hydrogen HE2 to O3 results in decomposition of the TI via TS II (displayed in D) to formation of a PDC (acylenzyme + alcohol).

			RC Values			$\Delta E_{QM/MM}$		ΔE_{QM}		ΔE_{MM}	
			TI	TS	MCC	TI → TS	MCC → TS	TI → TS	MCC → TS	TI → TS	MCC → TS
RIII	500	F	-0.34	-1.84	-3.34	1.65	9.55	3.95	6.84	-2.30	2.71
		B	-0.34	-1.71	-3.34	3.09	8.69	1.44	5.21	1.64	3.49
	600	F	-0.9	-	-3.2	-	9.13	-	1.12	-	8.01
		B	-0.93	-1.68	-3.30	1.00	7.83	1.92	4.91	-0.93	2.92
	700	F	-0.80	-1.68	-3.55	2.00	7.58	3.91	5.36	-1.92	2.22
		B	-0.80	-1.80	-3.43	1.74	6.53	1.64	4.04	0.09	2.49
	800	‡	-0.9	-2.0	-3.3	3.28	8.34	6.77	5.24	-3.49	3.10
	1000	F	-0.4	-	-3.3	-	9.70	-	7.57	-	2.13
		B	-0.4	-	-3.3	-	9.41	-	7.29	-	2.11
RIV	500	F	-0.08	-1.70	-3.20	2.01	15.08	1.84	13.79	0.16	1.29
		B	-0.1	-	-3.2	-	13.48	-	11.42	-	2.06
	600	F	-1.04	-1.54	-3.29	0.68	16.13	2.01	10.83	-1.33	5.30
		B	-1.0	-	-3.3	-	16.25	-	13.45	-	2.80
	700	F	-0.1	-	-3.3	-	18.16	-	9.94	-	8.22
		B	-0.1	-	-3.2	-	12.78	-	8.38	-	4.40
	1000	F	-0.1	-	-4.2	-	13.10	-	4.04	-	9.06
		B	-0.1	-	-3.2	-	12.10	-	7.59	-	4.51
RV	500	F	-1.02	-2.02	-3.27	3.48	17.00	7.37	6.46	-3.89	10.54
		B	-0.9	-	-3.3	-	13.90	-	10.63	-	3.28
	600	F	-0.1	-	-3.2	-	12.14	-	9.12	-	3.02
	700	F	-0.0	-	-3.4	-	15.56	-	14.13	-	1.43
	900	‡	-1.1	-1.8	-3.2	1.56	6.02	3.97	6.31	-2.41	-0.29
	900 †	‡	-1.0	-1.9	-3.3	2.40	5.70	2.70	9.60	-0.30	-3.90
	1000	F	-0.2	-	-3.2	-	15.16	-	10.35	-	4.80
SI	500	F	-0.93	-1.81	-3.31	1.35	12.82	3.80	8.98	-2.45	3.84
		B	-0.93	-1.68	-3.31	1.45	10.04	2.72	8.65	-1.27	1.39
	600	F	-0.18	-1.93	-3.80	2.20	6.73	3.25	8.65	-1.04	-1.93
	700	F	-0.24	-1.99	-3.37	2.28	13.08	4.12	8.02	-1.83	5.06
		B	-0.12	-1.87	-3.24	3.00	9.49	3.02	9.08	-0.02	0.40
	800	‡	-1.0	-1.9	-3.0	2.58	3.44	4.97	2.45	-2.39	0.99
	900	‡	-1.0	-1.8	-3.1	2.40	6.00	3.65	6.79	-1.25	-0.79
	1000	F	-0.94	-1.94	-4.08	2.64	8.14	5.90	7.86	-3.26	0.28
SII	500	F	-1.4	-	-3.2	-	6.23	-	2.15	-	4.07
		B	-1.4	-	-3.2	-	6.14	-	1.93	-	4.21
	600	F	-1.3	-	-3.3	-	10.77	-	3.46	-	7.31
		B	-1.3	-	-3.3	-	8.54	-	2.77	-	5.77
SIV	500	F	-0.14	-1.64	-3.39	1.54	10.36	3.12	9.85	-1.59	0.51
		B	-0.14	-1.51	-3.39	2.20	9.92	1.84	8.45	0.36	1.47
	600	F	-0.22	-1.97	-3.22	3.68	8.51	5.10	7.68	-1.42	0.83
		B	-0.22	-1.59	-3.34	1.74	10.17	0.01	10.68	1.73	-0.51
	700	F	-0.3	-	-3.4	-	14.09	-	9.65	-	4.44
		B	-0.27	-1.52	-3.40	0.55	12.54	0.85	10.41	-0.30	2.13
	1000	F	-0.2	-	-3.2	-	13.30	-	4.23	-	9.07
		B	-0.17	-1.80	-3.17	3.69	6.85	3.53	6.66	0.16	0.20

† With QM partition QMII, using the preceeding TS geometry as input for optimisation. Published in [72].

Table 4.7: Energy barriers for reaction step [I]. All values are extracted from PES scans, except those marked as ‡, which are derived from direct searches for a TS. All RC values in , Å energies in kcal/mol.

inreases the barrier TI → TS. The interaction of the QM-atoms with the charges in the MM region is the main cause for the large stabilisation, according to electrostatic perturbation calculations (see chapter 4.2.5). The MM contribution counteracts the QM energies and thus destabilises the TI. This can be concluded from slope complementarity during steps TI → TS

			RC Values			$\Delta E_{QM/MM}$		ΔE_{QM}		ΔE_{MM}	
			TI	TS	PDC	TI → TS	PDC → TS	TI → TS	PDC → TS	TI → TS	PDC → TS
RIII	500	F	-0.77	-2.02	-3.77	5.77	15.67	10.94	12.63	-5.17	3.05
		B	-0.77	-1.65	-3.65	4.68	12.51	8.23	10.11	-3.55	2.41
	600	F	-0.30	-1.93	-3.68	4.62	13.22	9.54	10.04	-4.92	3.18
		B	-0.30	-1.68	-3.68	5.50	14.16	11.79	11.29	-6.29	2.86
	700	F	-0.30	-2.05	-4.18	7.27	12.17	12.31	9.69	-5.04	2.48
		B	-0.30	-1.93	-3.55	5.44	9.73	8.57	7.24	-3.14	2.50
	800	‡	-0.8	-2.0	-3.7	7.60	12.18	11.12	9.75	-3.52	2.44
	900	‡	-0.8	-1.9	-3.7	4.97	13.07	8.86	9.61	-3.88	3.46
	1000	F	-0.69	-1.82	-4.19	4.32	17.72	7.56	9.02	-3.24	8.70
		B	-0.69	-1.82	-3.69	5.21	12.26	9.01	10.61	-3.80	1.65
RIV	500	F	-1.0	-	-3.7	-	12.73	-	6.81	-	5.92
	600	F	-0.18	-1.81	-4.06	2.66	13.20	7.51	9.07	-4.85	4.13
	700	F	-1.0	-	-3.3	-	11.14	-	5.37	-	5.76
		B	-0.96	-1.83	-3.33	2.52	8.24	5.60	7.69	-3.08	0.56
	1000	F	-0.98	-1.85	-4.35	1.95	13.39	5.25	5.43	-3.30	7.96
		B	-1.0	-1.9	-3.5	2.04	10.85	5.13	8.73	-3.09	2.12
RV	500	F	0.20	-1.55	-4.55	6.93	13.93	13.40	11.96	-6.48	1.97
		B	0.20	-2.17	-4.42	7.44	13.12	13.32	11.71	-5.89	1.40
	600	F	-1.0	-1.8	-3.5	1.85	8.74	4.46	7.23	-2.61	1.51
		B	-1.0	-2.0	-3.2	2.26	7.41	4.78	7.49	-2.51	-0.09
	700	F	-0.94	-1.81	-3.56	2.52	10.64	4.57	10.66	-2.06	-0.02
	900	‡	-0.9	-1.8	-3.3	2.84	9.84	6.12	8.12	-3.28	1.72
	1000	F	-1.00	-1.75	-3.87	1.65	12.76	3.88	11.41	-2.23	1.35
SI	500	F	-0.2	-	-4.5	-	21.18	-	14.14	-	7.05
		B	-0.29	-1.67	-4.42	1.42	18.54	5.42	16.86	-4.00	1.67
	600	F	-1.02	-1.64	-4.39	0.76	16.47	3.13	14.04	-2.37	2.44
		B	-1.0	-1.4	-4.3	0.62	15.49	2.33	13.39	-1.70	2.10
	700	F	-0.95	-1.57	-4.45	0.40	22.82	3.28	13.33	-2.87	9.49
		B	-1.0	-	-4.3	-	16.96	-	11.80	-	5.16
	800	‡	-1.1	-1.6	-3.8	0.82	15.30	2.38	10.43	-1.56	-4.87
	1000	F	-0.2	-	-4.2	-	15.09	-	8.57	-	6.52
		B	-0.2	-	-4.0	-	16.42	-	10.46	-	5.96
SII	500	F	0.3	-	-3.6	4.87	24.60	7.89	12.43	-3.02	12.17
		B	0.3	-	-3.6	3.08	18.73	7.64	12.58	-4.56	6.15
	600	F	0.2	-	-4.1	3.33	21.35	6.73	14.81	-3.40	6.55
		B	0.2	-0.8	-3.8	4.01	19.54	6.28	14.15	-2.27	5.39
SIV	500	F	-0.82	-1.57	-5.07	1.26	16.99	3.32	13.59	-2.06	3.40
	600	F	-0.9	-	-3.6	-	15.59	-	3.60	-	12.00
		B	-0.93	-1.55	-3.55	0.57	12.11	3.07	9.77	-2.49	2.34
	700	F	-0.9	-	-3.7	-	21.57	-	8.76	-	12.81
		B	-0.90	-1.65	-3.65	0.67	15.81	2.59	12.00	-1.93	3.81
	1000	F	-0.9	-	-4.4	-	25.24	-	10.61	-	14.62
		B	-0.93	-1.68	-4.31	1.41	14.96	3.91	16.81	-2.50	-1.85

Table 4.8: Energy barriers for reaction step [II]. All values are extracted from PES scans, except those marked as ‡, which are derived from direct searches for a TS. All RC values in , Å energies in kcal/mol.

(also noticed from tables 4.7 and 4.8). Steric repulsion and strain with the MM environment keep the stabilisation of the TI at bay, which seems sensible, since the intermediate must not be overstabilised for a resonable conversion of reactants. For the half-steps MCC → TS and PDC → TS QM and MM contributions act in the same direction, and thus enlarge the

acylation barrier. On average, the MM contribution to the QM/MM energy is well below the QM contribution in both steps [I] and [II]. An exception to this rule is step [I] in binding mode SI, where all MCC → TS barriers are dominated by the MM contribution.
Barriers calculated from TS optimisations sometimes yield atypically low energy barriers, for example in RV900‡ , SI800‡, and SI900‡ (only [I]). In all three cases typical barriers calculated from the energy paths are several kcal/mol higher. From RIII800‡, on the contrary, a barrier typical for the RIII binding mode is computed. A possible reason for the shallow barriers in the indicated cases is that the energy minimisations from the perturbed TS exit in a plateau region of the PES.

Character Of Binding Modes

RI and SIII. Both binding modes belong to the class of *nonproductive* conformations, since the configuration of the bond around O3 is shuch that prevents efficient ester hydrolysis. It was not possible to obtain a stable TI for any of these binding modes. **RII.** All snapshots except RII700 yielded unstable TI geometries. Therefore geometries of RII were not pursued any further in reaction path studies. Free energy calculations (chapter 5) show that the acylation barrier is too high to be competitive with more efficient binding modes. **RIII.** A water molecule is located below the TI and forms a third hydrogen bond to the oxyanion O15. In the reaction path optimisation for step [I] the water stays coordinated to O15, while [II] it becomes coordinated to O3 in [II]. This latter change in coordination leads to a discontinuity in the PES (figure 4.11). In the back scans of step [II] the discontinuities were found closer to the TI; a hysteresis effect. This emphasizes the need to account for realistic relaxation of the solvent environment, e.g. by MD simulations (see chapter 5). The energy surfaces of RIII belong to class C) in figure 4.8. The barriers for step [II] are the highest of the entire set of binding modes, well above 4 kcal/mol, in RIII700 and RIII800 even above 7 kcal/mol. **RIV.** The potential surfaces are largely discontinous for step [I] and the stabilisation of the TI is at most 2 kcal/mol. The shape of the barriers is flat and not very well defined, which is why an assignment of TSs is hardly possible. Attempts to locate a TS directly, from RIV800 and RIV900, failed. Like RIII the acylation barrier is of type C) in figure 4.8. **RV.** The shapes of barriers are similar to RIV. 'B' scans, and scans of step [II] are significantly smoother than 'F' scans of [I]. **SI.** The profiles are somewhat rugged for step [I] in the 'F' direction, but rather smooth for the 'B' scans. The contribution of the MM part to the QM/MM energy is reduced significantly for [I] in the backward direction and is even almost constant. In TI → TS of step [II] the TI is hardly stabilised, mainly due to the QM stabilisation energy being the lowest for [II], together with SIV. The 'B' scans of step [II] did not always provide a

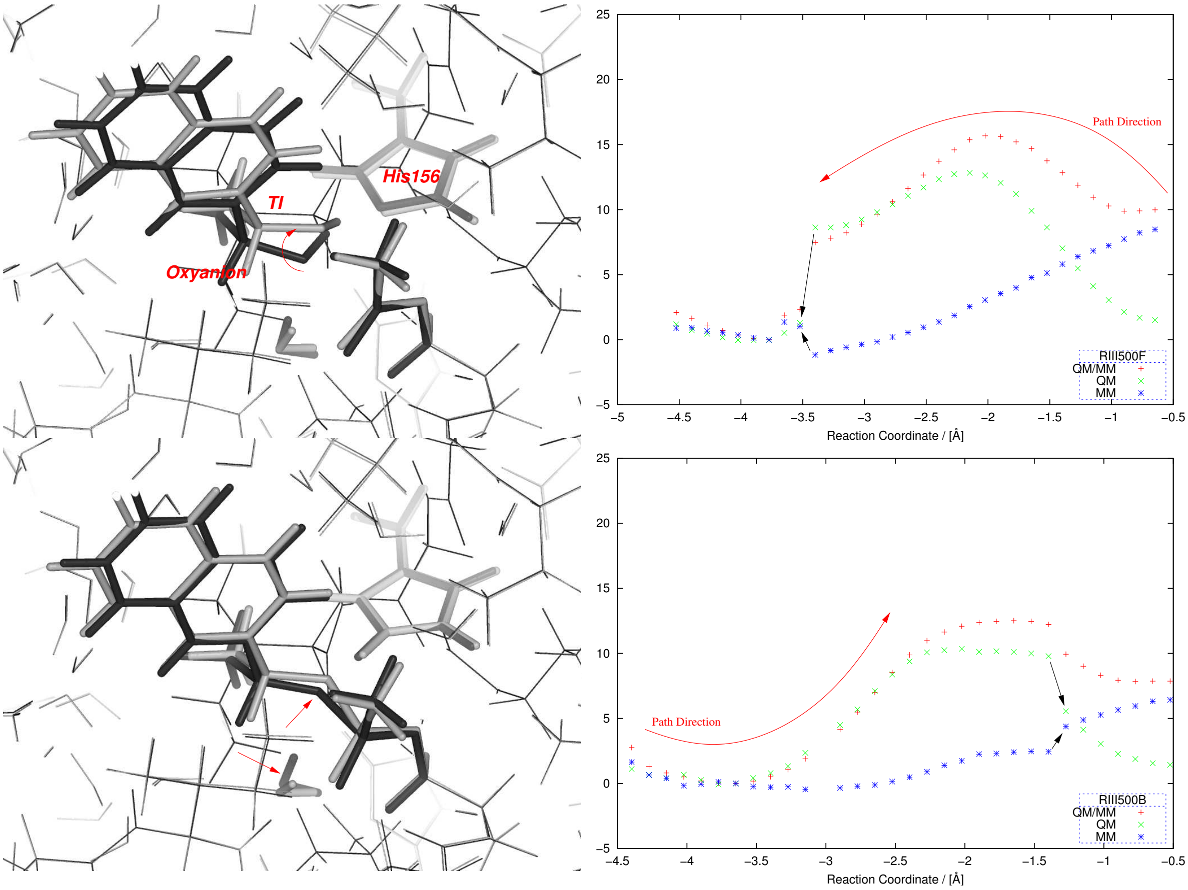


Figure 4.11: Example of PES fragmentation in step [II] of RIII500F (top) and RIII500B (bottom). In the example at the top, fragmentation occurs due to abrupt rotation about the functional group of the leaving alcohol close to the end of the reaction path. The reaction path at the bottom was calculated from the last optimised structure of the upper path, its directionality is therefore reversed. It is smooth in the place where the former is not, but features a little discontinuity again close the end. Now there is little rotation about the alcohol group, but more rotation of a nearby water molecule.

smooth PES (see figure 4.12). The shape of the PES of acylation is as seen in A) of figure 4.8. **SII.** With binding mode SII, only configurations SII500 and SII600 produced a stable TI. The values of the RC at the TI in step [I] are the lowest recorded for any binding mode, and mark the geometries of SII as being the most advanced in direction of decomposition of the TI. Consequently, in step [II], therefore, the recorded values of the RC are the *largest* obtained from any snapshot. The barriers for decomposition via reaction channel [II] are found to be above 3 kcal/mol, on average, and hence the second largest after RIII . The acylation profile is thus of type C). **SIV.** This binding mode is similar to SI, it differs mainly by a rotation of the naphthyl ring of about 180°, and therefore delivers results which are very similar to SI.

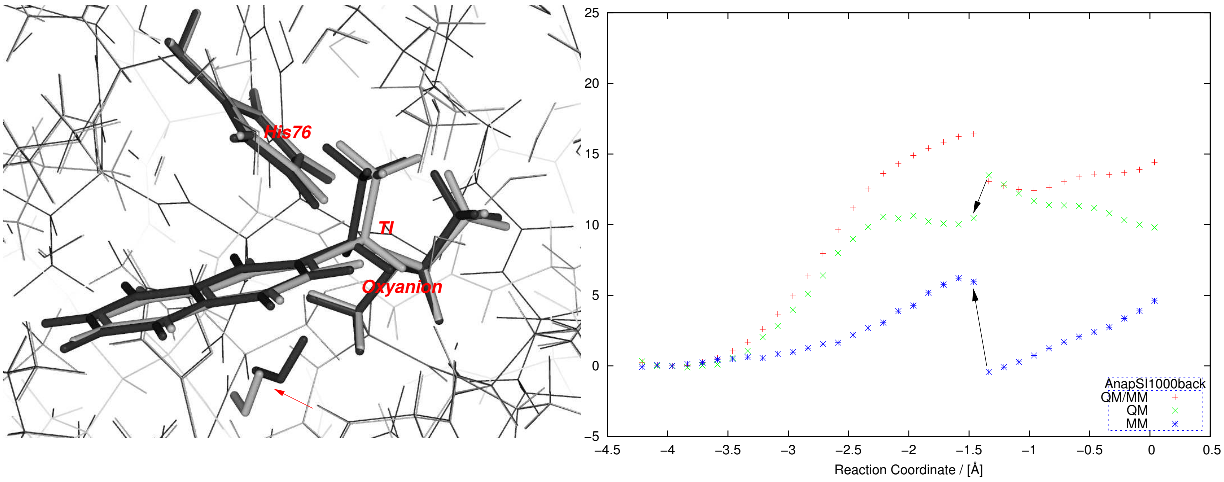


Figure 4.12: Example of PES fragmentation, the discontinuity occurs as a water molecule shifts position from one optimisation to the next. Example taken from reaction path scan of step [II] of SI600B.

4.2.4 The Tightness of Binding

In a QM/MM investigation of acetylcholinesterase an empirical descriptor was found to correlate well with the barrier height[236]. This descriptor was defined as a sum of key distances between key active site residues and the ligand, and was considered to describe the 'tightness of binding'. Lower potential energy barriers for $\mathrm{MCC} \rightarrow \mathrm{TS}$ were found for the substrate that was more tightly bound in the MCC. Since our reaction path optimisations proceeded in the reverse order of the cited publication[236], from the TI towards the MCC, it seemed more sensible to evaluate the tightness of binding at the TI. We used the distances Ile12-O15, Met78-O15, C14-O17, and O17-HE2 in an attempt to find correlations of the $\mathrm{MCC} \rightarrow \mathrm{TS}$ barrier height with geometric parameters of the TI.

The result is shown in form of a diagram in figure 4.13, the tightness of binding is the sum of the foregoing key distances. The shape of the graph does not reveal any pattern, least of all correlation. Investigation of the barrier of $\mathrm{TI} \rightarrow \mathrm{TS}$ in the same fashion did not produce any interesting pattern, either. The same is true for the individual distances, Ile12-O15, Met78-O15, C14-O17, and O17-HE2, which show no sign of correlation with the barrier heights.

These findings imply that the barriers for the various snapshots are rather insensitive to the exact geometry of the oxyanion hole (Ile12-O15, Met78-O15) and the orientation of His156 (O17-HE2).

			TI	TS	MCC																								
			O17-HE2			O3-HE2			NE2-HE2			C14-O17			C14-O3			Ile12-O15			Met78-O15			NE2-O17-HE2			NE2-O3-HE2		
RIII	500	F	2.26	1.37	1.02	1.82	2.67	3.34	1.05	1.17	1.68	1.55	2.04	2.68	1.55	1.41	1.35	1.85	1.82	1.82	2.08	2.25	2.71	16.2	7.0	8.1	6.0	16.5	20.9
		B	2.25	1.51	1.04	1.83	3.00	3.57	1.05	1.10	1.62	1.53	2.13	2.75	1.54	1.39	1.35	1.85	1.81	1.83	2.09	2.34	2.86	16.3	6.9	6.3	6.0	12.9	19.4
	600	F	1.75	1.39	1.03	2.24	2.65	3.02	1.06	1.16	1.66	1.62	2.03	2.55	1.49	1.41	1.35	1.85	1.85	1.83	2.08	2.20	2.45	10.0	6.8	10.1	13.8	16.5	22.1
		B	1.75	1.44	1.02	2.26	2.73	3.17	1.06	1.13	1.69	1.62	1.99	2.63	1.49	1.41	1.35	1.86	1.85	1.84	2.09	2.21	2.53	10.3	7.2	11.0	13.5	14.9	21.6
	700	F	1.85	1.45	1.01	2.24	2.69	3.52	1.05	1.14	1.74	1.60	1.99	2.83	1.50	1.42	1.35	1.87	1.85	1.83	2.11	2.22	2.71	10.8	7.1	10.2	12.2	15.2	20.3
		B	1.84	1.48	1.02	2.31	3.11	3.51	1.05	1.12	1.69	1.59	2.17	2.76	1.50	1.38	1.34	1.88	1.84	1.83	2.15	2.32	2.71	10.3	6.1	9.3	12.1	12.6	19.8
	800	‡	1.82	1.30	1.04	2.29	2.68	3.32	1.05	1.22	1.64	1.61	2.05	2.70	1.49	1.41	1.36	1.89	1.88	1.92	2.04	2.15	2.66	11.2	7.5	7.5	11.7	16.8	19.1
	1000	F	2.18	1.42	1.02	1.90	2.78	3.22	1.05	1.15	1.69	1.55	2.07	2.64	1.54	1.40	1.35	1.92	1.88	1.85	2.01	2.16	2.44	15.6	6.5	9.3	7.6	15.1	20.7
		B	2.17	1.45	1.03	1.91	2.99	3.28	1.05	1.13	1.67	1.55	2.12	2.65	1.54	1.39	1.35	1.92	1.87	1.85	2.02	2.20	2.46	15.3	6.1	9.4	7.7	13.4	20.4
RIV	500	F	2.49	1.35	1.02	1.66	2.48	2.95	1.07	1.20	1.71	1.49	1.86	2.51	1.62	1.45	1.35	1.86	1.84	1.78	1.89	1.99	2.18	17.0	7.0	10.7	3.0	18.0	22.0
		B	2.49	1.41	1.02	1.65	2.89	2.95	1.07	1.16	1.71	1.50	2.08	2.51	1.62	1.39	1.35	1.85	1.81	1.78	1.93	1.98	2.16	16.7	7.1	11.4	4.0	13.7	21.7
	600	F	1.65	1.43	1.01	2.40	2.54	3.13	1.07	1.15	1.71	1.61	1.82	2.59	1.51	1.45	1.35	1.86	1.87	1.82	1.91	1.96	2.20	7.3	6.0	9.5	14.7	16.7	21.8
		B	1.71	1.42	1.02	2.58	2.96	3.13	1.06	1.14	1.72	1.69	2.06	2.58	1.47	1.39	1.35	1.86	1.84	1.81	1.91	1.98	2.20	8.4	6.1	10.5	12.3	13.6	21.8
	700	F	2.49	1.43	1.02	1.71	2.95	3.04	1.06	1.15	1.73	1.50	2.10	2.61	1.60	1.38	1.35	1.89	1.86	1.81	1.89	1.97	2.21	15.6	5.6	9.5	4.8	14.0	22.3
		B	2.49	1.45	1.03	1.70	2.99	3.02	1.07	1.14	1.69	1.50	2.13	2.54	1.60	1.38	1.35	1.90	1.85	1.82	1.90	1.99	2.17	15.6	5.8	8.9	5.0	13.4	21.5
	1000	F	2.47	1.41	0.99	1.68	2.91	3.46	1.07	1.15	2.00	1.51	2.11	3.22	1.60	1.38	1.35	1.83	1.79	1.79	1.90	2.01	3.06	16.2	6.1	17.1	4.3	14.5	27.9
		B	2.46	1.41	1.02	1.70	2.95	3.05	1.06	1.15	1.68	1.50	2.12	2.57	1.60	1.38	1.35	1.83	1.79	1.77	1.90	2.01	2.27	16.2	6.0	8.7	4.5	14.2	23.0
RV	500	F	1.63	1.21	1.02	2.82	2.84	3.23	1.07	1.31	1.68	1.58	1.93	2.61	1.54	1.45	1.35	1.90	1.90	1.83	1.85	1.93	2.16	4.1	4.1	8.0	12.5	16.9	20.6
		B	1.77	1.39	1.02	2.58	3.09	3.20	1.05	1.16	1.70	1.62	2.13	2.59	1.48	1.38	1.35	1.89	1.86	1.83	1.87	1.97	2.15	7.9	4.7	9.5	12.0	13.3	20.6
	600	F	2.45	1.38	1.02	1.72	2.94	3.05	1.06	1.17	1.71	1.49	2.07	2.54	1.61	1.39	1.35	1.92	1.88	1.84	1.91	1.96	2.14	15.5	6.2	9.4	5.1	13.9	21.5
	700	F	2.53	1.44	1.01	1.71	3.00	3.22	1.07	1.15	1.77	1.50	2.08	2.65	1.58	1.38	1.35	1.92	1.86	1.80	1.88	1.96	2.17	16.0	5.8	9.3	3.2	13.2	21.0
	900	‡	1.63	1.31	1.03	2.62	2.73	3.11	1.07	1.22	1.69	1.61	1.93	2.57	1.50	1.43	1.38	1.92	1.94	1.94	2.02	2.15	2.41	5.5	4.5	8.3	13.5	16.8	23.3
	1000	F	2.38	1.40	1.02	1.67	2.78	2.95	1.07	1.16	1.67	1.51	2.06	2.54	1.60	1.39	1.35	1.83	1.80	1.77	1.93	2.04	2.25	16.1	5.8	8.8	5.2	15.5	23.4
SI	500	F	1.73	1.35	1.02	2.41	2.67	3.18	1.06	1.20	1.72	1.60	1.96	2.61	1.52	1.44	1.37	1.90	1.88	1.80	1.93	2.03	2.47	7.7	5.8	8.2	13.6	16.7	22.5
		B	1.73	1.44	1.02	2.43	2.80	3.20	1.07	1.15	1.71	1.60	1.98	2.62	1.52	1.43	1.37	1.92	1.90	1.81	1.93	2.04	2.47	7.8	5.9	8.4	13.4	14.6	22.5
	600	F	2.38	1.24	1.00	1.69	2.56	3.37	1.07	1.27	1.83	1.49	1.90	2.97	1.64	1.45	1.36	1.85	1.82	1.84	1.92	2.06	3.41	17.3	5.3	12.2	3.7	19.3	25.5
	700	F	2.31	1.23	1.02	1.75	2.55	3.02	1.06	1.29	1.72	1.49	1.93	2.67	1.64	1.45	1.36	1.89	1.87	1.76	1.97	2.10	3.03	16.4	5.4	10.2	5.2	19.8	23.9
		B	2.43	1.37	1.03	1.68	2.75	3.00	1.07	1.17	1.66	1.48	2.07	2.61	1.67	1.41	1.37	1.90	1.85	1.76	2.03	2.42	3.02	16.7	5.3	9.0	3.6	16.3	23.7
	800	‡	1.66	1.27	1.03	2.37	2.58	2.77	1.07	1.25	1.64	1.60	1.95	2.40	1.52	1.44	1.39	1.88	1.86	1.80	1.95	2.05	2.32	7.2	5.4	7.6	14.8	18.9	25.2
	900	‡	1.69	1.31	1.02	2.31	2.49	2.75	1.07	1.22	1.71	1.59	1.88	2.44	1.52	1.45	1.37	1.92	1.91	1.84	1.90	1.98	2.27	7.7	5.9	10.8	15.0	18.9	25.3
	1000	F	1.72	1.28	1.00	2.37	2.58	3.56	1.06	1.25	1.88	1.59	1.96	3.19	1.52	1.43	1.35	1.91	1.89	1.78	1.96	2.08	3.41	7.8	5.9	13.6	14.1	18.7	26.0
SII	500	F	1.53	1.34	1.03	2.86	2.95	3.03	1.10	1.19	1.65	1.82	2.05	2.53	1.45	1.42	1.37	1.90	1.89	1.88	1.91	1.97	2.20	5.5	5.0	7.9	13.0	14.8	23.0
		B	1.53	1.37	1.03	2.85	2.95	3.02	1.10	1.17	1.64	1.83	2.09	2.54	1.45	1.41	1.37	1.90	1.89	1.87	1.93	2.01	2.23	5.7	4.9	8.1	13.1	14.8	23.1
	600	F	1.56	1.44	1.02	2.77	2.82	3.08	1.09	1.14	1.70	1.73	1.81	2.58	1.48	1.47	1.37	1.83	1.82	1.81	1.85	1.88	2.17	6.5	5.9	8.7	13.1	14.2	25.0
		B	1.56	1.35	1.02	2.79	2.91	3.03	1.09	1.18	1.72	1.74	2.06	2.57	1.48	1.42	1.37	1.82	1.81	1.81	1.89	1.98	2.20	6.3	5.5	9.8	13.1	15.3	25.6
SIV	500	F	2.41	1.43	1.01	1.78	2.53	3.14	1.05	1.15	1.76	1.49	1.92	2.64	1.55	1.44	1.36	1.75	1.80	1.76	2.02	2.04	2.44	14.6	7.1	10.3	6.1	16.7	22.9
		B	2.42	1.50	1.01	1.73	2.73	3.15	1.06	1.11	1.76	1.49	1.90	2.65	1.62	1.43	1.36	1.82	1.80	1.75	1.93	2.02	2.44	16.0	7.1	10.4	4.4	13.9	22.8
	600	F	2.35	1.26	1.02	1.74	2.46	2.83	1.06	1.27	1.72	1.50	1.96	2.52	1.61	1.43	1.36	1.84	1.84	1.78	1.95	2.04	2.37	15.4	5.9	9.7	5.9	20.7	25.6
		B	2.36	1.48	1.02	1.68	2.75	3.06	1.07	1.12	1.74	1.50	1.96	2.61	1.64	1.42	1.36	1.85	1.84	1.77	1.96	2.06	2.48	16.5	6.1	9.8	5.3	14.6	23.9
	700	F	2.30	1.36	1.01	1.78	2.81	3.22	1.06	1.19	1.75	1.51	2.06	2.65	1.62	1.41	1.36	1.85	1.84	1.78	1.93	2.07	2.40	16.0	6.0	9.8	5.9	15.9	22.5
		B	2.30	1.49	1.01	1.79	2.76	3.23	1.06	1.12	1.75	1.51	1.89	2.66	1.63	1.44	1.36	1.86	1.86	1.78	1.94	2.00	2.44	16.1	6.4	9.7	6.0	14.0	22.4
	1000	F	2.39	1.36	1.02	1.71	2.55	2.91	1.06	1.19	1.70	1.50	1.98	2.50	1.59	1.42	1.36	1.83	1.82	1.76	1.94	2.05	2.38	15.8	6.4	9.3	5.0	18.1	24.7
		B	2.38	1.38	1.03	1.73	2.68	2.94	1.06	1.17	1.68	1.50	2.01	2.52	1.59	1.40	1.36	1.84	1.82	1.76	1.93	2.07	2.42	15.7	6.4	9.4	5.1	16.5	24.5

Table 4.9: Values of geometric parameters during step [**I**] of the acylation reaction. The values for each parameter are given in the order (TI,TS,MCC).

			TI	TS	PDC																								
			O17-HE2			O3-HE2			NE2-HE2			C14-O17			C14-O3			Ile12-O15			Met78-O15			NE2-O17-HE2			NE2-O3-HE2		
RIII	500	F	2.26	2.66	3.10	1.82	1.23	1.00	1.05	1.31	1.99	1.55	1.42	1.37	1.55	1.94	2.79	1.85	1.90	1.97	2.08	2.14	1.89	16.2	22.2	33.9	6.0	3.2	11.9
		B	2.28	2.71	3.06	1.81	1.55	1.01	1.05	1.11	1.93	1.54	1.42	1.37	1.53	2.09	2.73	1.85	1.87	1.95	2.06	1.88	1.89	16.4	18.0	32.9	5.4	1.7	11.0
	600	F	1.75	2.62	3.14	2.24	1.26	1.00	1.06	1.29	1.97	1.62	1.43	1.37	1.49	1.89	2.71	1.85	1.92	1.98	2.08	2.10	1.92	10.0	21.8	32.6	13.8	3.7	13.1
		B	1.75	2.79	3.10	2.24	1.52	1.00	1.06	1.12	2.00	1.63	1.42	1.37	1.49	2.08	2.67	1.86	1.90	1.97	2.06	1.92	1.93	10.1	17.2	33.9	13.7	0.6	14.6
	700	F	1.85	2.70	3.22	2.24	1.23	0.99	1.05	1.32	2.35	1.60	1.42	1.36	1.50	1.97	2.81	1.87	1.95	1.96	2.11	2.18	1.93	10.8	21.7	40.6	12.2	3.4	23.5
		B	1.86	2.90	3.06	2.24	1.47	1.01	1.05	1.14	1.93	1.60	1.40	1.37	1.50	2.26	2.63	1.87	1.92	1.95	2.13	1.94	1.93	11.0	17.6	32.7	12.0	1.5	12.7
	800	‡	2.34	2.68	3.09	1.84	1.23	1.00	1.05	1.31	1.98	1.54	1.42	1.37	1.54	1.95	2.69	1.91	1.97	1.97	2.03	2.10	1.96	16.2	22.0	33.5	5.1	3.8	15.4
	900	‡	2.48	2.68	3.13	1.77	1.29	1.00	1.06	1.26	1.95	1.53	1.43	1.37	1.55	1.92	2.72	1.90	1.94	1.95	2.00	2.07	1.93	16.1	20.6	32.1	4.8	3.7	12.9
	1000	F	2.18	2.64	3.22	1.90	1.32	0.98	1.05	1.23	2.45	1.55	1.44	1.36	1.54	1.91	2.72	1.92	1.96	1.96	2.01	2.05	1.94	15.6	20.6	42.4	7.6	3.3	29.7
		B	2.20	2.81	3.10	1.90	1.48	1.00	1.05	1.14	2.04	1.56	1.42	1.37	1.54	2.15	2.65	1.92	1.94	1.97	2.02	1.91	1.92	15.6	17.8	34.6	7.3	0.6	16.2
RIV	500	F	2.49	2.70	3.07	1.66	1.33	0.99	1.07	1.22	2.04	1.49	1.43	1.36	1.62	1.90	2.61	1.86	1.88	1.92	1.89	1.92	1.88	17.0	19.5	34.3	3.0	1.8	16.7
	600	F	1.65	2.67	3.13	2.40	1.32	0.98	1.07	1.22	2.30	1.61	1.43	1.36	1.51	1.91	2.74	1.86	1.92	1.96	1.91	1.92	1.88	7.3	20.1	40.6	14.7	1.9	21.1
	700	F	2.49	2.71	2.88	1.71	1.32	1.00	1.06	1.22	1.87	1.50	1.42	1.37	1.60	1.93	2.46	1.89	1.93	1.97	1.89	1.94	1.95	15.6	18.8	30.4	4.8	3.4	7.5
		B	2.50	2.71	2.85	1.71	1.33	1.00	1.06	1.21	1.88	1.50	1.42	1.37	1.60	1.95	2.45	1.89	1.94	1.97	1.89	1.94	1.96	15.5	18.6	31.1	4.9	3.8	7.1
	1000	F	2.47	2.68	3.25	1.68	1.30	0.98	1.07	1.24	2.33	1.51	1.42	1.36	1.60	1.91	3.01	1.83	1.86	1.93	1.90	1.93	1.91	16.2	19.8	40.5	4.3	3.1	11.1
		B	2.45	2.66	2.90	1.68	1.32	1.01	1.07	1.22	1.90	1.51	1.42	1.36	1.59	1.95	2.59	1.83	1.87	1.91	1.88	1.92	1.91	16.3	19.8	33.4	4.5	3.7	7.0
RV	500	F	1.63	2.55	3.02	2.82	1.63	1.00	1.07	1.09	1.99	1.58	1.41	1.35	1.54	2.09	3.55	1.90	2.00	2.03	1.85	1.88	1.95	4.1	18.6	36.5	12.5	3.9	22.7
		B	1.64	2.79	2.97	2.81	1.45	1.00	1.07	1.14	1.97	1.57	1.38	1.35	1.54	2.47	3.46	1.89	2.05	2.03	1.85	1.90	1.94	3.9	19.5	36.5	12.7	5.2	22.5
	600	F	2.45	2.70	2.87	1.72	1.34	1.00	1.06	1.21	1.90	1.49	1.42	1.36	1.61	1.95	2.55	1.92	1.96	1.97	1.91	1.97	1.99	15.5	18.7	31.6	5.1	3.1	7.1
		B	2.45	2.69	2.81	1.72	1.30	1.02	1.06	1.24	1.78	1.49	1.41	1.37	1.61	2.01	2.43	1.93	1.97	1.98	1.92	1.98	2.00	15.4	19.5	29.8	5.1	3.4	6.1
	700	F	2.53	2.71	3.04	1.71	1.30	0.99	1.07	1.24	2.01	1.50	1.43	1.36	1.58	1.87	2.54	1.92	1.95	1.93	1.88	1.94	1.91	16.0	19.5	31.9	3.2	1.8	12.0
	900	‡	2.42	2.63	2.84	1.74	1.32	1.01	1.06	1.24	1.85	1.50	1.42	1.36	1.57	1.89	2.47	1.96	2.03	2.16	1.95	1.98	1.92	17.1	21.1	32.8	2.4	2.4	10.2
	1000	F	2.38	2.63	2.90	1.67	1.34	0.99	1.07	1.19	2.09	1.51	1.43	1.36	1.60	1.90	2.77	1.83	1.86	1.91	1.93	1.94	1.87	16.1	19.1	37.8	5.2	2.1	7.1
SI	500	F	1.73	2.66	3.15	2.41	1.40	1.00	1.06	1.18	2.24	1.60	1.42	1.36	1.52	2.01	3.30	1.90	1.97	1.94	1.93	1.96	1.90	7.7	19.0	39.6	13.6	1.8	7.7
		B	1.76	2.63	3.08	2.30	1.45	1.00	1.07	1.15	2.21	1.59	1.43	1.36	1.53	1.97	3.21	1.91	1.95	1.94	1.93	1.95	1.91	9.3	18.5	39.6	13.5	2.1	7.6
	600	F	2.38	2.58	3.08	1.69	1.42	1.00	1.07	1.16	2.05	1.49	1.43	1.35	1.64	1.91	3.34	1.85	1.87	1.92	1.92	1.94	1.90	17.3	19.6	37.4	3.7	1.8	5.0
		B	2.38	2.66	3.03	1.70	1.39	1.00	1.07	1.18	2.04	1.49	1.40	1.36	1.65	2.10	3.24	1.85	1.89	1.92	1.92	1.94	1.91	17.3	20.1	37.5	3.6	1.7	5.1
	700	F	2.31	2.58	3.12	1.75	1.49	1.00	1.06	1.12	2.12	1.49	1.43	1.36	1.64	1.95	3.33	1.89	1.92	1.94	1.97	1.99	1.92	16.4	18.3	37.9	5.2	1.9	6.5
		B	2.29	2.65	3.05	1.77	1.47	1.00	1.06	1.14	2.09	1.49	1.40	1.36	1.66	2.16	3.23	1.89	1.93	1.94	1.98	1.96	1.92	16.6	19.0	37.9	5.3	1.9	6.5
	800	‡	2.43	2.64	2.96	1.65	1.48	1.00	1.08	1.13	1.92	1.48	1.43	1.36	1.65	1.91	2.92	1.90	1.93	1.96	1.93	1.93	1.90	17.1	18.1	33.8	3.3	1.5	7.7
	1000	F	1.72	2.71	3.06	2.37	1.38	1.00	1.06	1.19	2.07	1.59	1.42	1.36	1.52	2.03	3.14	1.91	1.98	2.07	1.96	1.96	1.93	7.8	19.3	36.7	14.1	1.7	5.0
		B	1.76	2.71	2.98	2.36	1.53	1.01	1.05	1.12	1.96	1.58	1.40	1.36	1.52	2.25	3.00	1.89	2.00	2.05	1.94	1.94	1.93	8.3	18.4	34.9	13.3	1.9	5.0
SII	500	F	1.53	2.63	2.78	2.86	1.52	1.01	1.10	1.13	1.80	1.82	1.41	1.36	1.45	2.21	2.78	1.90	2.02	1.99	1.91	1.92	1.93	5.5	18.0	30.5	13.0	5.1	8.5
		B	1.56	2.62	2.74	2.86	1.54	1.01	1.09	1.12	1.79	1.77	1.40	1.36	1.46	2.23	2.79	1.95	2.01	1.97	1.94	1.93	1.92	4.5	17.8	31.2	13.3	5.4	9.2
	600	F	1.56	2.64	3.00	2.77	1.56	1.00	1.09	1.11	1.93	1.73	1.40	1.36	1.48	2.26	3.12	1.83	1.94	1.94	1.85	1.89	1.92	6.5	17.9	33.7	13.1	6.0	13.9
		B	1.55	2.63	2.90	2.77	1.56	1.01	1.10	1.11	1.85	1.74	1.40	1.36	1.48	2.26	2.96	1.82	1.94	1.93	1.87	1.91	1.92	6.6	18.1	32.7	13.2	5.6	12.1
SIV	500	F	2.41	2.59	3.41	1.78	1.43	0.99	1.05	1.16	2.34	1.49	1.45	1.36	1.55	1.84	3.72	1.76	1.83	1.92	2.02	1.94	1.95	14.6	18.7	39.4	6.3	1.5	4.9
	600	F	2.35	2.71	2.94	1.74	1.38	1.01	1.06	1.18	1.86	1.50	1.42	1.36	1.61	2.00	2.70	1.84	1.88	1.93	1.95	1.97	1.90	15.4	18.5	30.8	5.9	1.9	4.6
		B	2.45	2.68	2.90	1.73	1.51	1.01	1.06	1.11	1.85	1.50	1.43	1.36	1.60	1.95	2.71	1.83	1.87	1.93	1.95	1.95	1.90	15.5	17.0	31.2	4.4	2.5	3.9
	700	F	2.30	2.71	2.94	1.78	1.34	1.00	1.06	1.21	1.95	1.51	1.41	1.37	1.62	2.03	2.69	1.85	1.89	1.94	1.93	1.96	1.89	16.0	19.5	33.6	5.9	1.7	7.6
		B	2.37	2.65	2.87	1.78	1.46	1.00	1.06	1.14	1.96	1.51	1.43	1.37	1.61	1.97	2.69	1.85	1.88	1.94	1.93	1.95	1.88	16.0	18.2	34.3	5.0	2.3	6.1
	1000	F	2.39	2.65	3.08	1.71	1.33	1.00	1.06	1.22	2.19	1.50	1.42	1.36	1.59	1.92	3.24	1.83	1.86	1.90	1.94	1.95	1.88	15.8	19.6	39.3	5.0	0.8	4.0
		B	2.38	2.63	2.99	1.72	1.40	1.00	1.06	1.17	2.17	1.50	1.43	1.36	1.59	1.92	3.14	1.83	1.85	1.90	1.93	1.95	1.88	15.7	18.5	39.4	5.2	1.9	3.8

Table 4.10: Values of geometric parameters during step [**II**] of the acylation reaction. The order of values for each parameter are given in the order (TI,TS,PDC).

			TI	TS	MCC									
				ω_1			ω_2			ω_3			ω_4	
RIII	500	F	-106.30	-120.41	-132.46	165.31	153.97	121.98	91.68	96.22	117.78	17.09	18.83	32.10
		B	-107.67	-135.97	-136.21	164.06	128.99	113.63	93.42	115.34	125.49	15.10	26.17	37.31
	600	F	-113.15	-121.41	-125.58	160.41	148.52	130.61	98.29	103.64	116.88	7.41	8.16	14.07
		B	-116.91	-128.00	-127.60	156.51	139.39	125.19	100.99	111.95	124.62	6.44	9.14	17.04
	700	F	-114.27	-122.43	-130.84	164.05	154.97	117.79	93.00	95.86	126.25	19.98	22.65	45.24
		B	-118.29	-136.47	-131.82	159.56	126.98	117.10	95.65	116.74	128.65	19.84	34.53	47.18
	800	‡	-117.25	-122.42	-140.15	168.52	161.44	127.20	89.22	89.54	104.13	24.76	29.86	44.81
	1000	F	-106.93	-120.57	-125.35	160.59	147.34	126.29	93.27	99.65	114.69	17.26	22.87	35.60
		B	-106.78	-129.91	-125.28	160.42	133.13	124.42	93.13	112.32	117.11	17.78	28.79	37.35
RIV	500	F	-78.57	-92.51	-112.52	163.80	157.13	103.16	126.46	132.64	163.15	-86.29	-81.80	-47.36
		B	-85.82	-117.50	-112.67	141.72	105.24	102.55	138.17	165.17	163.20	-67.60	-50.02	-46.08
	600	F	-83.88	-87.18	-107.76	159.60	155.63	102.06	124.14	126.43	162.34	-116.23	-114.59	-85.53
		B	-107.85	-113.71	-107.61	121.59	109.06	100.79	155.94	161.69	162.94	-96.84	-89.94	-82.65
	700	F	-82.37	-117.36	-110.31	150.77	102.99	100.22	137.48	166.92	162.95	-87.80	-66.16	-60.67
		B	-82.77	-118.29	-112.04	148.44	101.85	99.29	138.87	168.00	164.55	-86.53	-64.05	-59.93
	1000	F	-79.57	-111.36	-88.15	150.31	104.90	93.17	142.34	170.02	167.04	-51.48	-23.30	-5.78
		B	-81.27	-112.38	-104.35	146.85	102.93	100.80	144.44	171.60	166.79	-50.55	-21.47	-15.82
RV	500	F	-86.33	-86.85	-110.16	175.11	173.47	104.48	109.12	108.80	159.37	25.09	26.99	74.19
		B	-109.13	-117.18	-109.79	127.50	109.08	104.15	151.06	159.35	159.49	61.74	70.97	77.07
	600	F	-81.57	-120.16	-114.65	150.32	106.89	101.41	119.86	155.16	154.03	87.62	102.28	104.67
	700	F	-83.48	-118.68	-113.66	156.80	114.81	105.59	117.68	153.80	157.61	80.37	89.91	97.53
	900	‡	-60.90	-59.20	-46.75	-176.12	-173.19	-159.70	98.52	93.55	81.37	62.59	65.74	72.73
	1000	F	-78.68	-109.42	-105.88	151.20	109.95	104.14	135.38	166.77	167.28	95.98	115.42	120.81
SI	500	F	-82.61	-86.23	-94.08	170.38	164.37	134.88	-140.53	-142.06	-124.61	161.03	164.87	172.24
		B	-86.29	-98.84	-93.04	164.46	146.66	135.55	-138.64	-127.32	-124.27	164.27	169.23	172.35
	600	F	-78.22	-82.38	-90.74	178.61	179.98	131.41	-137.37	-143.79	-104.70	159.28	160.01	154.97
	700	F	-76.45	-83.19	-104.55	176.24	176.87	133.72	-138.46	-141.86	-94.93	138.32	138.62	137.98
		B	-76.74	-108.88	-102.54	174.93	144.35	135.46	-132.43	-99.73	-96.68	143.25	140.21	137.15
	800	‡	-89.92	-90.80	-87.08	174.87	176.30	172.01	-134.76	-136.06	-135.34	158.26	156.34	154.15
	900	‡	-83.97	-87.86	-94.61	168.92	166.55	145.21	-135.42	-135.07	-118.87	148.98	149.21	152.51
	1000	F	-84.58	-87.36	-94.45	179.07	178.82	133.16	-137.80	-139.12	-98.37	171.46	171.21	164.63
SII	500	F	-107.00	-106.43	-98.20	68.48	68.65	72.57	-156.08	-157.54	-166.06	162.22	162.34	163.85
		B	-106.53	-105.04	-97.64	70.21	71.29	73.22	-157.54	-160.71	-167.91	161.29	161.96	164.59
	600	F	-104.28	-105.18	-95.52	62.21	61.26	66.43	-165.19	-166.13	-175.39	159.38	155.09	149.46
		B	-104.83	-103.40	-95.12	64.05	66.10	69.27	-166.92	-170.83	-177.97	150.07	148.20	149.44
SIV	500	F	-69.49	-96.22	-105.56	174.41	151.94	124.09	-144.07	-132.08	-106.73	-17.66	-5.04	-8.20
		B	-82.44	-111.26	-105.09	163.16	134.21	123.94	-134.46	-110.91	-106.48	-3.61	-2.17	-8.66
	600	F	-76.24	-85.93	-102.13	173.96	167.41	134.44	-139.28	-138.70	-104.31	-28.66	-24.19	-22.16
		B	-82.18	-108.85	-102.89	164.89	134.24	127.59	-134.42	-108.55	-103.98	-21.38	-14.84	-20.48
	700	F	-79.63	-99.01	-98.91	160.54	134.94	121.10	-139.81	-124.55	-114.00	-18.46	-8.83	-11.70
		B	-80.94	-104.26	-98.12	156.80	132.40	120.31	-137.58	-118.29	-114.17	-15.39	-9.65	-10.44
	1000	F	-81.02	-93.42	-99.15	166.41	147.02	130.55	-134.15	-129.40	-111.76	-5.94	-8.07	-11.29
		B	-81.96	-104.00	-97.01	164.10	135.86	130.09	-132.53	-113.50	-111.56	-3.29	-8.55	-12.60

Table 4.11: Values of dihedral angles during step [**I**] of the acylation reaction. The values for each parameter are given in the order (TI, TS, MCC) as demonstrated for ω_1.

			TI	TS	PDC									
				ω_1			ω_2			ω_3			ω_4	
RIII	500	F	-106.30	-102.33	-79.47	165.31	160.83	-163.34	91.68	92.45	44.64	17.09	15.25	31.82
		B	-105.57	-91.63	-80.21	166.36	-178.82	-164.54	92.44	69.79	46.15	15.92	24.06	31.79
	600	F	-113.15	-104.11	-87.16	160.41	159.70	-178.73	98.29	98.83	66.41	7.41	5.15	12.33
		B	-113.73	-93.48	-87.69	159.54	174.06	-179.96	99.13	79.70	66.79	7.15	12.75	12.66
	700	F	-114.27	-103.59	-86.55	164.05	160.68	-169.79	93.00	92.25	45.32	19.98	20.75	41.18
		B	-114.55	-92.61	-88.09	163.30	178.68	-174.80	92.50	66.48	51.58	21.13	31.65	39.51
	800	‡	-108.12	-103.48	-86.66	171.46	167.87	-167.25	87.33	85.96	45.50	25.45	26.23	42.59
	900	‡	-102.61	-99.97	-85.89	163.94	160.34	179.57	92.08	91.57	55.24	15.38	14.86	28.44
	1000	F	-106.93	-101.67	-84.46	160.59	158.90	176.41	93.27	92.36	60.83	17.26	16.56	31.30
		B	-105.92	-91.66	-84.71	161.28	171.46	176.81	91.30	73.54	59.99	18.73	23.88	31.64
RIV	500	F	-78.57	-77.05	-72.26	163.80	161.53	175.17	126.46	126.63	114.27	-86.29	-85.57	-86.17
	600	F	-83.88	-77.08	-70.55	159.60	155.91	166.76	124.14	123.83	105.89	-116.23	-108.86	-102.57
	700	F	-82.37	-80.38	-82.24	150.77	148.26	140.50	137.48	136.93	142.47	-87.80	-86.89	-83.56
		B	-81.50	-81.69	-83.81	150.17	145.78	139.92	137.06	138.28	143.78	-87.43	-85.22	-84.33
	1000	F	-79.57	-79.12	-91.23	150.31	147.04	125.31	142.34	143.93	175.60	-51.48	-50.73	-39.17
		B	-81.11	-83.61	-89.80	147.69	140.90	131.21	144.45	149.03	164.37	-50.77	-47.69	-41.96
RV	500	F	-86.33	-82.16	-66.05	175.11	172.46	-175.27	109.12	115.60	101.34	25.09	27.65	14.10
		B	-86.75	-77.73	-67.43	174.16	176.40	-175.26	110.91	109.66	101.37	27.84	28.30	14.78
	600	F	-81.57	-80.03	-85.76	150.32	146.60	137.16	119.86	118.73	124.89	87.62	88.63	88.38
		B	-81.77	-82.84	-85.92	149.06	143.63	138.18	120.63	121.70	125.11	88.50	88.98	87.24
	700	F	-83.48	-81.80	-74.18	156.80	154.16	164.62	117.68	114.80	101.08	80.37	80.87	76.34
	900	‡	-65.03	-65.31	-65.51	170.74	169.96	179.24	114.67	113.70	105.35	71.03	72.47	68.96
	1000	F	-78.68	-77.79	-74.63	151.20	149.13	158.16	135.38	134.69	116.77	95.98	97.37	113.08
SI	500	F	-82.61	-73.31	-58.96	170.38	177.38	-150.66	-140.53	-136.58	-169.67	161.03	155.21	139.28
		B	-82.07	-73.75	-59.98	172.52	179.05	-150.89	-139.64	-137.92	-168.86	161.45	156.08	139.30
	600	F	-78.22	-76.43	-61.00	178.61	178.85	-152.45	-137.37	-137.83	-168.99	159.28	158.51	142.26
		B	-76.02	-73.38	-61.52	-178.91	-176.45	-152.53	-140.41	-142.91	-168.72	157.85	154.41	141.92
	700	F	-76.45	-74.16	-60.87	176.24	176.37	-148.73	-138.46	-137.67	-165.99	138.32	139.11	130.84
		B	-74.70	-73.11	-61.90	-179.61	-176.55	-149.09	-138.96	-140.81	-165.19	144.74	143.08	130.88
	800	‡	-77.18	-75.63	-65.84	178.87	179.20	-161.61	-135.64	-136.54	-159.65	141.80	142.16	140.86
	1000	F	-84.58	-76.56	-63.00	179.07	179.87	-151.35	-137.80	-136.59	-168.31	171.46	166.87	148.67
		B	-83.50	-73.32	-65.32	-179.03	-168.71	-153.86	-139.15	-148.98	-164.86	170.82	156.62	149.22
SII	500	F	-107.00	-90.64	-82.98	68.48	97.29	96.16	-156.08	-174.05	-179.93	162.22	148.35	148.38
		B	-104.41	-88.22	-84.78	74.61	96.08	94.84	-156.69	-175.90	179.96	157.12	147.99	149.83
	600	F	-104.28	-90.86	-86.19	62.21	84.25	81.41	-165.19	177.53	175.69	159.38	150.13	143.31
		B	-104.59	-91.17	-86.59	63.25	86.33	83.56	-165.78	179.36	175.49	151.88	143.71	143.12
SIV	500	F	-69.36	-79.13	-49.38	174.36	165.54	-144.33	-144.13	-135.92	175.05	-17.68	-5.40	-35.35
	600	F	-76.24	-75.45	-70.76	173.96	173.29	-167.80	-139.28	-138.39	-152.02	-28.66	-30.56	-40.19
		B	-77.02	-75.40	-70.22	173.90	176.41	-164.01	-137.89	-140.48	-155.62	-28.91	-31.50	-42.39
	700	F	-79.63	-75.20	-72.50	160.54	162.86	-177.79	-139.81	-138.58	-147.37	-18.46	-19.93	-38.16
		B	-77.69	-74.44	-71.34	162.10	169.66	-169.29	-140.10	-141.29	-154.06	-19.06	-27.45	-42.23
	1000	F	-81.02	-78.22	-61.19	166.41	167.29	-146.80	-134.15	-135.21	176.47	-5.94	-8.40	-44.51
		B	-77.63	-73.88	-62.31	167.65	172.83	-147.43	-137.97	-141.80	177.15	-12.48	-21.88	-45.23

Table 4.12: Values of dihedral angles during step [**II**] of the acylation reaction. The values for each parameter are given in the order (TI, TS, PDC) as demonstrated for ω_1.

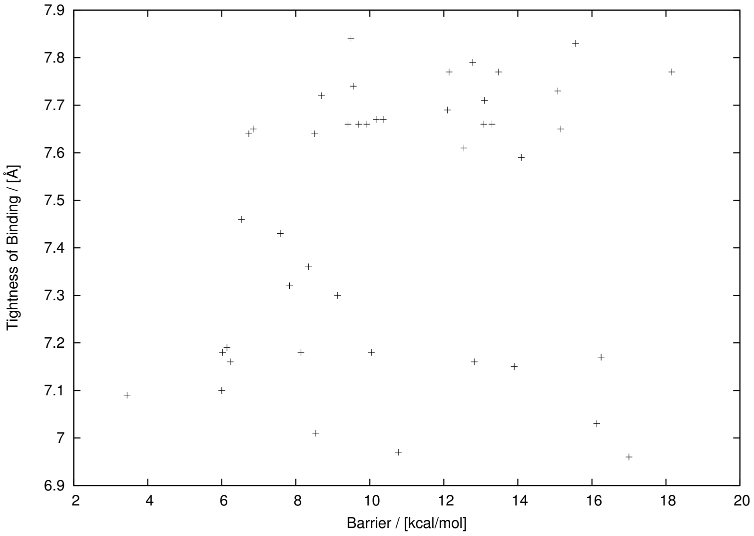


Figure 4.13: Potential energy barrier for MCC → TS in step [I] versus the tighness of binding[236] descriptor.

4.2.5 Electrostatic Perturbation

The electrostatic impact of amino acids residues on reaction barriers can be estimated by a perturbation scheme[12][50]. In this procedure one successively deletes the MM partial charges on the side chains of individual amino acids. Each of the charge sets thus obtained is used to re-evaluate the electronic energies of the TI and the Michaelis complex. The electron densities are allowed to relax in the modified charge field. For each such charge field the calculation is carried out for both the TS and product/reactant and the target quantity $\Delta\Delta\ \mathrm{E}^{\ddagger} = (\mathrm{E}_{\mathrm{TS}}^{\mathrm{pert.}} - \mathrm{E}_{\mathrm{Product/Reactant}}^{\mathrm{pert.}}) - (\mathrm{E}_{\mathrm{TS}} - \mathrm{E}_{\mathrm{Product/Reactant}})$ is evaluated. This method may be seen as an in silico electrostatic equivalent of alanine scanning mutagenesis[145]. We applied this scheme to RVS900, for which the first TS apart from ANRU453 could be located. The scan was done on all 175 amino acid side chains resolved in the x-ray structure of the enzyme excluding Ser77, Asp133 and His156, which belong to the catalytic triad and are within the QM-region (QM II). This analysis was part of a recent publication[72].

Our calculations have identified five amino acid positions that have a pronounced effect (>1 kcal/mol) on the reaction barrier (Fig. 2). Four of these (Lys44, Asp43, Asp40 and Arg142) represent ionizable groups located on the protein surface. Shielding of these charges, e.g. by counterions from the surrounding solution under physiological conditions should diminish the influence of those residues. To test this hypothesis we have added counterions close to the charged sites of the groups above and re-evaluated the barrier. We found that the contributions drop below 1 kcal/mol for each group, and consequently, we do not consider them as "hot spots". The remaining position identified in the QM/MM-scan was residue His76 which is located below the Ser77 (Fig. 1) and His156, and can form a hydrogen bond to

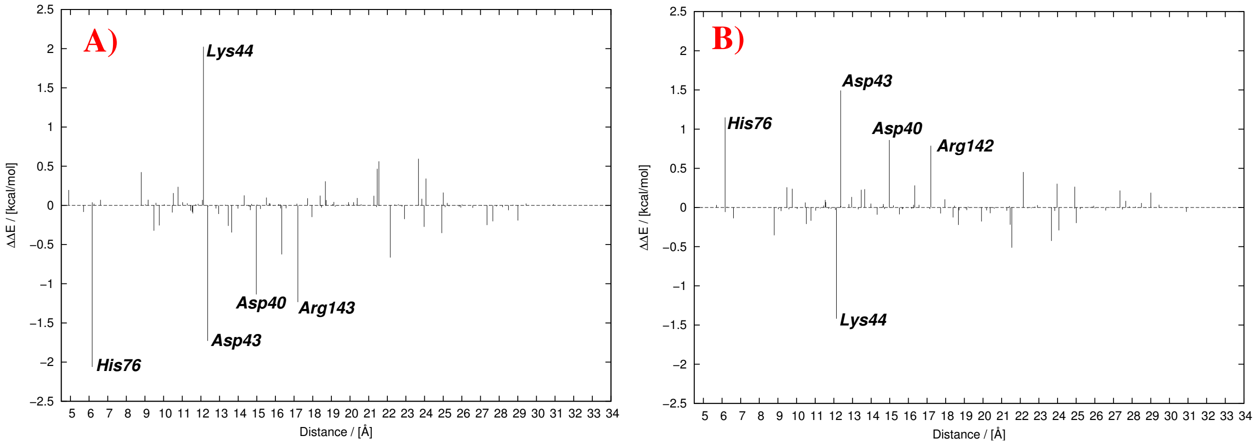


Figure 4.14: Modulation of the reaction barrier going from A) the Michaelis complex to the TS, and B) from the TI to the TS by individual amino acid side chains. The distance is measured from the tetrahedral carbon in the TI to the geometric center of the individual amino acid side chains. Large contributions are labeled. Negative values indicate residues that raise the barrier, and positive values those which decrease it.

the backbone oxygen of the active site histidine (His156). This histidine is singly protonated in our model setup, and its overall charge is therefore zero. Due to its position, orientation and polarity it may exert an important role during ester hydrolysis.

The diagrams in figure 4.14 show an interesting phenomenon. While the electronic effect of His76 is to *raise* the energy barrier for nucleophilic attack in step [I] ($\mathrm{MCC} \rightarrow \mathrm{TI}$), it *lowers* the barrier for the decomposition of the TI (via $\mathrm{TI} \rightarrow \mathrm{MCC}$). These finding suggest that His76 should have an adverse effect on the rate of acylation, however, this may easily be offset by the structural influences of His76, which we do not capture by this estimative method.

Independent of the computational analysis our experimental partners[8] identified position 76 to be a 'hot spot' for enantioselectivity by saturation mutagenesis of the entire sequence space of BSLA in combination with an activity screening using enantiomerically pure (R)- and (S)-NAPHAC as model substrates. They found stable and active mutations His76Ala and His76Leu, with enantioselectivities of $\mathrm{E} = 8.5$ (S) and $\mathrm{E} = 6.8$ (R). While the wild type enzyme exhibits almost no activity for the S enantiomer, both mutations feature an increased activity towards (S)-NAPHAC. The remarkable finding of an inversion of enantioselectivity after the His76Leu exchange indicates that the residue must impart subtle conformational effects since the character of the mutated side chains, which are both unpolar, is identical. A rational explanation of the structural and energetic implications of the mutations would require a more detailed study that goes beyond the current perturbational analysis.

---

[8] Collaboration with Prof. Jaeger, Jülich.

4.3 Enantioselectivity

Is the experimentally determined enantioselectivity of BSLA towards the NAPHAC substrate reproduced on the basis of the calculated potential energy barriers? To answer this question we computed the enantioselectivity between pairs of binding modes of (R)-NAPHAC and (S)-NAPHAC. We chose the binding modes with lowest effective potential energy barriers for this purpose, since the acylation reaction will proceed mainly via these channels. This is an approximation, since under experimental conditions several binding modes could coexist and contribute to the overall rate of conversion. This could, in principle, be accounted for *in silico* as well, but would make sense only if precise populations of the individual binding modes were known.

Binding modes RIII, RIV, RV, and SII belong to type C), whereas SI and SIV have a shape shown in A). The barriers for the individual steps are reported in tables 4.7 and 4.8. Qualitative effective barriers based on the formulae in section 4.2.2 lead to the conclusion that RIV and RV are the least proficient binding modes of the current set (RIV: 15.9 kcal/mol, RV: 16.3 kcal/mol)[9]. Therefore the R enantiomer is represented by RIII (RIII: 12.7 kcal/mol). The S enantiomer is catalytically more proficient with SI being the fastest reacting binding mode of the whole set (SI: 10.7 kcal/mol[10]). The value for SIV is very similar to SI (SIV: 10.8 kcal/mol), but SII has a higher barrier (11.7 kcal/mol).

For the evaluation of **E** we compare RIII and SI, which are the fast reacting binding modes of each enantiomer. An approximate **E** value can be calculated (see appendix E) as $\mathbf{E} = \exp\left(-\frac{\Delta \mathrm{E}^{\ddagger}_{\mathrm{R}} - \Delta \mathrm{E}^{\ddagger}_{\mathrm{S}}}{\mathrm{RT}}\right)$ using the qualitative effective barriers of the binding modes as activation energies. The $\Delta\Delta \mathrm{E}^{\ddagger}_{\mathrm{R-S}}$ difference is 2.2 kcal/mol, which, at room temperature (300 K), amounts to **E**=39 (S). If this approximate theoretical value is contrasted with the experimental value of **E**=140 (R), we arrive at an error of 5.2 kcal/mol in $\Delta\Delta \mathrm{E}^{\ddagger}$[11].

For a quantitative *in silico* evaluation of **E**, a simulation of the kinetic resolution experiment was performed using the program in appendix D (pages D.3-D.5). The parameters for the stochastic simulation runs are provided in table 4.14. 25 runs were performed for each simulation and the results were averaged. The concentrations of enzyme and water were set to 5 and 10000000, reactants were input as a racemic mixture of 5000 molecules of both species R and S. The simulation temperature was set to 300 K. For the binding and unbinding steps, *ad hoc* parameters were introduced. The computed enantioselectivity was independent of

---

[9] Energy values in parentheses denote here average values over all available scans. The individual values for the effective barriers are given in table 4.13

[10] excluding SI600F and SI800‡ in table 4.13 from the average due to artifacts (see section 4.2.3).

[11] Computed as $\Delta\Delta \mathrm{E}_{\mathrm{Error}} = \Delta\Delta \mathrm{E}^{\ddagger}_{\mathrm{Experiment}} + \Delta\Delta \mathrm{E}^{\ddagger}_{\mathrm{Theory}}$.

			$\Delta E_{QM/MM}$	ΔE_{QM}	ΔE_{MM}
RIII	500	F	13.67	13.83	-0.16
		B	10.28	11.99	-1.71
	600	F	13.75	10.66	3.09
		B	12.33	14.77	-2.44
	700	F	12.86	13.76	-0.90
		B	10.23	10.97	-0.74
	800	‡	12.66	9.59	3.07
	1000	F	14.02	15.13	-1.11
		B	14.61	16.30	-1.69
RIV	500	F	15.08	13.79	1.29
	600	F	18.12	16.33	1.79
	700	F	18.16	9.94	8.22
		B	15.30	13.98	1.32
	1000	F	15.05	9.29	5.76
		B	14.14	12.72	1.42
RV	500	F	20.44	12.49	7.95
		B	21.35	23.95	-2.61
	600	F	13.99	13.58	0.41
	700	F	18.08	18.70	-0.63
	900	‡	7.30	8.46	-1.16
	1000	F	16.80	14.23	2.57
SI	500	F	12.82	8.98	3.84
		B	10.04	8.65	1.39
	600	F	6.73	8.65	-1.93
	700	F	13.08	8.02	5.06
		B	9.49	9.08	0.40
	800	‡	3.44	2.45	0.99
	1000	F	8.14	7.86	0.28
SII	500	F	11.09	10.04	1.05
		B	9.22	9.57	-0.35
	600	F	14.10	10.19	3.91
		B	12.55	9.05	3.50
SIV	500	F	10.36	9.85	0.51
	600	F	8.51	7.68	0.83
		B	10.17	10.68	-0.51
	700	F	14.09	9.65	4.44
		B	12.66	12.15	0.51
	1000	F	13.30	4.23	9.07
		B	6.85	6.66	0.20

Table 4.13: Qualitative effective barriers of the acylation reaction for individual snapshots (see section 4.2.2.).

these parameters as long as the binding step was essentially barrierless, and the unbinding step had a barrier smaller barrier than that of acylation. The values were chosen such that binding and unbinding steps occurred with similar frequency. This speeds up the stochastic simulation considerably. Identical sets of parameters were applied for binding, unbinding and deacylation of both enantiomers.

Average barriers for steps [I] and [II] of the acylation reaction were computed from average values for the snapshots of the individual binding modes. In the case of binding mode RIII the average barrier for MCC → TS in step [I] is 8.5 kcal/mol. In the average for TI → TS of step [I] RIII600F was excluded due to a spurious discontinuity in the early stages of the reaction path scan. This yields a value of 1.6 kcal/mol. At TI → TS of step [II] averaging all ten values gives a barrier of 5.5 kcal/mol. In binding mode SI values of SI800‡ and SI900‡ are outliers (see section 4.2.3), and SI600F is excluded due to the discontinuity of the PES. Then one obtains an average barrier of 10.7 kcal/mol. For the barrier TI → TS of step [I] the average value is 2.3 kcal/mol from all snapshots, and in TI → TS of step [II] the estimate is 0.8 kcal/mol. In addition to binding mode SI we let RIII compete with SII, the least proficient binding mode of (S)-NAPHAC. For SII few values are available due to the instability of tetrahedral intermediates of this binding mode after 600 ps. The average value for MCC → TS in step [I] is 7.9 kcal/mol, 0.0 kcal/mol in TI → TS of step [I], and 3.8 kcal/mol for TI → TS of step [II].

					R-Enantiomer			S-Enantiomer		E
	Bind.	Unbind.	Deacyl.	MCC → TS	TI → TS [I]	TI → TS [II]	MCC → TS	TI → TS [I]	TI → TS [II]	
RIII vs. SI										
1	1E-05	6.0	12.0	8.5	1.6	5.5	10.7	2.2	0.8	**15.9 (S)**
RIII vs. SII										
2				8.5	1.6	5.5	7.9	0.0	3.8	**3.3(S)**

Table 4.14: Parameters and results of the stochastic simulation runs. Input parameters are barriers in kcal/mol. The preferred enantiomer is indicated after the **E** value. Notice that the order of parameters is different from the input of the program. For simulation 1 the correct order (plus temperature) is: 300.0 1E-05 6.0 1E-05 6.0 10.7 2.2 8.5 1.6 0.8 5.5 12.0. Exchange of the enantiomers leads to a computed **E** value smaller than 1 in this case. This number must then be inverted to yield a correct **E**.

Results of the simulations are shown in table 4.14. The experimentally determined enantioselectivity, an **E** value of 140 pro R, cannot be reproduced. The **E** value of 15.9 (simulation 1 in table 4.14), corresponds to an apparent $\Delta\Delta E^{\ddagger}$ difference of 1.7 kcal/mol pro S (from formula 1.1 in chapter 1). The estimated error (see above) of our calculation compared to the experiment would thus amount to 4.7 kcal/mol, which is close to the value of 5.2 kcal/mol computed above using approximate effective barriers.

Why can the experimental enantioselectivity not be reproduced? First, we make the approximation of using potential energy rather than free energy differences. For apolar substrates, as NAPHAC can be considered due to its hydrophobic side chain, differential free entropies can have a large contribution. In their paper Ottosson et al.[155] presented a mutant[12] of lipase from *Candida antarctica* where the differential entropy term $T\Delta\Delta S^{\ddagger}$ was measured to be 7.7 kcal/mol. This value is a rather extreme example, but without experimental data of the BSLA-NAPHAC system it cannot be excluded that entropy plays a significant role.

Secondly, the reaction path scans sometimes lead to irregular potential energy surfaces, which are in part a result of insufficient relaxation of the solvent water. Especially for RIII this poses a problem, since at step [II] the expectedly beneficial water molecule below the TI cannot dissociate properly, and thereby interferes with the dissociation of the product alcohol, which causes a 'jumpy' surface. The aberration occurs in all reaction paths of RIII and raises the energy barrier TI → TS of step [II] by an estimated 2 kcal/mol due to the discon-

---

[12] Mutant W104H, experimental temperature 296 K, see table 1 in this paper.

tinuity alone. The contribution of reorganisation of the solvent and a proper dissociation of the alcohol cannot be estimated in a simple manner.

4.4 Conclusion

The activation barriers computed in this chapter by geometry optimization techniques do not account for entropic contributions. It would be possible to include the entropic terms for the QM region in harmonic approximation through a normal mode analysis, but such a treatment would miss the entropic contributions of the protein environment which may be substantial. Including the latter in the normal mode analysis quickly becomes prohibitive computationally so that entropic effects are difficult to capture with the approach taken in this chapter.

Another problem concerns the local character of geometry optimizations in condensed phase systems. The reaction coordinate drives the system from reactants to products, and the environment is expected to relax adiabatically. In the condensed phase, the surrounding cannot reorganize freely, but is guided to the closest minimum which may lead to some congestion as one proceeds along the reaction path. Such a frustrated system may relax instantaneously on a later point of the path and produce a seemingly discontinuous PES (hysteresis). In our system, water was sometimes subject to such frustration. This comes as no surprise, considering that the reaction studied took place on the water-accessible surface of the enzyme, and that the treated steps include partial dissociation of ligand and product.

Such spurious effects demonstrate the need for proper relaxation of the solvent degrees of freedom. One solution would be to perform the optimisations in implicit solvent. This would only yield realistic results if directionalities of hydrogen bonds were not important. The example of RIII where water plays a structural role, is counter-indicative of implicit solvation. A different approach to modeling the solvent was taken by Dinner et al. [50], who did not treat bulk solvent explicitly as it *'is unlikely to yield a realistic dielectric as the solvent cannot reorganize extensively during energy minimization.'* In their scheme they used a very limited number of structurally important water molecules during geometry optimisation. Charges of ionic groups were scaled[194] to avoid distortions of their geometries. A continuum electrostatics methods was subsequently applied to calculate solvent-corrected effective energies. This approach may ameliorate the continuity of calculated PESs by allowing the structurally important water molecules to relax more freely, but will offer only a partial solution to the problems outlined above.

A more rigorous solution is to avoid geometry optimization as such and to move to a MD-based treatment with sampling of all relevant structural degrees of freedom along the reaction path. This will overcome the limitations of local relaxation methods and will also include entropic effects in an appropriate manner. The results of such free energy calculations are presented in the following chapter.

Chapter 5

The Acylation Reaction: Free Energy Calculations

In the preceding chapter we studied the PES of the acylation reaction for a number of binding modes using a simple geometric RC. It was noticed that insufficient relaxation and solvent reorganisation were limiting factors of accuracy and that entropic contributions were not included, which are potentially important or may even be the sole cause for enantioselectivity[159][155].

A natural approach to the problem of enantioselectivity should thus focus on free energies of activation, $\Delta G^{\ddagger}$, rather than potential energies of activation, $\Delta E^{\ddagger}$. To this end, we use a QM/MM setup as established in the last chapter to perform umbrella sampling calculations. This MD-based approach involves configurational averaging and should thus avoid artefacts due to non-relaxation.

In the first section of this chapter a detailed account of the computational methodology is given. This is followed by a discussion of the accuracy of our calculations and possible sources of error. We open the presentation of our results with the analysis of the $\Delta G^{\ddagger}$ values. These are translated into kinetic rate constants and used in the stochastic model of a kinetic resolution experiment to evaluate enantioselectivity *in silico*. Thereafter we analyse the variation of the structural parameters during the course of the simulation and assess their importance with respect to the discrimination of enantiomers. A comparison with the results of chapter 4 is done where appropriate. Finally, we try to estimate the free energy of binding and compute a two-dimensional free energy surface of the acylation reaction.

5.1 Details of Simulations

In the present section we account for strategic and technical details of the umbrella sampling calculations. More general specifications are considered first.

As in chapter 4, snapshots of the trajectories were generated at the MM level as described in chapter 3 and were used as input for the calculations. We employed a QM/MM potential, where the QM fragment was represented by SCC-DFTB (see chapter 2) and the MM fragment by Charmm22. The size of the QM fragment was equivalent to QMI. The QM region was terminated by link atoms, and the contribution of the Coulomb interaction between QM and MM fragments was handled by electrostatic embedding. No cutoffs were applied to QM/MM electrostatics. We used CHARMM version 31b1 in QM/MM simulations. The reaction coordinate was implemented in exactly the same way as in Chemshell to ensure compatibility with our earlier results. The simulation temperature was always set to 300 K, and all internal degrees of freedom of TIP3 water were constrained to their standard values[107]. Bonds of heavy atoms with hydrogen in the protein were not constrained, since this would have interfered with the hydrogen transfer in the QM fragment. This forced us to adopt a smaller time step than in chapter 3; a value of 0.5 fs was considered appropriate[130]. As before some parts of the enzyme were fixed during the simulations. For a consistent comparison of $\Delta G^{\ddagger}$ values one should perform simulations with identical sets of moving atoms. To this end the initial mobile selection of protein atoms as defined in the MD simulations of chapter 3 was used, which included 900 atoms of the enzyme and approximately 1500 water molecules.

Computational Strategy

All simulations consisted of two parts. In the first part the aim was to prepare snapshots of the MM dynamics for the production run. Thus they were allowed to equilibrate during an initial MD of 40 ps. The bias potentials were set to the initial values of the umbrella sampling calculations.

Production run calculations were executed sequentially. In each iteration atomic positions and velocities of the previous simulation window were used to initiate the next one, and the initial window was started from the equilibration run. Window potentials were modified by incrementing the equilibrium position S of the restraint (equation 4.1).

The position of the RC was written to separate files for each window. These output files additionally contained the current value of S and the force constant k of the bias potential. The automatic WHAM procedure was programmed to be applicable in a convenient fashion to the output generated by CHARMM 31b1.

Choice of Biasing Parameters

The essential parameters of an umbrella sampling simulation encompass the force constant k, the interval of S to be sampled, and the increment added to S iteratively. In umbrella sampling, the force constant k should be chosen large enough to allow even sampling over the entire range of the RC. Substantial overlap between neighbouring histograms of the RC must also be guaranteed. This places an upper bound on the restraint, which generates narrower distributions at higher values. To some extent this can be compensated by a smaller stepsize of S. In production runs we used force constants of 45, 50 and 75 kcal/(molÅ^2). The initial value S of the simulation was 0.0Å and the final value was -4.4Å. Experience showed that the acylation reaction usually ended close to -3.5 Å with formation of a MCC or PDC, but a slightly longer simulation was considered appropriate. The window increment was set to -0.1 Å, and proved capable of producing strongly overlapping distributions. A single simulation thus generated 45 windows.

Additional Restraints

The RC chosen in chapter 4, (see equation 4.1), cannot be used to study both steps [I] and [II] in a single simulation run. Four atoms are affected by this RC. The restraint is between C14, O17, HE2 and NE2 in step [I], and between C14, O3, HE2, and NE2 in step [II] (assignment of labels: Figure 4.1, page 61). Bond C14-O3 in step [I] is therefore not affected by the RC, in the same way that C14-O17 is unrestrained in step [II]. With geometry optimisations as performed in chapter 4, this approach worked without complications. When the first umbrella sampling calculations were run, it became obvious that the system frequently dissociated at the unrestrained C14-O17 or C14-O3 bond in the early stages of simulation. This seems natural, considering that during a few thousand MD steps the probablility of attaining kinetic energy sufficient to break the labile bonds of the TI is appreciable. Systems evolving this way could not be used to determine free energies of activation and their data had to be discarded. This was not acceptable, since umbrella sampling calculations are computer intensive. Therefore a second restraint was introduced at the formerly unrestrained C-O bond of the TI. A functional form as given in equation 2.7 was used. The force constant k was determined empirically and set to a value where premature dissociation was no longer a problem. When k was set to 150 kcal/(molÅ^2) the simulations were stable.

How Long to Sample

The iterative approach to cover the interesting range of the reaction coordinate does not explicitly specify equilibration periods. One assumes notwithstanding that any system requires some time to adapt to each new position, into which it is forced by the bias potential, on each iteration. These initial slices of data in each window are then usually discarded, and the post-equilibration time slice is used to compute free energy differences. One will typically try to limit equilibration to short periods and maximise the amount of analysable data. The required amount of these data is in turn determined by the target accuracy of $\Delta G^{\ddagger}$, which usually increases with the volume of data. How to analyse the accuracy of free energy simulations is discussed in a separate section (*vide infra*).

Equilibration can in principle be studied within each separate window. To this end statistical methods can be applied[180]. In a recent paper from our group statistical tests were used to establish equilibration in a thermodynamic integration of *p*-hydroxybenzoate hydroxylase[187]. These tests require data which follow a normal distribution. For umbrella sampling calculations using harmonic bias potentials with a small force constant, this is generally not fulfilled. In the large force constant limit, which is required by umbrella integration, the distributions become normal which enables standard statistical tests to be applied.

A popular alternative to this very consistent approach is to use equilibration intervals of fixed length[169][147][174][169][24]. By variation of the interval length in the evaluation of free energy differences, one can get a first impression of the size of expected errors due to nonequilibration. In WHAM this may be combined with the statistical method of *bootstrap* sampling, which is described in the next chapter, to gain confidence in the calculated $\Delta G^{\ddagger}$ values.

Simulation windows of 7.5 and 25 ps were tested, which comprise 15000 and 50000 single MD steps. The grand total of sampling time was therefore 337.5 ps in the small simulation setup (45 windows), and 1125 ps in the large setup. With the QM/MM model introduced above, the small setup consumed approximately four days on a 2 GHz AMD Opteron CPU, and the large setup took 13 days. Several sizes of the equilibration interval were tested, ranging from 2.5 ps to 24.75 ps. The consequences of our choice of simulation parameters were studied on the basis of model RIII.

5.2 Accuracy of Free Energy Simulations

The accuracy of simulations is limited by several factors. There are two types of bottlenecks which are described in the following two subsections.

5.2.1 Statistical Accuracy

The MD simulations are used to evaluate the free energy as a function of the distribution over a RC, as $\mathrm{G}(\varphi) = -\mathrm{k_B T ln P}(\varphi)$. Here φ is the reaction coordinate (or a more general order parameter), and $\mathrm{P}(\varphi)$ is the distribution function of the reaction coordinate. The true distribution is not known, of course, but one would like to have an estimate of the statistical accuracy of the computed free energy.

In thermodynamic integration the PMF is evaluated at discrete points along the RC. The forces on the RC are assumed to be normally distributed for an equilibrated system, and are thus amenable to standard treatments of statistical error analysis.

How can one still perform an analysis of the statistical accuracy of the results without the prerequisite of normality of the sampling distributions ? An unparametric method which expects little from the sampled system is the *bootstrap* method, which has been developed by Efron[52][53][54]. A bootstrap sample is a sample of size n, which is drawn with replacement from an original set of data, for example, the time series of RC values. The bootstrap data vector is not a permutation of the original data since it contains a fraction of repeated values. All samples are drawn using a random-number generator to access the vector of original data. If we have $\mathbf{y} = (\mathrm{x_1, x_2, x_3, ..., x_n})$, then one bootstrap sample may be $\mathbf{y}^* = (\mathrm{x_1^* = x_3, x_2^* = x_5, x_3^* = x_5, x_4^* = x_n, ..., x_n^* = x_1})$. This sample replaces the original data in the statistics afterwards (e.g. the WHAM), and by repeating the procedure B times, a mean and a standard error can be calculated.

$$\hat{\sigma} = \left(\frac{\sum_{\mathrm{b}=1}^{\mathrm{B}} [\hat{\theta}(\mathbf{y}^*(\mathrm{b})\ ) - \hat{\theta}^*(\cdot)]^2}{\mathrm{B}-1} \right), \hat{\theta}^*(\cdot) = \frac{\sum_{\mathrm{b}=1}^{\mathrm{B}} \hat{\theta}(\mathbf{y}^*(\mathrm{b})\ )}{\mathrm{B}}. \tag{5.1}$$

Here $\hat{\theta}(\mathbf{y}^*(\mathrm{b})\ )$ is the quantity of interest, calculated with a bootstrap sample, and $\hat{\theta}^*(\cdot)$ the mean from B iterations. For a large number of such samples, the sampling distribution is expected to approach normality by virtue of the central limit theorem[231], justifying the specification of a mean and a standard error $\hat{\sigma}$. Efron[53] recommends to perform on the order of 200 bootstrap resamples to estimate the standard error for each value on the RC. The error in the barriers can then be evaluated according to Gauss' law of error propagation as $\hat{\sigma}_{\mathrm{A}\to\mathrm{B}} = \sqrt{\hat{\sigma}_{\mathrm{A}}^2 + \hat{\sigma}_{\mathrm{B}}^2}$, where A and B are the points on the surface for which a barrier is calculated.

Although the original sampling distributions are not normal and the standard analysis techniques[180] cannot be applied to the raw data, we are in a position to estimate standard deviations at each point of the reaction coordinate by construction of a normal distribution using bootstrap resampling. Grossfield has introduced bootstrap sampling in the context of WHAM[86].

5.2.2 Systematic Errors

A well-known problem in MD simulations is insufficient sampling[130]. While it was argued in chapter 3 that in the bound state of the TI the ligand was confined to a limited number of stable states, its degrees of freedom are expected to increase on dissociation of the TI to form the MCC. Figure 5.1 offers a visualisation to illustrate this concept. One would essentially have to trace all transformations from the binding modes of the TI to the most probable configurations in the MCC to find the pathways which prove most competitive in a kinetic resolution experiment, (i.e. those with low free energy barriers). This prescription applies likewise to step [II]. However, one cannot guarantee that all possible conformations are sampled and missing transitions between TI and MCC will contribute to the systematic error. In the limit of an infinite simulation all configurations and their statistical weights would be known, all pathways of dissociation would be considered, and the systematic error due to sampling would be zero.

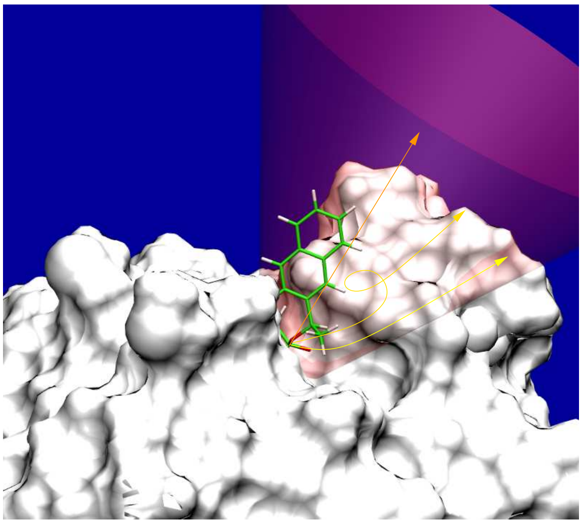


Figure 5.1: Schematic of the TI and illustrative dissociation pathways.

Another problem that affects the results of simulations is linked to the QM/MM potential and its component QM and MM contributions. The force field is expected to be of limited accuracy, as is the semiempirical SCC-DFTB method. Furthermore, when the QM and MM fragments are connected by bonds, force field terms are retained across the boundary[191], which model the system faithfully in its initial state, but are usually not adapted to the product state. Therefore they may exert an artificial bias in the advanced stages of the reaction and in the product.

Other systematic errors may arise from structural problems induced by wrongly assigned protonation states of the amino acid side chains or insufficient hydration of cavities in the protein.

5.3 Activation Free Energy Barriers of Acylation

We first describe how the barriers were extracted from the free energy surfaces, and then justify our choice of parameters for simulations and analysis.

5.3.1 Interpretation of the Surfaces

In chapter 4, a stable TI was a prerequisite to start a minimum energy path calculation and to define the first stationary point for assignment of barriers. In the procedure used here, a stable TI need not exist, since the restraint keeps the system in position and drives it over the entire range of the RC. Close to RC values of 0.0 Å and above, the free energy assumes artificially large and increasing values due to the second restraint, introduced to prevent dissociation of the second C-O bond (*vide supra*). For a surface of type D) in figure 4.8 (page 77), this leads to problems. Usually, the surface was not quite flat and a slight ascent leading to the artificial region (see section 5.1: 'Additional Restraints') around 0.0 Å was noticed. In this case, the barrier MCC→TS was assigned on the basis of geometric considerations: We located the value of the RC with equal probabilities to dissociate via [I] or [II]. As figure 5.2 shows, this happens around RC values of 0.5 Å. The step TI→TS was assigned a barrier of 0.0 in such cases. For surfaces of type A), B), or C), the assignment of barriers was not problematic: first the MCC was located, then the TS as maximum between MCC and the point with RC value of 0.5 Å. The TI was the free energy minimum located at larger values of the RC than the TS. The differences between these values were identified as the required activation barriers.

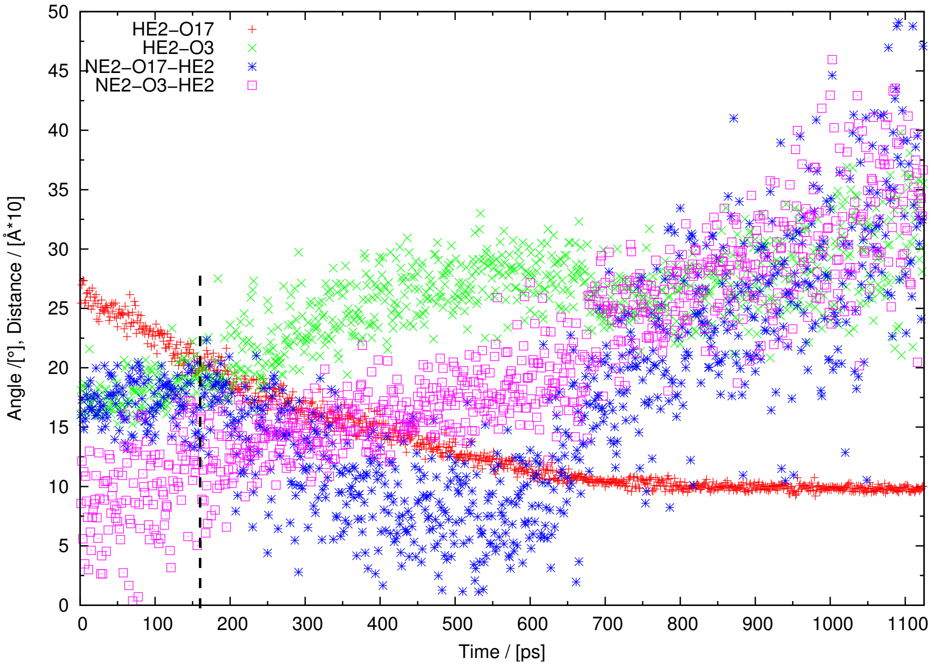


Figure 5.2: Evolution of geometric patterns in umbrella sampling calculations (from simulation of RIII1000). At a RC value of 0.5 Å, after 150 ps of simulation time, the bond lengths HE2-O17 and HE2-O3 become similar.

5.3.2 Choice of Parameters

We have studied the dependence of the computed free energy barriers on the parameters used in the simulations and their analysis, in order to derive a standardised optimum set of parameters. All tests were conducted on model RIII for step [I] of the acylation reaction and with simulations starting from several different snapshots extracted from the associated MM trajectory.

Two setups of simulations were tested, a small one, with 7.5 ps of sampling time per window, and a large one where the dynamics in each window was propagated for 25 ps. Force constants of 45 and 75 kcal/(mol$\mathring{A}^2$) were employed in the window potential.

The WHAM procedure requires a temperature, a bin width, and a convergence parameter to be specified. The temperature was always set equal to the simulation temperature, and the convergence criterium for the free energy constants $\mathrm{F_i}$ (see chapter 2) was set to a value of 0.00001 kcal/mol. For tighter convergence criteria no improvement of the accuracy of the free energy barriers was observed. The WHAM procedure required on the order of 1000-4000 iterations to reach convergence. Although this is not rate limiting for the calculation of a single barrier, the repetitions needed for bootstrap sampling (*vide supra*) can increase the effort considerably.

The importance of the bin width has been a matter of debate[59], and we tested two different

Equilibration Interval	**MCC→TS**	**TI→TS**
Large setup: 25 ps windows		
Bin width 0.1 Å; force constant 75 kcal/(molÅ^2)		
5.0	15.74±0.75 (0.02)	0.75±0.18 (0.03)
12.5	15.87±0.78 (0.03)	0.78±0.19 (0.05)
20.0	15.92±0.80 (0.05)	0.82±0.26 (0.07)
24.75	16.07±0.88 (0.22)	0.90±0.44 (0.30)
Bin width 0.025 Å; force constant 75 kcal/(molÅ^2)		
5.0	15.92±0.71 (0.02)	0.90±0.20 (0.03)
Umbrella Integration		
5.0	15.79±0.80 (-)	0.73±0.23 (-)
Small Setup: 7.5 ps windows		
Bin width 0.1 Å; force constant 75 kcal/(molÅ^2)		
5.0	16.36±2.55 (0.07)	0.73±0.33 (0.10)
Bin width 0.1 Å; force constant 45 kcal/(molÅ^2)		
5.0	15.31±1.40 (0.05)	0.91±0.32 (0.07)

Table 5.1: The influence of different parameters on the free energy barriers. Activation energies are given in kcal/mol, and equilibration intervals in picoseconds. The values denoted ± are standard deviations computed from the set of snapshot simulations, and the values in parentheses are mean values of the statistical error, as estimated using the bootstrap method. For further explanation see text.

bin widths (0.1 Å and 0.025 Å) to check their impact on the converged activation free energy. A comparison with umbrella integration was performed. With our large simulation setup (25 ps per window) several fixed equilibration intervals were tested (5 ps, 12.5 ps, 20 ps, 24.75 ps).

For each choice of simulation parameters 13 simulations starting from different snapshots were performed (12 in the case of the small setup with a force constant of 75 kcal/(molÅ^2)). The trajectory data were evaluated by the WHAM procedure. Our own implementation of WHAM was used. The umbrella integration was performed using J Kästner's original implementation. Results of the statistics are found in table 5.1. The organisation of this table is as follows. The upper part is dedicated to the results of the simulations with the large setup, and the lower part to those with the small setup. In the case of the longer simulations, we varied the equilibration interval, bin width, and evaluation method. If not indicated otherwise, WHAM is used, umbrella integration was applied in one case to validate the results of the WHAM method. For the short simulations only the force constant was varied. The data presented in the columns of the table are the mean values of the activation free energy barriers of step [I] of the acylation reaction calculated from the available sets of 13 (12) simulations, and the associated standard deviations. The values in parentheses are mean values of the statistical error of individual simulations as estimated using the bootstrap Monte Carlo method with 200 samples.

Focusing on the variation of the equilibration period an increase in the value of mean barriers on extension of the equilibration time is realised. In a similar fashion the standard deviations also increase. The mean statistical errors are tiny compared to the standard deviations, a deterioration of statistical accuracy can nevertheless be detected, and for the smallest set of data it increases fourfold. The standard deviation of the barriers mirrors in part systematic errors introduced by incomplete sampling. The mean statistical error estimated from each trajectory is small in contrast and may be neglected with typical sample sizes.

Due to the standard deviation being lowest at 5 ps equilibration, this became our standard equilibration interval, and all other variations in simulation parameters and data analysis were tested on its basis. The bin width was determined empirically. We considered a size of 0.1 Å useful for the present analysis. However, we contrasted this reasonable choice with an extreme choice, where the bin width was set to 0.025 Å. A comparison of the values in table 5.1 shows that this indeed makes a difference of 0.18 kcal/mol in activation free energy barriers. Umbrella integration is independent of the bin width parameter, a comparison with the activation barriers that were derived using this method may therefore serve as a guideline to choose it wisely. The table reveals that umbrella integration produces a mean barrier closer to the one produced with a bin width of 0.1 Å, which was from then on used in all further analyses.

The small setup produced large standard deviations which were considered unacceptable in the context of a study of enantioselectivity. Although the use of a smaller force constant, k=45kcal/(molÅ^2), leads to a smaller standard deviation, it is still about twice as large as in the large setup. Therefore we used the larger setup as the standard in the determination of activation barriers. The low value of k sometimes led to an artificial dissociation as described before, despite the presence of an additional restraint. At k=75 kcal/(molÅ^2) this was not found to be a problem, and this value was used as the default force constant in the window potential. The standard values employed in the simulations are summarised in table 5.2.

Parameters of simulation			
QM-method	SCC-DFTB	MM-method	Charmm22
Temperature	300 K	Number of windows	45
Stepsize	0.5 fs	Sampling per window	25 ps
k of window potential	75 kcal/(molÅ^2)		
Parameters of analysis by WHAM			
Bin width	0.1 Å	Temperature	300 K
Equilibration	10000 steps (5 ps)	Bootstrap	200 samples

Table 5.2: Standard values of umbrella sampling simulations and their analysis. This set of parameters was used unless noted otherwise.

5.3.3 Results

The same set of binding modes as in chapter 4 was used. In addition, we included SVI and RVII, for which no minimum energy paths were computed in chapter 4. The number of simulations per binding mode varied, and emphasis was put on the best performing models identified in chapter 4. SI and RIII were chosen as the main representative of the S and the R-enantiomer, respectively. RI and SIII were included in test calculations, but all attempts failed early in the simulations due to numerical instabilities in the propagation of the trajectory, caused by large repulsive energies of the distorted geometries of the TI. This resulted from application of a bias on the misconfigured tetrahedral group.

The results of all calculations are summarised in table 5.4. The mean values were computed from all barriers available for the respective binding mode. With RIII and SI, we performed 13 simulations for step [I], and the barriers computed from all of them entered the average value and standard deviation in table 5.4. Approximately six months of computer time on a 2.0 GHz Opteron machine were needed for 13 simulations. A total of 23.4 million data points contributed to the average of $\Delta G^{\ddagger}$ of RIII in 5.4. For SI, only twelve simulations entered the averages in table 5.4. SI1000 was excluded since the system took an unaccounted "shortcut" by a concerted motion of binding pocket and the TI (see figure 5.4), which led to an artificially low activation barrier of 13.36 kcal/mol.

An impression of the shapes of the free energy surfaces and the very reasonable convergence of our simulations towards a mean value is conveyed by figure 5.3. The statistical error calculated using the bootstrap sampling was found to be negligibly small (see also section 5.3.2). The extrema are found at the same positions as on the PES, the MCC is located at RC≈ -3.5 Å, and the TS in RIII at RC≈ -1.6 Å. Interestingly, SI exhibits a flat surface of type D) (figure 4.8, page 77), whereas in chapter 4 it was of type A).

The character of the surface of SIV also changes from type D) in optimisations to A) in free energy calculations, but this would be expected owing to the similarity of SIV with SI, which

		$\Delta G^{\ddagger}$		
Model	Snapshot	MCC→TS	TI→TS [I]	TI→TS [II]
RIII	500	14.49	0.74	1.48
	550	16.32	0.51	
	600	16.62	0.59	2.12
	650	14.77	0.66	
	700	15.48	1.10	2.28
	750	15.73	0.68	
	800	15.34	0.66	1.77
	850	16.18	0.51	
	900	16.55	0.74	1.94
	950	15.83	0.87	
	1000	15.29	0.95	2.38
	1050	15.09	0.91	
	1100	16.91	0.86	1.37
SI	500	16.33	0.00	0.0
	550	18.02	0.00	
	600	18.08	0.21	0.2
	650	16.58	0.00	
	700	18.27	0.00	0.00
	750	17.62	0.00	
	800	17.89	0.07	
	850	18.23	0.00	
	900	17.90	0.00	0.1
	950	17.20	0.00	0.0
	1050	17.99	0.27	
	1100	17.02	0.40	0.4

Table 5.3: Barriers (kcal/mol) for all individual simulations in RIII and SI. These binding modes were studied most extensively.

Model	Number of simulations [I] / [II]	Mean $\Delta G^{\ddagger}$±standard deviation MCC→TS	TI→TS [I]	TI→TS [II]
RII	2[1] / 1	20.20±2.31	0.00	0.00
RIII	13 / 7	15.75±0.75	0.75±0.18	1.91±0.39
RIV	3 / 2	19.86±0.63	0.00	0.00
RV	2 / 2[1]	18.86±0.18	0.00	0.00
RVII	5 / 3	15.91±1.49	0.84±0.74	1.30±0.27
SI	12 / 6	17.59±0.66	0.08±0.14	0.14±0.23
SII	9 / 3[2]	17.56±0.60	0.03±0.05	0.00
SIV	3[2] / 3	17.87±1.18	0.09±0.10	0.00
SVI	2 / 3	20.56±2.33	0.00	0.70±1.21

[1] 30000 steps per window, k=50 kcal/(molÅ^2)

[2] 15000 steps per window, k=45 kcal/(molÅ^2)

Table 5.4: Mean $\Delta G^{\ddagger}$ values (kcal/mol) of the acylation reaction and standard deviations (kcal/mol) calculated from the respective sets of barriers.

differ mostly by a rotation of about 180° in ω_4. RIII and RVII are related to each other in the same sense, and thus it is no surpise that they both are of type C), in optimisations as well as simulations. RV changed from type C) to D), RIV and SII remain of type D).

The resulting $\Delta G^{\ddagger}$ values (table 5.4) are considerably higher than the $\Delta E^{\ddagger}$ values calculated from geometry optimisations. This fact should be attributed to more effective relaxation, as discussed below. A comparison with experimental data of ester hydrolysis in chymotrypsin shows that $\Delta G^{\ddagger}$ should be in the range of 13-19 kcal/mol[138]. Our barriers are in this range. Their standard deviations are considered reasonably small. With our most extensively studied binding modes, RIII and SI, the standard deviations are 0.75 kcal/mol and 0.66 kcal/mol, respectively. For SI and RIII all individual barriers are given in table 5.3. It is reassuring that simulations, which were started from different snapshots and were propagated for a comparably long time lead to similar values of $\Delta G^{\ddagger}$ implying a reasonable convergence of parameters in this high dimensional system. When quantitative $\Delta G^{\ddagger}$ values are needed, it does not suffice to rely on a single calculation, as the presence of an outlier in SI (SI1000) demonstrates.

5.4 Enantioselectivity

Enantioselectivity was evaluated as in chapter 4, using the kinetics program (Appendix D) for the simulation of the acylation reaction under the conditions of kinetic resolution. In so doing, we imposed similar conditions as in the experimental setup to determine the enantioselectivity, where the reactants compete with each other for the enzyme. This situation is not accounted for in our atomistic QM/MM models, where only one substrate molecule

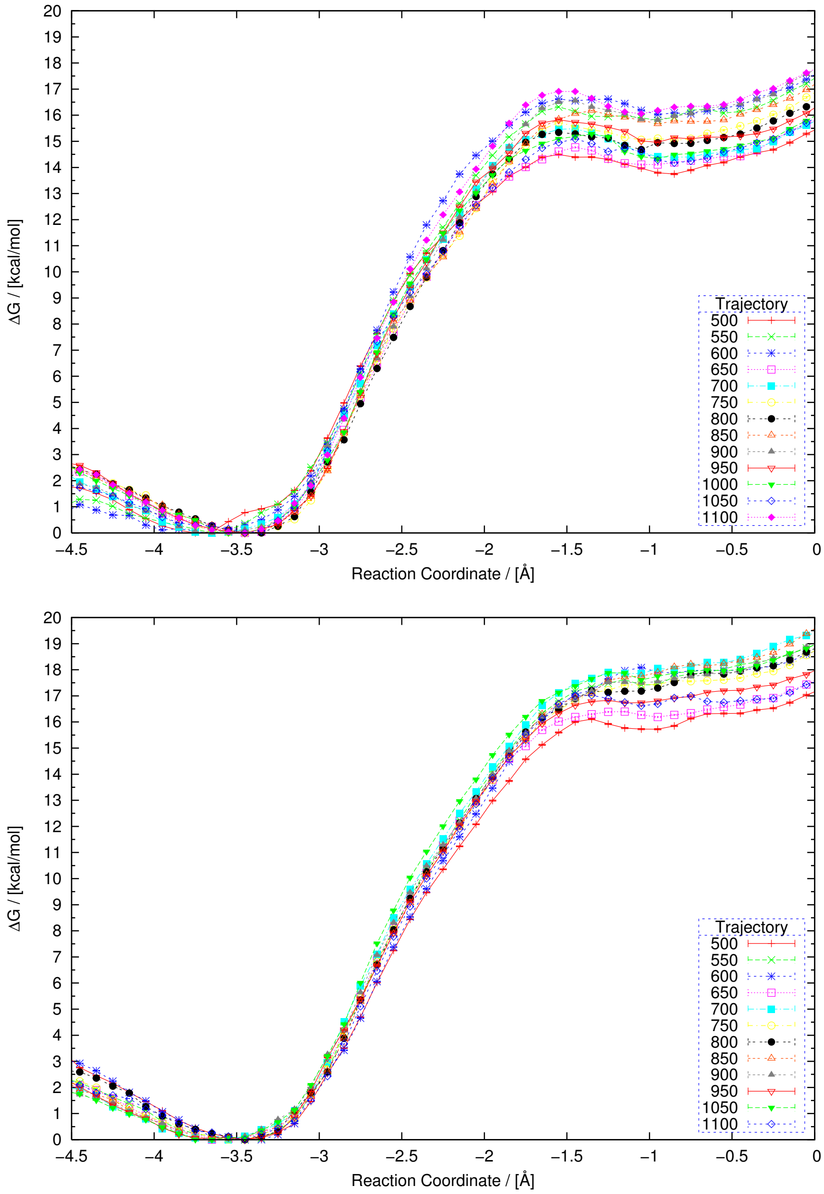


Figure 5.3: Free energy barriers of the first step of acylation for the RIII (top) and SI (bottom) systems. The free energy simulations were started from snapshots of the corresponding MM trajectory, at the times (picoseconds) specified in the legends. Each data point represents the mean value of 200 bootstrap samples. Error bars are attached to every point and represent standard deviations calculated as calculated by bootstrap sampling.

is present at any time. The computed $\Delta G^{\ddagger}$ values thus serve as input parameters for the non-atomistic kinetic model.

The $\Delta G^{\ddagger}$ values for the steps corresponding to k_1 (k_3) and k_{-1} (k_{-3}) (see appendix D) were chosen as 10^{-5} kcal/mol and 14 kcal/mol, respectively. The unbinding of MCC can be assigned an arbitrary value, unless it becomes rate-limiting, in which case enantioselectivity is eliminated. It proved practical to use values close to the acylation barrier since this speeded up the simulations significantly. The reason for this behaviour is rooted in the kinetic treat-

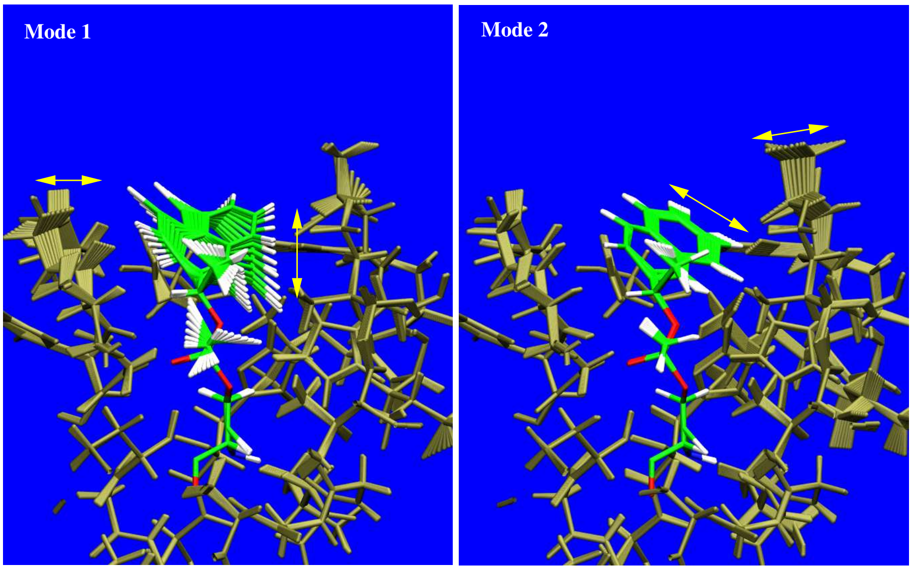


Figure 5.4: Principal component analysis of the atomic fluctuations in the umbrella sampling simulation of SI1000[3][74]. Shown are visualisations of the first two eigenvectors of the variance-covariance matrix, with correlated motions of the atoms. Yellow arrows indicate the most important oscillations of TI and its surrounding. A similar motion was found also for SI500 and SI950, which suggests that these modes are characteristic of binding mode SI. In SI1000, there is an inelastic stretching along the first two eigenvectors around a value of -0.8 Å on the RC, causing the system to leave its previous equilibrium position. The free energy difference for this transformation is unacounted for since this corresponds to a rare event. One could take it into account by extending the umbrella sampling to two dimensions, constructing a bias potential using linear combinations of the principal component eigenvectors[87].

Competing Models		E-value	Apparent $\Delta\Delta G^{\ddagger}$
RIII	SI	5.70 (R)	1.04
RIII	SII	5.01 (R)	0.97
RIII	SIV	8.02 (R)	1.25
RIII	SVI	794.0 (R)	4.00
RVII	SI	10.97 (R)	1.44
RVII	**SII**	**9.72 (R)**	**1.36**
RII	SII	85.93 (R)	2.67
RV	SII	8.99 (S)	1.32
RIII	RVII	1.92 (RVII)	0.39
SI	SII	1.14 (SII)	0.08

Table 5.5: Enantioselectivities resulting from the use of $\Delta G^{\ddagger}$ values, calculated by umbrella sampling calculations, within the kinetic model. The preferred enantiomer is placed in parentheses after the E-value. The apparent $\Delta\Delta G^{\ddagger}$ values were calculated using the relation $\Delta\Delta G^{\ddagger} = \mathrm{RT}\ln(\mathrm{E})$. The definitive value for **E** is typed in boldface.

Model	$\Delta G^{\ddagger}$
RII	20.2
RIII	16.91
RIV	19.86
RV	18.86
RVII	16.37
SI	17.65
SII	17.56
SIV	17.87
SVI	21.26

Table 5.6: Qualitative effective barriers $\Delta G^{\ddagger}$ (in kcal/mol) of the acylation reaction (see section 4.2.2) computed from the values in table 5.4.

ment, which becomes ineffective when the individual steps have very different rates. Deacylation was assigned a barrier of 12 kcal/mol.

As in chapter 4 we compare pairs of binding modes of (R)-NAPHAC and (S)-NAPHAC. Using qualitative arguments (see chapter 4.2.2) we computed effective $\Delta G^{\ddagger}$ values from the data in table 5.4, which are presented in table 5.6.

RVII, RIII, SII, SI and SIV are identified as the five fastest binding modes by the qualitative effective free energy barriers (MCC→TS) in table 5.6. The results with kinetic model of appendix D, which are presented in table 5.5, verify quantitatively that RVII is indeed the best binding mode of (R)-NAPHAC. In competition with RIII we notice that the relative rate of conversion of RVII is larger than that of RIII[1]. Similarly, we find that SII is faster than SI. The direct comparison of the fastest binding modes of each enantiomer, SII and RVII, yields $\mathbf{E} \approx 10$ in favour of RVII, which our best estimate of the enantioselectivity factor.

Comparison of the slow RII, RV, and SVI modes against the more proficient binding modes of the respective other enantiomer show apparently large enantioselectivities. This is misleading, of course, because the fast-reacting species will dominate the enantiopreference.

The order of enantiomers is thus correctly predicted by our model. However the absolute experimental value of $\mathbf{E}$=140[72] is not reproduced quantitatively. The large discrepancy between the experimental and theoretical value $\mathbf{E}$, 140 versus 10, corresponds to a difference in $\Delta\Delta G^{\ddagger}$ of 1.6 kcal/mol.

5.5 Structural Observations

While the umbrella sampling can be performed in a black box fashion to compute free energies of activation, it is useful to consider fluctuations of structural parameters and compare

---

[1] This is formally measured here in terms of $\mathbf{E} = \frac{k_{RVII}}{k_{RIII}}$ (see appendix E).

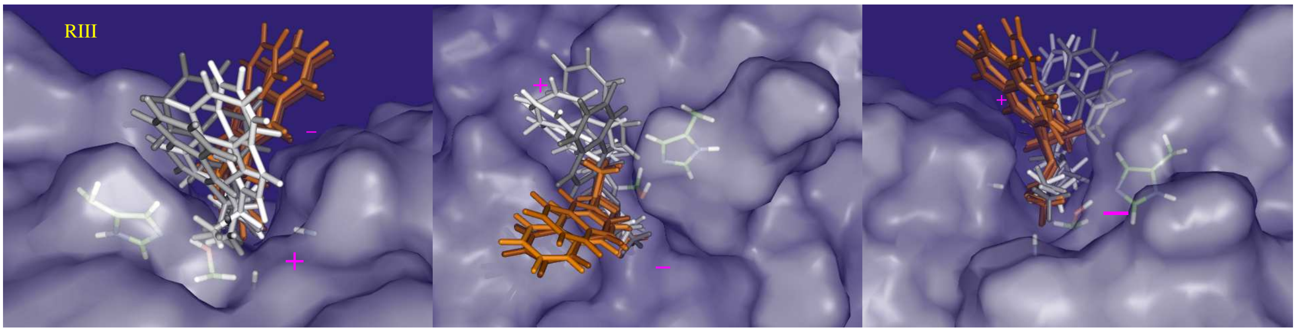


Figure 5.5: Comparison of model geometries for the MCC of the acylation reaction. Geometry optimised models are shown in orange, and umbrella sampling derived structures are coloured grey.

them to the results obtained in chapter 4. With the exception of additional hydration of O15 in RIII (RVII), similar patterns exist in all simulations. Since these could not be related to enantioselectivity they are not discussed for each simulation separately. Instead, we choose RIII as a representative case.

Michaelis Complexes

In the geometry optimisations it was observed that the reaction TI→MCC proceeded as expected with respect to the transformation of the bonding pattern of the TI and the hybridisation of C14 while the naphthyl ring remained in essentially the same orientation (see figure 3.8, page 44). Figure 5.5 demonstrates that the final position of the ligand in the MCC is quite different in the free energy simulations. In the MD based model of the acylation reaction, the ligand tilts towards the surface of the enzyme in the anterior region of BSLA. The location of the ester group is comparable for both methods. Similar observations apply to other binding modes as well.

The tilting of the NAPHAC in the case of RIII could have serious consequences for the free energy surface, especially with regard to displacement of water in the vicinity of the active site. RIII and RVII share the common feature of a water molecule coordinated to O15 in the TI. The fluctuations of the local water molecules were inspected for RIII, again to compare to the optimisation results.

Hydration of O15

To analyse the hydration of O15, we divided the space around this atom into several (spherical) shells. The first extended from 0-2.5 Å, the second from 2.5-3.5 Å, and the third from 3.5-4.5 Å. We defined a hydration index by counting the water molecules within these shells along the RC. The RC was divided into bins, and the number of water molecules in each bin

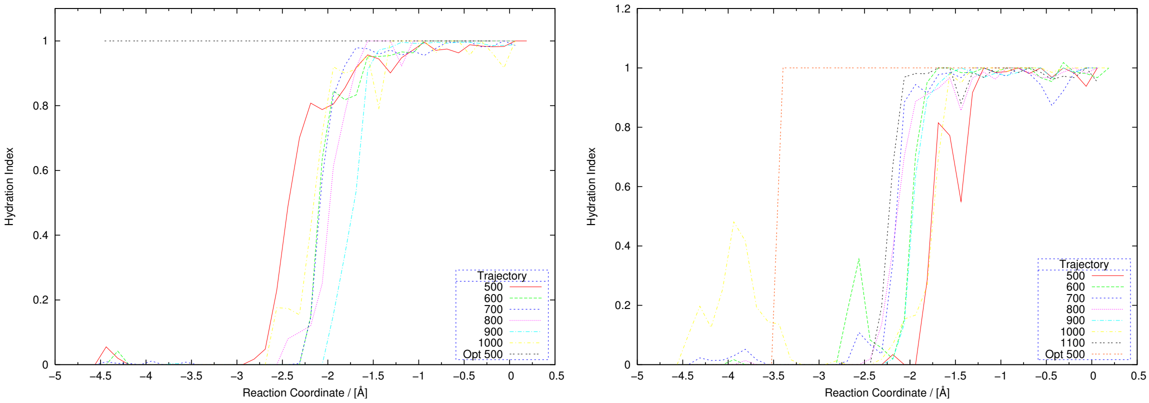


Figure 5.6: Hydration index of model RIII in different umbrella sampling simulations, in comparison to a reaction path optimisation, which is termed ”Opt”. The left picture visualises the third hydration shell, which extends from 3.5 Å to 4.5 Å in step [I], and the right picture visualises the same shell in step [II].

was divided by the total number of times the system visited this bin. Thus fractional water occupation numbers were obtained.

First we consider the hydration in umbrella sampling calculations, and compare both steps [I] and [II] (see figure 5.6). The first two water shells were found to be empty during the simulations. The additional water spent most of the time in shell 3, the hydration index being always close to one around the TI. This hydration persisted well beyond the TS, to about -1.6 Å on the RC. After passing the TS, the hydration index faded to zero. Only in RIII1000 of step [II], O15 regaines a nonzero hydration index by the end of the simulation.

In the geometry optimisations, the first two shells were also found to be empty. Shell 3 was occupied over the entire course of the reaction in step [I], as seen in figure 5.6. In step [II] a water molecule is present in shell three, which moves out of the active site at the PDC. This effect was noticed to lead to a discontinuity in the PES in chapter 4.

The presence of additional hydrogen bonds may be a cause for enantiodiscrimination[22]. In this previous study[22], the microscopic realisation during the acylation reaction could not be investigated since a MM model was used. Here we see some details of the mechanism. During formation of the TI from the MCC, the water molecule slips into the active site and can already exert its stabilising effect at the TS. It thus acts to lower $\Delta\Delta G^\ddagger$ *and* stabilise the TI. The enthalpic effect on the TI is expected to be larger than at the TS, since the charge on O15 will be fully developed only at the TI. The lowering of $\Delta\Delta G^\ddagger$ in [I] seems to come at a price, since appreciable barriers in step [II] are present only for RIII and RVII. A small barrier is also noted for SVI, but here the additional water cannot counteract the large intramolecular strain of the TI in this binding mode (see last column in table 3.6).

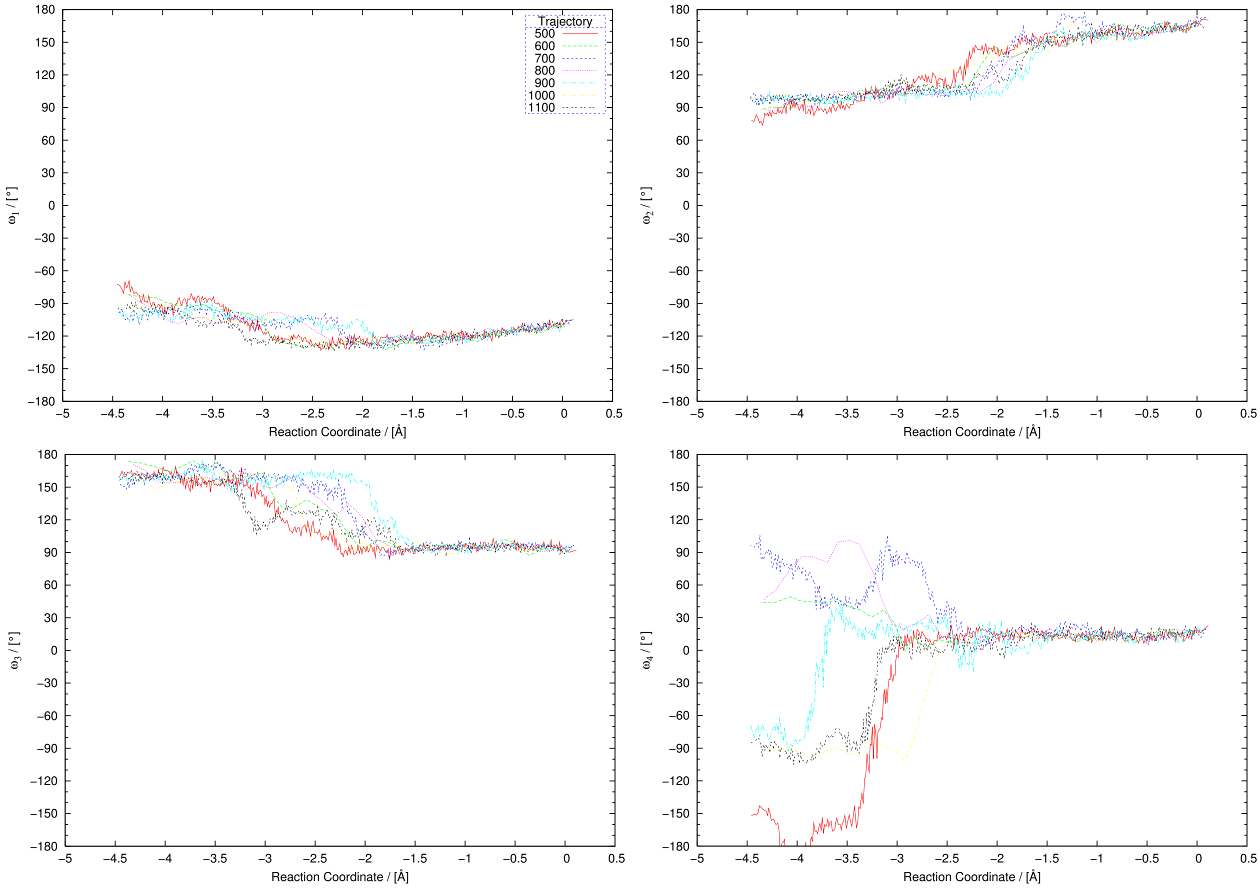


Figure 5.7: Evolution of torsion angles ω_1-ω_4 in umbrella sampling calculations on model RIII in different snapshot simulations. The snapshots were extracted from the MM trajectory of binding mode RIII (see chapter 3), at the time (in picoseconds) which is specified in the legend. Each curve belongs to a single simulation.

Fluctuations of Torsions

In chapter 4 it was realised that torsion angles ω_1-ω_4 did not provide an intuitive picture of the energetics of the acylation reaction. Their often significant changes during the reaction paths mainly reflect the change in hybridisation at C14. Here we inspect the fluctuations occuring in an MD simulation on the basis of model RIII.

Figure 5.7 shows the evolution of torsions ω_1 to ω_4 along the RC in several simulations. For torsions ω_2 and ω_3 a transition between two states is observed in the interval between -1.5 Å and -3.5 Å on the RC. In the case of ω_4 the behaviour is less ordered, since individual trajectories end at values for ω_4 of $90°$ and $-90°$ on the one hand, and $-150°$ and $45°$ on the other hand. These pairs correspond to almost equivalent states with respect to the rotation of the naphthyl ring. Although the final states are very similar for the torsions ω_1-ω_3 in the set of snapshot simulations, they are attained at significantly different values of the RC. Since free energy is a function of state, however, the values of $\Delta\Delta G^{\ddagger}$ should be independent of the actual paths. The rather tight band of free energy value for the different simulations in figure 5.3 demonstrates that this seems indeed to be the case.

Dissociation Pathways

One of the most interesting questions in enzyme catalysis is the process of substrate binding and its free energy surface. Although advanced procedures are necessary to calculate accurate free energies of binding, it would be instructive to have a qualitative view of the process. In principle, umbrella sampling can be used to obtain important contributions to the ΔG of binding.

As reasoned in chapter 4, the MCC is formed on coordination of O15 with Ile12 of the oxyanion hole. The umbrella sampling simulations show that this distance stays approximately constant over the whole range of the RC (figure 5.8), even for those values below -3.5 Å, where the MCC is assumed to decompose, as the increase in free energy suggests. Only the distance to Met78 is observed to change for values below -3.5 Å. Upon binding of the substrate, this distance also assumes a reasonably stable value. The substrate is thus seen to coordinate first to Ile12 and then slide further into the binding pocket as the distance to Met78 decreases.

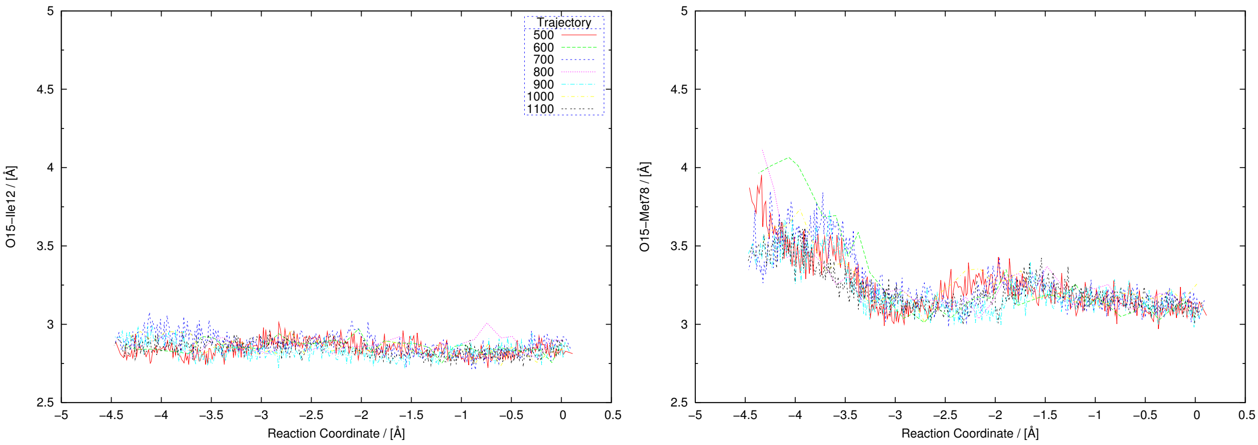


Figure 5.8: Evolution of the distances of O15 to the nitrogen atoms of the oxyanion hole in umbrella sampling calculations on binding mode RIII.

We have attempted to estimate the change in free energy upon binding in an umbrella sampling calculation. The simple distance RC of equation 2.6 promised to be more useful here than our standard RC (equation 2.8). Such a RC is disadvantageous close to the TI where hydrogen transfer between O17 and NE2 occurs concomitant with the approach of atoms C14 and O17, but it is considered to be reasonable close to the MCC, where HE2 is always bound tightly to O17.

The biased distance was chosen to be C14-O17, and the simulation was initiated from the last configuration in window 30 of the umbrella sampling simulation of RIII500. This corresponds to a value of -2.9 Å in terms of the standard RC. Our sampling was executed in 81 windows between 2.1 Å and 10.1 Å, and 15000 MD steps were computed in each window. The force constant k was set to 75 kcal/(molÅ^2).

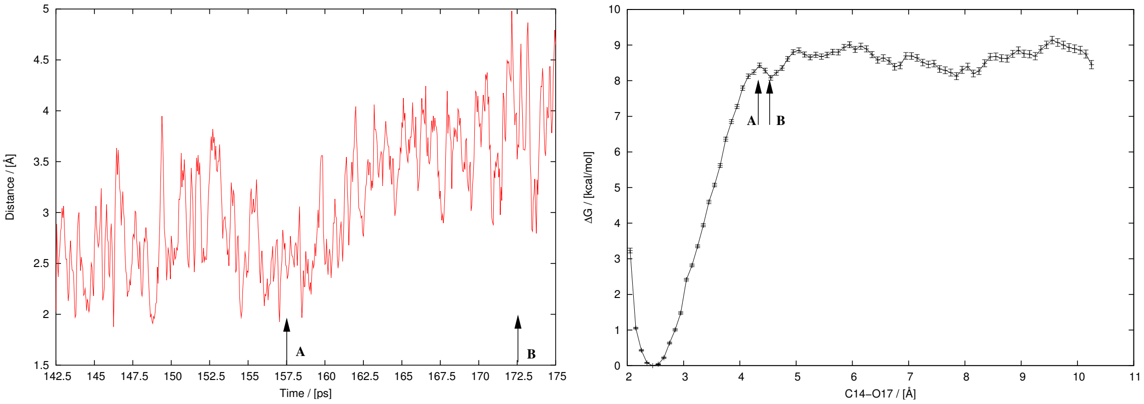


Figure 5.9: Plot of the distance between O15 and the hydrogen atom of the oxyanion hole residue Ile12 on the left hand side, free energy profile for the binding process on the right hand side with error bars (200 bootstrap samples). Positions of the arrows indicate the point in time at which snapshots were taken for visualisation in figure 5.10.

The free energy profile for this process is shown in figure 5.9. It features a deep well with a minimum around 2.5 Å. This distance C14-O17 characterises the MCC. Escape from the well requires a $\Delta G^{\ddagger}$ of about 9.2 kcal/mol and is completed at a C14-O17 distance of 5 Å. An incisive event takes place at 4.2 Å, where an inflection point is found on the profile. This event is reflected in an increase in the Ile-O15 distance. To visualise this stage of the dissociation process two snapshots were extracted from the simulation corresponding to the period before the event, and thereafter. These configurations show that NAPHAC slips out of the active site, as the contact to the oxyanion hole is finally lost. In the final configuration of the simulation (figure 5.11), the contacts with the enzyme are almost broken, which is why one should not expect convergence beyond this point.

The free energy profile further shows that entry of the NAPHAC into the active site of BSLA can proceed without any barrier, thus justifying our choice of very large k_1 (k_2) in the kinetic model (Appendix D).

5.6 Two-Dimensional Free Energy Surfaces

As an extension of our one-dimensional free energy profiles we created a two-dimensional surface of the acylation reaction using a two-dimensional window potential. The choice of

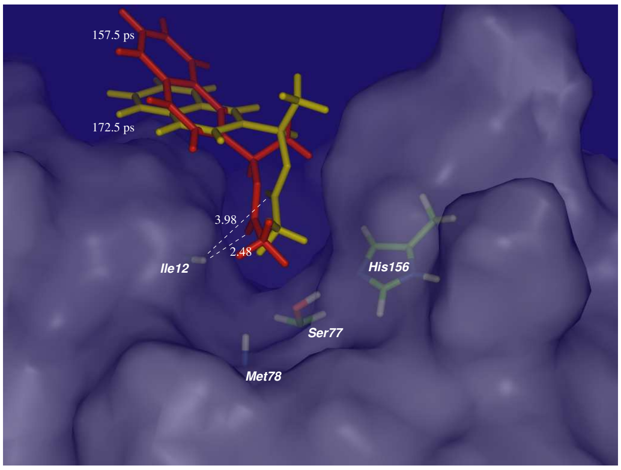


Figure 5.10: Two snapshots from the simulation of the dissociation event. Two NAPHAC molecules located close to the active site are shown in red and yellow colour. The structures correspond to an earlier (red) and later (yellow) configuration on the dissociation pathway. The red structure belongs to position A on the RC in figure 5.9, and the yellow configuration to position B in the same plot. The red structure is still bound to the oxyanion hole (Ile12-O15), but this coordination is lost in the yellow configuration. For further explanation see text.

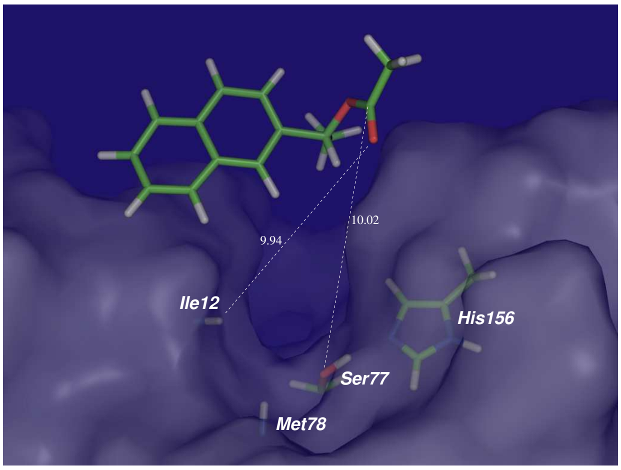


Figure 5.11: Geometry of the final snapshot in the simulation of the dissociation event. The NAPHAC has almost lost contact with the enzyme.

the potentials enables a continuous and therefore seamless calculation, which complements the one-dimensional representations.

In one dimension, a RC of the type shown in equation 2.7 was applied as $r_{O17-C14} - r_{C14-O3}$. The restraint in the other dimension was of the simple harmonic distance type (equation 2.6) and acted on $r_{NE2-HE2}$. As for all other calculations, the TI was found to be the optimum starting point. At the TI, both bonds C14-O17 and C14-O3 are essentially equivalent and $r_{O17-C14} - r_{C14-O3}$ assumes a mean value of 0.0 Å. In the grid calculations, $r_{O17-C14} - r_{C14-O3}$ was initialised at 0.0 Å, and scans proceeded towards the positive and negative axis in separate simulations, which were executed in the same fashion as the one-dimensional scans.

RC $r_{\mathrm{NE2-HE2}}$ remained at fixed positions in these scans and was incremented by an external Python script, which dispatched the jobs in an autonomous fashion. The grid was computed for the interval [-1.6:1.6]Å for $r_{\mathrm{O17-C14}} - r_{\mathrm{C14-O3}}$ and for [1.0:2.0]Å for $r_{\mathrm{NE2-HE2}}$, in both cases using a stepsize of 0.1 Å. A force constant of 400 kcal/(molÅ^2) was used for $r_{\mathrm{NE2-HE2}}$, and 50 kcal/mol for $r_{\mathrm{O17-C14}} - r_{\mathrm{C14-O3}}$. 20000 MD steps were performed in each window at a stepsize of 0.5 fs. The total sampling time in the 363 windows was thus 3.63 ns. A Grossfield implementation of the WHAM algorithm was used to compute the two-dimensional free energy surface[86], the result is visualised in figure 5.12.

The most probable reaction path can clearly be identified in both views of the free energy surface. It proceeds from the MCC in the region [2.0:1.2] over a barrier at [1.2:0.5], identified as the TI, towards the PDC ([2.1:-1.65]). The shape of the free energy surface is approximately symmetric, as expected. From the scan we obtain a $\Delta G^{\ddagger}$ value of 11.5 kcal/mol, which is too low. The TI is stabilised by 1.5 kcal/mol.

Two-dimensional surfaces are too expensive computationally for repeated calculations and have large requirements concerning disk space. The test here served illustrative purposes and validates the assumptions about the shape of the free energy surface, different variations of which were considered in figure 4.8. With 3.63 ns of sampling, the free energy surface is considered not to be very accurate. An estimate of the systematic error could be obtained by repeating the calculation many times which is beyond the scope of this work.

5.7 Conclusion

We performed free energy calculations to determine individual activation barriers $\Delta G^{\ddagger}$ of the acylation reaction catalysed by BSLA with both enantiomers of NAPHAC as substrate. The enantioselectivity was then computed in a kinetic model using the activation barriers as input parameters.

Our results give the correct order of enantioselectivity with NAPHAC. The absolute value of the selectivity **E**, however, was not reproduced quantitatively. The experimentally determined enantioselectivity is on the order of **E**=140 in favour of (R)-NAPHAC, whereas the theoretically determined value is around **E**=10. The absolute error in terms of $\Delta\Delta G^{\ddagger}$ is 1.6 kcal/mol.

Our present study indicates that quantitative QM/MM free energy simulations are still of limited utility in enzyme engineering, e.g., in the optimisation of enzymes with regard to enantioselectivity. Restrictions arise from the remaining uncertainty in the calculated values of $\Delta\Delta G^{\ddagger}$ due to systematic errors, introduced mainly by an incomplete sampling of

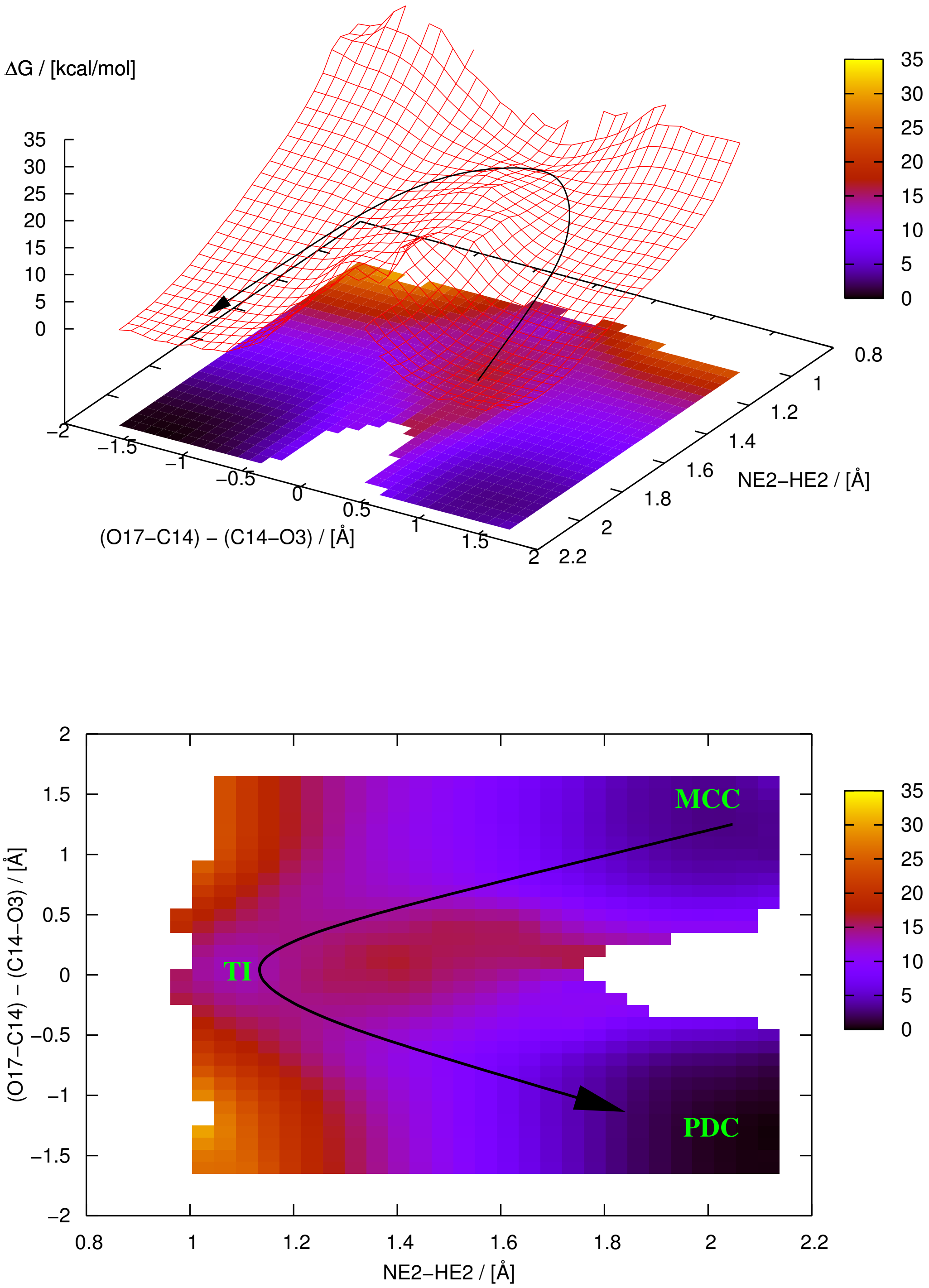


Figure 5.12: Free energy surface in two dimensions. Two restraint potentials were used at the same time. Arrows indicate the direction in which the reaction proceeds, starting from the MCC.

configuration space, and the limited accuracy of the QM/MM potential. Moreover, such calculations consume large amounts of computer time to evaluate enantioselectivity in each single case. In order to guide experimental work in practice, one would need to study either many mutants of the enzyme or many different substrates. This is at present still a too ambitious task for computational chemistry, and cannot be considered to become a standard procedure in the near future.

Chapter 6

A Force Field for the Tetrahedral Intermediate

6.1 Biomolecular Force Fields

Force fields are a necessity in the study of large systems such as enzymes. Although it is nowadays possible to perform quantum mechanical calculations on systems containing several thousands of atoms, the range of these methods is limited to the evaluation of selected geometries and not suited to the exploration of conformational space by MD which requires the evaluation of energies and gradients for large numbers (>1000000) of molecular configurations. Furthermore the force fields used in the simulation of biomolecules are targeted towards a limited class of molecules and are highly optimised to reproduce their characteristics well. Their high degree of specialisation is both a benefit and a curse. On the one hand we can expect to obtain better results than with low level MO or DFT methods, on the other hand it makes extensions of the existing set of parameters a tedious task that has to be tackled each time new ligands or unusual modifications of amino acids or nucleic acids are introduced.

The methodology and philosophy of parametrisation shows significant differences for most force fields, and the ones specialised on biomolecules are no exception. To guarantee proper integration into the existing set the development of new parameters should follow the published methodologies. The force field to be extended in this work is Charmm22[108] for amino acids (Note that nucleic acids are contained in a different set of parameters[66]). Other important force fields for biomolecular science are OPLS-AA[110][106], Amber[160] and Gromos[209][185].

Functional form

The potential energy of biomolecular force fields is represented using an empirical functional form that is a sum of several types of contributions, which in turn are sums over energy expressions. The potential is pairwise additive, which allows simple decomposition of energy contributions and the study of interactions of subsets of a molecular system. It contains both bonded and nonbonded contributions, and the contributing terms are determined by the connectivity pattern. The potential energy function of Charmm22 is as follows:

$$\begin{aligned}
\mathrm{E} &= \sum_{\mathrm{bonds}} \mathrm{K_r}(\mathrm{r}-\mathrm{r}^0)^2 + \sum_{\mathrm{angles}} \mathrm{K}_\theta(\theta-\theta^0)^2 \\
&+ \sum_{\mathrm{UB}} \mathrm{K_r}(\mathrm{r}-\mathrm{r}^0)^2 + \sum_{\mathrm{dihedrals}} \mathrm{K}_\phi(1+\cos(\mathrm{m}\phi-\phi^0)) \\
&+ \sum_{\mathrm{impropers}} \mathrm{K}_\varphi(\varphi-\varphi^0)^2 + \sum_{\mathrm{coulomb}} \frac{\mathrm{q_i q_k}}{\mathrm{Dr_{ik}}} \\
&+ \sum_{\mathrm{VdW}} \epsilon_{\mathrm{ik}} \left[\left(\frac{\mathrm{R_{min,ik}}}{\mathrm{r_{ik}}} \right)^{12} - 2 \left(\frac{\mathrm{R_{min,ik}}}{\mathrm{r_{ik}}} \right)^{6} \right] \qquad (6.1)
\end{aligned}$$

All bonded interactions except the dihedral term are modeled using a harmonic approximation. $\mathrm{K_r}$ and K_θ are bond stretching and bond angle bending force constants, r^0 and θ^0 denote the equilibrium bond distance and equilibrium bond angle. The torsion energy is represented by a cosine function, K_ϕ is the force constant, m the periodicity and ϕ^0 the phase angle. Dihedral angles can be treated as Fourier series to synthesize complicated torsional profiles and this technique was also used for the TI. The third term in equation 6.1 is the Urey-Bradley term, which is a distance constraint between the first and third atom of a bond angle, K_r is the force constant and r^0 the equilibrium distance. The fifth term represents the improper torsion to maintain planarity of planar groups. Here K_φ is the force constant for the improper torsion and φ^0 the improper dihedral angle (usually set to 0). Nonbonded interactions are described by a Coulomb and Lennard-Jones potential (last two terms above). The charges are located on the atomic centers.

The topology defines the members of a molecular fragment and the connectivity of atoms within each such fragment. Proteins are molecules built from basic building blocks, the amino acids, and for each amino acids there exists an entry in the *residue topology file*. Any molecule for which a calculation is to be run has to be *imported* into the force field driver program in this manner.

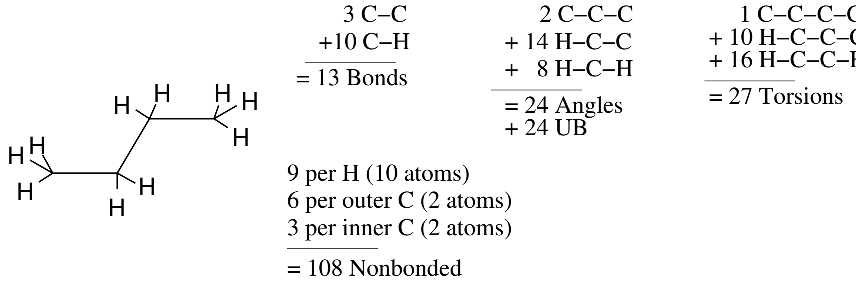


Figure 6.1: The butane molecule as an example for the assignment of force field parameters.

Force field terms are assigned based on the following rules:

1. Bond stretching: All pairs of bonded atoms.
2. Angle bend and Urey-Bradley: Between all pairs of bonds that share one common atom.
3. Dihedrals: All combinations of atoms attached to one end of a central bonded pair of atoms and the other end.
4. Impropers: Are not automatically assigned and must be requested in the residue topology definition.
5. Nonbonded: Between all pairs of atoms that are separated by at least three bonds or are located in different, unbonded residues.

The parameters in the force field expression are determined on the basis of predefined atom types with the exception of charges, which can be redefined for each residue. All parameters contain a geometric component specifying the minima of interactions and a force component specifying the stiffness or steepness of the valleys. Each atom type requires specification of VdW parameters to enable proper treatment of nonbonded interactions. The existence of bonds prompts the presence of bond parameters for all combination of atom types found in any residue definition. Similarly, parameters are required for all other bonded terms that are known to be present from the bonding pattern.

6.2 Parametrisation Procedure

In the following we describe the setup of a new residue and the derivation of new parameters which fit into the existing Charmm22 set. Parametrisations of other molecules within the Charmm22 framework are described elsewhere[172][157].

Our focus was on the development of new charge and torsion parameters. Charge parameters are important since the Coulomb interaction contributes significantly to the nonbonded energy and is a long range potential. Coulomb interactions are computed at QM level in QM/MM calculations and a seamless match of the MM and QM potential surfaces is needed for an efficient transition from MM preparation to QM production stage. Therefore the optimisation of charges should be given due consideration.

New charge parameters are determined for the Charmm22 force field using a standard recipe. The principle idea is to reproduce *ab initio* interaction energies between model compounds and with water molecules with the force field. The charges, which are centered on the atoms, are adjusted until a close match with the *ab initio* derived values is obtained. Thus, water serves as a device to measure interaction capabilities: the interaction energy between other, non-solvent fragments is assumed to be reliable when using the same parameters as with water. For biomacromolecular systems water is the natural choice for a mediator molecule, due to its small size and ubiquity. During the fitting procedure (as well as in simulations performed later) water is constrained to TIP3[107] geometry, where the OH distance is set to $0.9572\,\text{Å}$, and the HOH angle fixed at $104.52°$. H-bonded complexes are built between the fragment to be parametrised and TIP3 water. The fragments are kept internally rigid and one or two nonbonded degrees of freedom are optimised. All calculations are done at HF/6-31G* level of theory. The following modifications are introduced[157] after the optimisation: In the case of neutral complexes the interaction energies are scaled by a factor of 1.16. In the case of both neutral and charged complexes, the hydrogen bond lengths are shortened by 0.2 Å. The geometry thus computed serves as input to the force field calculation, which must reproduce the *ab initio* interaction energy. As an initial guess of the charge parameters, Mulliken charges provide useful starting points.

For the required VdW parameters, we adopted suitable values from the Charmm22 set of parameters. Development of these parameters is involved since no standard recipe is available as in the case of the Coulomb interactions. Often they have been parametrised against thermodynamic properties derived by molecular simulation[134][111]. Attempts to obtain parameters using rare gas atoms as probes of the VdW surface[233][232] are not yet established as a standard tool. Nevertheless VdW parameters are important, also within the QM/MM framework[173] and should be chosen with care.

Torsion angles are the critical degrees of freedom in conformational searches since they can often assume a broad spectrum of values and have complex potential energy surfaces. For parametrisation, relaxed PES scans are performed at the *ab initio* level along the torsions of interest. Optimisations must then be repeated at the MM level to reproduce the *ab initio* data.

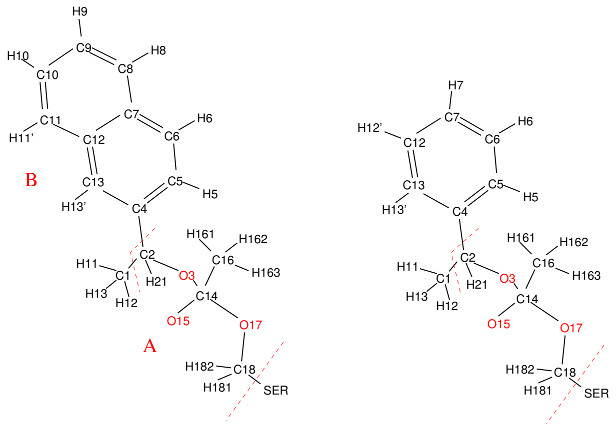


Figure 6.2: The TI for naphtylethyl ester (left) and benzylethyl ester (right). Red lines indicate the fragmentation pattern and identify the tetrahedral moiety *A* as a logical core unit which can be reused, and *B* as an exchangeable part. This trivial example demonstrates that no changes have to be to the core, only the *side chain* belonging to the substrates' alcohol moiety is modified.

Torsion parameters are then adjusted until the profiles match. In most cases several periodicities m (equation 6.1) were mixed to obtain a better fit. In addition, one always has to keep in mind that energy terms other than torsions contribute to the latter. The Coulomb interaction, for example, can have a fairly large influence on the parametrisation. Thus it is a good idea to derive charges first, and then torsions. Manual adjustment of torsion parameters is likely to consume on the order of several days in the absence of fitting experience.

Bonds and bond angles are rather inflexible degrees of freedom in comparison to torsion angles, and they have simple potential surfaces characterised by single minima. We found it unnecessary in most cases to reparametrise force constants of bonds and bond angles. Nonetheless, we adjusted the equilibrium values of these terms to match values extracted from *ab initio* geometry optimisations of small model compounds.

One important aspect of the parametrisation was reusability. Since we were interested in the simulation of a range of TI's derived from acetic acid esters we wanted to keep most of the parameters if possible and fit only for the new fragments. The Charmm22 parametrisation philosophy accommodates this demand owing to the decomposition scheme that is used (see figure 6.2). In Charmm22 a target molecule is first decomposed into small fragments. This is similar to thinking in *functional groups*, which is common to chemistry. We classify the central moiety of the TI as fragment A in figure 6.2. Fragment B is the alcohol moiety of the TI and should be exchangeable. A trivial example is the exchange of the aromatic naphthyl ring system with a phenyl ring. While this changes the number of atoms it does not necessitate any adjustments in subunit A. Similary we built a TI of 1,4-diacetoxycyclopentene (DIAC), the parametrisation and modelling of which is described in appendix A. Core fragment A was thus excised, saturated at C2 and C18, and used to determine charge parameters via complexes with water, which are shown in figure 6.3.

The torsion parameters require additional considerations since they pass the boundary between fragments A and B in figure 6.2, and also between A and the backbone of the serine side chain. Therefore they also pass the QM/MM boundary, which causes some of them to be retained as coupling terms. A reasonable parametrisation is likely to enhance the quality of results at the QM/MM stage. Our standard set of torsion parameters enables the coupling of different types of fragments to A.

6.3 Results

All model compounds were geometry optimised using DFT at the BLYP/6-31+G* level in the case of charged species and BLYP/6-31G* in the case of uncharged species. For further refinement all structures were reoptimised at the B3LYP level using the same basis set as before. The torsion PESs were generated from MP2/6-311+G* single point energies at the DFT optimized geometries. Gaussian03[71] was used for all calculations.

Atomic Charge Parameters

New parameters were determined for fragment A in figure 6.2. We decided not to optimise all charges, but only those of tetrahedral carbon C14, oxyanion O15, and oxygen atoms O3 and O17. The methyl groups around carbon atoms C2, C16, and C18 were considered to be well represented by the incremental system that is used in Charmm22 to determine charges of most aliphatic groups. Here one assigns a charge of +0.09 to each hydrogen, and the sum of increments of -0.09 to aliphatic carbon atoms for each hydrogen atom attached to it.

Thus only four partial charges remained for assignment. The charges on oxygen atoms O3 and O17 were treated as being equal since these atoms can be considered equivalent at the MM level. Hence three charges were varied in the TI whose total charge was fixed at -1. The complexes used in this procedure are shown in figure 6.3, the corresponding optimised geometrical parameters are given in table 6.1, and the interaction energies are listed in table 6.2. Our final set of charges is specified in table 6.3.

Apart from interaction energies we used the value of the dipole moment and the dipole moment vector to control the quality of our parametrised charge distribution compared to the *ab initio* derived values. This comparison was done for the isolated fragment A. The results shown in figure 6.4 demonstrate that the computed dipole moments are of comparable size and that the dipole moment vectors point into the same direction; this is not necessarily

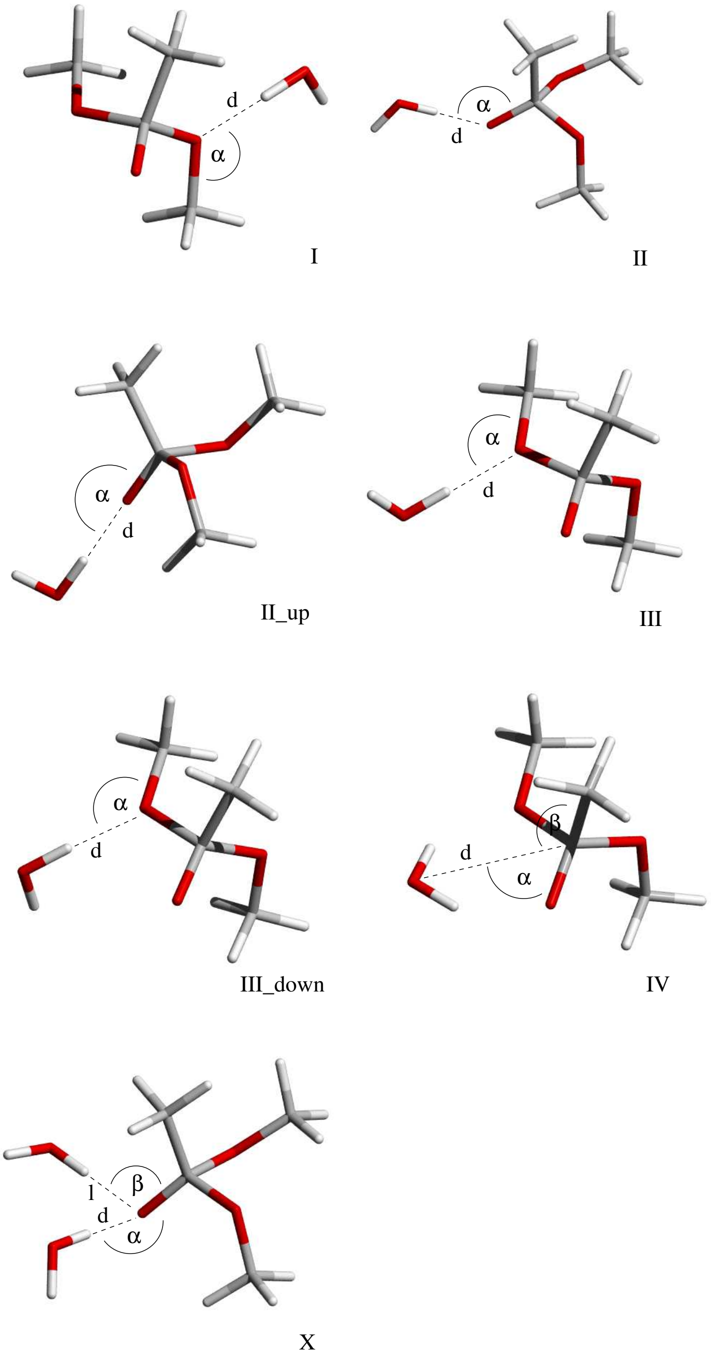


Figure 6.3: Complexes used in the optimization of atomic charges in the TI for the Charmm force field. Degrees of freedom that were optimized are indicated.

the case for all charge distributions which fit the interaction energies. A working force field model should reproduce the dipole properties to generate sensible QM/MM starting gometries. The differences in interaction energies (table 6.2) between the MM and QM representations of the target molecule are of the same order as found in other studies[157]. It should be kept in mind that complex continuous charge distributions are approximated by a sim-

Complex (figure 6.3)	d / [Å]	l / [Å]	$\alpha/[°]$	$\beta/[°]$
I	1.69		114.81	
II	1.59		126.07	
II_up	1.69		179.93	
III	1.70		114.8	
III_down	1.71		119.25	
IV	3.15		56.31	104.36
X	1.63	1.66	113.91	113.57

Table 6.1: Geometrical parameters optimized at HF/6-31G* level, the resulting distances have been reduced by 0.2 Å [157].

Complex (figure 6.3)	Charmm22	HF/6-31G*	Deviation
I	-9.88	-11.19	+1.31
II	-16.13	-16.81	+0.68
III	-9.37	-12.13	+2.77
III_down	-11.80	-12.92	+1.12
IV	-19.95	-19.03	-0.91
II_up	-15.64	-15.39	-0.25
X	-28.20	-30.40	+2.21

Table 6.2: Complexation energies (kcal/mol) of the model TI with water in TIP3 geometry. Complex X contains *two* molecules of water.

ple, discrete, and atom-centered charge distribution. Wennmohs and Schindler have shown that a discrete charge distribution can be enhanced using non atom-centered charges[226]. Especially interactions with oxyanion O15 may be improved using such a more complex representation, which, however, would not be consistent with the Charmm22 force field and was therefore not attempted.

Internal Parameters

Parametrisation of the intramolecular force terms was carried out against torsion potentials, normal modes and harmonic vibrational frequencies derived from *ab-initio* calculations.

Atom (→ figure 6.2)	Type	Reference	Charmm22	Mulliken HF/6-31G*
C14	CTET	CTL1	0.55	0.98
O15	OC	OC	−0.75	-0.83
O3	OSL	OSL	−0.40	-0.73
O17	OSL	OSL	−0.40	-0.71
C2/C16/C18	CIL3	CTL3	−0.27	-0.12/-0.49/-0.12
H$ ($=21,22,...)	HAL	HAL	0.09	≈ 0.1

Table 6.3: Parametrised partial charges (in e). Charmm notation is used in the second and third column, see file 'par_all27_prot_lipid.prm'.

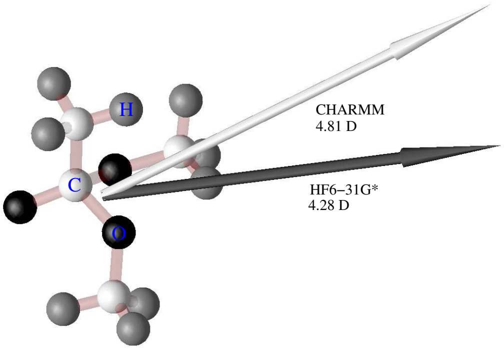


Figure 6.4: Dipole moment vectors, originating at the center of charge (standard orientation in Gaussian03).

Torsions ω_1-ω_4 are characteristic of the TI and require parametrisation. We usually did not compare them against *ab initio* data over the entire interval $[-180° : 180°]$, but only over a sensible range of angles. We used different model compounds and always chose the simplest molecular representation to avoid spurious effects caused by peripheral degrees of freedom. These models were derived from fragment A of figure 6.2, and are shown in the diagram of each individual torsion PES.

Fragment A was the model of choice for ω_1 and ω_2, (see figures 6.5 and 6.6). An additional methyl group was required to parametrise ω_3 (figure 6.7). This torsion influences the coupling of TI and the backbone of serine. The PES for the rotation of the phenyl or naphthyl ring is mapped by torsion ω_4 (figure 6.8). In figures 6.5-6.8 we present the original *ab initio* PES, the PES as optimised using our new Charmm22 force field, and the *ab initio* energies calculated at the stationary geometries of Charmm22. In this way we show that the energies and intramolecular forces of our new force field compare favourably to those of the *ab initio* model.

Our final parameters are given in table 6.4. Torsions ω_1 and ω_2 had to be synthesised using three cosine functions with different periodicities n. As the graphs in figures 6.5 and 6.6 show, both ω_1 and ω_2 are modelled rather accurately by Charmm22. For ω_3 and ω_4, the force field and *ab initio* curves are similar overall, but there are also qualitative differences at several positions. Figure 6.7 reveals weaknesses around $-90°$ down to $-180°$, and from $60°$ to $120°$. A minimum at $-180°$ is reproduced by the force field; in the *ab initio* models this minimum is flat, however, and extends up to $-90°$. The minimum predicted by the force field is narrow, and the system has to spend 2 kcal/mol to reach a value of $90°$. One

$V_{torsion} = K_a(1 + cos(n\varphi - \delta))$				
Torsion	Reference	K_a	n	δ
OSL CTET OSL CIL$	OBL CL OSL CTL3	**1.0**	**1**	**0.00**
OSL CTET OSL CIL$ ($ = 1,2,3)	OBL CL OSL CTL3	**0.8**	**2**	**0.00**
OSL CTET OSL CIL$	OBL CL OSL CTL3	**0.3**	**3**	**0.00**
CIL3 CTET OSL CIL$	X CTL1 OSL X	**2.8**	**1**	**0.00**
CIL3 CTET OSL CIL$ ($ = 1,2,3)	X CTL1 OSL X	**0.4**	**2**	**0.00**
CIL3 CTET OSL CIL$	X CTL1 OSL X	**1.4**	**3**	**0.00**
OC CTET OSL CIL$	OBL CL OSL CTL3	**-0.3**	1	**0.00**
CIL3 CIL1 CA CA	CT3 CT2 CA CA	**0.1**	2	**0.00**
OSL CIL1 CA CA	CT3 CT2 CA CA	**-0.3**	2	**0.00**
CIL3/CA/CT1 CIL1/2 OSL CTET	X CTL1 CTL2 X	**0.3**	3	0.00
CTET OSL CIL$ HAL	X CTL1 CTL3 X	**0.1**	3	0.00
OSL CTET CIL3 HAL	X CTL1 CTL3 X	0.2	3	0.00
OC CTET CIL3 HAL	X CTL1 CTL3 X	0.2	3	0.00
X CIL2 CIL3 X	X CTL2 CTL3 X	0.16	3	0.00
X CT1 CIL2 X	X CT1 CT2 X	0.2	3	0.00
X CIL1 CIL3 X	X CT1 CT3 X	0.2	3	0.00
NH1 C CT1 CIL2	NH1 C CT1 CT2	0.0	1	0.00
CIL2 CT1 NH1 C	CT2 CT1 NH1 C	1.8	1	0.00
O C CT1 CIL2	O C CT2 CT2	1.4	1	0.00
H NH1 CT1 CIL2	H NH1 CT1 CT2	0.0	1	0.00

$V_{improper} = K_a(\varphi - \varphi_0)^2$			
Improper	Reference	K_a	φ_0
CPT CPT CA CA	-	100.0	0.000

Table 6.4: Torsions: New parameters printed in bold letters.
Wildcards are labeled 'X'.
References taken from 'par_all27_prot_lipid.prm'.

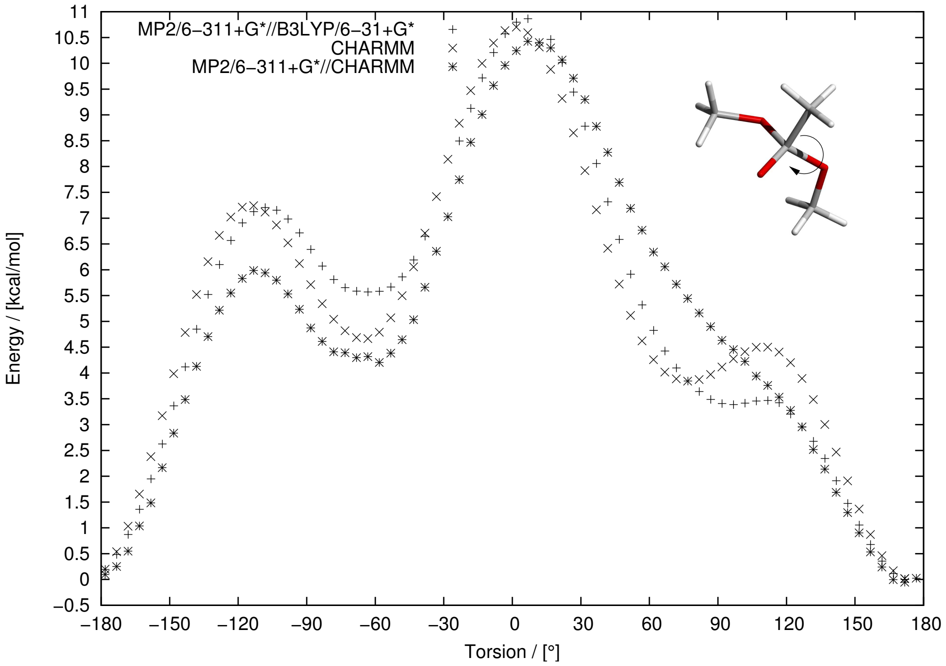


Figure 6.5: PES of the indicated torsion of the model compound. MP2//B3LYP used as reference in the parametrisation of ω_1.

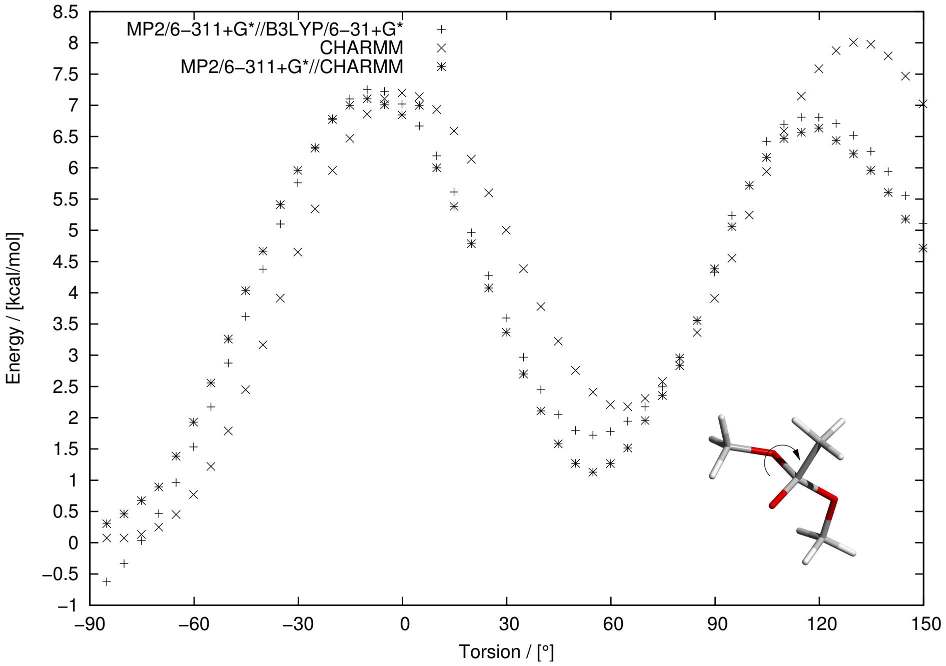


Figure 6.6: PES of the indicated torsion of the model compound. MP2//B3LYP used as reference in the parametrisation of ω_2.

should therefore expect a bias for the *trans* configuration with Charmm22. Furthermore our force field fails to reproduce a second minimum predicted by the QM model at 75°. Interestingly, both features are correctly reproduced in QM calculations at the force field optimised geometries (figure 6.7). This suggests a participation of electronic effects which cannot be modelled with the simple functional form of Charmm22 but require a quantum mechanical treatment. The same is true for torsion ω_4 (figure 6.8). Our force field predicts a maximum at $-40°$, where the *ab initio* generated PES has a local minimum. In Charmm22 this torsion is dominated by the Coulomb interaction, and the torsion potential is used only as a small corrective term. The distances between O15 and H13′, and between O15 and H5 are correlated with the position of the barrier. The attractive interaction between these groups is the main component of the torsional profile at the MM level. In the QM calculation we noticed an angular dependence of charges (Mulliken) for the atoms of the phenyl ring. Such polarisation effects cannot be reproduced by the Charmm22 force field, and attempts to impose the correct shape of the PES on the MM model using a very strong torsion potential failed. Still the force field predicts the lower lying minimum of ω_4 correctly, and the higher lying minimum is not expected to be significantly populated in practice. Important is the existence of a barrier which inhibits a free rotation of the aromatic ring, and this feature of the PES is captured by the force field.

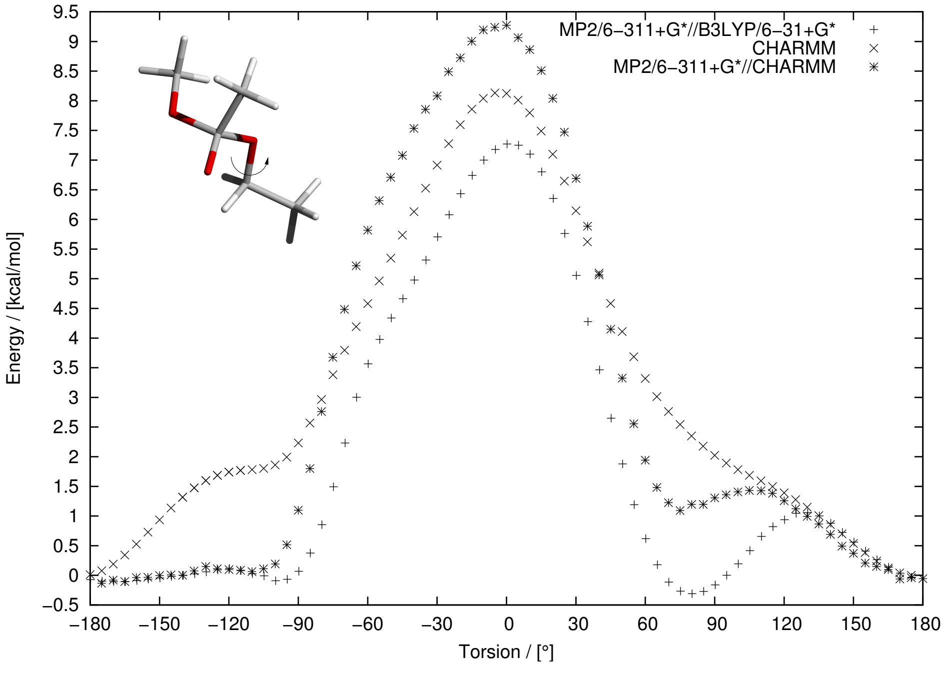


Figure 6.7: PES of the indicated torsion of the model compound. MP2//B3LYP used as reference in the parametrisation of ω_3.

Parametrisation of bonds and bond angles was performed manually, in the same spirit as that of the torsion potentials. The reference compound was fragment A. We found that in most cases it is not necessary to adjust the force constants of the bond and bond angle terms for an accurate modelling of geometries. An exception are the bonds between C14 and O3 or O17. The standard Charmm force constants of alcohol or carbonyl C=O bonds were found

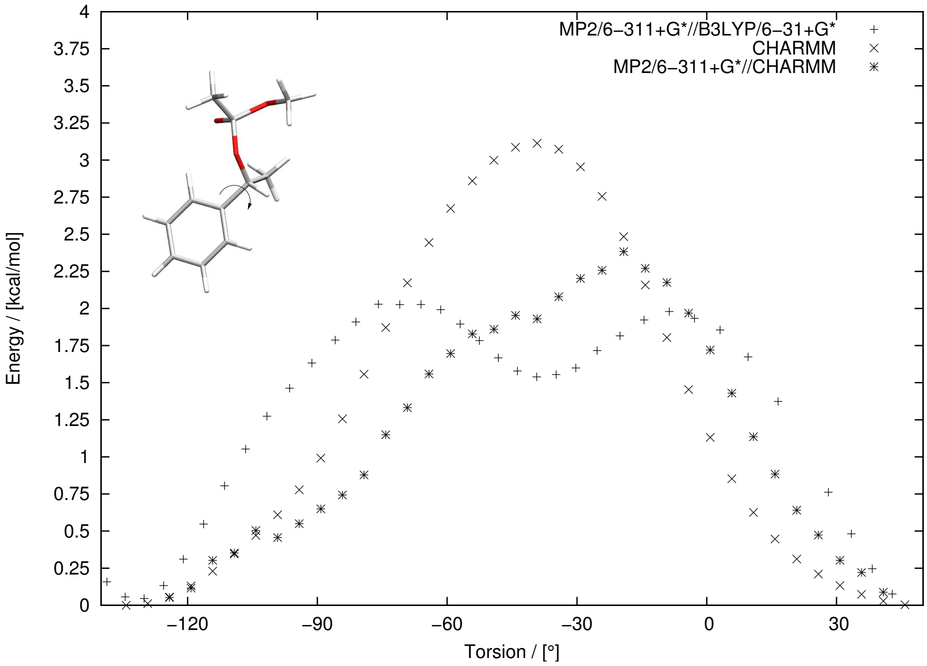


Figure 6.8: PES of the indicated torsion of the model compound. MP2//B3LYP used as reference in the parametrisation of ω_4.

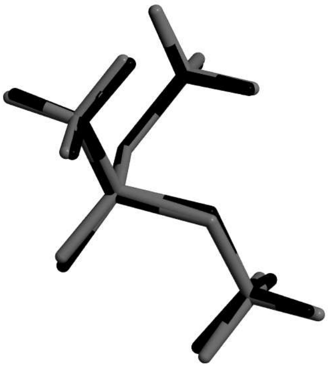


Figure 6.9: Superposition of B3LYP/6-31+G* optimised reference geometry with Charmm22 minimised geometry.

to be too large, and needed to be reduced drastically in the TI. This is justified intuitively considering that the intermediate is an unstable compound and its decomposition is initiated by a breaking of either C14-O3 or C14-O17. The equilibrium positions r^0 and θ^0 were adapted to match the *ab initio* values. After all internal parameters were assigned (see tables 6.4-6.6) we studied the internal force balance by comparing *ab initio* and Charmm22 generated normal modes at the individually optimised minima of the reference compound which are very similar (figure 6.9). To assess their similarity, we computed the scalar products of the ab initio and force field normal modes and collected them into a matrix. In case of a close match of internal molecular forces this matrix should be diagonal. Similarly, the frequencies from the two normal mode calculations should lie on a straight line when plotted against

$V_{bond} = K_a(x - x_0)^2$			
Bond	Reference	K_a	x_0
CIL3 HAL	CTL3 HAL	322.00	1.111
CIL2/3 OSL	CTL2/3 OSL	340.0	1.43
CIL2 CIL3	CTL2 CTL3	222.500	1.528
CIL1 CIL3	CTL1 CTL3	222.500	1.538
CIL1 OSL	CTL1 OSL	340.0	1.43
CIL1/2 HAL	CTL1/2 HAL	309.00	1.111
CA CIL1	CA CTL1	230.000	1.4900
CTET CIL3	CTL1 CTL3	222.5	**1.55**
CTET OC	CT2 OC	450.0	**1.27**
CTET OSL	CTL1 OSL	**200.0**	**1.52**
CIL2 CT1	CT2 CT1	222.500	1.5380

Table 6.5: Bond parameters: New parameters printed in bold letters.
References taken from ′par_all27_prot_lipid.prm′.

$V_{Angle} = K_a(\theta - \theta_0)^2 + K_b(x - x_0)^2$					
Angle	Reference	K_a	θ_0	K_b	x_0
HAL CIL3 HAL	HAL CTL3 HAL	35.500	108.40	5.40	1.80200
HAL CIL3 OSL	HAL CTL3 OSL	60.0	109.5		
HAL CIL2 CIL3	HAL CTL2 CTL3	34.600	110.10	22.53	2.17900
OSL CIL1 CA	OSL CTL1 CTL3	75.700	110.10		
HAL CIL1 CA	HA CT2 CA	49.300	107.5000		
CIL3 CIL1 CA	CT3 CT2 CA	51.800	107.5000		
CIL1 CA CA	CT2 CA CA	45.800	122.3000		
HAL CIL3 CIL2	HAL CTL3 CTL2	34.600	110.10	22.53	2.179
HAL CIL2 HAL	HAL CTL2 HAL	35.500	109.00	5.40	1.80200
HAL CIL2 OSL	HAL CTL2 OSL	60.000	109.5		
OSL CIL2 CIL3	OSL CTL2 CTL3	75.700	110.10		
HAL CIL3 CIL1	HAL CTL3 CTL1	33.430	110.10	22.53	2.179
HAL CIL1 CIL3	HAL CTL1 CTL3	34.500	110.10	22.53	2.179
OSL CIL1 CIL3	OSL CTL1 CTL3	75.700	110.10		
HAL CIL1 OSL	HAL CTL1 OSL	60.000	109.5		
OSL CTET OC	OSL CL OBL	90.000	109.6	**0.0**	**0.0**
OC CTET CIL3	OC CT2 CT3	65.000	**109.6**		
OSL CTET OSL	OSL CL OBL	90.000	109.6	**0.0**	**0.0**
CIL$ OSL CTET ($ = 1,2,3)	CT2 OS CD	60.000	109.60	**0.0**	**0.0**
OSL CTET CIL3	OSL CTL1 CTL3	75.000	110.10		
CTET CIL3 HAL	HAL CTL1 CTL3	34.500	110.10	**0.0**	**0.0**
NH1 CT1 CIL2	NH1 CT1 CT2	70.000	113.5		
HB CT1 CIL2	HB CT1 CT2	35.000	111.0		
CIL2 CT1 C	CT2 CT1 C	52.000	108.0000		
OSL CIL2 CT1	OH1 CT2 CT1	75.700	110.1000		
HAL CIL2 CT1	HA CT2 CT1	33.430	110.10	22.53	2.17900
CA CPT CA	CY CPT CA	160.000	130.6000		

Table 6.6: Angle parameters: New parameters are printed in bold letters.
References taken from ′par_all27_prot_lipid.prm′.

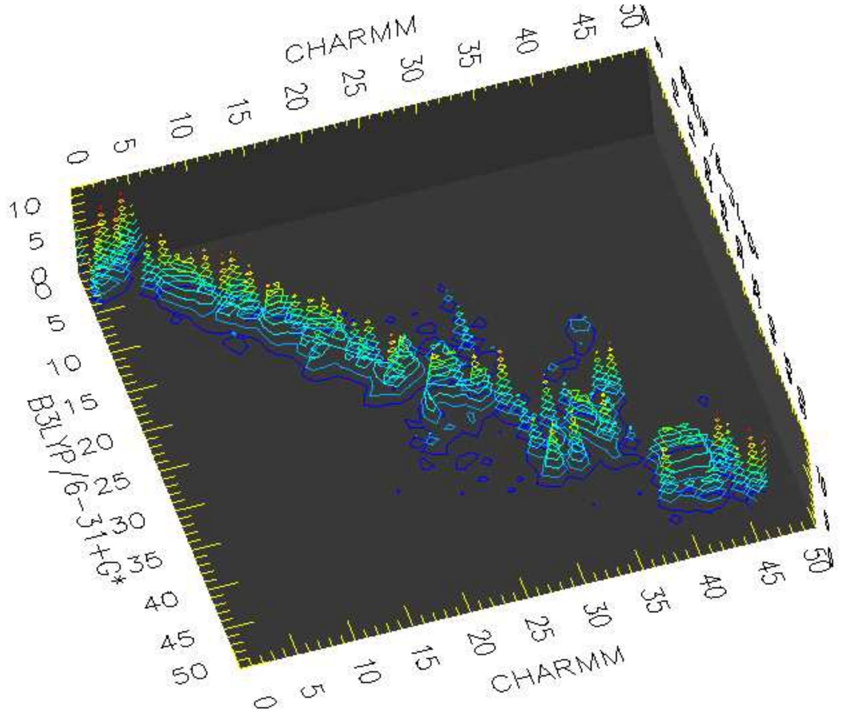


Figure 6.10: Projection matrix of normal mode vectors (see text).

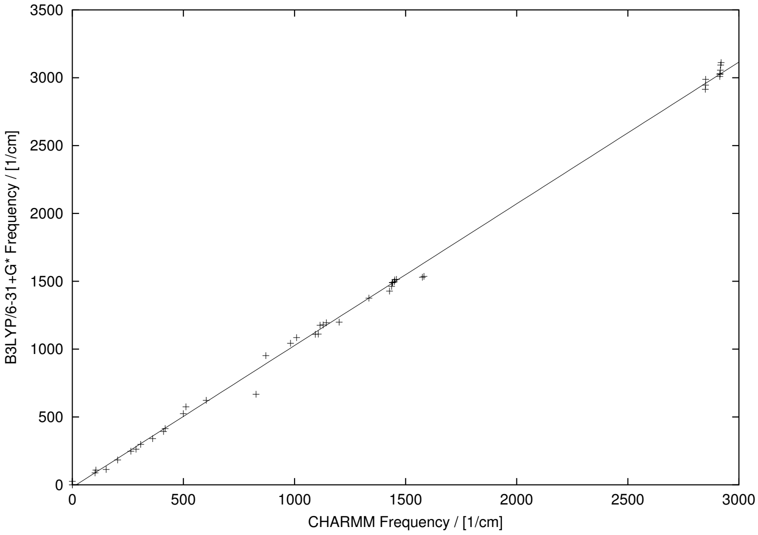


Figure 6.11: Comparison of frequencies with best fit straight line ($y = -18.20 + 1.04x$).

each other. The results are displayed in figures 6.10 and 6.11 which show a close correspondence between the *ab initio* model and the Charmm22 force field. How these two measures of similarity may be used in automated parametrisation is explained in the next paragraph.

Accelerating Optimisation

The optimisation of force field parameters is quite time consuming, and it seems impossible to parametrise a larger number of target molecules manually. Automatted fitting procedures against first and second derivatives of *ab initio* data, generated by random distortions of the target molecules have been reported[135], and proved to be convenient. It should also be possible to introduce an automatted procedure in the present context. If we were to design a cost function (CF) that could be minimised using an optimisation procedure like simulated annealing[116] it would contain the normal modes, the harmonic frequencies, and the deviation (as function of the gradient) of two structures.

$$
\begin{aligned}
\mathrm{CF} &:= \mathrm{w}_1 \sum_{\mathrm{i}} (\nu_{\mathrm{i}_{\mathrm{ref}}} - \nu_{\mathrm{i}_{\mathrm{ff}}})^2 + \mathrm{w}_2 \sum_{\mathrm{i}} \sum_{\mathrm{k} \neq \mathrm{i}} |\tilde{\mathrm{r}}_{\mathrm{i}_{\mathrm{ref}}} \tilde{\mathrm{s}}_{\mathrm{k}_{\mathrm{ff}}}| + \mathrm{w}_3 \mathrm{RMSD} \\
\mathrm{RMSD} &: \text{Root mean square deviation of computed gradient.} \\
\nu_{\mathrm{i}_{\mathrm{ref}}}, \nu_{\mathrm{i}_{\mathrm{ff}}} &: \text{QM reference- and MM frequencies.} \\
\mathrm{w}_1, \mathrm{w}_2, \mathrm{w}_3 &: \text{Weighting factors.} \\
\tilde{\mathrm{r}}_{\mathrm{i}_{\mathrm{ref}}}, \tilde{\mathrm{s}}_{\mathrm{k}_{\mathrm{ff}}} &: \text{QM reference- and MM normal mode vectors.}
\end{aligned}
$$

The parameter set could thus be conveniently optimised in a Monte Carlo scheme. A thorough check of the performance of the parameters in practical situations has to be performed subsequently.

6.4 Force Field Validation

The quality of our new Charmm22 force field was assessed by

- comparing QM/MM and MM geometries.
- using molecular dynamics simulations: The stability of the TI complex should be guaranteed in the range of nanoseconds.

The first item on the list can be evaluated using the large number of reaction path calculations which are presented in chapter 4. For each reaction path that was computed using the distinguished coordinate method, initial geometry optimisations were performed at the MM level and subsequently at the QM/MM level. Under ideal circumstances, if intramolecular and intermolecular forces were equal with both methods, the structures derived from the minimisations should also be identical. To compare MM and QM/MM optimised geometries, we have chosen both intramolecular and intermolecular structural features of the TI as probes.

	Median	Min.	Max.
Distance			
O3-HE2	0.27	0.03	0.48
O17-HE2	0.20	0.01	0.52
NE2-HE2	0.04	0.00	0.10
C14-O3	0.07	0.00	0.14
C14-O15	0.02	0.00	0.03
C14-O17	0.02	0.01	0.30
ILE12-O15	0.07	0.00	0.15
MET78-O15	0.13	0.00	0.28
Angle			
NE2-O17-HE2	3.68	0.36	6.57
NE2-O3-HE2	3.94	0.04	8.55
Torsion			
ω_4	3.54	0.21	39.11
ω_3	4.59	0.50	44.74
ω_2	3.17	0.04	49.43
ω_1	14.63	2.83	39.28

Table 6.7: Median, minimum and maximum differences of structural parameters from MM and QM/MM geometry optimised snapshots. Differences of bond distances are given in Angstrom, angles and torsions in degrees. The statistics was computed for 75 pairs of structures from the set of optimised reaction paths.

Table 6.7 shows the results obtained from 75 pairs of MM and QM/MM geometries. The distances show rather small deviations, the largest ones being encountered for O3-HE2, O17-HE2, and MET78-O15 which appear in hydrogen bonds and are expected to show larger variations than covalent bonds. ILE12-O15 is an exception and performs noticeably better than the other hydrogen bonds. Angles and torsions show tolerable median deviations of typically $3°$ - $5°$, with a much larger value of almost $15°$ for torsion of ω_1.

An important point is that the comparisons refer to the condensed phase and the torsions are not "pure" as during the parametrisation, which was performed *in vacuo*. Torsion angle ω_1 is more rigidly fixed in the enzyme and must comply with the steric and electronic requirements of the environment more than the other torsions. It is modulated by relocation of the oxyanion O15 that is triggered on switching from MM to QM/MM. Overall, we found that the force field is capable of producing useful input geometries for QM/MM calculations. This was realised in the optimisations at the QM/MM level, which were unsuccessful until the parametrisation of the Charmm22 force field was performed.

A different check of the parametrisation was provided by the MD simulations of the TI. The central moiety around the tetrahedral carbon atom C14 of the TI needs to be stable during the full duration of these simulations, and this has been found to be true not only in the present project (see chapter 3), but also in an MM study of a different lipase (from *Pseudomonas aeruginosa*) where the origin of the enantioselectivity was analysed qualitatively at the MM level[22]. Hence the derived parameters for the TI appear to be well integrated into the Charmm22 force field.

Chapter 7

Summary

Our aim was to understand enantioselectivity in enzymatic ester hydrolysis using a detailed and realistic representation of the catalytic system. We chose lipase A of *Bacillus subtilis* as the catalyst and 1-(2-naphthyl)-ethyl-acetate as model substrate for our studies. Lipase A selectively hydrolyses the R enantiomer of the substrate with an E-value of 140.

The computational model consisted of the enzyme, the substrate, and a considerable amount of water to simulate the typical conditions found in the condensed phase. Discrimination of the enantiomers was argued to occur during the acylation reaction. We therefore focused on a precise calculation of all reaction barriers for this chemical transformation. A QM/MM potential was employed to enable an efficient and yet precise handling of the computational model.

A number of complementary techniques were employed, and the consecutive use of results obtained with each of those was essential to achieve the final goal of modelling enantioselectivity. An important first component was the parametrisation of a force field for the tetrahedral intermediate. To this end a thorough examination of the tetrahedral intermediate *in vacuo* was mandatory in order to satisfactorily implement this procedure on the basis of *ab initio* data. This already necessitated an intensive examination of all important molecular degrees of freedom of this key chemical species.

We studied the conformational degrees of freedom of the tetrahedral intermediate in the active site of the enzyme; the complete model of the catalytic system was built and treated at the MM level using force field parameters derived in the preceding step. This led to the discovery of a number of stable binding modes of the tetrahedral intermediate. At the MM level it was difficult to unambigously identify the most probable binding modes.

With the binding modes obtained from the MM study, the acylation reaction was modelled at the QM/MM level. Two methodologically different approaches were tested, the first one was based on the potential energy surface of the acylation reaction, and the second one on the free energy surface. In studying the potential energy surface we validated the electronic structure component of our QM/MM potential, and gathered structural and energetic information about important stationary points. These were identified as the tetrahedral intermediate, two transition states, the Michaelis complex and an acylenzyme-alcohol complex (product complex). A simple geometric reaction coordinate was implemented and enabled large-scale production of reaction paths to determine activation energies for different binding modes. The approach suffered from artefacts and the resulting statistics were found insufficient for a precise evaluation of enantioselectivity. As such, we decided to study the acylation reaction with molecular dynamics methods and evaluate its free energy surface.

In umbrella sampling calculations sufficiently accurate $\Delta G^{\ddagger}$ values were computed and used as input for a stochastic model of the acylation reaction under the conditions of a kinetic resolution experiment. This type of experimental setup is typically used to determine the enantioselectivity starting from a racemate of the reactants. The stochastic model of the chemical system allowed us to capture competitive effects due to the presence of both enantiomers, and thus complemented our atomistic calculations, where only one enantiomer was present at any time. Results of this simulation showed that 1-(R)-(β-naphthyl)-ethyl-acetate was preferred over the S-enantiomer, which is in accordance with experimental findings. The estimated apparent $\Delta\Delta G^{\ddagger}$ difference between R and S substrate was 1.36 kcal/mol, which translates to an $\mathbf{E}$-value of 10. Although this seems vastly different from the experimental result of E=140 at first glance, it should be recalled that $\mathbf{E}$ grows exponentially with $\Delta\Delta G^{\ddagger}$, and the discrepancy is only 1.6 kcal/mol on the free energy scale (E=140, $\Delta\Delta G^{\ddagger}$=2.96 kcal/mol).

The preferred enantiomers in ester hydrolysis have an additional water molecule in the vicinity of the tetrahedral intermediate, which stabilises the oxyanion by an additional hydrogen bond. This hydrogen bond is formed during the first step of acylation upon formation of the tetrahedral intermediate and causes a decrease in the free energy barrier of nucleophilic attack. The benefit of the additional water is partially alleviated in the second step of the acylation reaction, where the tetrahedal intermediate dissociates to form an acylenzyme. Here it is found that binding modes with an additional water in the active site have a higher free energy barrier for this step.

Our present study indicates that quantitative QM/MM free energy simulations are still of limited utility in enzyme engineering, e.g., in the optimisation of enzymes with regard to enantioselectivity. Restrictions arise from the remaining uncertainty in the calculated values of $\Delta\Delta G^{\ddagger}$ due to systematic errors, introduced mainly by an incomplete sampling of configuration space, and the limited accuracy of the QM/MM potential. Moreover, such calculations consume large amounts of computer time to evaluate enantioselectivity in each single case. In order to guide experimental work in practice, one would need to study either many mutants of the enzyme or many different substrates. This is at present still a too ambitious task for computational chemistry, and cannot be considered to become a standard procedure in the near future.

This thesis has contributed to the following publications:

1) J Kästner, H M Senn, S Thiel, N Otte, W Thiel, QM/MM Free-Energy Perturbation Compared to Thermodynamic Integration and Umbrella Sampling: Application to an Enzymatic Reaction, *J. Chem. Theor. Comput.* (**2006**), 2, 452-461.
2) S A Funke, N Otte, T Eggert, M Bocola, K-E Jaeger, W Thiel, Combination of computational prescreening and experimental library construction can accelerate enzyme optimization by directed evolution, *Protein Engineering* (**2005**), 18, 509-514
3) M Bocola, N Otte, K-E Jaeger, M T Reetz, W Thiel, Learning from Directed Evolution: Theoretical Investigations into Cooperative Mutations in Lipase Enantioselectivity, *ChemBioChem* (**2004**), 5, 214-223
4) S A Funke, A Eipper, M T Reetz, N Otte, W Thiel, G van Pouderoyen, B W Dijkstra, K-E Jaeger, T Eggert Directed Evolution of an Enantioselective Bacillus subtilis Lipase, *Biocatal. Biotransform.* (**2003**), 21, 67-73

Acknowledgement

I sincerely thank my supervisor Prof. Thiel for the freedom he gave me in conducting my research, and for the continuous financial support.

I am indebted to Prof. Marian for kindly accepting to act as a co-referee for this thesis.

I would like to acknowledge the fruitful collaboration with our experimental partners, Prof. Reetz at the MPI Mülheim, and Prof. Jaeger and Dr. Eggert from the Institut für Molekulare Enzymtechnologie in Jülich.

I am grateful to Tell Tuttle, Hans-Martin Senn, Anja Metzelthin and Maja Parac for proofreading my thesis.

A great thanks goes to all my colleagues at the MPI, for the sociable atmosphere, and many interesting scientific discussions. Especially I would like to thank Marco Bocola, Jürgen Breidung, Axel Koslowski, Hans-Martin Senn and Stephan Thiel for sharing their insights.

I thank the computer department of the MPI Mülheim, especially Horst Lenk, who helped solving many hardware and software problems.

I express my deep gratidute to Anja and my family, for their constant support and encouragement.

Appendix A

The Meso-Diacetate

A special class of substrates for lipases are meso-compounds, which can theoretically be transformed at 100% yield, while chiral reactants can be transformed at a mere 50% maximum yield of enantiopure products[201]. The use of meso-compounds thus minimises waste and allows more efficient production of precursors for the pharmaceutical industry. One such substrate, which was tested for directed evolution of BSLA is meso-1,4-diacetoxy-cyclopentene (DIAC) (see figure A.7). Enzymatic hydrolysis leads to the production of the monoester, which is a chiral compound[1]. Wildtype BSLA shows only weak enantiopreference for DIAC (ee = 38%). Using directed evolution enantioselectivity was increased considerably, with mutations at position 18 in the amino acid sequence being most fruitful. An asparagine is located at this position in the wildtype enzyme, and exchange for serine raised enantioselectivity to ee $\approx 85\%$. Interestingly enantiopreference was reverted as a side effect of this mutation[60]. The overall effect is too small for QM/MM methods to predict safely, therefore we confined our attempts to molecular modelling at the MM level. A special challenge was the parametrisation of the force field for the DIAC substrate. While we reused our parameters of the TI as far as the tetrahedral moiety was concerned, some degrees of freedom of the five-membered ring proved difficult to model by the simple form of the Charmm22 force field. The next section describes the parametrisation procedure, thereafter we present the molecular modelling of the TI of DIAC.

A.1 Parametrisation

FigureA.2 shows the compound 3-acetoxy-cyclopentene which was employed to derive parameters for the angles specified. Internal force balance was checked by comparing ab-initio

---

[1] For unknown reasons hydrolysis stops after mono-hydrolysis and does not proceed to the alcohol as might be expected[216].

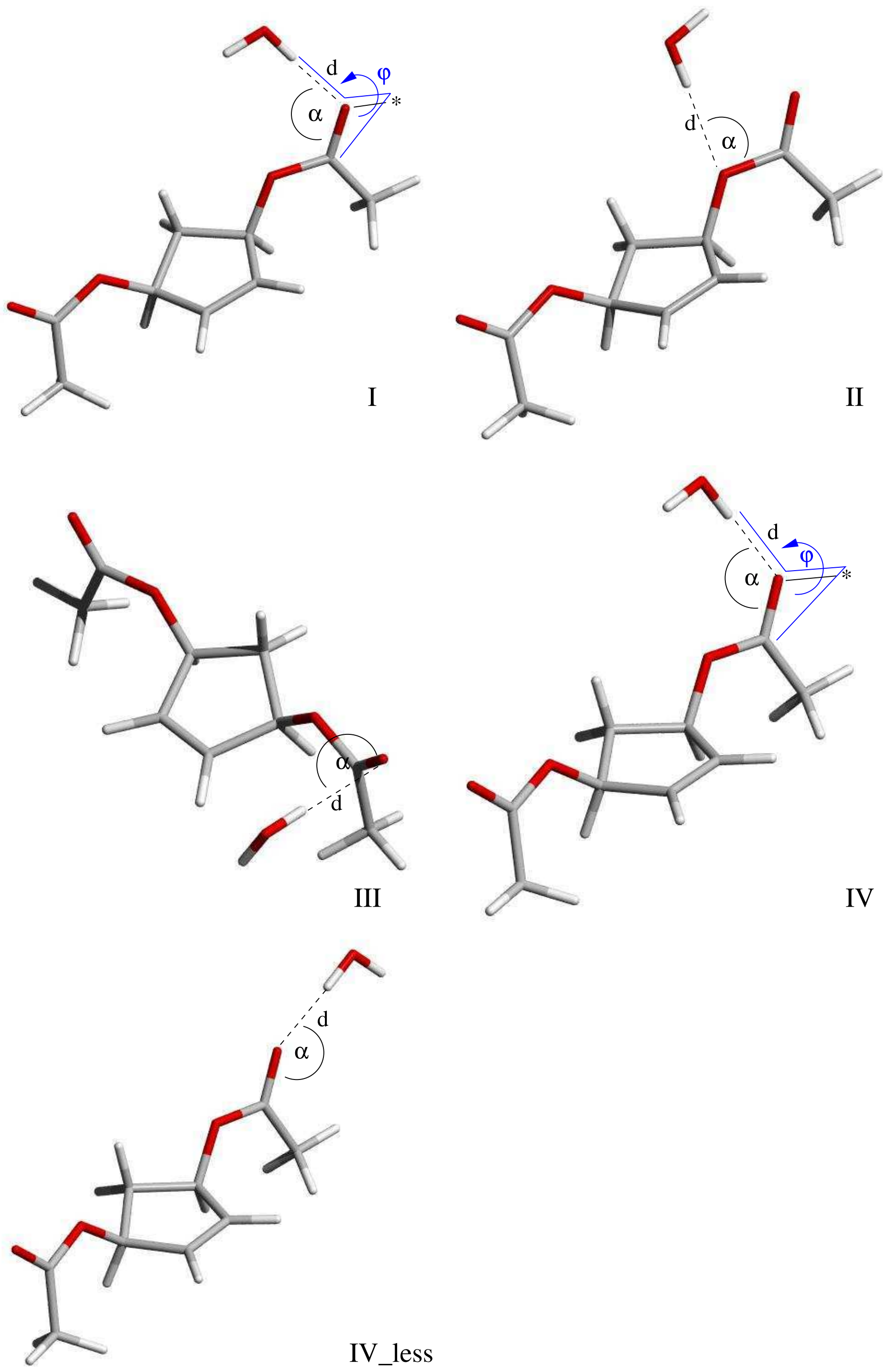


Figure A.1: Complexes with water, used to determine the charges in DIAC.

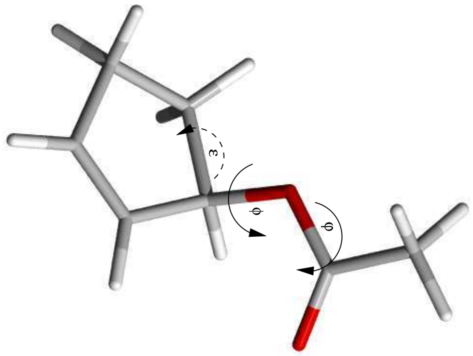


Figure A.2: 3-acetoxy-cyclopentene and dihedral angles. This model compound was used in the parametrisation of torsions which are important for DIAC. Here, ϵ is a descriptor of ring puckering, ϕ and φ are regular torsions studied at DFT level.

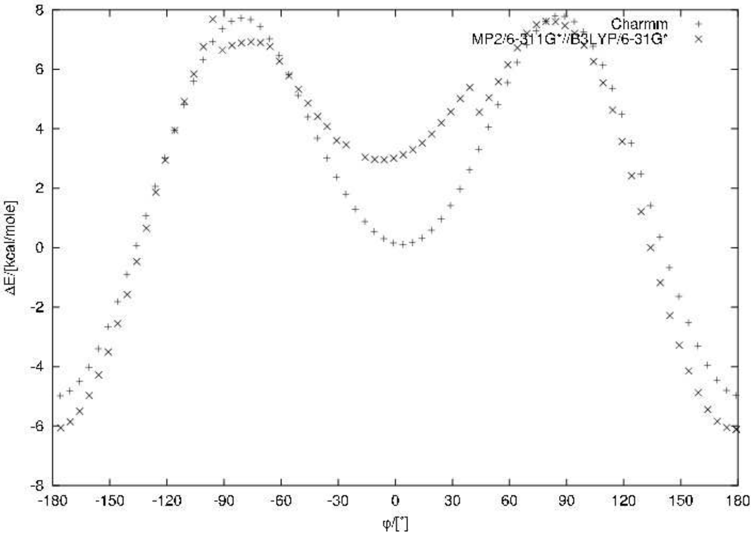


Figure A.3: Rotation about the angle φ in 3-acetoxy-cyclopentene (see figure A.2).

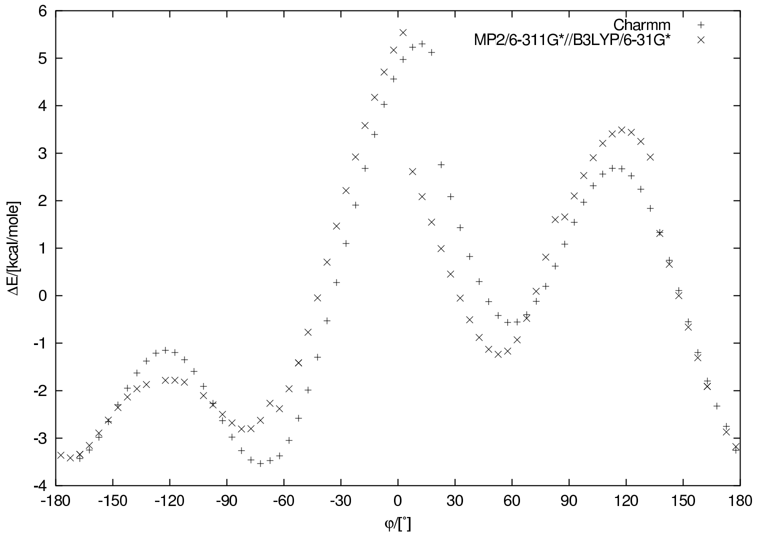


Figure A.4: Rotation about the angle ϕ in 3-acetoxy-cyclopentene (see figure A.2).

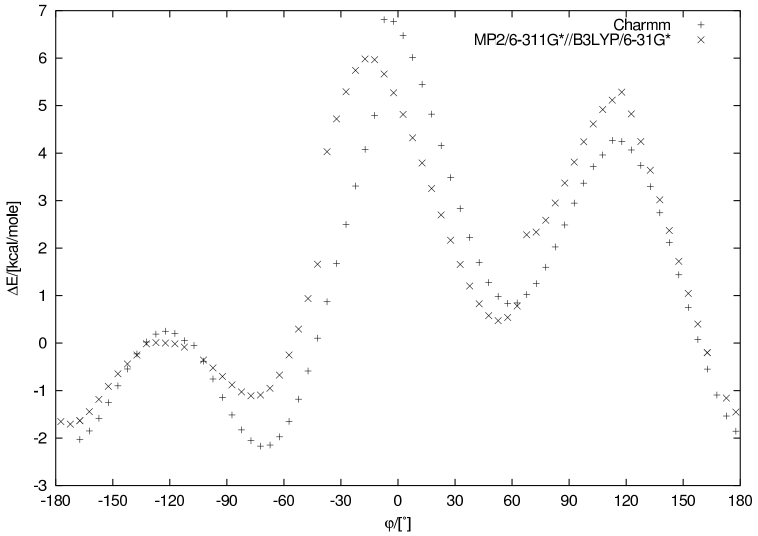


Figure A.5: Rotation about the angle ϕ in 3-acetoxy-cyclopentene (see figure A.2) in reverse direction.

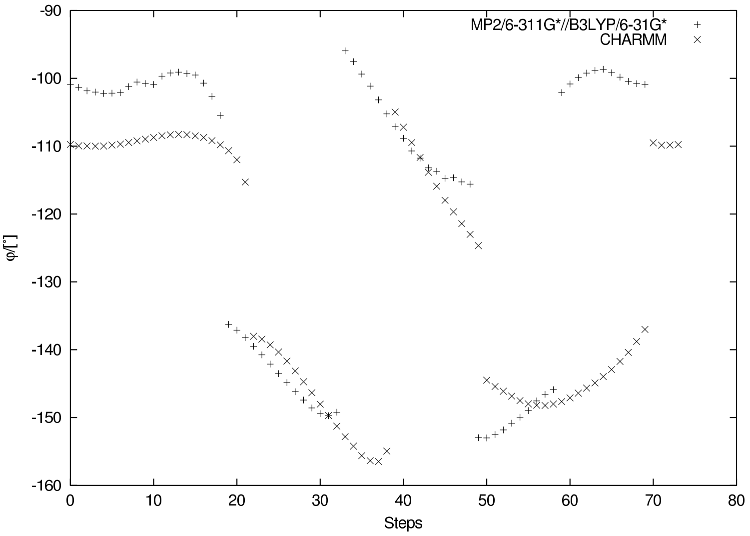


Figure A.6: Puckering of the ring as described by torsion ϵ in 3-acetoxy-cyclopentene (see figure A.2). ϵ was measured as function of dihedral ϕ.

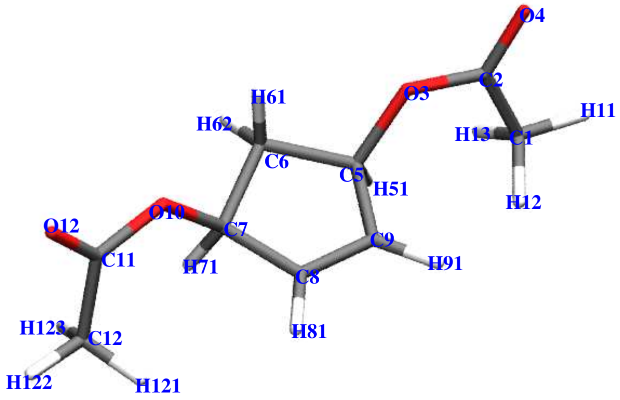


Figure A.7: Meso-1,4-diacetoxy-cyclopentene (DIAC) with atoms labeled in Charmm notation.

normal modes to those generated with the Charmm22 force field, similar to chapter 6. These comparisons indicate that there is no compelling need to adjust force constants for bond stretching or angle bending. We decided nevertheless to fine-tune angle bending force constants and positions of θ_0 to allow for puckering of the cyclopentene ring. Special attention was given to the dihedral angles, since their large amplitude motion generates very specific energy landscapes. Charges were adopted by chemical analogy from existing Charmm22 parameters. They provided sufficiently accurate interaction energies for water-DIAC complexes (see figure A.1) compared with *ab initio* calculations, and were therefore left unaltered (see table A.1).

Atom (see figure A.7)	Type	Reference	Charmm22
C1	CTL3	CTL3	-0.04
C2	CL	CL	0.63
O3	OSL	OSL	-0.34
O4	OBL	OBL	-0.52
C5	CXL	CTL2	-0.09
C6	CXL	CTL2	-0.18
C7	CTL2	CTL2	-0.18
C8	CEL1	CEL1	-0.15
C9	CEL1	CEL1	-0.15
H61	HAL2	HAL2	0.09
H62	HAL2	HAL2	0.09
H71	HAL2	HAL2	0.09
H72	HAL2	HAL2	0.09
H81	HEL1	HEL1	0.15
H91	HEL1	HEL1	0.15
H51	HAL2	HAL2	0.09
H11	HAL3	HAL3	0.09
H12	HAL3	HAL3	0.09
H13	HAL3	HAL3	0.09
H121	HAL3	HAL3	0.09
H122	HAL3	HAL3	0.09
H123	HAL3	HAL3	0.09
C12	CTL3	CTL3	-0.04
C11	CL	CL	0.63
O10	OSL	OSL	-0.34
O12	OBL	OBL	-0.52

Table A.1: Partial charges (in e). Charmm notation is used in the second and third column, see file 'par_all27_prot_lipid.prm'.

The puckering of the cyclopentene ring was most difficult to reproduce with the force field. Such motions can easily be triggered by a rotation about the dihedral angle ϕ. Consequently we recorded the puckering as dihedral ϵ, in dependence of dihedral angle ϕ (see figure A.2). Manual parametrisation was laborious (60 hours of manual optimization), and proceeded in cycles of readjustment of torsion and angle bending contributions to optimise the puckering behaviour. The result is fairly satisfactory. Ring puckering cannot be reproduced accurately, but the choice of force constants and equilibrium values for the neighbouring angles and dihedrals is chosen such as to generate the desired floppiness. Special care had to be taken in the assignment of the dihedral angle ϕ since its influence on the puckering was quite dominant. This dihedral is composed of four different angle functions. The main features of the dihedral potential energy surfaces match quite well with the ab-initio calculations, however, the barrier heights could not be reproduced so well in all cases. The barrier for rotation about the dihedral angle φ is too high by approximately 2.5 kcal/mol, while those for the dihedral ϕ (see figure A.2) are off by 1-2 kcal/mol. Since the *ab initio* barriers are high already, these rotations will be rare events in any case. The force field stresses this circumstance, but retains the essential features of the potential curve.

A.2 Molecular Modelling

The DIAC substrate has two prochiral carbon atoms. Cleavage of one of the ester groups leads two possible diastereomers, 1S4R and 1R4S. Two essential binding modes must be differentiated, as shown in figure A.8. TIs were modelled for the mutant Asn18Ser, the model setup proceeded in the same manner as in chapter 3. This revealed that a hydrogen bond to Ser18 is very unlikely, as it implies a distorted and energetically unfavourable conformation of the DIAC regarding torsions φ and ϕ. More likely binding modes are displayed in figure A.9, they do not exhibit any interaction with the diacetate. The structural cause of enantioselectivity therefore does not seem to be rooted in hydrogen bonding of the DIAC with Ser18. A different explanation can be provided considering the network of water in the anterior region of BSLA.

Shown in figure A.9 is a TI leading to 1R4S product. The geometry for 1S4R is nearly identical. This comes as no surprise, considering the geometry of the substrate. The asymmetry arises structurally from a puckering of the cyclopentene ring, rendering one of the enantiomers sterically less favourable. The local perturbation introduced by the asparagine to serine mutation is assumed to lead to subtle changes of the water and hydrogen bond network in the vicinity of the tetrahedral intermediate. Two hypotheses may be proposed based

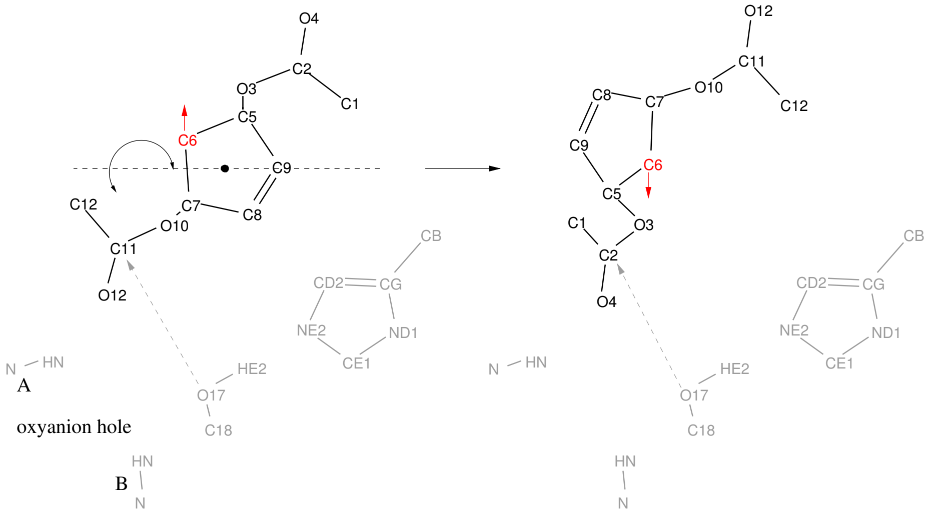


Figure A.8: Schematic representation of the two basic binding modes of the DIAC substrate in the active site of BSLA, leading to monoester products in either 1R4S (left hand side) or 1S4R (right hand side) configuration. Notice the orientation of the ring, its tip points upwards or downwards. Rotation about the indicated axis transforms one pro-diastereotopic conformation into the other one. The atoms are labeled according to Charmm notation. The oxyanion hole is represented by the backbone amide groups of Ile12 (A) and Met78 (B).

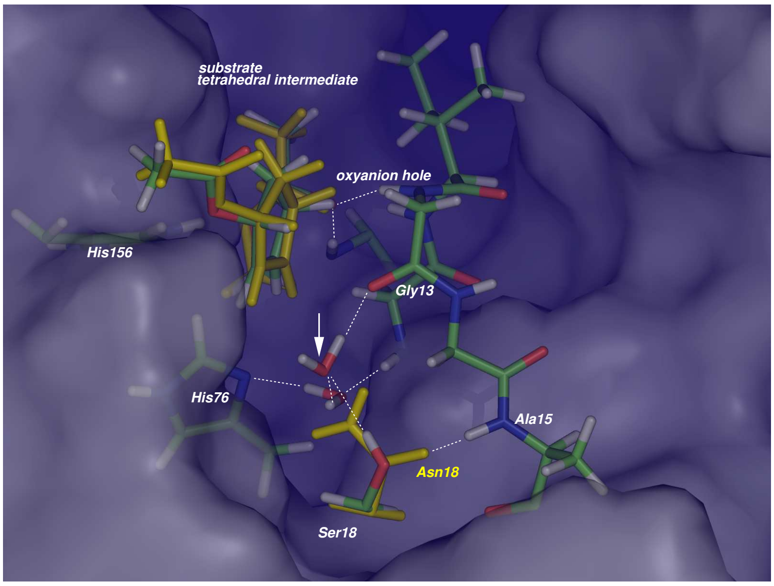


Figure A.9: Coloured atoms are from the mutant N18S, yellow atoms belong to wildtype BSLA. The hydrogen network in the active site is delicate and changed on mutation. The marked water molecule slips into this position in the mutant N18S. For an explanation please refer to the text.

on structural features. The side chain oxygen of Asn18 forms a hydrogen bond to Ala15, which is lost in the N18S mutation. The serine side chain adopts an orientation collinear with the amide group of the asparagine (O-H bond vector in serine vs. C-N vector of asparagine). Serine is sterically less demanding, and allows an additional water molecule to slip in between itself and the tetrahedral intermediate. This water molecule in turn interacts with a localised water (which is manifest as crystal water 35 in structure 1R50, chain B). Another hydrogen bond is formed by the new water and the backbone of Gly13 and thus may replace the lost hydrogen bond at Ala15. However, these changes could equally well modify the structure or dynamics of the coil formed by amino acids 10-15, which includes the oxyanion hole. These effects may be quantified by measuring the mobility of water at the active site, for example using NMR spectroscopy. How they may translate into differences in activation free energy for the enantiomers remains unclear at present.

$V_{bond} = K_a(x - x_0)^2$			
Bond	Reference	K_a	x_0
CXL CXL	CTL2 CTL2	222.500	1.530
CXL CEL1	CEL1 CTL2	365.000	1.502
CXL HAL2	CTL2 HAL2	309.00	1.111
CXL OSL	CTL2 OSL	340.0	1.43

Table A.2: Bond parameters for DIAC: New parameters printed in bold letters. References taken from 'par_all27_prot_lipid.prm'.

$V_{Angle} = K_a(\theta - \theta_0)^2 + K_b(x - x_0)^2$					
Angle	Reference	K_a	θ_0	K_b	x_0
CL OSL CXL	CL OSL CTL1	**80.0**	**116.0**	**30.0**	**2.43**
OSL CXL CEL1	OSL CTL1 CTL2	**58.00**	**117.0**		
OSL CXL CXL	OSL CTL2 CTL2	**28.00**	**106.0**	**30.0**	**2.5**
CXL CXL CXL	CTL2 CTL2 CTL2	58.350	**106.0**		
HAL2 CXL CXL	HAL2 CTL2 CTL2	26.500	110.10	22.53	**2.22**
HAL2 CXL OSL	HAL2 CTL2 OSL	60.0	**106.5**		
CEL1 CXL CXL	CEL1 CTL2 CTL2	**100.00**	**103.0**		
HAL2 CXL CEL1	HAL2 CTL2 CEL1	45.00	111.50		
HAL2 CXL HAL2	HAL2 CTL2 HAL2	**35.500**	**107.00**		
CEL1 CEL1 CXL	CEL1 CEL1 CTL2	48.00	**112.0**		
HEL1 CEL1 CXL	HEL1 CEL1 CTL2	40.00	**122.00**		

Table A.3: Angle parameters for DIAC: New parameters are printed in bold letters. References taken from 'par_all27_prot_lipid.prm'.

$V_{torsion} = K_a(1 + cos(n\varphi - \delta))$				
Torsion	Reference	K_a	n	δ
CEL1 CEL1 CXL OSL	CEL2 CEL1 CTL3 HAL3	0.05	3	180.00
HEL1 CEL1 CXL OSL	CEL2 CEL1 CTL3 HAL3	0.05	3	180.00
CXL CXL OSL CL	CTL2 CTL2 OSL CL	**0.40**	**3**	**0.00**
		0.08	**4**	**180.00**
CEL1 CXL OSL CL	-	**0.08**	**4**	**180.00**
		0.40	**3**	**0.00**
CXL CXL CXL OSL	-	**0.05**	**6**	**0.00**
		0.12	**3**	**180.00**
		0.2	**4**	**180.00**
		0.2	**2**	**0.00**
X CXL CXL X	X CTL2 CTL2 X	**0.00**	3	0.00
X CXL OSL X	X CTL2 OSL X	0.00	3	0.00
CEL1 CEL1 CXL HAL2	CEL1 CEL1 CTL2 HAL2	**0.00**	3	0.00
HEL1 CEL1 CXL HAL2	HEL1 CEL1 CTL2 HAL2	**0.95**	**3**	**0.00**
		0.015	6	0.00
CEL1 CEL1 CXL CXL	CEL1 CEL1 CTL2 CTL2	**0.3**	3	0.00
CEL1 CXL CXL CXL	-	**0.3**	**3**	**0.00**
HEL1 CEL1 CXL CXL	HEL1 CEL1 CTL2 CTL2	**0.9**	3	0.00
CTL3 CL OSL CXL	-	**4.0**	**2**	**180.0**
$V_{improper} = K_a(\varphi - \varphi_0)^2$				
Improper	Reference	K_a	φ_0	
CEL1 HEL1 CEL1 CXL	-	90	0	0.00

Table A.4: Torsions in DIAC: New parameters printed in bold letters.
Wildcards are labeled ’X’.
References taken from ’par_all27_prot_lipid.prm’.

Appendix B

Study of Mutants

In experimental as well as in independent theoretical work[72] the residue His76 was found to have significant effects on the activity and enantioselectivity of BSLA. Interesting mutants characterised by our experimental partners had alanine and leucine substitutions of His76. Variant His76Ala displayed reversed enantioselectivity for NAPHAC compared to the wild-type enzyme, which means it was selective for the S enantiomer, at E=8.5. Mutant His76Leu was R selective but with a reduced enantioselectivity compared to the wildype, at E=6.8.

We decided to build the His76Ala mutant in order to rationalise the experimental findings. The setup of the model was performed as described in chapter 3. A good way to introduce mutations was offered by the freeware utility Pymol[46]. This program provides a rotamer library that allows one to pick probable conformations of the mutated side chain. In the case of alaine the choice was rather simple as the methyl group was chosen to be at the position of the CG atom of the histidine.

The His76Ala trajectories share an interesting feature. We observed that water enters the vacant site of His76. Considering the amount of space occupied by a histidine residue compared to alanine it would seam feasible to have two entrant water molecules, and this is what we found in our simulations. It seems noteworthy that this did not happen during hydration at the setup stage of the MD, but during the simulations due to diffusion events which were observed for all mutant models. The hydrogen bond network created by the additional water could in principle replace the histidine according to conformational considerations. The replacement of His76 therefore does not lead to immediate structural consequences for the TI. It can be speculated that the exchange causes alterations in the water network around the active site, or other subtle structural changes which we cannot capture in our model. Regarding the His76Leu variant, which introduces one of the bulkiest amino acid side chains, we estimate that this substitution fills the space occupied by His76, but does not mediate

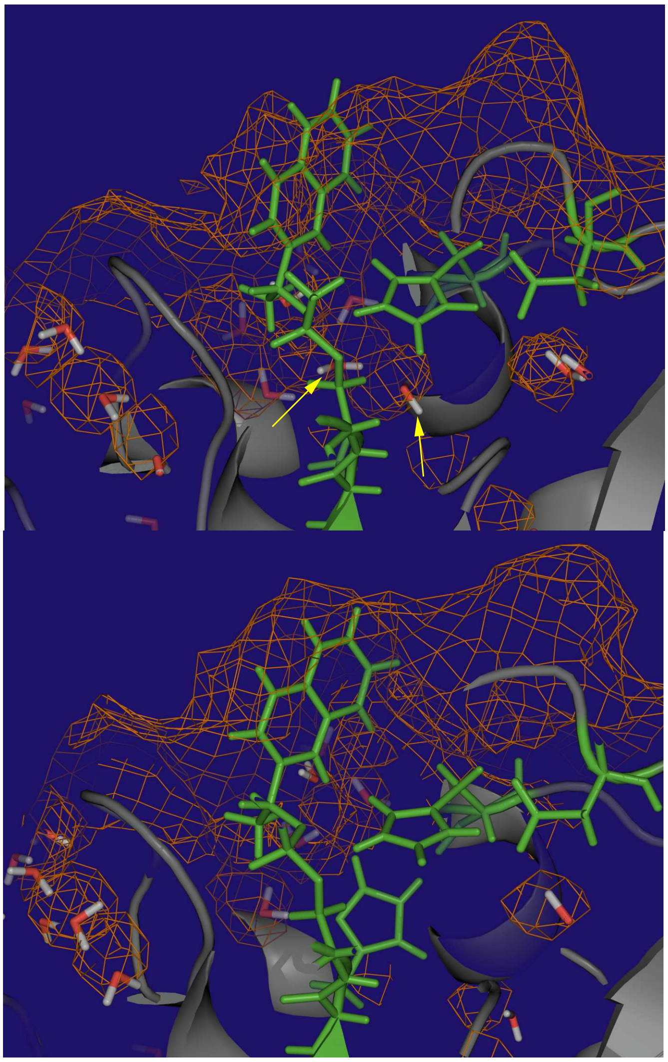


Figure B.1: Structural changes compared to the wildtype enzyme in mutation His76Ala. The wildtype is displayed at the bottom, the mutant in the top figure. Both structures correspond to the final (1.1 ns) snapshots of MD simulations at the MM level. The grid in orange colour visualises the presence of water molecules averaged over the trajectory. Some water molecules are clearly visible as isolated entities. They correspond to localised water on the surface and in the interior of the enzyme. Such water molecules are drawn as stick models.. Yellow arrows indicate the position of water molecules that enter the cavity created by the 'shrunk' amino acid in position 76. In the lower figure this space is occupied by His76.

its hydrogen bonds. This should have consequences both to the fixation of a special crystal water (see figure 3.2 in chapter 3), and for the equilibrium position of His156, towards the side chain of which a hydrogen bond of His76 is directed.

A more precise evaluation of the enantioselectivity would necessitate the same approaches as described in the main text.

Appendix C

Gas Phase Models of Lipases

Our QM/MM study of the reaction mechanism in the lipase was preceded by detailed QM calculations on a model system. The latter should contain as few fragments as possible while retaining a realistic view of the reactive event. The model thus constructed was treated by quantum mechanics alone, in order to gain experience with the reaction and possible reaction coordinates that could be used in subsequent QM/MM energy path calculations. The model system proved to be more complicated than the full QM/MM system reported in the main text. Owing to the flatness of the PES in the region of interest, it was very difficult to locate the transition states. In this appendix we describe our QM model calculations and compare the results with computational data from the literature[100][235].

Building the Model

As in chapter 4 the study starts at the TI, and the same crystal coordinates were used to build the model as in chapter 3. In the first step the crystal coordinates were transferred to Insight2000 and all atoms were deleted that were considered unneccessary. At positions where open valencies were created by these deletions hydrogen atoms were inserted in the direction of the bond to the former atom. The phosphorous atom of the inhibitor molecule was replaced with a carbon atom. The model system consisted of a formic acid molecule mimicking ASP133, an imidazole in place of the HIS156 side chain, and a complex of the tetrahedral intermediate of acetic acid methylester and the SER77 side chain, plus part of its backbone and that of the neighbouring MET78 which is part of the oxyanion hole. Furthermore the backbone amide of ILE12 and the carbonyl function of GLY11 were retained as a formamide molecule. In the study by Hu et al. [100] water molecules were used as reduced model of the oxyanion hole. The amides in the present study were included to build a more realistic model. In order to retain the relative orientations of the fragments towards

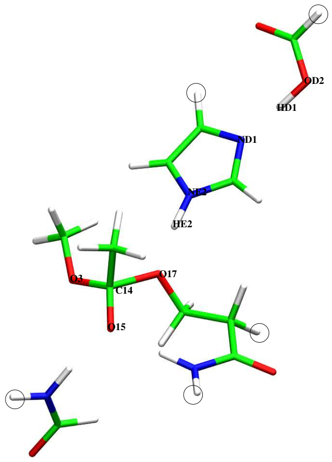


Figure C.1: Structure of the model system in TI geometry at the BLYP/SVP+ level. All atoms that were frozen during optimization are marked.

each other constraints were introduced. Some atoms, which are marked in figure C.1 were frozen. Such constraints were not used in other studies [100]. It is clear that without the full model of the enzyme the problem of enantioselectivity can not be adressed. Therefore we did not choose a chiral substrate, but a very simple ester. The resulting system is called "model100"

Computational Study

All computations were done using Chemshell/Turbomole in DFT calculations at the BLYP/SVP+ level. Diffuse functions were included to improve the description of the anion. Use of the BLYP functional reduced the computational cost by a factor of five to ten relative to B3LYP due to the efficient RI (resolution-of-identity) implementation in Turbomole. During geometry optimisation soft degrees of freedom such as rotations around hydrogen bonds proved problematic. These occured between the aspartate mimick, imidazole and the substrate complex, and also between the oxyanion hole residues and the substrate complex. As can be seen in figure C.1 the imidazole is in a position that facilitates breakdown of the TI and release of the substrate ester if the reaction coordinate is followed back to the *reactants* ('enzyme + ester', MCC). Evolving the system towards the *products* ('acylenzyme + ester', PDC) would require rotation of the imidazole. Since the two reactions are essentially of the same type, we only studied the first one at the QM level and delayed the investigation of the second one to the QM/MM stage.

The oxyanion hole residues form hydrogen bonds (figure C.1) to the oxyanion, with similar distances. The proton ND1 that was initially assigned to the imidazole migrated to the formic acid. Forcing the proton to stay with the imidazole caused the TI to be destabilized

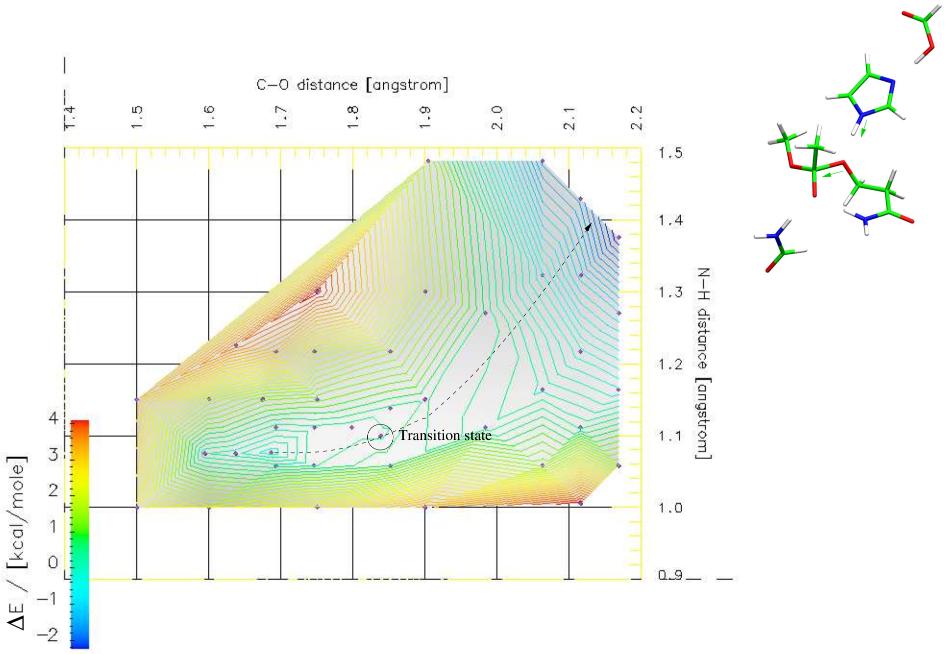


Figure C.2: Two-dimensional PES constructed from model100.

and the substrate ester to be released. The same happened when formic acid was not present to stabilise the positively charged HIS. Hu et al.[100] on the other hand found a stable TI in the latter case. They left all residues to optimize freely and so better stabilization could be attained at geometries, however, that deviate appreciably from the available experimental crystal structure. In our earlier studies, we used the NE2-HE2 bond length as a simple RC along which relaxed potential energy surface scans were performed. This resulted in a discontinous PES similar in shape to that in the study by Nehmuhkin et al.[148] who also used a one-dimensional RC in their investigation of serine proteases. The barrier height obtained in this manner was always too high and a transition state structure could not be located when using any of the structures in the peak region as input to transition state search algorithms. The RC was subsequently assumed to be a combination of two distances, namely the NE2-HE2 distance as before augmented with the C14-O17 leaving coordinate. A relaxed two-dimensional PES scan was performed, starting from the TI. The grid points with fixed NE2-HE2 and C14-O17 distances were chosen manually and the PES was determined only in the most interesting region around the transition state. The result is seen in figure C.2. Starting at the TI the reaction is first mainly guided by the C14-O17 distance, while the NE2-HE2 bond length varies less. HD1 is transferred to ND1 only after the TS is passed. If the NE2-HE2 bond is stretched initially, as in our early attempts, the system is forced into a high energy region of the potential surface (north-west of transition state). With the NE2-HE2 bond length constrained, a passage trough the transition state is then impossible. Better re-

	model100	Zhang[235]	Hu[100]	
NE2-HE2	1.08	1.07	1.05	TI
	1.10	1.22	1.12	transition state
	1.70	1.77	1.81	free ester and enzyme
O17-C14	1.71	1.56	1.53	
	1.84	1.70	1.84	
	2.91	2.59	3.25	
OD2-HD1	1.10	1.66	1.38	
	1.10	1.68	1.48	
	1.48	1.68	1.67	

Table C.1: Comparison of key geometries of three theoretical studies. The coordinates of model100 were derived from crystals of *bacillus subtilis* lipase A, the calculations of other authors are based on models derived from structures of acetylcholinesterases. Computations on our model were done at BLYP/SVP+ level with frozen atoms, Hu used B3LYP/6-31+G* (no atoms frozen) and Zhang employed QM/MM techniques to treat the whole enzyme plus solvation shell, the MM part being represented by the Amber force field and the QM part by HF/3-21G calculations.

model100	Zhang[235]	Hu[100]	
0.15	0.8	1.1	TI → TS
-7.98	-13.4	-16.3	TS → enzyme + ester

Table C.2: Comparison of relative energies (in kcal/mol). Computational levels: model100 BLYP/SVP+ level, Zhang B3LYP/6-31+G*//RHF/3-21G (large QM region), Hu B3LYP/6-31+G*//B3LYP/6-31+G*.

sults could be obtained when more flexible **restraint** reaction coordinates were employed in place of **constrained** distances.

Figure C.3 contains the key geometrical parameters at the stationary points of the reaction. At the transition state HD1 is still assigned to formic acid. In the system studied by Hu[100] the proton remains bound to the imidazole all the time, but the authors noted that the proton transfer between formic acid and the imidazole in the TI proceeded almost barrierless ($\sim$ 0.5 kcal/mol). In our case the HD1 atom which was located at the formic acid in the TI shuttled to the imidazole ring during the reaction. This is a feature of the charge-relay mechanism proposed by Hunkapiller et al.[101]. In the QM/MM model of Zhang et al.[235] the proton stays coordinated to the histidine, and there is a steep barrier for proton transfer to the acid (a GLU side chain). These QM/MM results indicate that our model system may be somewhat artificial. Also, in BSLA there are crystal waters close to the ASP residue of the catalytic triad and there is an indication of hydrogen bonding from the backbone of nearby residues. These interactions are expected to stabilize the ASP anion, which is naked in our model system.

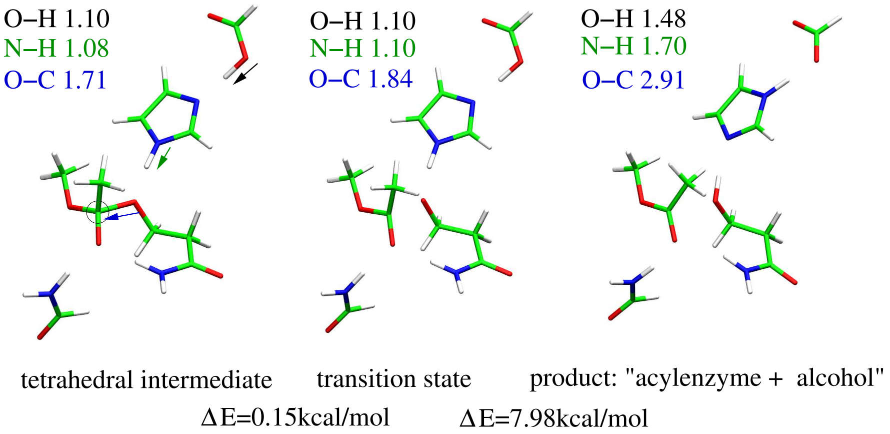


Figure C.3: Reactant, transition state structure and product from BLYP/SVP+ calculations. Reactant and product were derived from the transition state by adding or subtracting increments of the normal mode vector with the negative eigenvalue followed by optimization. The reactant is a model of the tetrahedral intermediate while the product corresponds to an enzyme-substrate complex. Therefore this study proceeds in the reverse direction of the initial step in the acylation reaction. Characteristic bond lengths are color coded. The tetrahedral carbon is marked.

A quantitative comparison with the literature data[100][235] is not possible, since different models and methods were employed. The closest match is the system of Hu[100], but the geometry is not the same and and the oxyanion hole is treated differently. In spite of these caveats, the TI is formed to be a very shallow minimum in all studies, with barriers of only 0.15 - 1.1 kcal/mol for TI→TS (table C.2). The position of HE2 is halfway between NE2 and O17 at the TS in the system of Zhang while the vacuum studies predict a position close to NE2. The equilibrium distances of C14-O17 in the product also differ quite a bit (table C.1).

Appendix D

Numerical Simulation of Chemical Kinetics

The standard procedure for studying the time evolution of reacting chemical systems is solving a set of coupled differential equations[88]. For N chemically active species a set of N differential equations must be treated. For many practical applications these sets of equations can not be solved analytically and numerical methods must be used. This approach is based on the *deterministic* formulation of chemical kinetics.

Gillespie[81][82] introduced a workable *stochastic* formulation which is briefly described in this chapter. Its main advantage for the present work is its elegant and straightforward implementation. Gillespie showed that the stochastic approach is valid whenever the deterministic is, and even in cases where the deterministic is not. His algorithm found practical application and by now (October 2005) there are 498 citations for his 1977 article[82], with a significant part in biological applications. Its usefulness for this discipline can be explained by the ability to treat systems with chaotic behaviour[82], which can be important in gene expression, as well as spatially inhomogeneous systems which, for example, is necessary when communication between cell compartments is to be studied.

Gillespie introduced the reaction probability density function to simulate the time evolution of a chemical reacting system. The function $\mathrm{P}(\tau, \mu)\mathrm{d}\tau$ expresses the probablility that the sytem in state $(\mathrm{X}_1, \mathrm{X}_2, \mathrm{X}_3, ..., \mathrm{X}_\mathrm{N})$, with concentrations X_i of molecular species $\mathrm{i} = 1, ...\mathrm{N}$, experiences a reactive event in the infinitesimal time interval $\mathrm{t} + \tau, \mathrm{t} + \tau + \mathrm{d}\tau$, and that the reaction taking place is of type R_μ, where μ is an integer from the set $\mu = 1, ..., \mathrm{M}$ of possible reactions. The working equation is

$$\mathrm{P}(\tau, \mu) = \mathrm{h}_\mu \mathrm{c}_\mu \exp\left[-\sum_{\nu=1}^{\mathrm{M}} \mathrm{h}_\nu \mathrm{c}_\nu \tau \right], \tag{D.1}$$

where h_μ is a function of the current concentrations X_i, and c_μ is a reaction parameter in the spirit of rate constants in the deterministic formulation.

For a set of initial concentrations and reactions the simulation according to D.1 is an iterative process that runs from time T_1 to T_2. First the values of $\mathrm{h}_\mu\mathrm{c}_\mu$ are calculated, stored, and summed up. A Monte Carlo scheme is then used to generate a random pair (τ, μ) according to the probability density function D.1. The time is advanced by τ and the occurence of reaction μ requires the adjustment of the participating molecular species. If, for example, we have $\mathrm{X} \rightarrow \mathrm{Y}$, we would decrease the molecular population of X by 1 and increase that of Y by the same number. The next iteration is then performed using the new concentration of molecules.

The Monte Carlo scheme required to sample from D.1[81][82] involves recasting the two-variable probability density function into a product of one-variable density functions. The transformation method[1] is applied to generate random deviates from the latter using uniformly distributed random numbers $(\mathrm{r}_1, \mathrm{r}_2)$. This can be done with any of the standard pseudo random number generators. The working expressions are

$$\tau = \left(\sum_{\mu=1}^{\mathrm{M}} \mathrm{h}_\mu \mathrm{c}_\mu \right)^{-1} \ln(\mathrm{r}_1^{-1}), \tag{D.2}$$

$$\sum_{\nu=1}^{\mu-1} \mathrm{h}_\nu \mathrm{c}_\nu < \mathrm{r}_2 \sum_{\mu=1}^{\mathrm{M}} \mathrm{h}_\mu \mathrm{c}_\mu \leq \sum_{\nu=1}^{\mu} \mathrm{h}_\nu \mathrm{c}_\nu \tag{D.3}$$

where μ is an integer, i.e. we compute the sum $\sum_{\nu=1}^{\mu} \mathrm{h}_\nu \mathrm{c}_\nu$ in a *for* loop until it is larger than the central term and take the index of the last added contribution to be μ.

As an example we consider the consecutive reaction

$$\mathrm{X} \underset{\mathrm{k}_{-1}}{\overset{\mathrm{k}_1}{\rightleftharpoons}} \mathrm{Y} \underset{\mathrm{k}_{-2}}{\overset{\mathrm{k}_2}{\rightleftharpoons}} \mathrm{Z}\,. \tag{D.4}$$

We need to find expressions for the quantities $\mathrm{h}_\mu\mathrm{c}_\mu$ appearing in formula D.2. We set $\mathrm{h}_\mu = \mathrm{X}_\mathrm{i}$, since in our example the reaction probabilities are solely dependent on the instantaneous concentrations of a single species i. If two reactants are needed in one step, e.g. for $\mathrm{X} + \mathrm{Y} \rightleftharpoons \mathrm{Z}$, we would set $\mathrm{h}_\mu = \mathrm{X}_\mathrm{i}\mathrm{X}_\mathrm{j}$. Further examples are given in Gillespie's article[81]. The program for such a simulation is rather simple and can be adapted in a straightforward manner to new reaction types. Figure D.2 shows the C++ program that was used to simulate reaction D.4. Generally one run of a simulation is not enough, since it is only one stochastic

---

[1] Nicely described in Numerical Recipes[163], chapter 7.2. It is also used with the bootstrap method in assessing the error bars of free energy calculations.

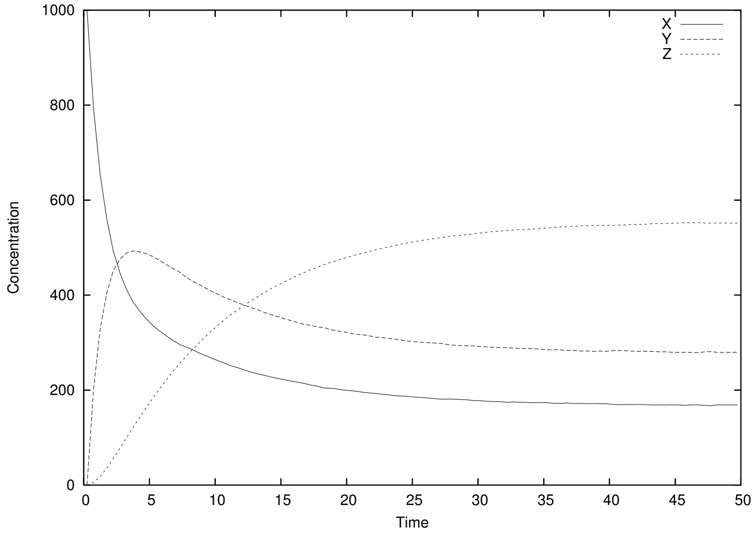


Figure D.1: The course of concentrations during a consecutive reaction (see text).

realisation of the process under study. Therefore 250 runs were executed in our example to obtain a statistics. The initial concentration of X is 1000 molecules, Y and Z are set to zero. The parameters c_i which correspond to rate constants k_1, k_{-1}, k_2, k_{-2} of the deterministic formulation given in D.4 are taken to be 0.5, 0.3, 0.1, and 0.05. The chosen time interval permits the reaction to attain a steady state. The binning interval was set to 0.5 time units and used to create a histogram for all species, which is shown in fig. D.1. The results are easily appreciated since this example appears in most textbooks on physical chemistry and chemical kinetics[7][128].

Numerical Simulation of Kinetic Resolution Experiments

The method presented here can be used to simulate the time evolution of a reacting system of enantiomers under conditions of kinetic resolution. In a sense one performs the experiment *in silico*. Activation barriers from quantum chemical calculations are the input parameters (see appendix E).

To this end a program was written, which is listed on pages 176-178. A comparison to the example program for simulation of a consecutive reaction (page 175) reveals a very similar structure and highlights the adaptability of Gillespie's algorithm to any kind of kinetic problem. The kinetics implemented is more detailed than that of equation 1.3 and the discussion in chapter 1, and includes the TI as a stable state, which adds four rate constants compared to the simpler kinetics. Please refer to chapter 4 for a discussion of the PES around the TI (see section 4.2.2). The corresponding kinetic equations and rate constants are given in the header of the program on page 176.

Although it is possible to affect enantioselectivity during the noncovalent steps of the enzymatic reaction (modulation of $\mathrm{k}_1, \mathrm{k}_{-1}, \mathrm{k}_2, \mathrm{k}_{-2}$) this is unlikely to be the case in lipases[57]. Accordingly, identical values were thus used for k_1 and k_2, as well as k_{-1} and k_{-2}. The parameter 0.00001 kcal/mol was chosen to represent k_1 (k_2), and 6 or 14 kcal/mol for k_{-1} (k_{-2}). This makes the binding step essentially barrierless, and adds a reasonable barrier for the reverse reaction. Experimental values for BSLA are not known, but from the estimated values of chapter 5 we conclude that the adopted values are realistic. The choice of a low (6 kcal/mol) and a high (14 kcal/mol) value for k_{-1} (k_{-2}) was made based on efficiency considerations. When the rates of individual reactions in the system are very different, most time is spent on simulation of the faster steps. These are the unbinding steps in the kinetic resolution, since the barriers of k_{-1} (k_{-2}) must be smaller than the effective barrier of the acylation step to retain enantioselectivity. The values were therefore chosen low enough to allow an enantioselective reaction to occur, while maintaining numerical efficiency. The entire simulations consumed on the order of 10-30 minutes on a pentium IV processor at 1.8 GHz, when 50 runs were computed for the sake of averaging the results.

The analysis of enantioselectivity was restricted to the covalent steps of the acylation reaction (rates $\mathrm{k}_3, \mathrm{k}_{-3}, \mathrm{k}_4, \mathrm{k}_{-4}, \mathrm{k}_5, \mathrm{k}_6$). Step [I] was treated as being fully reversible. Step [II] (figure 1.2) was presumed irreversible due to water being in large excess of alcohol under the experimental conditions of ester hydrolysis by lipases. Deacylation was treated as a single irreversible step (k_7) and assigned a barrier of 12 kcal/mol. A precise value was not needed because the acylation reaction is irreversible via k_5 and k_6, which makes the deacylation step non-enantiodiscriminative.

Program output. Please refer to the program header on page 176.

Compilation notes. On a linux machine use the gnu c++ compiler and type "g++ programName.cc -o programName". The program utilises the random number generator ran2[163] that can be retrieved from http://nr.com. Instead of including ran2 (##include "ran2.c") from a file you may paste the function in front of the main program.

```
#include <iostream>
#include <iomanip>
#include "math.h"
#include "ran2.c"


/*  Sample program for simulation of a consecutive reaction X <> Y <> Z. Simulation based on
    the theory by D Gillespie, J. Comp. Phys., 22:403-434 (1976).
    The ran2 pseudo random number generator from "Numerical Recipes" (www.nr.com) is used.
*/

//seed for random number generator
long seed = 31423523;

using namespace std;

//Number of stochastic simulations
int runs = 250;
/* Number of possible reactions, back-reactions
   are treated as separate reactions
*/
int ingred = 4;

//Dump interval
const double printit = 0.5;

double z1,z2,z2a0,sum;
int mu;

int main(void){

  for (int s = 0; s < runs;s++){

    //start time
    double T = 0.0;
    //end time
    double T2 = 50.0;
    //concentrations of molecules at T
    double X[] = {1000.0,0.0,0.0};
    //Expressions for the rate parameters
    double C[] = {0.5,0.3,0.1,0.05};
    //Array holding the h_mu*c_mu's
    double A[] = {0.0,0.0,0.0,0.0};
    // sum_mu h_mu*c_mu
    double A0;
    double tcount = 0.0;

    while( T < T2){

      A[0] = C[0]*X[0];
      A[1] = C[1]*X[1];
      A[2] = C[2]*X[1];
      A[3] = C[3]*X[2];
      A0 = A[0] + A[1] + A[2] + A[3];

      z1 = ran2(&seed);
      z2 = ran2(&seed);

      T = T + log(1.0/z1)/A0;
      if (T ≥ tcount){
       cout << setw(15) << T << setw(15) << X[0] << setw(15) << X[1]
            << setw(15) << X[2] << endl;
       tcount += printit;
      }

      z2a0 = z2*A0;
      sum = 0.0;
      for (int i = 0; i < ingred;i++){
        mu = i;
        sum += A[i];
        if (sum ≥ z2a0) break;
      }

      switch(mu){
        case 0:
          X[0] -= 1.0;
          X[1] += 1.0;
          break;
        case 1:
          X[0] += 1.0;
          X[1] -= 1.0;
          break;
        case 2:
          X[2] += 1.0;
          X[1] -= 1.0;
          break;
        case 3:
          X[2] -= 1.0;
          X[1] += 1.0;
          break;
      }

    }

  }

};
```

Figure D.2: C++ program for stochastic simulation of a consecutive reaction.

```
#include <iostream>
#include <fstream>
#include <iomanip>
#include "math.h"
#include "ran2.c"
#include "stdlib.h"

/*

    Programmed by N. Otte in 2006.
    Program for simulation of the following kinetic resolution experiment, where R and S are chiral esters.
    The chiral moiety is the alcohol, thus a common intermediate is formed by the end of the acylation reaction:

    E + S <> E*S <> R*TI  ->  EX + H2O  ->  E + P2
                         -S1  ^
                              |
                              /
                             /
    E + R <> E*R <> S*TI  --/
                         -R1


    We simulate the following set of reactions:

              k1
    1. E + S  <>  E*S
              k-1

              k2
    2. E + R  <>  E*R
              k-2


            k3
    3. E*S  <>  S*TI
            k-3

            k4
    4. E*R  <>   R*TI
            k_4


             k5
    5. S*TI   ->   EX + S1


             k6
    6. R*TI   ->   EX + R1



                  k7
    7. EX + H2O   ->    E + P2

    We therefore have 12 chemical species and 11 rate constants.


    The program expects as command line input:

     1. Reaction temperature.
     2. Reaction barriers in kcal/mol corresponding to the 11 rate constants presented above.
        The program calculates the rate constants according to the Arrhenius equation.

    Output:

      A text file named "results.plt". The output is in columns with the following content:

        1. Extent of reaction (0-1)
        2. Concentration S1
        3. Concentration R1
        4. ee from S1 and R1
        5. E


      The stochastic simulation is run n times (5 is default, please adjust variable "runs"
      below to change this), and the output in file "results.plt" is sequential.
      At the end of the simulation the program writes out an averaged E value. The E value is
      measured and averaged in the interval of reaction extent 0.3-0.5, where it is most stable.

      Note that it is assumed species S is the preferred enantiomer. If this is not the case
      the ee and E values will have to be inverted.

      Feel free to adjust initial concentrations found in array X[]. The current values were
      found to give reliable and results in a timesaving fashion.

    Stochastic simulation based on
    the theory by D T Gillespie, J. Comp. Phys., 22:403-434 (1976).
    The ran2 pseudo random number generator from "Numerical Recipes" (www.nr.com) is used.

*/


/* seed for random number generator */
long seed = 31423523;

const double boltz = 1.3806505e-23;
const double avogadro = 6.0221415e23;
```

Figure D.3: C++ program for stochastic simulation of kinetic resolution. Part **A**.

```
const double toJoule = 4.184;
using namespace std;

/* Number of stochastic simulation runs */
int runs = 5;
/* Number of possible reactions, back-reactions
   must be treated as separate reactions.
*/
int ingred =11;

/* Dump interval */
const double printit = 1.0;
double z1,z2,z2a0,sum;
int mu;
int valu = 0;
double summit = 0.0;


int main(int argc, char *argv[]){

  if (argc < 13){
  cout << "Please enter the following arguments:" << endl
       << "1. Reaction temperature [K]" << endl
       << "2. Barriers [kcal/mol] for all 11 reaction steps." << endl;
   exit(1);
  }

  ofstream outdata;
  outdata.open("results.plt",ios::out);

  //factor RT for T = 25°C
  double temp = atof(argv[1]);
  double RT = ( temp * boltz * avogadro ) / (toJoule*1000.0);

  double a,b,c,d,e,f,g,h,i,j,k;
  double scale = 1.0;
  a = exp(-atof(argv[2])/RT)*scale;
  b = exp(-atof(argv[3])/RT)*scale;
  c = exp(-atof(argv[4])/RT)*scale;
  d = exp(-atof(argv[5])/RT)*scale;

  e = exp(-atof(argv[6])/RT)*scale;
  f = exp(-atof(argv[7])/RT)*scale;
  g = exp(-atof(argv[8])/RT)*scale;
  h = exp(-atof(argv[9])/RT)*scale;
  i = exp(-atof(argv[10])/RT)*scale;
  j = exp(-atof(argv[11])/RT)*scale;
  k = exp(-atof(argv[12])/RT)*scale;


  for (int s = 0; s < runs;s++){

    /* start time */
    double T = 0.0;
    int counter = 0;

    /* Order of species in contration array: E, S, R, E*S, E*R, S*TI, R*TI,
     EX, S1, R1, H20, P2 */
    double X[] = {5.0, 5000.0, 5000.0, 0.0, 0.0, 0.0 ,0.0, 0.0, 0.0, 0.0, 1000000.0, 0.0 };

    /* The array of rate parameters */
    double C[] = {a,b,c,d, e,g,f,h,i,j,k};
    /* Array holding the h_mu*c_mu's */
    double *A = new double[11];
    double initR = X[1];
    double initS = X[2];
    double initconcentration = X[1]+X[2];
    double A0,newext,ee;
    double oldext =-5.0;
    double tcount = 0.0;

     while( oldext < 0.5 ){

      A[0] = C[0]*X[0]*X[1];
      A[1] = C[1]*X[3];

      A[2] = C[2]*X[0]*X[2];
      A[3] = C[3]*X[4];

      A[4] = C[4]*X[3];
      A[5] = C[5]*X[4];

      A[6] = C[6]*X[5];
      A[7] = C[7]*X[6];

      A[8] = C[8]*X[5];
      A[9] = C[9]*X[6];

      A[10] = C[10]*X[10]*X[7];

      A0 = 0.0;
      for (int i = 0; i < ingred; i++){
        A0 = A0 + A[i];
      }

      z1 = ran2(&seed);
      z2 = ran2(&seed);
```

Figure D.4: C++ program for stochastic simulation of kinetic resolution. Part **B**.

```
      T = T + log(1.0/z1)/A0;
      counter++;
      newext = 1.0 - (X[1]+X[2])/initconcentration;
      if ( (newext - oldext) > 0.01 ){
          ee = (X[8]  - X[9]) / (X[8]  + X[9]);
          outdata  << setw(15) << newext  << setw(15) << X[8]
                << setw(15) << X[9]  <<  setw(15) << ee*100.0
                << setw(15) << log(1.0-newext*(1.0+ee))/log(1.0-newext*(1.0-ee))
                << setw(15) << endl;
       tcount += printit;
       oldext = newext;
       if (oldext > 0.3 ∧ oldext < 0.5){
         valu++;
         summit += log(1.0-newext*(1.0+ee))/log(1.0-newext*(1.0-ee));
       }
      }

       z2a0 = z2*A0;
      sum = 0.0;
      for (int i = 0; i < ingred;i++){
        mu = i;
        sum += A[i];
        if (sum ≥ z2a0) break;
      }

      switch(mu){
        case 0:
          X[0] -= 1.0;
          X[1] -= 1.0;
          X[3] += 1.0;
          break;
        case 1:
          X[0] += 1.0;
          X[1] += 1.0;
          X[3] -= 1.0;
          break;
        case 2:
          X[0] -= 1.0;
          X[2] -= 1.0;
          X[4] += 1.0;
          break;
        case 3:
          X[0] += 1.0;
          X[2] += 1.0;
          X[4] -= 1.0;
          break;
        case 4:
          X[3] -= 1.0;
          X[5] += 1.0;
          break;
        case 5:
          X[4] -= 1.0;
          X[6] += 1.0;
          break;
        case 6:
          X[3] += 1.0;
          X[5] -= 1.0;
          break;
        case 7:
          X[4] += 1.0;
          X[6] -= 1.0;
          break;
        case 8:
          X[5] -= 1.0;
          X[7] += 1.0;
          X[8] += 1.0;
          break;
        case 9:
          X[6] -= 1.0;
          X[7] += 1.0;
          X[9] += 1.0;
          break;
        case 10:
          X[7] -= 1.0;
          X[10] -= 1.0;
          X[0] += 1.0;
          X[11] += 1.0;
          break;
      }

    }

    /* write last step */
    outdata  << setw(15) << newext  << setw(15) << X[8]
                << setw(15) << X[9]  <<  setw(15) << ee*100.0
                << log(1.0-newext*(1.0+ee))/log(1.0-newext*(1.0-ee))
                << endl;
    delete A;
  }

 cout << "Averaged E value: " << summit/static_cast<double>(valu) << endl;
 outdata.close();

};
```

Figure D.5: C++ program for stochastic simulation of kinetic resolution. Part **C**.

Appendix E

Kinetic Resolution: A Closer Look

In this chapter we investigate how to connect our theoretical results to experimental findings. To this end basic concepts from the analysis of kinetic resolution experiments are introduced. Thorough overviews of this field are provided by Kagan and Fiaud[109] as well as Sih and Wu[193]. We assume an irreversible enzymatic reaction that is described as:

$$\begin{aligned} \mathrm{E}+\mathrm{R} &\overset{k_1}{\rightarrow} \mathrm{E}+\mathrm{A} \\ \mathrm{E}+\mathrm{S} &\overset{k_2}{\rightarrow} \mathrm{E}+\mathrm{B} \end{aligned} \tag{E.1}$$

where E is the concentration of enzyme and R and S that of the chiral reactants. A and B are chiral products. Constants k_1 and k_2 are *apparent* rate constants, that can be composed from a number of rates of more elementary steps. For the *analysis* of an experiment, be it a real world, or a numerical experiment, no rate constant is needed and there can be more than one step for each enantiomer. To be able to *perform* a numerical experiment one assumes knowledge of the underlying kinetics and elementary steps, of course. The point to note is that the following analysis of enantioselectivity remains identical, irrespective of the number of steps in E.1. All parameters that we need are the concentrations of reactants and products, measured at intervals along the course of the reaction. In the real word chromatographic methods can be used, while in numerical simulation of the reactions all concentrations are easily at hand. The **enantiomeric excess (e.e.)** can be calculated in both reactants and products. In terms of E.1 we can calculate the e.e. in the products as:

$$\mathrm{e.e.}[\%] = \frac{[\mathrm{S}]-[\mathrm{R}]}{[\mathrm{S}]+[\mathrm{R}]} \cdot 100[\%] \tag{E.2}$$

The e.e. must be a positive number, therefore one may have to exchange S and R, or simply take the absolute value of E.2. This expression to evaluate the e.e.'s is used by the kinetic resolution simulation program in the preceding appendix. Please note that the e.e. depends on the degree of conversion of the reactants. The progression of curves in figure 1.6 on page 14 confirms that the e.e. deteriorates quickly after 50% conversion of reactants.

A concentration-free parameter of enantioselectivity is **E**. In chapter 1 **E** was defined in terms of rate constants. Alternatively one can use a definition in terms of concentrations:

$$\mathbf{E} = \frac{\ln[1 - \mathrm{c}(1 + \mathrm{e.e.}_{\mathrm{products}})]}{\ln[1 - \mathrm{c}(1 - \mathrm{e.e.}_{\mathrm{products}})]} \tag{E.3}$$

$$\mathbf{E} = \frac{\ln[(1 - \mathrm{c})(1 - \mathrm{e.e.}_{\mathrm{reactants}})]}{\ln[(1 - \mathrm{c})(1 + \mathrm{e.e.}_{\mathrm{reactants}})]} \tag{E.4}$$

Expression E.3 is to be used with the e.e. in the products, and E.4 with the e.e. in the reactants. Variable c specifies the extent of the reaction in terms of the conversion of reactants and assumes values between 0 and 1. **E** assumes values ≥ 1. If values < 1 are encountered the value must be inverted.

A comparison of experimentally and theoretically derived enantioselectivies requires a common parameter. The obvious choice is E. From our calculations of the acylation reaction potential energy and free energy barriers for the elementary steps of the acylation reaction were derived. How then to make a connection from these barrier parameters to **E** ?

Exact Method

Obviously $\Delta\mathrm{E}^{\ddagger}$ and $\Delta\mathrm{G}^{\ddagger}$ can be transformed into rate constants using the Eyring equation $\mathrm{k} = \frac{\mathbf{k}\mathrm{T}}{\mathrm{h}}\exp(-\frac{\Delta\mathrm{G}^{\ddagger}}{\mathrm{RT}})^{1}$. These rate constants can then be used as input to solve the kinetic system of equations, or an alternative is to use a stochastic simulation of the type described in the preceding appendix[2]. In this manner we introduce competition between the enantiomers and/or binding modes. Ultimately **E** is a measure of the relative proficiency of the enantiomers under competitive conditions. From the molecular distributions in the output of the program **E** is easily computed according to equation E.3.

Approximate Method

An approximate **E** value can be calculated from the *absolute* barriers using the Eyring equation

$$\mathbf{E} = \exp(-\frac{\Delta\mathrm{G}^{\ddagger}_{\mathrm{R}} - \Delta\mathrm{G}^{\ddagger}_{\mathrm{S}}}{\mathrm{RT}}) = \exp(-\frac{\Delta\Delta\mathrm{G}^{\ddagger}_{\mathrm{R-S}}}{\mathrm{RT}}) \tag{E.5}$$

---

[1] $\Delta E^{\ddagger}$ is often also used in place of $\Delta\mathrm{G}^{\ddagger}$.

[2] The program that was used to evaluate **E** throughout this text starts on page 176.

These $\Delta G^{\ddagger}$ values do not correspond to elementary steps of the reaction, but represent the activation barrier of the rate-determining step[230]. Blue arrows in figure 4.8 on page 77 indicate the rate-determining barriers of the acylation reaction for all possible combinations of two barriers.

Once calculated the $\mathbf{E}$ value can be used to reassess the *apparent* (effective) $\Delta\Delta G^{\ddagger}$ value from equation (E.5). If the approximate method was used to calculate $\mathbf{E}$ this simply gives back the input parameters, of course. With the numerically computed 'exact' value of $\mathbf{E}$ the apparent $\Delta\Delta G^{\ddagger}$ value contains contributions from all elementary steps, in a way not easily assessed by qualitative arguments.

Appendix F

The Empirical Valence Bond Model

In studies of chemical systems the surrounding of a reacting fragment can have an important influence on the chemical transformation. When effects that originate from the directionality of interactions are considered important an atomistic treatment of the whole condensed phase system may be the only realistic way to incorporate these contributions. In such models we can account for reorganisation during chemical reactions only through MD or Monte Carlo simulations, for example free energy simulations as described in chapter 2. The success of such techniques depends to a large degree on the ability to execute very large numbers of simulation steps. This is presently not possible with either pure *ab initio* or QM/MM methods that use *ab initio* representations of the QM region. Semiempirical QM methods are applicable to limited classes of molecules and usually have problems treating non-standard systems. In these cases the empirical-valence-bond (EVB) model may provide an interesting alternative.

EVB is conceptually derived from valence bond models, however, electronic degrees of freedom are not treated explicitly. The method uses classical force fields to represent the resonance structures. Transfer from one resonance structure to the other is brought about by a mixing of the individual potentials using parametrised exchange integrals. In a two-state EVB model we have one exchange term H_{12} which is often taken to be a Gaussian of some distance parameter. The EVB model can be derived writing

$$\mathrm{E} = \frac{\int \varphi \hat{\mathrm{H}} \varphi \mathrm{d}\tau}{\int \varphi \varphi \mathrm{d}\tau}, \tag{F.1}$$

where φ is a linear combination

$$\varphi = \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{c_i} \chi_{\mathbf{i}}, \tag{F.2}$$

and insertion of this expression into F.1 yields

$$E = \frac{\sum_{i,k}^{n} c_i^* c_k \int \chi_i \hat{H} \chi_k d\tau}{\sum_{i,k}^{n} c_i^* c_k \int \chi_i \chi_k d\tau}. \tag{F.3}$$

For the two-state model the energy expression reads

$$E = \frac{c_1^2 H_{11} + c_2^2 H_{22} - 2c_1c_2H_{12}}{c_1^2 S_{11} + c_2^2 S_{22} + 2c_1c_2S_{12}} \tag{F.4}$$

The overlap element S_{12} is taken to be zero and everything is parametrised into the element H_{12} itself. S_{11} and S_{22} are equal to one. With the requirement that $c_1^2 + c_2^2 = 1$ and the abbreviation $a_1 = \sqrt{c_1}$ and $a_2 = \sqrt{c_2}$ we obtain

$$E = a_1 H_{11} + a_2 H_{22} - 2\sqrt{a_1 a_2} H_{12}. \tag{F.5}$$

This can be used as a mapping potential in the calculation of free energy differences by FEP, or to drive the system from one resonance structure ('state') to the other in geometry optimisations, similar to the minimum energy reaction path method. The coefficients a_1 and a_2 serve as constraints in the Hamiltonian. The energy *without* constraint is calculated as the solution of the characteristic polynomial that is obtained from the determinant of the 2x2 EVB matrix

$$\begin{vmatrix} H_{11} - E & H_{12} \\ H_{12} & H_{22} - E \end{vmatrix}. \tag{F.6}$$

The energy expressions H_{11} and H_{22} contain all force field terms that are needed for a separate simulation of the individual resonance structures (with a few modifications described below), plus an additional constant energy offset that is added to either of them.

The lower root of the polynomial is

$$E = \frac{1}{2}\left[H_{11} + H_{22} - 2\sqrt{(H_{11} - H_{22})^2 + 4H_{12}^2}\right], \tag{F.7}$$

and the gradient, needed for optimisation and MD, is expanded as

$$\frac{dE}{d\mathbf{r}} = \frac{1}{2}\left[\frac{dH_{11}}{d\mathbf{r}} + \frac{dH_{22}}{d\mathbf{r}} - \frac{(H_{11} - H_{22})\frac{dH_{11}}{d\mathbf{r}} - (H_{11} - H_{22})\frac{dH_{22}}{d\mathbf{r}} + 4H_{12}\frac{dH_{12}}{d\mathbf{r}}}{\sqrt{(H_{11} - H_{22})^2 + 4H_{12}^2}}\right]. \tag{F.8}$$

The simulation of bond forming and breaking requires a different bond potential than the simple harmonic form to allow for dissociation. Morse potentials of the type

$$\mathrm{E} = \mathrm{D} \cdot [1 - \exp(-\beta \cdot (\mathrm{r} - \mathrm{r}_0))]^2 . \tag{F.9}$$

prove to be a practical choice and offer enough flexibility.

The VdW potential used in Charmm22 is of the standard 6-12 Lennard Jones type, which is too repulsive at short distances. Between dissociating atoms a 'soft-core' VdW potential of the form

$$\mathrm{E} = \mathrm{a} * \exp(-\mathrm{b} * \mathrm{r}) \tag{F.10}$$

is used. The multiplicative factor a scales the potential and b determines its width.

Finally the functional form of H_{12} must be chosen. In his extensive EVB work Warshel mainly used Gaussians of simple atomic distances to parametrise the exchange element H_{12}. Presently H_{12} is represented by the function

$$\mathrm{H}_{12} = \mathrm{a}_1 \exp[-\mathrm{b}_1(\mathrm{r}_{12} - \mathrm{r}_{34} - \mathrm{r}_{56} - \mathrm{S}_1)^2] + \mathrm{a}_2 \exp[-\mathrm{b}_2(\mathrm{r}_{12} - \mathrm{r}_{34} - \mathrm{r}_{56} - \mathrm{S}_2)^2], \tag{F.11}$$

which expresses H_{12} in terms of a three-component RC. The relevant distances r_{12}, r_{34}, and r_{56} were chosen as in chapters 4 and 5, i.e. they refer to the core region of the reacting system: $\mathrm{r}_{12} = \mathrm{r}_{\mathrm{O17-HE2}} \vee \mathrm{r}_{\mathrm{O3-HE2}}$, $\mathrm{r}_{34} = \mathrm{r}_{\mathrm{NE2-HE2}}$, $\mathrm{r}_{56} = \mathrm{r}_{\mathrm{C14-O17}} \vee \mathrm{r}_{\mathrm{C14-O3}}$. The use of two Gaussians allows a finer tuning of the PES via scaling of the potential through $\mathrm{a}_1, \mathrm{a}_2$, and adjustment of widths $\mathrm{b}_1, \mathrm{b}_2$, and positions $\mathrm{S}_1, \mathrm{S}_2$. This more complex form of the exchange integral has interesting consequences. The dependence of H_{12} on the RC mixes in contributions of the dissociated state for configurations where HE2 comes close to either O3 or O17, which leads to a stretching of the bond C14-O3 or C14-O17. In an EVB model it is thus possible to model the bond elongation upon hydrogen bonding of HE2 with O17 or O3 (see chapter 4.2.1). Since a two-state EVB model is used, dissociation of the TI can only be towards MCC or PDC in any single simulation limiting this effect to *either* C14-O3 or C14-O17. A three-state EVB representation could serve as a global model of the acylation reaction.

All parameters of the modified force field terms, the exchange element H_{12}, and the energetic offset may be chosen such that reasonable transition geometries and energies are obtained. To be in a position that allows such parametrisation obviously presupposes detailed knowledge of the mechanism before the EVB study can be performed. Therefore the EVB model cannot be used to elucidate new reaction mechanisms as everything is parametrised with existing knowledge. Its strengths are in studying differential effects of the environment, such as those which occur upon mutations in enzymes or play a role for enantioselectivity. The

Morse	D	b	r_0	EVB State
O3(O17)-HE2	102.0	2.35	0.96	2
C14-O3(O17)	93.0	2.09	1.52	1
Soft-core VdW	a	b		
O3(O17)$\cdots$HE2	60.0	2.5		1
C14$\cdots$O3(O17)	1300	2.5		2
NE2$\cdots$HE2	60.0	2.5		2
Exchange Potential H_{12}				
	a_1	a_2	b_1	
	50.0	0.0	0.8	
	b_2	S_1	S_2	
	-	-2.0	-	
Energy Gap State 1↔2			109.0	

Table F.1: Parameters for a two-state EVB simulation of the acylation reaction in lipases. Charge parameters of the TI were adopted from our Charmm22 parametrisation, those of the esters in either MCC or PDC from top_all22_model.inp and top_all27_prot_lipid.inp. For step [II] of the acylation reaction the above parameters of O17 can be exchanged for their O3 counterparts.

groups of Warshel[222] and Hammes-Schiffer[225] were able to relate specific mutations to experimental findings by FEP calculations using the EVB potential.

Simulation Parameters

In our simulations state 1 was the TI, and state 2 the PDC. Most parameters were derived from values given by Warshel[217]. Parameters of Morse potentials were adopted almost unaltered, as were the soft-core VdW interactions. Most of the parametrisation concentrated on the coupling term H_{12}. Variation of parameters was done with the intention of obtaining geometries in approximate accordance with previous QM/MM geometry optimisations. Potential barriers were only expected to be roughly of the same order as in QM/MM optimisations since the study was comparative. Charges of the esters were set as for methyl-acetate (see top_all22_model.inp), as were the atoms types. Histidine charges and parameters were set to the typical values found in top_all27_prot_lipid.inp. For state 1 we selected topology HSP (doubly protonated His), and HSD (singly protonated at ND1) for state 2.

With the resulting set of parameters (table F.1) we attempted an optimisation of the system ANRU453, which also served as a test system in chapter 4. To this end the modified CHARMM program was used. We employed the mapping potential of equation F.5 to drive the system from the initial (TI) to the final (PDC) configuration in discrete steps of 0.05 of

Geometrical feature	TI	TS	PDC
Bond Distance			
O3-HE2	1.69	1.26	0.98
NE2-HE2	1.04	1.24	1.95
C14-O3	1.54	1.73	2.87
C14-O15	1.27	1.25	1.22
C14-O17	1.51	1.48	1.33
ND1-HD1	1.03	1.03	1.02
OD2-HD1	1.61	1.69	1.71
ILE2-O15	1.80	1.82	2.04
MET78-15	1.94	1.96	1.90
Angle			
NE2-O3-HE2	5.61	2.45	7.87
OD2-ND1-HD1	6.13	8.00	6.64
Torsion			
ω_4	83.8	95.3	113.0
ω_3	121.6	120.5	67.3
ω_2	158.6	153.1	-171.7
ω_1	-66.9	-71.5	-65.5

Table F.2: Geometrical parameters at the stationary points of optimisations with the two-state EVB model of the BSLA.

the coefficients a_1 and a_2. At each step a geometry optimisation was executed. The results are given in table F.2.

From a qualitative point of view the model reproduced the typical development of geometric features over the reaction path that were described in chapter 4. The optimised distances and bond angles agree almost quantitatively with the QM/MM derived structures. Torsions have larger differences, and the potential barriers are too high: we computed 4.4 kcal/mol for TI → TS of step [II], and 20.0 kcal/mol for the transformation $\mathrm{PDC} \rightarrow \mathrm{TS}$. Therefore the EVB model does not fully describe the energetic and geometric features that were found in the QM/MM calculations. FEP simulations with EVB were tested as an option, once more using mapping potential F.5. This did not produce the correct ordering of energies of RIII and SI in calculations of step [I]. One drawback of the FEP-EVB method is that free energy differences of the surrounding with itself are not taken into account in the determination of ΔG, since only the forces on the region of pseudo-QM atoms in the EVB core are included. A recently proposed approach[197] that combines FEP calculations with WHAM may be used to obtain more realistic free energy differences.

The EVB approach is flexible enough, at least in principle, to be parametrised for complex chemical systems. In the present implementation its computational cost is twice that of a standard force field calculation. When fast (semiempirical) MO or DFT methods are available, which allow extended MD simulations at the QM/MM level, they are preferable over

EVB, since normally a parametrisation can be avoided and the handling of calculations is much simpler. For other systems, where MD simulation can give valuable insights into a molecular process, but where only *ab initio* methods capture the consequences of a complicated electronic structure realistically, EVB is an interesting alternative. It may be extended to more than two states to obtain more realistic representations of molecular systems. Schmitt and Voth devised the multistate-EVB method to simulate water in the condensed phase[182]. A recent application to condensed phase systems was reported by Brancato and Tuckerman[28].

Implementation

The EVB model was included in the CHARMM program version c28b2. Our implementation is based on an EVB routine that was supplied by Konrad Hinsen, and had been used with CHARMM as of version c24b1. This routine was employed by Hinsen and Roux in published work[98] on quantum dynamics of a small molecule (acetylacetone) in the gas phase. We had to extend the code by subroutines which build the interaction lists of the individual resonance structures with the remainder of the protein system automatically. In this context the EVB model proved to have some similarities with the QM/MM philosophy. Atoms which experience changes in charge, VdW parameters or bond list form a 'QM' region, the frontier is at the first atom which does not experience such a direct modification. All energy terms over the boundary are retained and enter H_{11} or H_{22}. Morse potentials, soft-core VdW interactions and the functional form of H_{12} were introduced in CHARMM as well as the possibility to perform FEP simulations with mapping potential F.5.

Appendix G

Cluster Analysis

Models of condensed phase molecular systems can provide realistic reaction barriers via simulation or geometry optimimisation. In either case the computational chemist must verify that certain boundary conditions, such as structural integrity of the model, are met and that a reaction coordinate, or more generally an order parameter, can be devised to differentiate reactants from products. The simulation or optimisation protocol is then implemented to produce potential or free energy barriers. In complex models these barriers will clearly depend on a number of geometrical parameters, which are possibly not identified by the time simulations are executed. In other cases one has an idea of the parametric dependence beforehand and wants to ensure a structural diversity with regard to these structural elements at the outset of the calculations.

In either case it can be helpful to attempt a reduction of data space, which is where multivariate statistial methods[94][200] come into play. Among these, cluster analysis has the capacity to identify groups of configurations with respect to arbitrary geometric criteria. Cluster analysis of conformations generated by MD has some tradition[112][206][189][42]. One recent application addressed protein folding[49].

Cluster analysis depends on the ability to calculate distances between items in data space. For distances between atoms the simple Euclidean metric can be used, but in a space of torsion angles the periodicity of the data has to be accounted for. In the first step distances between all items of the data set are computed. One must then decide which clustering method to use. The conceptually simplest techniques are the *hierarchical* methods[94], but it is also possible to perform clustering with the help of neural nets[112]. Here we used *single linkage* and *complete linkage* hierarchical clustering. A cluster analysis is performed in sequential steps. Initially we have N distance clusters. At each following step two clusters are merged into a new cluster, so that N ← N - 1. The distance of the new clusters all other objects is re-evaluated, and another round of clustering executed. One always merges clusters with the closest distance, but how the distance between the clusters is computed differentiates single linkage from complete linkage clustering. In single linkage clustering the distance between two clusters is taken to be the smallest distance between individual data points of clusters, and in complete linkage the largest distance. The distances of clusters are thus assigned based on single item distances, for example between two atoms, or between two torsions. A reduction of the data is successful only when few significant clusters remain at the final clustering steps.

A clustering program was written and tested on a system where the results could be visualised easily. As test data we used our trajectory data, and as objects to be clustered the torsion angles $\omega_1 - \omega_4$. A survey of the raw data helped to develop a feeling for the cluster structure and to settle our expectations. Figure G.1 shows the data set projected on dihedral space in two two-dimensional scatter plots. The dihedral angles are those which define the orientation of the TI in the active site (see Figure 3.3). The left plot reveals a rather narrow bandwidth of torsion ω_1, which clearly allows the identification of three groups along torsion ω_2. A further divison of clusters 1 and 2 may be possible and would lead to a total number of five clusters. The scatter plot on the right hand side for torsions ω_3 and ω_4 has a less tidy structure and offers more possibilities for decomposition. Although six clusters are indicated, it seems equally justified to further subdivide clusters 1, 5 and 6, while number 4 may even be split into three parts. The maximum number of clusters thus created would be eleven. Since the structure of ω_3 versus ω_4 is more interesting than that of ω_1 versus ω_2 cluste analysis was tested on this reduced data set. We applied single linkage and complete linkage hierarchical clustering, the output of which is visualised at different steps of the clustering procedure in figure G.2. Obviously both methods have a tendency to stress extremes: The single linkage method exaggerates the proximity of clusters, complete linkage the distance between them. This became very apparent in the final steps of the clustering procedure, which are visualised in the topmost plots in G.2 and show the results for the remaining six clusters. Figure G.2 reveals that neither single linkage nor complete linkage clustering can reproduce the subjective structure of six groups predicted by visual inspection of the raw data. When twelve or sixteen clusters remained, the single linkage method identified the minimum number of six clusters, not counting outliers. In a selection process very small groups may be identified as outliers and dropped according to a size-dependent cutoff criterium. Complete linkage leads to eleven and fifteen larger clusters. Neither method separated cluster 5 of figure G.1.

Cluster analysis can be helpful only when the procedure is used in *unsupervised* fashion,

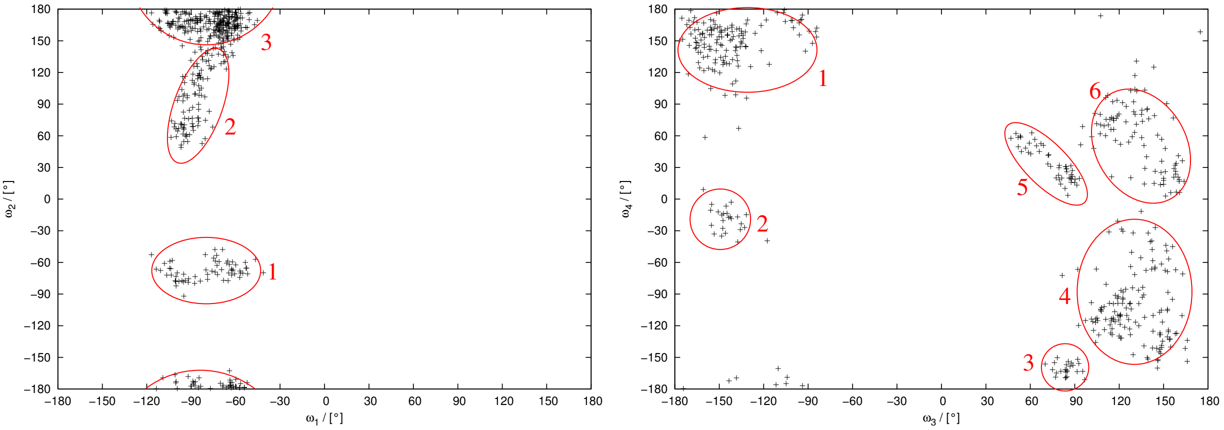


Figure G.1: Two-dimensional torsional diagrams created from the entire set of trajectories of binding modes of (R)-NAPHAC and (S)-NAPHAC (see chapter 3). Data points correspond to snapshot configurations which were extracted every 50 ps.

which means that the analyst can rely on the results without having to reinspect the data visually each time. Techniques that allow judgment of the goodness of a cluster solution are available in [94]. For smaller data sets of limited dimensionality visual inspection seems to be a more robust alternative. Since the binding modes studied here formed natural clusters in torsion space owing to their long-term stability, a cluster analysis was not required. It should be a helpful technique in less stable multidimensional systems, but for explorative usage it requires input of a large number of allegedly important parameters. A reliable analysis of cluster solutions requires prior experience with the method.

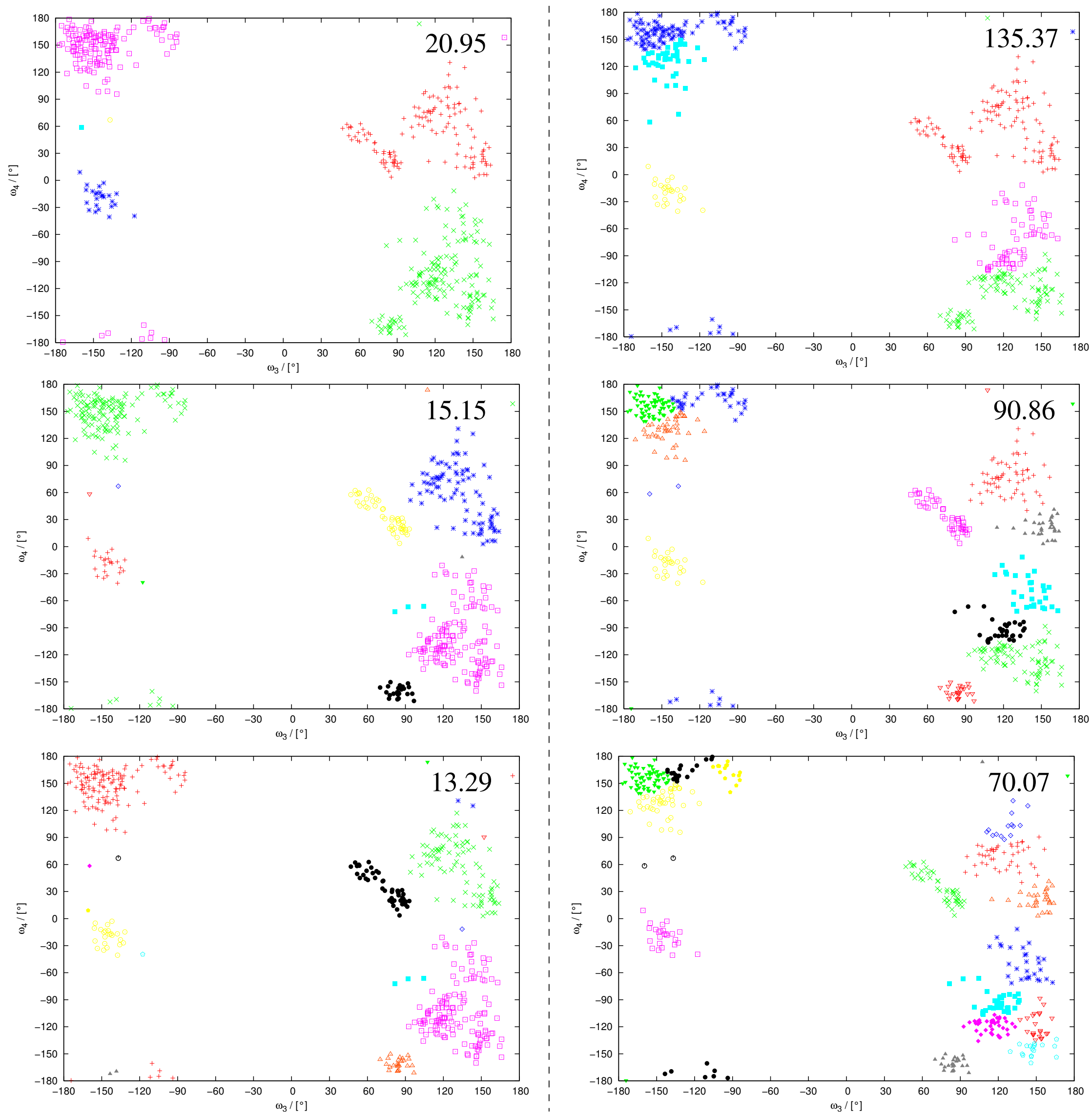


Figure G.2: Two-dimensional torsional diagrams created from the entire set of trajectories of binding modes of (R)-NAPHAC and (S)-NAPHAC (see chapter 3). Data points correspond to snapshot configurations which were extracted every 50 ps. Comparison of single linkage (left hand side) and complete linkage (right hand side) clustering in two-dimensional dihedral space. The number of clusters is 6, 12, 16 (from the top), the number in the upper-right corner of each diagram is the distance (in degrees) at which the last fusion of clusters took place.

Appendix H

Umbrella Sampling in Practice

In this appendix we present a short guide to successful umbrella sampling calculations, and an analysis program that can be used for the evalution of the simulation data with the WHAM procedure (for a theoretical discussion see chapter 2). The analysis program is written in the Python language, which is integrated in all standard installations of Linux distributions. Simply copy and paste the code from this document into a single file using your favourite text editor.

You are in the position to do an umbrella sampling simulation after choosing a sensible RC and the level of theory needed to capture the desired amount of detail in the studied system. The validity of a RC may be tested by minimum energy reaction path scans in combination with separate TS optimisations (chapter 4). A well behaved RC should lead smoothly from the reactants to the products via a TS. The electronic structure calculation for the QM part should be *fast* and on the same timescale as that of the MM part in a QM/MM approach. At present this means that the use of semiempirical methods is strongly indicated. The number of MD steps needed for an accurate evaluation of reaction free energies will vary with the type of reaction, but as a guideline you should be able to perform a minimum of 100000 steps in the MD. The necessary amount of time can be estimated from the preceding validation study of the RC as the amount of time spent for one energy and gradient evalation times the desired number of MD steps.

Once these matters are settled an umbrella sampling calculation can be accomplished. Starting from the reactants the RC is subdivided into simulation 'windows' in ascending (or descending) order of values S of the RC. The RC, which is most probably of the harmonic type described in chapter 2 (equations 2.6-2.8), is enforced using a bias of strength k which will be much smaller than in minimum energy reaction path calculations. To pull the system over the barrier region at dense and uniform sampling of all values of the RC, the force constant k must be chosen such that the bias potential at least matches the slope of the free energy surface. A larger force constant is often not harmful, but try to keep it as low as possible to be least invasive to the system. Useful values should be in the range of 5-100 $\mathrm{kcal}/(\mathrm{mol}\cdot\text{Å}^2)$. The final structure and velocities of any simulation window are used as input to initialise the MD in successive windows, which is also the right time for adjustment of the equilibrium position S of the bias potential. When sampling intervals on the RC are sufficiently small, the system can usually be expected to be close to equilibrium in each new window. For a typical covalent reaction a spacing of 0.05Å-0.25Å of the RC should lead to success. RCs with simple distances as in equation 2.6 would require a denser spacing than those composed of several distances (equations 2.7-2.8). For a first test use a fixed number of 10 windows, and adjust the spacing to cover the RC over the desired region.

After the simulation is completed one must check the coverage of the RC. There must be no intermittent gaps in the distributions. Ideally one would want to achieve uniform sampling over the entire range of the RC. While this is never achieved in practice, the number of points in the least and most sampled regions should be similar. To see if this is the case use a plotting program and display your data as global histogram over the RC. If gaps or sparsely sampled intervals are detected, adjust the force constant or simulate for longer times in those regions.

Use of the python program is described in its header. In addition to the Python program, which is listed in the following, we have a C++ implementation of the WHAM algorithm. This program also features the bootstrap method (see chapter 5.2.1), and was used in the evaluation of activation free energies throughout this thesis.

```
import os,sys
from string import *
from math import *
from copy import *

#Version 1.0
#11.05.2005 by Nikolaj Otte


#Umbrella sampling free energy evaluation program that provides WHAM data processing. Free energy output
#is provided by two alternative routes, as raw umbrella sampling data (unbiased free energy curve fragments),
#where the user still has to the additional work of adjusting the curve fragments, and as complete
#free energy curves by the fully automatted WHAM.
#It is a good idea and also very instructive to check your free energy
#surface via both routes.


#Read the following papers as an introduction to (chemical, reaction coordinate driven)
#Umbrella sampling and WHAM:

#S H Northrup, et. al., Proc. Natl. Acad. Sci USA 79:4035-4039 (1982)
#J Kottalam, D A Case, J. Am. Chem. Soc. 110:7690-7697 (1988)
#E M Boczko, C H Brooks, J. Phys. Chem. 97:4509-4513 (1993)
#B Roux, Comp. Phys. Comm. 91:275-282 (1995)
#S Kumar, et al., J. Comp. Chem., 13:1011-1021 (1992)

#The original Umbrella sampling paper (Though probably not the best treat for chemists):

#G M Torrie, J P Valleau, J. Comp. Phys. 23:187-199 (1977)


#---------------- What does the script do and what what is expected by the user ---------------------

#UMBRELLA SAMPLING output is placed in directory umbr and consists of individually unbiased
#free energy fragments. For a smooth free energy curve over the entire reaction coordinate
#those fragments must be arranged in one diagram, and continuosity is achieved by applying
#vertical offsets to each individual fragment. Those offsets are the undetermined  free
#energy constants that are now determined, in a graphical, eye-balling approach, by you.

#WHAM is an automatted procedure to compute a free energy surface from biased molecular simulations
#(MD or MC). The biasing procedure and the output needed are identical to Umbrella sampling. For the
#systems we are studying it is really an only an automatted Umbrella sampling data evaluation. One serious
#advantage over the manual free energy curve generation is that ALL data points are utilised,
#whereas points in the dangling ends of the individual free energy fragments in the graphical approach are
#discarded. That results in higher efficiency of WHAM. The (quadratic) error progression is another
#disadvantage of the graphical approach. The WHAM is implemented as in ROUX95.


#------------------------------------------------------------------------------------------------------
#-------------------------------------------I N P U T-------------------------------------------------
#------------------------------------------------------------------------------------------------------


#Input files must reside in one directory that contains nothing else than simulation output
#to be utilised in the free energy calculation.

#input format:
#one directory with all free energy output files. scripts will iterate on all
#files in directory and crash if there is any other, nonconforming file there.
#file format:
#plain files, with three columns,
#        1. reaction coordinate value (x),
#        2. force constant value (k),
#        3. reaction coordinate minimum (S)
#
# that means restraint is assumed to be in harmonic format:
#
#  Restr. = k*(x - S)^2
#

# usage: python umbrel.py name_of_directory_with_files temperature min_x max_x number_of_bins


#Feel free to adjust these ...

#Maximum number of iterations in WHAM
niter = 1000
#Convergence criterium for Free Energy Constants in WHAM
criter = 0.01
#change your energy unit if you want something else than kcal/mol
kb = 0.001982923700

#------------------------------------------------------------------------------------------------------
#----------------------------------------O U T P U T--------------------------------------------------
#------------------------------------------------------------------------------------------------------

#The program assumes a global histogram that is composed of the individual histograms from the different
#simulations. It prints the first and last bins of the individual histograms on the global histogram. If
#you find that in the output ther ar "empty" histograms, this means your reaction coordinate was not
#chosen wide enough and data is lost. The reaction coordinate should then be extended.


#--(A)--
#Summed histogram of counts of values on the reaction coordinate: "globalhistogram.dat"
#Check that all parts of the reaction coordinate are sampled equally well (highest to lowest
# sampled regions within same dimension of counts) ! Watch the transition state region ! Simply ADD simulation
#data to the "data" directory by sampling again the less visited regions and increase the force constants of the
#bias potential.
```

Figure H.1: WHAM program in Python. Part A.

```
#--(B)--
#Unbiased Free energy fragments reside in directory "umbr".

#--(C)--
#WHAM computed Free energy curve is in "freenergy.plt"


#NOTE: reaction coordinate values are in angstroms, free energy output is in kcal/mol.




#---------------------------------------------------------------------------------------------------
#-------------------------------------- PROGRAM SECTION --------------------------------------------
#---------------------------------------------------------------------------------------------------






#--------------------------------------- INITIALISATION --------------------------------------------

try:
  fileresid = sys.argv[1]
  T =  float(sys.argv[2])
  min_x = float(sys.argv[3])
  max_x = float(sys.argv[4])
  binno = int(sys.argv[5])
except IndexError:
  print "usage: python umbrel.py name_of_directory_with_files temperature min_x max_x number_of_bins"
  print "PLEASE READ THE PROGRAM HEADER !"
  sys.exit(1)

try:
  filelist = os.listdir(fileresid)
except OSError:
   print "ERROR: Directory with free energy output is not existent."
   sys.exit(1)

print "I will work on %d output files !" % (len(filelist))


# now produce histograms (in x) and store force constants (k) and spring locations (S)

multihist = []
globalhist = []
indices = []
filno = len(filelist)
incr = (max_x - min_x) / float(binno)
rangelist = []
focolist = []
slist = []
countslist = []
totcount = 0.0
averagebias = []
emptylist = []
zerobin = []
#kb =
kt = kb*T

#wham arrays and constants
fact = []
nt = []
frecosav = []
freco = []
prob = []

def bias(k,x,S):
  m = x - S
  return k*m*m


#--------------------------------INITIALISE LISTS, FILL THE HISTOGRAMS ------------------------------


#the list that contains the reaction coordinate values. Will be used to map the histogram values.
for s in range(binno):
  rangelist.append(min_x + (0.5 + float(s))*incr )

for s in range(filno):
  hist = []
  for i in range(binno):
    hist.append(0.0)
    if s < 1:
      frecosav.append(0.0)
      freco.append(0.0)
      fact.append(1.0)
      nt.append(float(filno))
      globalhist.append(0.0)
      prob.append(0.0)
      zerobin.append(0.0)
  multihist.append(deepcopy(hist))
  indices.append([0,binno-1])
  focolist.append(0.0)
  slist.append(0.0)
  countslist.append(0.0)
  averagebias.append(0.0)
  emptylist.append(1)
```

Figure H.2: WHAM program in Python. Part B.

```
#Create histograms from the individual files. Note that values outside the
#predefined reaction coordinate span are omitted (not counted either, to avoid erroneous normalisation).
ps = 0
for s in range(filno):
  input = open(fileresid + "/" + filelist[s]).readlines()
  if ps > 9:
    print "%s Input files read" % (s)
    ps = 0
  ps += 1
  for i in range(len(input)):
    chop = split(input[i])
    if i < 1:
      chop[1] = float(chop[1])
      chop[2] = float(chop[2])
      focolist[s] = chop[1]
      slist[s] = chop[2]
    val = float(chop[0]) - min_x
    index =  int( floor( val / incr ) )
    if ¬ (index < 0 ∨ index > binno-1) :
      multihist[s][index] += 1.0;
      totcount += 1.0;
      countslist[s] += 1.0;

#get indices for first and last bins on global scale

for s in range(filno):
  for i in range(binno):
    if multihist[s][i] > 0:
      indices[s][0] = i
      emptylist[s] = 0
      break
  for i in range(binno-1,-1,-1):
    if multihist[s][i] > 0:
      indices[s][1] = i
      break



print "Minimum and Maximum Indices of histograms on global histogram:"
print "%12s%6s%6s" % ("Histogram:", "Lower","Upper")
for s in range(filno):
  if emptylist[s] < 1:
    print ("%12d%6d%6d" % (s+1,indices[s][0],indices[s][1]))
  else:
    print ("%12d  %6s" % (s+1,"Empty"))



#print global (summed histogram) to file globalhistogram.dat for evaluation
out = open("globalhistogram.dat",’w+’)
for s in range(filno):
  for i in range(indices[s][0],indices[s][1]+1):
    globalhist[i] += multihist[s][i]

for s in range(binno):
  out.write("%8.3f %d\n" % (rangelist[s],globalhist[s]) )

out.close()


#---------------------------------- Umbrella Sampling Part (simple histogram reweighting)-------------

#calculate average bias (ensemble average of the bias in each ’bin’)
for s in range(filno):
  scratchhis = []
  if emptylist[s] < 1:
    k = focolist[s]
    S = slist[s]
    for i in range(indices[s][0],indices[s][1]+1):
      averagebias[s] += exp( bias(k,rangelist[i],S) / kt) * (multihist[s][i] / countslist[s])


out.close()
#print the free energy fragments to files in directory umbr
try:
  os.listdir("./umbr")
except OSError:
  os.system("mkdir umbr")


for s in range(filno):
  if emptylist[s] < 1:
    k = focolist[s]
    S = slist[s]
    out = open( "umbr/win%d" % (s),’w+’ )
    for i in range(indices[s][0],indices[s][1]+1):
      if multihist[s][i] > 0:
        val = exp( bias(k,rangelist[i],S) / kt) * multihist[s][i] / averagebias[s]
        out.write("%8.3f %8.3f\n" % (rangelist[i], -kt*log(val) ) )
    out.close()



#-------------------------------------------- WHAM ---------------------------------------------
```

Figure H.3: WHAM program in Python. Part C.

```
#now the wham part, proceeding like Roux95

notconverged = 1
iteration  = 1
sc = 0

print
print "------------------------------------"
print "-------------- WHAM ---------------"
print

print "%-10s%5s%-20s" % ("Iteration:","   ","Free Energy Constants")
print

while notconverged ∧ iteration < niter:

  if sc > 8:
    print "%-10d%5s%-20s" % (iteration,"   ","--------------------")
    for s in range(filno):
      print "%-15s%-10d:%10.3f" % (" ",s+1,frecosav[s])
    sc = 0

  for s in range(binno):
    nom = 0.0
    denom = 0.0
    prob[s] = 0.0
    for i in range(filno):
#       if emptylist[i] < 1:
      nom = nom + multihist[i][s]
      for j in range(filno):
        bf = exp( ( -bias(focolist[j],rangelist[s],slist[j]) + frecosav[j]) / kt)
        denom = denom + bf * countslist[j]
    prob[s] = nom / denom
    if nom > 0.0:
      zerobin[s] = 0
    else:
      zerobin[s] = 1


  #now calculate new free energy constants
  for s in range(filno):
    val = 0.0
    for i in range(binno):
      bf = exp( - bias(focolist[s],rangelist[i],slist[s])  / kt)
      val = val + bf*prob[i]
    freco[s] = -log(val) * kt

  #bound the constants by shifting about constant 0
  subs = freco[0]
  for s in range(filno):
    freco[s] = freco[s] - subs

  notconverged = 0
  for s in range(filno):
    dev = abs(freco[s] - frecosav[s])
    if dev > criter:
      notconverged = 1

  if notconverged < 1:
    print "WHAM converged after %d iterations !" % ( iteration )

  #save current free en
  for s in range(filno):
    frecosav[s] = freco[s]


  iteration += 1
  sc += 1

#----------------------------------------- END OF WHAM LOOP -------------------------------------


if (notconverged > 0 ):
  print "Convergence of WHAM to cutoff %f not achieved within %d iterations !" % (criter,niter)
  sys.exit(1)

print "Free energy constants:"
for s in range(filno):
  print "%3d: %8.3f" % (s,freco[s])


out = open("freeenergy.plt",'w+')

#set smallest value as 0.0
offs = 100000000000.0
for s in range(binno):
  if(zerobin[s] < 1):
    prob[s] = -kt*log( prob[s] )
    if prob[s] < offs:
      offs = prob[s]

for s in range(binno):
  if (zerobin[s] < 1):
    out.write( "%8.3f %8.3f\n" % (rangelist[s],prob[s]-offs)  )
out.close()
```

Figure H.4: WHAM program in Python. Part D.

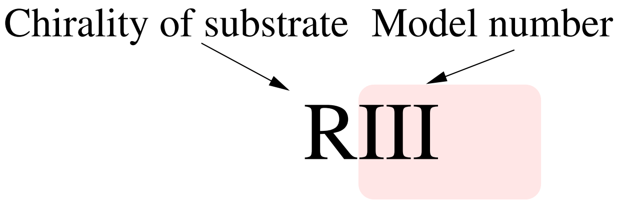
Appendix I

Abbreviations

ABNR	Adopted basis Newton-Raphson optimisation method
BSLA	Lipase A of *Bacillus subtilis*
CHARMM	**Ch**emistry-**A**t-**H**arvard-**M**olecular-**M**echanics[30]
	A molecular modelling software package
Charmm22	Biomolecular force field[108]
DIAC	meso-1,4-diacetoxycyclopentene
FEP	Free energy perturbation
HIS156,ASP133 etc.	Three letter amino acid names with primary structure information
MC	Monte-Carlo method
MCC	Michaelis complex
MD	Molecular dynamics method
MODEL100	Gas phase model (Appendix C)
NAPHAC	1-β-Naphthylethylester of acetic acid
NE2,HE2,O3,O15 etc.	Atom names, see figure 3.3 on page 39
PDC	Product complex, i.e., the complex of acylenzyme + alcohol
PES	Potential energy surface
PMF	Potential of mean force: Free energy along a reaction coordinate
SD	Steepest descent optimisation method
QMI,QMII,QMIII	Quantum mechanical regions, definition in figure 4.1
TI	Tetrahedral intermediate
TS	Transition state
VdW	Van-der-Waals
WHAM	**W**eighted-**H**istogram-**A**nalysis-**M**ethod[126]
[I]	First covalent step in the acylation reaction (see figure 1.2)
[II]	Second covalent step in the acylation reaction (see figure 1.2)

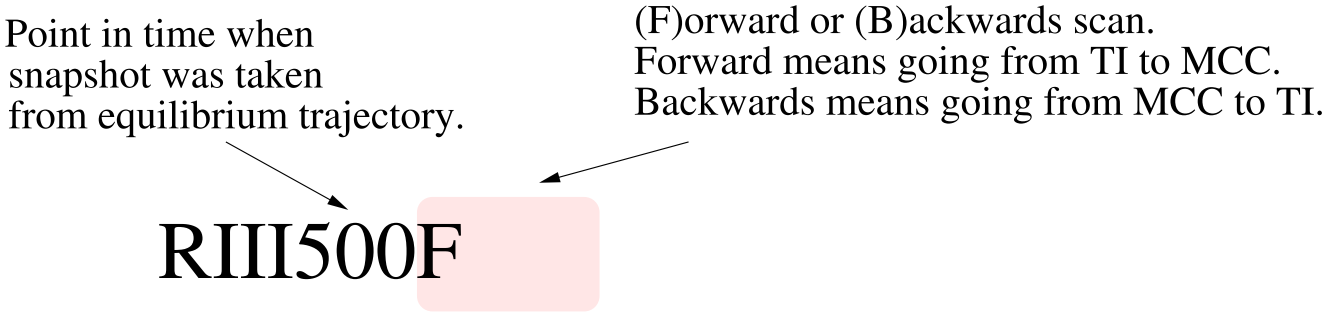
Naming scheme of computer models

Binding modes of the TI:



Enantiomers are named according to Cahn-Ingold-Prelog rules. The numbering of the binding modes is purely chronological, and starts at I for both enantiomers.

PES scans and umbrella sampling simulations:


Point in time when
snapshot was taken
from equilibrium trajectory.

(F)orward or (B)ackwards scan.
Forward means going from TI to MCC.
Backwards means going from MCC to TI.

Bibliography

[1] P Acharya, E Rajakumara, R Sankaranarayanan, and N M Rao. Structural Basis of Selection and Thermostability of Laboratory Evolved *bacillus subtilis* Lipase. *J. Mol. Biol.*, 341:1271–1281, 2004.

[2] R Ahlrichs, M Bär, M Häser, H Horn, and C Kölmel. Electronic structure calculations on workstations: the program system Turbomole. *Chem. Phys. Lett.*, 162:165ff, 1989.

[3] A Amadei, A B M Linssen, and H J C Berendsen. Essential dynamics of proteins. *Proteins*, 17:412–425, 1993.

[4] J B Anderson. Predicting Rare Events In Molecular Dynamics. In *Advances in Chemical Physics*, volume XCI, pages 381–431. John Wiley & Sons, 1995.

[5] I Antes and W Thiel. On the Treatment of Link Atoms in Hybrid M ethods. In J Gao and M A. Thompson, editors, *Combined Quantum Mechanical and Molecular M echanical Methods*, volume 712 of *ACS Symposium Series*, pages 50–65. American Chemical Society, Washington, DC, 1998.

[6] I Antes and W Thiel. Adjusted Connection Atoms for Combined Quantum Mechanical and Molecular Mechanical Methods. *J. Phys. Chem. A*, 103:9290–9295, 1999.

[7] P W Atkins. *Physical Chemistry*. Oxford University Press, Oxford, 1994.

[8] W W Bachovchin. Contributions of NMR spectroscopy to the study of hydrogen bonds in serine protease active sites. *Magn. Reson. Chem.*, 39:S199–S213, 2001.

[9] D Bakowies and W Thiel. Hybrid Models for Combined Quantum Mechanical and Molecular Mechanical Approaches. *J. Phys. Chem.*, 100:10580–10594, 1996.

[10] P Bala, P Grochowski, K Nowinsky, B Lesyng, and J A McCammon. Quantum-Dynamical Picture of a Multistep Enzymatic Process: Reaction Catalyzed by Phospholipase A_2. *Biophys. J.*, 79:1253–1262, 2000.

[11] A K Balls and H N Wood. Acetyl Chymotrypsin And Its Reaction With Ethanol. *J. Biol. Chem*, 219:245–256, 1956.

[12] P A Bash, M J Field, R C Davenport, G A Petrsko, D Ringe, and M Karplus. Computer Simulation and Analysis of the Reaction Pathway of Triosephosphate Isomerase. *Biochemistry*, 30:5826–5832, 1991.

[13] J E Basner and S D Schwartz. How Enzyme Dynamics Helps Catalyze a Reaction in Atomic Detail: A Transition Path Sampling Study. *J. Am. Chem. Soc.*, 127:13822–13831, 2005.

[14] A D Becke. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A*, 38:3098–3100, 1988.

[15] M L Bender. Oxygen Exchange as Evidence for the Existence of an Intermediate in Ester Hydrolysis. *J. Am. Chem. Soc.*, 73:1626–1629, 1951.

[16] M L Bender, G E Clement, C R Gunter, and F J Kezdy. The Kinetics of α-Chymotrypsin Reactions in the Presence of Added Nucleophiles. *J. Am. Chem. Soc.*, 86:3697–3703, 1964.

[17] M L Bender and F J Kezdy. The Current Status of the α-Chymotrypsin Mechanism. *J. Am. Chem. Soc.*, 86:3704–3714, 1964.

[18] D L Beveridge and F M DiCapua. Free Energy Via Molecular Simulation: Applications to Chemical and Biomolecular Systems. *Annu. Rev. Biophys. Biophys. Chem.*, 18:431–491, 1989.

[19] S R Billeter, A J Turner, and W Thiel. Linear Scaling Geometry Optimisation and Transition State Search in Hybrid Delocalized Internal Coordinates. *Chem. Phys. Phys. Chem.*, 2:2177–2186, 2000.

[20] S A Bizzozero, A W Kaiser, and H Dutler. A Numerical Method for Aquisition and Processing of Steady-State Kinetic Data Contained in an Entire Progression Curve. *Eur. J. Biochem.*, 33:292–300, 1973.

[21] D M Blow, J J Birktoft, and B S Hartley. Role of Buried Acid Group in the Mechanism of Action of Chymotrypsin. *Nature*, 221:337–340, 1969.

[22] M Bocola, N Otte, K-E Jaeger, M T Reetz, and W Thiel. Learning from Directed Evolution: Theoretical Investigations into Cooperative Muations in Lipase Enantioselectivity. *ChemBioChem*, 5:214–223, 2004.

[23] M Bocola, M T Stubbs, C Sotriffer, B Hauer, T Friedrich, K Dittrich, and G Klebe. Structural and energetic determinants for enantiopreferences in kinetic resolution of lipases. *Protein Engineering*, 16:319–322, 2003.

[24] E M Boczko and C L Brooks. Constant-Temperature Free Energy Surfaces for Physical and Chemical Processes. *J. Phys. Chem.*, 97:4509–4513, 1993.

[25] P G Bolhuis. Transition path sampling of β-hairpin folding. *Proc. Natl. Acad. Sci. USA*, 100:12129–12134, 2003.

[26] P G Bolhuis, D Chandler, C Dellago, and P L Geissler. Transition Path Sampling: Throwing Ropes Over Rough Mountain Passes, in the Dark. *Annu. Rev. Phys. Chem.*, 53:291–318, 2002.

[27] F Bordusa. Proteases in Organic Synthesis. *Chem. Rev.*, 102:4817–4867, 2002.

[28] G Brancato and M E Tuckerman. A polarizable multistate empirical valence bond model for proton transport in aqueous solution. *J. Chem. Phys.*, 122:224507, 2005.

[29] D Branduardi, F L Gervasio, A Cavalli, M Recanatini, and M Parrinello. The Role of the Peripheral Anionic Site and Catio-π Interactions in the Ligand Penetratinmo of the Human AChE Gorge. *J. Am. Chem. Soc.*, 127:9147–9155, 2005.

[30] B R Brooks, R E Bruccoleri, B D Olafson, D J States, S Swaminathan, and M Karplus. CHARMM - A Program For Macromolecular Energy, Minimization, And Dynamics Calculations. *J. Comp. Chem.*, 4:187–217, 1983.

[31] C L Brooks and M Karplus. Deformable stochastic boundaries in molecular dynamics. *J. Chem. Phys.*, 79:6312–6325, 1983.

[32] P R Carey and P J Tonge. Unlocking the Secrets of Enzyme Power Using Raman Spectroscopy. *Acc. Chem. Res.*, 28:8–13, 1995.

[33] P Carter and J A Wells. Dissecting the catalytic triad of serine proteases. *Nature*, 332:564, 1988.

[34] D Chandler. *Introduction to Statistical Mechanics*. Oxford University Press, 1987.

[35] D Chandler. Interfaces and the driving force of hydrophobic assembly. *Nature*, 437:640–647, 2005.

[36] Y-T Chang and W H Miller. An Empirical Valence Bond Model for Constructing Global Potential Energy Surfaces for Chemical Reactions of Polyatomic Molecular Systems. *J. Phys. Chem.*, 94:5884–5888, 1990.

[37] T Clark, J Chandrasekhar, G W Spitznagel, and P v R Schleyer. Efficient diffuse function-augmented basis sets for anion calculations. III. The 3-21+G basis set for first-row elements, Li-F. *J. Comp. Chem.*, 4:294–301, 1983.

[38] W W Cleland and M M Kreevoy. Low-Barrier Hydrogen Bonds and Enzymic Catalysis. *Science*, 264:1887–1890, 1994.

[39] M D Collins, G Hummer, M L Quilin, B W Matthews, and S M Gruner. Cooperative filling of a nonpolar protein cavity observed by high-pressure crystallography and simulation. *Proc. Natl. Acad. Sci. USA*, 102:16668–16671, 2005.

[40] G Colombo, G Ottolina, G Carrea, and K M Merz. Modelling the enantioselectivity of subtilisin in water and organic solvents: insights from molecular dynamics and quantum mechanical/molecular mechanical studies. *Chem. Comm.*, pages 559–560, 2000.

[41] G Colombo, S Toba, and K M Merz. Rationalization of the Enantioselectivity of Subtilisin in DMF. *J. Am. Chem. Soc.*, 121:3486–3493, 1999.

[42] F Cordes, M Weber, and J Schmidt-Ehrenberg. Metastable Conformations via Successive Perron-Cluster Cluster Analysis of Dihedrals. Technical Report ZIB-Report 02-40, Konrad-Zuse-Zentrum für Informationstechnik Berlin, 2002.

[43] M Cossi, V Barone, R Cammi, and J Tomasi. Ab initio study of solvated molecules: a new implementation of the polarizable continuum model. *Chem. Phys. Lett.*, 255:327–335, 1996.

[44] Q Cui and M Karplus. Triosephosphate Isomerase: A Theoretical Comparison of Alternative Pathways. *J. Am. Chem. Soc.*, 123:2284–2290, 2001.

[45] M Cygler, P Grochulski, R J Kazlauskas, J D Schrag, F Bouthillier, B Rubin, A N Serreqi, and A K Gupta. A Structural Basis for the Chiral Preferences of Lipases. *J. Am. Chem. Soc.*, 116:3180–3186, 1994.

[46] W L DeLano. The PyMOL Molecular Graphics System. DeLano Scientific LLC, San Carlos, CA, USA. http://www.pymol.org.

[47] C Dellago, P G Bolhuis, and D Chandler. On the Calculation of rate constants in the transition path ensemble. *J. Chem. Phys.*, 110:6617–6625, 1999.

[48] C Dellago, P G Bolhuis, F S Csaijka, and D Chandler. Transition path sampling and the calculation of rate constants. *J. Chem. Phys.*, 108:1964–1977, 1998.

[49] F Ding, W Guo, N V Dokholyan, E I Shakhnovich, and J E Shea. Reconstruction of the src-SH3 Protein Domain Transition State Ensemble using Multiscale Molecular Dynamics Simulations. *J. Mol. Biol.*, 350:1035–1050, 2005.

[50] A Dinner, G M Blackburn, and M Karplus. Uracil-DNA glycosylase acts by substrate autocatalysis. *Nature*, 413:752–755, 2001.

[51] R Ditchfield, W J Hehre, and J A Pople. Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules. *J. Chem. Phys.*, 54:724–728, 1971.

[52] B Efron. Bootstrap Methods: Another Look at the Jacknife. *Ann. Stat.*, 7:1–26, 1979.

[53] B Efron and R Tibshirani. Bootstrap Methods for Standard Errors, Confidence Intervals and Other Measures of Statistical Accuracy. *Stat. Sci.*, 1:54–75, 1986.

[54] B Efron and R Tibshirani. Statistical Data Analysis in the Computer Age. *Science*, 253:390–395, 1991.

[55] K Eichkorn, O Treutler, H Öhm, M Häser, and R Ahlrichs. Auxiliary basis sets to approximate Coulomb potentials. *Chem. Phys. Lett.*, 242:652ff, 1995.

[56] M Elstner, D Porezag, G Jungnickel, J Elsner, M Haugk, T Frauenheim, S Suhai, and G Seifert. Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys. Rev. B*, 58:7260–7268, 1998.

[57] T Ema. Mechanism of enantioselectivity of lipases and other synthetically useful hydrolases. *Curr. Org. Chem.*, 8:1009–1025, 2004.

[58] B Engels. Personal Communication, 2003.

[59] J Kästner *et al.* Bridging the Gap Between Thermodynamic Integration and Umbrella Sampling Provides a Novel Analysis Method: ”Umbrella Integration”. *J. Chem. Phys.*, 123:144104/1–5, 2005.

[60] K-E Jaeger *et al.* Meso-diacetate modelling, in collaboration with prof. jaeger, working title unknown, to be published.

[61] K Eurenius, D C Chatfield, and B R Brooks. Enzyme Mechanisms with Hybrid Quantum and Molecular Mechanical Potentials. I. Theoretical Considerations. *J. Comp. Chem.*, 60:1189–1200, 1996.

[62] E Fermi. *Z. Physik*, 48:73, 1928.

[63] A Ferrenberg. *Optimal Use of Monte Carlo Data*. PhD thesis, Carnegie Mellon University, 1989.

[64] A Fersht. *Structure and Mechanism in Protein Science*. W H Freeman and Company, New York, 1999.

[65] M J Field, P Bash, and M Karplus. A Combined Quantum Mechanical and Molecular Mechanical Potential for Molecular Dynamics Simulations. *J. Comput. Chem.*, 11:700–733, 1990.

[66] N Foloppe and A D MacKerell. All-Atom Empirical Force Field for Nucleic Acids: 1) Parameter Optimization Based on Small Molecule and Condensed Phase Macromolecular Target Data. *J. Comp. Chem.*, 21:86–104, 2000.

[67] D Frenkel and B Smit. *Understanding Molecular Simulation*. Academic Press, London, 2002.

[68] P A Frey. Characterization of a low barrier hydrogen bond in the active site of chymotrypsin. *J. Mol. Struct.*, 615:153–161, 2002.

[69] P A Frey, S A Whitt, and J B Tobin. A Low Barrier Hydrogen Bond in the Catalytic Triad of Serine Proteases. *Science*, 264:1927–1930, 1994.

[70] R A Friesner and V Guallar. Ab Initio Quantum Chemical and Mixed Quantum Mechanics/Molcular Mechanics (QM/MM) Methods for Studying Enzymatic Catalysis. *Annu. Rev. Phys. Chem.*, 56:389–427, 2005.

[71] M J Frisch, G W Trucks, H B Schlegel, G E Scuseria, M A Robb, J R Cheeseman, J A Montgomery, Jr., T Vreven, K N Kudin, J C Burant, J M Millam, S S Iyengar, J Tomasi, V Barone, B Mennucci, M Cossi, G Scalmani, N Rega, G A Petersson, H Nakatsuji, M Hada, M Ehara, K Toyota, R Fukuda, J Hasegawa, M Ishida, T Nakajima, Y Honda, O Kitao, H Nakai, M Klene, X Li, J E Knox, H P Hratchian, J B Cross, V Bakken,

C Adamo, J Jaramillo, R Gomperts, R E Stratmann, O Yazyev, A J Austin, R Cammi, C Pomelli, J W Ochterski, P Y Ayala, K Morokuma, G A Voth, P Salvador, J J Dannenberg, V G Zakrzewski, S Dapprich, A D Daniels, M C Strain, O Farkas, D K Malick, A D Rabuck, K Raghavachari, J B Foresman, J V Ortiz, Q Cui, A G Baboul, S Clifford, J Cioslowski, B B Stefanov, G Liu, A Liashenko, P Piskorz, I Komaromi, R L Martin, D J Fox, T Keith, M A Al-Laham, C Y Peng, A Nanayakkara, M Challacombe, P M W Gill, B Johnson, W Chen, M W Wong, C Gonzalez, and J A Pople. Gaussian 03, Revision C.02. Gaussian, Inc., Wallingford, CT, 2004.

[72] S A Funke, N Otte, M Bocola, K-E Jaeger, and W Thiel. Combination of computational prescreening and experimental library construction can accelerate enzyme optimization by directed evolution. *Protein Engineering*, 18:509–514, 2005.

[73] J Gao. Hybrid Quantum and Molecular Mechanical Simulations: An Alternative Avenue to Solvent Effects in Organic Chemistry. *Acc. Chem. Res.*, 29:298–305, 1996.

[74] A E Garcia. Large-Amplitude Nonlinear Motions in Proteins. *Phys. Rev. Lett*, 68:2696–2700, 1992.

[75] M Garcia-Viloca, J Gao, M Karplus, and D G Truhlar. How Enzymes Work: Analysis by Modern Rate Theory and Computer Simulations. *Science*, 303:186–195, 2004.

[76] P R Gerber. Charge distribution from a simple molecular orbital type calculation and non-bonding interaction terms in the force field MAB. *J. Comput.-Aided Mol. Des.*, 12:37–51, 1998.

[77] P R Gerber and K Müller. MAB, a generally applicable molecular force field for structure modelling in medicinal chemistry. *J. Comput.-Aided Mol. Des.*, 9:251–268, 1994.

[78] J A Gerlt and P G Gassman. Undersrtanding the Rates of certain Enzyme-Catalyzed Reactions: Proton Abstraction from Carbon Acids, Acyl-Transfer Reactions, and Displacement Reactions of Phosphodiesters. *Biochemistry*, 32(45):11944–11951, 1993.

[79] F L Gervasio, A Laio, and M Parrinello. Flexible Docking in Solution Using Metadynamics. *J. Am. Chem. Soc.*, 127:2600–2607, 2005.

[80] A Ghanem and H Y Aboul-Enein. Application of Lipases in Kinetic Resolution Experiments. *Chirality*, 17:1–15, 2005.

[81] D T Gillespie. A General Method for Numerically Simulating the Stochastic Time Evolution of Coupled Chemical Reactions. *J. Comp. Phys.*, 22:403–434, 1976.

[82] D T Gillespie. Exact Stochastic Simulation of Coupled Chemical Reactions. *J. Phys. Chem.*, 81:2340–2361, 1977.

[83] H Goldstein. *Classical Mechanics*. Addison-Wesley, Reading (MA), 1980.

[84] W Greiner, L Neise, and H Stöcker. *Theoretische Physik Bd. 9: Thermodynamik und Statistische Mechanik*. Harri Deutsch, Thun und Frankfurt am Main, 1993.

[85] P Grochowski, B Lesyng, P Bala, and J A McCammon. Density Functional Based Parametrization of a Valence Bond Method and Its Applications in Quantum-Classical Molecular Dynamics Simulations of Enzymatic Reactions. *Int. J. Quant. Chem.*, 60:1143–1164, 1996.

[86] A Grossfield. An implementation of WHAM: the Weighted Histogram Analysis Method, (manual dated 2003). Department of Biochemistry & Molecular Biophysics, Washington University School of Medicine, http://dasher.wustl.edu/alan/wham/.

[87] H Grubmüller. Predicting slow structural transitions in macromolecular systems: conformational flooding. *Phys. Rev. E*, 52:2893–2906, 1995.

[88] H Gutfreund. *Kinetics For The Life Sciences*. Cambridge University Press, Cambridge UK, 1995.

[89] K C Haddad, J L Sudmeier, D A Bachovchin, and W W Bachovchin. α-Lytic protease can exist in two separately stable conformations with different His^{57} mobilities and catalytic activities. *Proc. Natl. Acad. Sci. USA*, 102:1006–1011, 2005.

[90] D Hamelberg and J A McCammon. Phosphorylation effects on cis/trans isomerization and the backbone conformation of serine-proline motifs: Accelerated molecular dynamics analysis. *J. Am. Chem. Soc.*, 127:1969–1974, 2005.

[91] D Hamelberg, J Mongan, and J A McCammon. Accelerated molecular dynamics: A promising and efficient simulation metjhod for biomolecules. *J. Chem. Phys.*, 120:11919–11929, 2004.

[92] D Hamelberg, T Shen, and J A McCammon. Relating kinetic rates and local energetic roughness by accelerated molecular-dynamics simulations. *J. Chem. Phys.*, 122(241103), 2005.

[93] S Hammes-Schiffer. Impact of Enzyme Motion on Activity. *Biochemistry*, 41(45):13335–13343, 2002.

[94] A Handl. *Multivariate Analysemethoden*. Springer, Berlin, 2002.

[95] P C Hariharan and J A Pople. The Effect oF d-Functions on Molecular Orbital Energies for Hydrocarbons. *Chem. Phys. Lett.*, 16:217–219, 1972.

[96] B S Hartley and B A Kilby. The reaction of p-nitrophenyl esters with chymotrypsin and insulin. *Biochem. J.*, 56:288, 1954.

[97] L Hedstrom. Serine Protease Mechanism and Specificity. *Chem. Rev.*, 102:4501–4523, 2002.

[98] K Hinsen and B Roux. Potential of mean force and reaction rates for proton transfer in acetylacetone. *J. Chem. Phys.*, 106:3567–3577, 1997.

[99] P Hohenberg and W Kohn. Inhomogeneous Electron Gas. *Phys. Rev.*, 136:B864–B871, 1964.

[100] C H Hu, T Brinck, and K Hult. Ab initio and Density Functional Theory Studies of the Catalytic Mechanism for Ester Hydrolysis in Serine Hydrolases. *Int. J. Quant. Chem.*, 69:90–103, 1998.

[101] M W Hunkapiller, S H Smallcombe, D R Whittaker, and J H Richards. Carbon Nuclear Magnetic Resonance Studies of the Histidine Residue in α-Lytic Protease. Implications for the Catalytic Mechanism of Serine Proteases. *Biochemistry*, 12(23):4732–4743, 1973.

[102] T Imai, R Hiraoka, A Kovalenko, and F Hirata. Water Molecules in a Protein Cavity Detected by Statistical-Mechanical Theory. *J. Am. Chem. Soc.*, 127:15334–15335, 2005.

[103] E F Jansen, M-D F Nutting, and A K Balls. Mode Of Inhibition Of Chymotrypsin By Diisopropyl Fluorophosphate. *J. Mol. Biol.*, 179:201–204, 1949.

[104] C Jarzynski. Non-equilibrium equality for free energy differences. *Phys. Rev. Lett.*, 78:2690–2693, 1997.

[105] F Jensen. *Introduction to Computational Chemistry*. John Wiley & Sons, 1999.

[106] W Jorgensen, D S Maxwell, and J Tirado-Rives. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.*, 118:11225–11236, 1996.

[107] W L Jorgensen, J Chandrasekhar, J D Madura, and R W Impey M L Klein. Comparison of Simple Potential Potential Functions for Simulating Liquid Water. *J. Chem. Phys.*, 79:926–935, 1983.

[108] A D McKerell Jr., D Bashford, M Bellott, R L Dunbrack Jr., J D Evanseck, M J Field, S Fischer, J Gao, H Guo, S Ha, D Joseph-McCarthy, L Kuchnir, K Kuczera, F T K Lau, C Mattos, S Michnick, T Ngo, D T Nguyend, B Prodhom, W W Reiher III, B Roux, M Schlenkrich, J C Smith, R Stote, J Straub, M Watanabe, J Wiorkiewicz-Kuczera, D Yin, and M Karplus. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*, 102:3586–3616, 1998.

[109] H B Kagan and J C Fiaud. Kinetic Resolution. volume 18 of *Topics in Stereochemistry*, pages 249–330. John Wiley, New York, 1988.

[110] G A Kaminski, R Friesner, J Tirado-Rives, and W Jorgensen. Evaluation and Reparameterization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. *J. Phys. Chem. B*, 105:6474–6487, 2001.

[111] G A Kaminski, H A Stern, B J Berne, R A Friesner, Y X Cao, R M Murphy, R Zhou, and T A Halgren. Development of a Polarizable Force Field For Proteins via *ab initio* quantum chemistry: Firste generation model and gas phase tests. *J. Comp. Chem.*, 23:1515–1531, 2002.

[112] M E Karpen, D J Tobias, and C L Brooks III. Statistical Clustering Techniques for the Analysis of Long Molecular Dynamics Trajectories: Analysis of 2.2-ns Trajectories of YPGDV. *Biochemistry*, 32:412–420, 1993.

[113] K Kawasaki, H Kondo, M Suzuki, S Ohgiya, and S Tsuda. Serine Of Bacillus Subtilis Lipase Determined At 1.3 A Resolution. *Acta Crystallogr. Sect. D*, 58:1168, 2002.

[114] T Ke, B Tidor, and A M Klibanov. Molecular-Modeling Calculations of Enzymatic Enantioselectivity Taking Hydration into Account. *Biotechnol. Bioeng.*, 57:741–745, 1998.

[115] K K Kim, H K Song, D H Shin, K Y Hwang, and S W Suh. The crystal structure of a triaglycerol lipase from *pseudomonas cepacia* reveals a highly open conformation in absence of a bound inhibitor. *Structure*, 5:173–185, 1997.

[116] S Kirkpatrick, C D Gelatt, and M P Vecchi. Optimization by Simulated Annealing. *Science*, 220:671–680, 1983.

[117] J G Kirkwood. Statistical Mechanics of Fluid Mixtures. *J. Chem. Phys.*, 3:300–313, 1935.

[118] A Klamt. Incorporation of solvent effects into density functional calculations of molecular energies and geometries. *J. Phys. Chem.*, 103:9312–9320, 1995.

[119] A M Klibanov. Improvin enzymes by using them in organic solvents. *Nature*, 409:241–246, 2001.

[120] A A Klyosov, N Van Viet, and I V Berezin. The Reactions of α-Chymotrypsin and Related Proteins with Ester Substrates in Non-aqueous Solvents. *Eur. J. Biochem.*, 59:3–7, 1975.

[121] W Kohn and L J Sham. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.*, 140:A1133–A1138, 1965.

[122] P Kollman. Free Energy Calculations: Applications to Chemical and Biochemical Phenomena. *Chem. Rev*, 93:2395–2417, 1993.

[123] P Kollman. Advances and Continuing Challenges in Achieving Realistic and Predictive Simulations of the Properties of Organic and Biological Molecules. *Acc. Chem. Res.*, 29:461–469, 1996.

[124] A A Kossiakoff and S Spencer. Direct Determination of the Protonation States of Aspartic Acid-102 and Histidine-57 in the Tetrahedral Intermediate of the Serine Proteases: Neutron Diffraction Structure of Trypsin. *Biochemistry*, 20:6462–6474, 1981.

[125] J Kottalam and D A Case. Dynamics of ligand escape from the heme pocket of myoglobin. *J. Am. Chem. Soc.*, 110:7690–7697, 1988.

[126] S Kumar, J M Rosenberg, D Bouzida, R H Swendsen, and P A Kollman. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comp. Chem.*, 13:1011–1021, 1992.

[127] S Kumar, J M Rosenberg, D Bouzida, R H Swendsen, and P A Kollman. Multidimensional Free Energy Calculations Using the Weighted Histogram Analysis Method. *J. Comp. Chem.*, 16:1339–1350, 1994.

[128] K Laidler. *Chemical Kinetics*. Harper & Row, 1987.

[129] A Laio and M Parrinello. Escaping Free-energy minima. *Proc. Natl. Acad. Sci. USA*, 99:12562–12566, 2002.

[130] A R Leach. *Molecular Modelling: Principles and Applications*. Longman, Harlow, 1996.

[131] C Lee, W Yang, and R G Parr. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B*, 37:785ff, 1988.

[132] H Lin, J Pu, T V Albu, and D Truhlar. Efficient Molecular Mechanics for Chemical Reactions: Multiconfiguration Molecular Mechanics Using Partial Electronic Structure Hessians. *J. Phys. Chem. A*, 108:4112–4124, 2004.

[133] H Lin and D G Truhlar. QM/MM: What have we learned, where are we, and where do we go from here? In *Proceedings of the 10th Electronic Computational Chemistry Conference (ECCC 10)*, 2005.

[134] A D MacKerell and M Karplus. Importance of Attractive van der Waals Contribution in Empirical Energy Function Models for the Heat of Vaporization of Polar Liquids. *J. Phys. Chem.*, 95:10559–10560, 1991.

[135] J R Maple, U Dinur, and A T Hagler. Derivation of force fields for molecular mechanics and dynamics from *ab initio* energy surfaces. *Proc. Natl. Acad. Sci. USA*, 85:5350–5354, 1988.

[136] J L Markley and I B Ibanez. Zymogen Activation in Serine Proteinases. Proton Magnetic Resonance pH Titration Studies of the Two Histidines of Bovine Chymotrypsinogen A and Chymotrypsin A-alpha. *Biochemistry*, 17:4627–4640, 1978.

[137] S Martí, Maite Roca, J Andrés, V Moliner, E Silla, I Tuñon, and J Bertrán. Theoretical insights in enzyme catalysis. *Chem. Soc. Rev.*, 33:98–107, 2004.

[138] K Martinek, A A Klyosov, N F Kazanskaya, and I V Berezin. The Free Energy-Reaction Coordinate Profile for α-Chymotryptic Hydrolysis of a Series of N-Acetyl-α-L-Amino Acid Methyl Esters. *Int. J. Chem. Kinet.*, 6:801–811, 1974.

[139] B W Matthews, P B Sigler, R Henderson, and D M Blow. Three-dimensional Structure of Tosyl-α-chymotrypsin. *Nature*, 214:652–656, 1967.

[140] M Mezei. Adaptive Umbrella Sampling: Self-consistent Determination of the Non-Boltzmann Bias. *J. Comp. Phys.*, 68:237, 1987.

[141] B Miehlich, A Savin, H Stoll, and H Preuss. Results obtained with the correlation energy density functionals of becke and Lee, Yang and Parr. *Chem. Phys. Lett.*, 157:200ff, 1989.

[142] E Miyako, T Maruyama, N Kamiya, and M Goto. A Supported Liquid Membrane Encapsulating a Surfactant-Lipase Complex for the Selective Separation of Organic Acids. *Chem. Eur. J.*, 11:1163–1170, 2005.

[143] K Möhle, H J Hofmann, and W Thiel. Description of Peptide and Protein Secondary Structures Employing Semiempirical Methods. *J. Comp. Chem*, 22:509–520, 2001.

[144] G Monard and K M Merz. Combined Quantum Mechanical/Molecular Mechanical Methodologies Applied to Biomolecular Systems. *Acc. Chem. Res.*, 32:904–911, 1999.

[145] K L Morrison and G A Weiss. Combinatorial alanine-scanning. *Curr. Opin. Chem. Biol.*, 5:302–307, 2001.

[146] E M Müller, A de Meijere, and H Grubmüller. Predicting unimolecular chemical reactions: Chemical flooding. *J. Chem. Phys.*, 116:897–905, 2002.

[147] K Nam, X Prat-Resina, M Garcia-Viloca, L S Devi-Kesavan, and J Gao. Dynamics of an Enzymatic Substitution Reaction in Haloalkane Dehalogenase. *J. Am. Chem. Soc.*, 126:1369–1376, 2004.

[148] A V Nemukhin, I A Topol, and S K Burt. Energy Profiles for the Rate-Limiting Stage of the Serine Protease Prototype Reaction. *Int. J. Quant. Chem.*, 88:34–40, 2002.

[149] I-S Ng and S-W Tsai. Partially Purified *carica papaya* Lipase: a Versatile biocatalyst for the hydrolytic resolution of (r,s)-2-arylpropionic thioesters in water-saturated organic solvents. *Biotechnol. Bioeng.*, 91:106–112, 2005.

[150] S H Northrup, M R Pear, C Y Lee, J A McCammon, and M Karplus. Dynamical theory of activated processes in globular proteins. *Proc. Natl. Acad. Sci. USA*, 79:4035–4039, 1982.

[151] IUPAC-IUB Commission on Biochemical Nomenclature. Abbreviations and Symbols for the Description of the Conformation of Polypeptide Chains. Tentative Rules (1969). *Biochemistry*, 9:3471–3479, 1970.

[152] R A Oosterbaan and M E van Adrichem. Isolation of acetyl peptides from acetylchymotrypsin. *Biochim. Biophys. Acta*, 27:423–425, 1958.

[153] J Ottosson, L Fransson, and K Hult. Substrate entropy in enzyme enantioselectivity: An experimental and molecular modeling study of a lipase. *Prot. Sci.*, 11:1462–1471, 2002.

[154] J Ottosson, L Fransson, J W King, and Karl Hult. Size as a parameter for solvent effects on *candida antarctica* lipase b enantioselectivity. *Biochim. et Biophys. Acta*, 1594:325–334, 2002.

[155] J Ottosson, J C Rotticci-Mulder, D Rotticci, and K Hult. Rational design of enantioselective enzymes requires considerations of entropy. *Protein Science*, 10:1769–1774, 2001.

[156] P L A Overbeeke, J A Jongejan, and J J Heinen. Solvent Effect on Lipase Enantioselectivity. Evidence for the Presence of Two Thermodynamic States. *Biotechnol. Bioeng.*, 70:278–290, 2000.

[157] J Pavelites, J Gao, P A Bash, and A D MacKerell. A Molecular Mechanics Force Field for NAD^+, NADH, and the Pyrophosphate Groups of Nucleotides. *J. Comp. Chem.*, 18(2):221–239, 1996.

[158] V T Pham and R S Phillips. Effects of Substrate Structure and Temperature on the Stereospecificity of Secondary Alcohol Dehydrogenase from *thermoanaerobacter ethanolicus*. *J. Am. Chem. Soc.*, 112:3629–3632, 1990.

[159] R S Philips. Temperature modulation of the stereochemistry of enzymatic catalysis: prospects for exploitation. *Tibtech*, 14:13–16, 1996.

[160] J W Ponder and D A Case. Force fields for protein simulation. *Adv. Prot. Chem.*, 66:27–85, 2003.

[161] G Van Pouderoyen. *Bacillus Subtilis* Lipase A With Covalently Bound SC/RC-IPG Phosphonate-Inhibitor. 2001.

[162] G Van Pouderoyen, T Eggert, K E Jaeger, and B W Dijkstra. The Crystal Structure Of *bacillus subtilis* Lipase: A Minimal α/β-Hydrolase Fold Enzyme. *J Mol Biol*, 309:215, 2001.

[163] W H Press, S A Teukolsky, W T Vetterling, and B P Flannery. *Numerical Recipes in Fortran 77*. Cambridge University Press, 1992.

[164] K Priya and A Chadha. Synthesis of hydrocinnamic esters by *pseudomona cepacia* lipase. *Enzyme Microb. Tech.*, 32:485–490, 2003.

[165] M Quack. Struktur und Dynamik chiraler Moleküle. *Angew. Chem.*, 101:588–604, 1989.

[166] R Radhakrishnan and T Schlick. Biomolecular free energy profiles by a shooting/umbrella sampling protocol, 'BOLAS'. *J. Chem. Phys.*, 121:2436–2444, 2004.

[167] R Radhakrishnan and T Schlick. Orchestration of cooperative events in DNA synthesis and repair mechanism unraveled by transition path sampling of DNA polymerase β's closing. *Proc. Natl. Acad. Sci. USA*, 101:5970–5975, 2004.

[168] R Radhakrishnan and T Schlick. Fidelity Discrimination in DNA Polymerase : Differing Closing Profiles for a Mismatched (G:A) versus Matched (G:C) Base Pair. *J. Am. Chem. Soc.*, 127:13245–13252, 2005.

[169] R Rajamani and J Gao. Balancing Kinetic and Thermodynamic Control: the Mechanism of Carbocation Cyclization by Squalene Cyclase. *J. Am. Chem. Soc.*, 125:12768–12781, 2003.

[170] M T Reetz and K-E Jaeger. Enantioselective Enzymes for Organic Synthesis Created by Directed Evolution. *Chem. Eur. J.*, 6:407–412, 2000.

[171] N Reuter, A Dejaegere, B Maigret, and M Karplus. Frontier Bonds in QM/MM Methods: A Comparison of Different Approaches. *J. Phys. Chem.*, 104:1720–1735, 2000.

[172] N Reuter, H Lin, and W Thiel. Green Fluoreszent Proteins: Empirical Force Field for the Neutral and Deprotonated Forms of the Chromophore. Molecular Dynamics Simulations of the Wild Type and S65T Mutant. *J. Phys, Chem. B*, 106:6310–6321, 2002.

[173] D Riccardi, G Li, and Q Cui. Importance of van der Waals Interactions in QM/MM Simulations. *J. Phys. Chem. B*, 108:6467–6478, 2004.

[174] B Roux. The calculation of the potential of mean force using computer simulations. *Comp. Phys. Comm.*, 91:275–282, 1995.

[175] A C Satterthwait and W P Jencks. The Mechanism of the Aminolysis of Acetate Esters. *J. Am. Chem. Soc.*, 96:7018–7031, 1974.

[176] A Schäfer, H Horn, and R Ahlrichs. Fully optimized contracted Gaussian basis sets for atoms Li to Kr. *J. Chem. Phys.*, 97:2571–2577, 1991.

[177] A Schäfer, C Huber, and R Ahlrichs. Fully optimized contracted Gaussian basis sets of triple zeta valence quality for atoms Li to Kr. *J. Chem. Phys.*, 100:5829–5835, 1994.

[178] N K Schaffer, S C May, and W H Summerson. Serine Phosphoric Acid From Diisopropylphosphoryl Chymotrypsin. *J. Biol. Chem.*, 202:67–76, 1952.

[179] G A Schenter, B C Garrett, and D G Truhlar. Generalized transition state theory in terms of the potential of mean force. *J. Chem. Phys.*, 119:5828–5833, 2003.

[180] S K Schiferl and D C Wallace. Statistical error in molecular dynamics averages. *J. Chem. Phys.*, 83:5203–5209, 1985.

[181] R D Schmid and R Verger. Lipasen: Grenzflächen-Enzyme mit attraktiven Anwendungen. *Angew. Chem.*, 110:1694–1720, 1998.

[182] U W Schmitt and G A Voth. Multistate Empirical Valence Bond Model for Proton Transport in Water. *J. Phys, Chem. B*, 102:5547–5551, 1998.

[183] J C Schöneboom, S Cohen, H Lin, S Shaik, and W Thiel. QM/MM Investigation of the Mechanism of C-H Hydroxylation of Camphor by Cytochrome $\mathrm{P450_{cam}}$: Theory Supports a Two-State Rebound Mechanism. *J. Am. Chem. Soc.*, 126:4017–4034, 2004.

[184] T Schulz, R D Schmid, and J Pleiss. Structural basis of stereoselectivity in *candida rugosa* lipase-catalyzed hydrolysis. *J. Mol. Model.*, 7:265–270, 2001.

[185] W. R. P. Scott, P. H. Hünenberger, I. G. Tironi, A. E. Mark, S. R. Billeter, J. Fennen, A. E. Torda, T. Huber, P. Krüger, and W. F. van Gunsteren. The GROMOS Biomolecular Simulation Program P ackage. *J. Phys. Chem. A*, 103(19):3596–3607, 1999.

[186] H M Senn, D O'Hagan, and W Thiel. Insight Into Enzymatic C-F Bond Formation from QM and QM/MM Calculations. *J. Am. Chem. Soc.*, 127:13643–13655, 2005.

[187] H M Senn, S Thiel, and W Thiel. Enzymatic Hydroxylation in p-Hydroxybenzoate Hydroxylase: A Case Study for QM/MM Molecular Dynamics. *J. Chem. Theory Comput.*, 1:494–505, 2005.

[188] S Shaik, D Kumar, S P de Visser, A Altun, and W Thiel. A Theoretical Perspective on Structure and Mechanism of Cytochrome P450 Enzymes. *Chem. Rev.*, 105:2279–2328, 2005.

[189] P S Shenkin and D Q McDonald. Cluster Analysis of Molecular Conformations. *J. Comp. Chem.*, 15:899–1916, 1994.

[190] P Sherwood. Hybrid Quantum Mechanics/Molecular Mechanics Approaches. volume 1 of *NIC Series*. John von Neumann Institute for Computing (NIC), 2000.

[191] P Sherwood, A H de Vries, M F Guest, G Schreckenbach, C Richard A Catlow, S A French, A A Sokol, S T Bromley, W Thiel, A J Turner, S Billeter, F Terstegen, S Thiel, J Kendrick, S C Rogers, J Casci, M Watson, F King, E Karlsen, M Sjøvoll, A Fahmi, A Schäfer, and C Lennartz. QUASI: A general purpose implementation of the QM/MM approach and its application to problems in catalysis. *J. Mol. Struct.*, 632:1–28, 2003.

[192] M Sierka and J Sauer. Finding Transition structures in extended systems: A strategy based on a combined quantum mechanics-empirical valence bond approach. *J. Chem. Phys.*, 112:6983–6996, 2000.

[193] C J Sih and S-H Wu. Resolution of Enantiomers via Biocatalysis. volume 19 of *Topics in Stereochemistry*, pages 63–97. John Wiley, New York, 1989.

[194] T Simonson, G Archontis, and M Karplus. Continuum treatment of long-range interactions in free energy calculations. Application to protein-ligand binding. *J. Phys. Chem. B*, pages 8349–8362, 1997.

[195] U C Singh and P A Kollman. A Combined *ab initio* Quantum Mechanical and Molecular Mechanical Method for Carrying out Simulations on Complex Molecular Systems: Applications to the $\mathrm{CH_3CL} + \mathrm{CL}^-$ Exchange Reaction and Gas-Phase Protonation of Polyethers. *J. Comp. Chem.*, 7:718–730, 1986.

[196] P T Snee, J Shanoski, and C B Harris. Mechanism of LIgand Exchange Studied using Transition Path Sampling. *J. Am. Chem. Soc.*, 127:1286–1290, 2005.

[197] M Souaille and B Roux. Extension to the weighted histogram method: combining umbrella sampling with free energy calculations. *Comp. Phys. Comm.*, 135:40–57, 2001.

[198] T P Straatsma, H J C Berendsen, and J P M Postma. Free energy of hydrophobic hydration: A molecular dynamics study of noble gases in water. *J. Chem. Phys.*, 85:6720–6727, 1986.

[199] J L Sussman, M Harel, F Frolow, C Oefner, A Goldman, L Toker, and I Silman. Atomic Structure of Acetylcholinesterase from Torpedo californica: A Prototypic Acetylcholine-Binding Protein. *SCIENCE*, 253:872–879, 1991.

[200] B G Tabachnick and L S Fidell. *Using Multivariate Statistics*. Allyn & Bacon, Boston, 2000.

[201] F Theil. *Enzyme in der organischen Synthese*. Spektrum, Akademischer Verlag, Heidelberg, 1997.

[202] W Thiel. Semiempirical Methods. volume 1 of *NIC Series*, pages 233–255. John von Neumann Institute for Computing (NIC), 2000.

[203] L H Thomas. *Proc. Cambridge Phil. Soc.*, 23:542, 1927.

[204] P J Tonge and P R Carey. Length of the Acyl Carbonyl Bond in Acyl-Serine Proteases Correlates with Reactivity. *Biochemistry*, 29:10723, 1990.

[205] P J Tonge and P R Carey. Forces, Bond Lengths, and Reactivity: Fundamental Insight into the Mechanism of Enzyme Catalysis. *Biochemistry*, 31:9122, 1992.

[206] A E Torda and W F van Gunsteren. Algorithms for Clustering Molecular Dynamics Configurations. *J. Comp. Chem.*, 15:1331–1340, 1994.

[207] G M Torrie and J P Valleau. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comp. Phys.*, 2:187–199, 1977.

[208] T S van Erp, D Moroni, and P G Bolhuis. A novel path sampling method for the calculation of rate constants. *J. Chem. Phys.*, 118:7762–7774, 2003.

[209] Wilfred F. van Gunsteren, Xavier Daura, and Alan E. Mark. GROMOS Force Field. In Paul von Ragué Schleyer, editor, *Encyclopedia of Computational Chemistry*, volume 2, pages 1211–1216. Wiley, Chichester, 1998.

[210] E Vedejs and M Jure. Efficiency in Nonenzymatic Kinetic Resolution. *Angew. Chem. Int. Ed.*, 44:3974–4001, 2005.

[211] J Villà, J Bentzien, À Gonzàlez-Lafont, J M Lluch, J Bertran, and A Warshel. Effective Way of Modelling Chemical Catalysis: Empirical Valence Bond Picture of Role of Solvent and Catalyst in Alkylation Reactions. *J. Comp. Chem.*, 8:607–625, 2000.

[212] J Villà and A Warshel. Energetics and Dynamics of Enzymatic Reactions. *J. Phys. Chem. B*, 105:7887–7907, 2001.

[213] D Voet and J Voet. *Biochemistry*. John Wiley & Sons, 1995.

[214] A F Voter. A method for accelerating the molecular dynamics simulation of infrequent events. *J. Chem. Phys,*, 106:4665–4667, 1997.

[215] G Vriend. WHAT IF: A molecular modeling and drug design program. *J. Mol. Graph.*, 8:52–56, 1990.

[216] Y F Wang, C S Chen, G Girdaukas, and C J Sih. Bifunctional chiral synthons via biochemical methods. iii. Optical purity enhancement in enzymic asymmetric catalysis. *J. Am. Chem. Soc.*, 106:3695–3696, 1984.

[217] A Warshel. *Computer Modeling of Chemical Reactions in Enzymes and Solutions*. John Wiley & Sons, New York, 1992.

[218] A Warshel. Computer Simulation Of Enzyme Catalysis. *Annu. Rev. Biophys. Biomol. Struct.*, 32:425–443, 2003.

[219] A Warshel, J Florián, M Štrajbl, and J Villà. Circe Effect versus Enzyme Preorganization: What Can Be Learned from the Structure of the Most Proficient Enzyme. *ChemBioChem*, 2:109–111, 2001.

[220] A Warshel and M Karplus. Calculation of π-π^* excited state conformations and vibronic structure of retinal and related molecules. *J. Am. Chem. Soc.*, 96:5677–5689, 1974.

[221] A Warshel and M Levitt. Theoretical Studies of Enzymic Reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme. *J. Mol. Biol.*, 103:227–249, 1976.

[222] A Warshel, G Naray-Szabo, F Sussman, and J K Hwang. How Do Serine Proteases Really Work. *Biochemistry*, 28:3629–3637, 1989.

[223] A Warshel, A Papazyan, and P A Kollman. On Low-Barrier Hydrogen Bonds and Enzyme Catalysis. *Science*, 269:102–104, 1995.

[224] A Warshel, F Sussman, and J K Hwang. Evaluation of Catalytic Free Energies in Genetically Modified Proteins. *J. Mol. Biol.*, 201:139–159, 1988.

[225] J B Watney, P K Agarwal, and S Hammes-Schiffer. Effect of Mutation on Enzyme Motion in Dihydrofolate Reductase. *J. Am. Chem. Soc.*, 125:3745–3750, 2003.

[226] F Wennmohs and M Schindler. Development of a Multipoint Model for Sulfur in Proteins: A New Parametrization Scheme to Reproduce High-Level Ab Initio Interaction Energies. *J. Comp. Chem.*, 26:283–293, 2005.

[227] C R Wescott, H Noritomi, and A M Klibanov. Rational Control of Enzymatic Enantioselectivity through Solvation Thermodynamics. *J. Am. Chem. Soc.*, 118:10365–10370, 1996.

[228] I H Williams and G M Maggiora. Use and abuse of the distinguished-coordinate method for transition state structure searching. *J. Mol. Struct.*, 89:365–378, 1982.

[229] Y Wu, J D Schmitt, and R Car. Mapping potential energy surfaces. *J. Chem. Phys.*, 121:1193–1200, 2003.

[230] S Yagisawa. Two types of rate-determining steps in chemical and biochemical processes. *Biochem. J.*, 263:985–988, 1989.

[231] T Yamane. *Statistik. Ein Einführendes Lehrbuch.* Fischer, Frankfurt, 1987.

[232] D Yin and A D Mackerell. Combined *ab initio*/Empirical Approach for the Optimization of Lennard-Jones Parameters.

[233] D Yin and A D Mackerell. Ab initio calculations on the use of helium and neon as probes of van der Waals surfaces of molecules. *J. Phys. Chem.*, 100:11946–11975, 1996.

[234] X Zhang and K N Houk. Why Enzymes Are Proficient Catalysts: Beyond the Pauling Paradigm. *Acc. Chem. Res.*, 38:379–440, 2005.

[235] Y Zhang, J Kua, and J A McCammon. Role of the Catalytic Triad and Oxyanion Hole in Acetylcholinesterase Catalysis: An ab initio QM/MM Study. *J. Am. Chem. Soc.*, 124:10572–10577, 2002.

[236] Y Zhang, J Kua, and J A McCammon. Influence of Structural Fluctuation on Enzyme Reaction Energy Barriers in Combined Quantum Mechanical/Molecular Mechanical Studies. *J. Phys. Chem. B*, 107:4459–4463, 2003.

[237] Y Zhang, T-S Lee, and W Yang. A pseudobond approach to combining quantum mechanical and molecular mechanical methods. *J. Chem. Phys.*, 110:46–54, 1999.

[238] Y Zhang, H Liu, and W Yang. Free energy calculation on enzyme reactions with an efficient iterative procedure to determine minimum energy paths on a combined ab initio QM/MM potential energy surface. *J. Chem. Phys.*, 112:3483–3492, 2000.

[239] R W Zwanzig. High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *J. Chem. Phys.*, 22:1420–1426, 1954.

Erklärung

Die hier vorgelegte Dissertation habe ich eigenständig und ohne unerlaubte Hilfe angefertigt. Die Dissertation wurde in der vorgelegten oder in ähnlicher Form noch bei keiner anderen Institution eingereicht. Ich habe bisher keine erfolglosen Promotionsversuche unternommen.

Mülheim an der Ruhr, den 2.6.2006

Nikolaj Otte

(Alexander-Nikolaj Otte)