

Propositions

accompanying the thesis

“Computational Study of Enzyme Enantioselectivity”

by Yu Zhou

1. The natural abundance of homochiral proteins and peptides, containing exclusively *L*-amino acids, over their heterochiral counterparts does not reflect their thermodynamic stability.

(This thesis)

2. The finding that chiral induction by the adjacent chiral centers located in the serine-substrate moiety of the tetrahedral intermediate of the serine-type hydrolase CaLB, does not quantitatively determine the enantioselectivity of the enzyme, suggests that there is room for further optimization of this property.

(This thesis)

3. In many reports of simulations using free energy perturbation methods to predict enzymatic enantioselectivity, the convergence problem has been underestimated.

(N.M. Micaelo *et al.*, *Biophys. J.* **2005**, 89, 999-1008; C.I. Sainz-Diaz, *et al.*, *J. Mol. Struct. (Theochem)* **1997**, 390, 225-237; K. Gruber, *Proteins* **2001**, 44, 26-31)

4. The term ‘fitting like a glove’ that is often used to exemplify the tight binding of substrates to enzymes should be replaced by ‘fitting like a mitten’.

(P.L.A. Overbeeke, *et al.*, *Chem. Phys. Lipids*, **1998**, 93, 81-93)

5. The application of synthetic *D*-amino acid polypeptides as therapeutics needs to be carried out with caution.

6. When discussing the importance of homochirality in relation to the origin of life on Earth, the overwhelming occurrence of *L*-amino acids and *D*-sugars should be considered to be accidental rather than deterministic.

(G. Wald, *Ann. N.Y. Acad. Sci.*, **1957**, 69, 352-368)

7. The significance of computational chemistry as an art is succinctly expressed in Pablo Picasso’s proposition that “Art is the lie that helps tell the truth”.

8. Respect towards Nature inhibited the ancient Chinese to develop science and technology.

9. A highly structured society, in which security is given more importance than personal freedom, leads to dependence and inability.

Stellingen

behorend bij het proefschrift

“Computational Study of Enzyme Enantioselectivity”

door Yu Zhou

1. De in de natuur aangetroffen overmaat van homochirale eiwitten en peptiden, welke uitsluitend *L*-aminozuren bevatten, ten opzichte van hun heterochirale tegenhangers is geen indicatie voor hun thermodynamische stabiliteit.

(Dit proefschrift)

2. Het feit dat chirale inductie, door de naast elkaar gelegen chirale centra in het serine-substraat-gedeelte van het tetraëdrisch intermediair van het serine hydrolase CaLB, de enantioselectiviteit van het enzym niet kwantitatief bepaalt, suggereert dat er ruimte is voor verdere optimalisatie van deze eigenschap.

(Dit proefschrift)

3. In veel beschrijvingen van simulaties die gebruik maken van vrije-energie perturbatiemethoden om de enantioselectiviteit van enzymen te voorspellen, wordt het convergentieprobleem onderschat.

(N.M. Micaelo *et al.*, *Biophys. J.* **2005**, *89*, 999-1008; C.I. Sainz-Diaz, *et al.*, *J. Mol. Struct. (Theochem)* **1997**, *390*, 225-237; K. Gruber, *Proteins* **2001**, *44*, 26-31)

4. De uitdrukking “passen als een handschoen”, die veelvuldig wordt gebruikt om de hechte binding tussen substraten en enzymen te illustreren, zou vervangen moeten worden door “passen als een want”.

(P.L.A. Overbeeke, *et al.*, *Chem. Phys. Lipids*, **1998**, *93*, 81-93)

5. Toepassing van synthetische *D*-aminozuur polypeptiden als therapeutische preparaten dient te gebeuren met grote zorg.

6. In discussies met betrekking tot het belang van homochiraliteit in relatie met het ontstaan van leven op aarde, moet het vrijwel uitsluitend voorkomen van *L*-aminozuren en *D*-suikers worden beschouwd als niet meer dan toeval.

(G. Wald, *Ann. N.Y. Acad. Sci.*, **1957**, *69*, 352-368)

7. De betekenis van computational chemistry als een vorm van kunst wordt treffend verwoord in Pablo Picasso's stelling dat “Kunst is de leugen die de waarheid helpt verwoorden”.

8. Respect voor de natuur heeft de ontwikkeling van wetenschap en technologie in het oude China geremd.

9. Een streng-gestruktureerde maatschappij, waarin meer belang wordt gehecht aan veiligheid dan aan persoonlijke vrijheid, leidt tot afhankelijkheid en onvermogen.

Computational Study of Enzyme Enantioselectivity

Thesis

YU ZHOU

2006

Computational Study of Enzyme Enantioselectivity

Proefschrift

ter verkrijging van de graad van doctor

aan de Technische Universiteit Delft,

op gezag van de Rector Magnificus Prof. dr ir J.T. Fokkema,

voorzitter van het College voor Promoties

in het openbaar te verdedigen op maandag 11 september 2006 om 17:30 uur

door

Yu ZHOU

Master of Biotechnology

geboren te Chongqing, China

Dit proefschrift is goedgekeurd door de promotoren:

Prof. dr. S.W. de Leeuw

Prof. dr. W.R. Hagen

Samenstelling van de promotiecommissie:

Rector Magnificus, voorzitter

Prof. dr. S.W. de Leeuw

Technische Universiteit Delft, promotor

Prof. dr. W.R. Hagen

Technische Universiteit Delft, promotor

Dr. ir. J.A. Jongejan

Technische Universiteit Delft, toegevoegd promotor

Prof. dr. W.F. van Gunsteren

Eidgenössische Technische Hochschule Zürich

Prof. dr. ir. J.A.M. de Bont

Technische Universiteit Delft

Prof. dr. ir. L.A.M. van der Wielen

Technische Universiteit Delft

Dr. C. Oostenbrink

Vrije Universiteit Amsterdam

The studies presented in this thesis were performed in the Section Enzymology of the Department of Biotechnology and the Section Physical Chemistry and Molecular Thermodynamics of the Department of Chemical Technology of the Delft University of Technology. The research was financially supported by the Board of Delft University of Technology.

ACKNOWLEDGEMENTS

No matter how good a thesis looks, it will never overshadow the great support of the people involved. When I listed the people I would like to acknowledge, I found they make a big team! Without the contribution from each of you, it would have been difficult for me to make the achievements so far.

I feel grateful to Jaap Jongejan for his four years supervision of my PhD work. Thank you for many inspiring contributions to my scientific research. You have shown me the important characteristics of a scientist: open-mindedness, curiosity, modesty, patience, persistence and exactness. The open discussions with you are always enjoyable and I will miss that informality in the future! Many thanks to Prof. Simon de Leeuw and Prof. Wilfred Hagen, my supervisors, for your constant support, guidance and trust! I very much like the open-mindedness in the groups led by you.

We had a wonderful collaboration with Prof. Wilfred van Gunsteren and his group at the ETH, Zurich. The work at the ETH, under the supervision of Wilfred and of Chris Oostenbrink really gave a big push to my project, from which I have benefited tremendously. Chris, later on moving to work at the Free University of Amsterdam (VU), kept on giving us great support. I appreciate your professionalism and your willingness to help others. You have been helping me with the free energy calculation with clear and patient explanations in person and through many emails. Aldo Jongejan from the VU also gave me great support with the calculation which led to a nice article. Many thanks!

I would like to give thanks to Jouke Heringa, Marcel van den Broek and Hans Kemper for their excellent support in setting up the high performance computing platform and solving the IT problems. I never felt that you were assisting me; instead, I always thought that we were a team. The help from Frank Sheldon with the Silicon Graphics at the very beginning of the project is also appreciated.

Many thanks to my colleagues Maarten Wolf, Jaap Flohil, Flavius Gligor and other members of the PCMT group and those of the Enzymology group, for the friendly environments you created.

Lies van der Meer, thank you so much for the consistent support during my stays in this country, especially during the most difficult period in 2004. Ger Aggenbach and Frieda von Boltog, thank you also for the support you gave me. Some words are short, but mean a lot! Sjaak and Jos Lisper, you are always trying your best to help me in the building. Your support, together with your sunny smiles and cheerful voices will stay as a nice memory in my mind.

I would also like to mention the three students who did their internships in my project: Sander van Pelt, Emmie Heeren and Eline van Maanen. Your work has been of great value for the project.

I cherish the friendship with some friends of mine: Song Bo and Jan, Li Xiaonan, Wu Liang, Huang Chengwei, Ren Penwei, Jana and Frank who made my life in Delft colorful and enjoyable; Roderick van der Graaf, with whom I can share the entrepreneurship; Zvika Frank, a fabulous dance therapist, a person who enriches the lives of many people; Wu Weiling, a person with such a big heart.

I also had a good time when serving as a board member of VCWI (the Society of Chinese Scholars and Engineers in the Netherlands) and CNLN (China – the Netherlands Life Science Network). Many great ideas were executed because of the wonderful teamwork in the clubs. Thanks, Yu sen, Xiu lan, Da wei, Zhenhua, Liao Yi, Chunming and many others!

I would like to give sincere thanks for the consistent support from my previous mentors in China: Prof. Wang Jingxing, Prof. Xiao Xiaopu and Dr. Dong Changjiang.

The inspirations from three entrepreneurs Jan Mellegers, Dirk Groenewegen and Frank Liang who introduced me to the business world are also appreciated.

Dear Mom and Dad, thank you for all the love and support you have been given to me. Without you, I could not have made the achievement so far. Mom, you set a good example for me how to fulfill responsibility for family and society. In my eyes, you are a common but an extraordinary woman. I have to say “thank you” to you, Boudewijn, for supporting me during the ups and downs in my life. You showed me the basics of life that help me to appreciate more many simple but precious things. I will also never forget the help from Zhang Keli, Liao Jiahui, Sunjun and Liu Yingfen, who have given me and my family so much support in difficult times.

CONTENTS

Acknowledgements	i
Summary	iv
Samenvatting	vi
Publications	viii
Chapter 1:	
General Introduction	1
Chapter 2:	
Computational prediction of enzyme enantioselectivity	14
Chapter 3:	
Computational study of the relative stability of homochiral and heterochiral alanine dipeptides. Effects of perturbation pathways and force-field parameters on free energy calculations	31
Chapter 4:	
Computational study of ground state chiral induction in small peptides: Comparison of the relative stability of selected amino acid dimers and oligomers in homochiral and heterochiral combinations	41
Chapter 5:	
Molecular modeling of the enantioselectivity of CaL B – Free Energy Calculation	56
Chapter 6:	
Conclusions and outlook	76
Curriculum Vitae	79

Computational Study of Enzyme Enantioselectivity

Yu Zhou

Over the past two decades the study of enzyme enantioselectivity has been intensified due to the increasing number of industrial applications of enzymes in the preparation of enantiomerically pure compounds. In the past, many efforts have been made to understand the mechanism of the enantioselectivity of enzymes and to modify the enantioselectivity to improve their performance. With the recent availability of x-ray structures of enzymes and the use of computer modeling, it is now possible to investigate enzyme enantioselectivity in even greater detail. In **Chapter 1** of this thesis we introduce the concepts and the implications of chirality, discuss its propagation in natural systems, and the possibilities for enzyme enantioselectivity prediction. A historical overview of computational studies of enzyme enantioselectivity is presented in **Chapter 2**, together with a more detailed explanation of the theory and the different strategies and approaches.

Molecular dynamics (MD) is a classical and powerful simulation method to explore the energy landscape of a biomolecular system, from which we can derive the free energy difference between two states. This free energy difference determines many thermodynamic properties of biomolecules including the relative stability of diastereomeric molecules and the substrate-enzyme complex during catalysis, which are related to homochirality propagation and enzyme enantioselectivity that has been investigated in this project. Prior to this thesis work, accurate prediction of enzyme enantioselectivity has been difficult due to the lack of efficient computational tools. Theoretically, quantum mechanical calculations are less biased, however, due to the computational costs it is virtually impossible to implement these methods for the total enzyme, while including explicit solvent. Molecular mechanics (MM) calculations are more time- and cost effective at the expense of additional assumptions and approximations. This is especially true when the methods rely on the possibility to relate the enantioselectivity of enzymes to the potential energy difference between the states on the reaction coordinate that determine the catalytic rate for either one of the enantiomeric substrates. Previously used computational methods gave correct qualitative predictions of the enantiopreference, but quantitative predictions of the enantioselectivity were commonly off by several orders of magnitude. The focus of this thesis is to develop a free energy calculation method based on MM/MD with the inclusion of entropy effects for the study of the relative stability of peptides and the prediction of enzyme enantioselectivity.

In molecular mechanics studies the free energy of a molecular system is quite sensitive to the force field employed. In order to evaluate the Gromos96 force field, the free energy calculations have been performed on small peptides. Integration pathways have been designed that produce minimal disturbance to the system, including the use of soft-atoms and low-energy intermediates. This leads to a smoother pathway with improved accuracy of free energy calculation. In **Chapter 3** dialanine peptide is used as the model molecule to investigate the relative stability of homochiral and heterochiral dialanine. The free energy calculation method and the thermodynamic integration protocol is validated using experimental data derived from peptide synthesis. The calculated free energy difference between homochiral and heterochiral dialanine peptide is 0.12 ± 0.03 kJ/mol, compared to the experimental value of 0.22 kJ/mol, derived from the kinetic resolution from peptide synthesis. With these results we formulated the hypothesis that the free energy difference between the homochiral and heterochiral alanine dipeptide might have played an important role in the evolutionary propagation of homochirality. This hypothesis was tested in the studies reported in **Chapter 4** where additional dipeptides (AS, AC, AV, AF, AK and AD), tripeptides (AVA) and a pentapeptide (AcGLSFA) were

investigated. The calculated free energy difference between the homochiral and heterochiral forms ranges from -0.9 to 0.5 kJ/mol (dipeptides) and from -2.4 to 2.3 kJ/mol (tripeptide) implying that homochiral peptides are not always favored over their heterochiral counterparts under the conditions tested. The preference appears to depend on size, character, and accessible conformational space of the peptides, as well as on temperature and solvent composition. As an indicator of the quality of the force field used, excellent agreement is found for the calculated and experimentally determined relative stabilities of the diastereomers of the all-*L* AcGLSFA and its diastereomer containing *D*-serine in the central position.

After validation of the protocol on the peptides, the computations were extended to the enantioselectivity prediction of the lipase CaLB, the results of which are described in **Chapter 5**. Two models: a truncated tetrahedral intermediate (TTI) model and a whole enzyme model were designed and tested in terms of accuracy and efficiency. The TTI model, simple and computationally inexpensive, serves as an efficient test to probe the enantioselectivity in enzymatic catalysis. Its usefulness can be rationalized as follows. If the enzyme accommodates the substrate in the (set of) conformations that are observed for the model compound that has the lower free energy of the two diastereomeric structures, the additional energetic costs are low. When other conformations are accommodated, however, the enzyme has to invest extra energy to overcome the local energy difference, which will reduce the reaction rate. Five different substrates, all secondary aliphatic alcohols but with different substituents, are studied both with the TTI model and the whole enzyme model. The TTI model gives correct predictions of the enantiopreference for three out of five substrates. The incorrect preference is obtained for substrates with bromine substituents. By comparison, the whole enzyme model not only gives the correct enantiopreference prediction for all the substrates, but also affords a quantitative estimate of the enantioselectivity that is remarkably close to experimentally determined values. Finally, we studied the enzyme mutant (Trp104Ala) for one of the substrates (1-phenylethanol) and were able to predict the enantioselectivity inversion from *R* to *S* (-11.5 → 4.0 kJ/mol), which has subsequently been observed in an experimental study of a highly similar substrate (-12.9 → 4.6 kJ/mol).

Een Computerstudie van Enzymenantioselectiviteit

Yu Zhou

Gedurende de laatste twee decennia is de bestudering van enzymenantioselectiviteit geïntensiveerd ten gevolge van een voortdurende toename van industriële toepassingen van enzymen in de bereiding van enantiomeer zuivere verbindingen. Er zijn veel pogingen gedaan om het mechanisme van de enantioselectiviteit van enzymen te begrijpen en om de enantioselectiviteit te wijzigen om verbeterde prestaties te bereiken. Met de beschikbaarheid van X-ray structuren van enzymen en computer modellering kunnen we enzymenantioselectiviteit op een meer rationele wijze onderzoeken. In **Hoofdstuk 1** van dit proefschrift introduceren we het concept van de chiraliteit, haar implicaties en propagatie, en de voorspelling van enantioselectiviteit. In **Hoofdstuk 2** wordt vervolgens een historisch overzicht gegeven van computer studies van enzymenantioselectiviteit en tevens een meer gedetailleerde uitleg van de theorie en de verschillende methoden en benaderingswijzen.

Moleculaire Dynamica (MD) is een klassieke en krachtige simulatiemethode om het energielandschap van een biomoleculair systeem te onderzoeken op grond waarvan we het vrije energieverval tussen twee toestanden kunnen afleiden. Dit vrije energieverval bepaalt veel thermodynamische eigenschappen van biomoleculen waaronder de relatieve stabiliteit van diastereoisomere moleculen en het substraat-enzymcomplex tijdens katalyse, welke gerelateerd zijn aan de in dit proefschrift onderzochte homochiraliteitspropagatie en enzymenantioselectiviteit. Vóór dit promotieonderzoek was nauwkeurige voorspelling van enzymenantioselectiviteit lastig vanwege een gebrek aan efficiënte computerprocedures. Theoretisch gezien zijn quantum-mechanische berekeningen minder 'biased' maar vanwege de hoge rekenkosten is het tot nu toe bijna onmogelijk om ze te implementeren met expliciet oplosmiddel en voor een heel enzym. Moleculaire Mechanica (MM) berekeningen zijn effectiever in termen van rekestijd en kosten, maar de nauwkeurigheid is minder. Dit is vooral het geval wanneer de methoden gebaseerd zijn op de mogelijkheid om de enantioselectiviteit van enzymen te relateren aan het potentiaalverschil van de toestanden op de reactiecoördinaat die de katalytische snelheid bepaalt voor een van de enantiomere substraten. Eerder gebruikte computermethoden zijn in staat tot correcte kwalitatieve voorspellingen van de enantiomere voorkeur, maar kwantitatieve voorspellingen van de enantioselectiviteit zitten er gewoonlijk verschillende orden van grootte naast. De focus van dit proefschrift is de ontwikkeling van een berekeningsmethode voor de vrije energie, gebaseerd op MM/MD met meenemen van entropieeffecten, ten behoeve van studies naar de relatieve stabiliteit van peptiden en van de voorspelling van enzymenantioselectiviteit.

De vrije energie van een moleculair systeem is nogal gevoelig voor welk 'force field' wordt gebruikt. Om het Gromos96 force field te evalueren zijn vrije energieberekeningen gedaan aan kleine peptiden. Integratiepaden zijn ontworpen welke minimale verstoring van het systeem bewerkstelligen, waarbij gebruik wordt gemaakt van 'zachte' atomen en laag-energetische intermediären. Dit leidt tot een meer geleidelijk pad met een meer nauwkeurige berekening van de vrije energie. In **Hoofdstuk 3** wordt het dialanine peptide gebruikt als modelmolecuul om de relatieve stabiliteit te onderzoeken tussen homochiraal en heterochiraal dialanine. De berekeningsmethode voor de vrije energie en het protocol voor de thermodynamische integratie zijn gevalideerd met behulp van experimentele peptide-synthesedata. Het berekende vrije energieverval tussen homochiraal en heterochiraal dialanine is 0.12 ± 0.03 kJ/mol, te vergelijken met de experimentele waarde van 0.22 kJ/mol afgeleid van de kinetische resolutie van peptidesynthese.

De hypothese wordt voorgesteld dat het vrije energieverschil tussen het homochirale en het heterochirale alanine dipeptide een belangrijke rol zou kunnen hebben gespeeld voor de evolutie van homochiraliteit. In **Hoofdstuk 4** wordt deze hypothese verder getest waarbij gebruik wordt gemaakt van meer dipeptiden (AS, AC, AV, AF, AK and AD), tripeptiden (AVA) en een pentapeptide (AcGLSFA). Het berekende vrije energieverschil tussen de homochirale en heterochirale vormen loopt van -0.9 tot 0.5 kJ/mol (dipeptiden) en van -2.4 to 2.3 kJ/mol (tripeptide) wat impliceert dat homochirale peptiden niet altijd de voorkeur hebben boven heterochirale peptiden onder de geteste condities. De voorkeur hangt af van de grootte, aard en toegankelijke conformatieruimte van de peptiden en van de temperatuur en de samenstelling van het oplosmiddel. Als een indicatie van de kwaliteit van het gebruikte force field wordt uitstekende overeenkomst gevonden tussen de berekende en de experimenteel bepaalde relatieve stabiliteiten van de diastereoisomeren van all-L AcGLSFA en het diastereoisomeer dat D-serine heeft in de middenpositie.

Na validering op peptide wordt vervolgens in **Hoofdstuk 5** toepassing van onze computermethode en ons protocol uitgebreid naar voorspelling van de enantioselectiviteit van het lipaseenzym CaLB. Twee modellen zijn getest in termen van nauwkeurigheid en efficiëntie: het ‘truncated tetrahedral intermediate’ (TTI) model en het hele enzymmodel. Het eenvoudige en qua rekentijd voordelige TTI model dient als een efficiënte test om enantioselectiviteit in enzymatische katalyse te sonderen. Hieraan ligt ten grondslag de gedachte dat, als er lokaal een vrije energieverschil bestaat tussen de twee diastereoisomeren, het waarschijnlijker is dat het enzym zich aan dit verschil aanpast. Anders moet het enzym extra energie investeren om het lokale energieverschil te overwinnen en dat is energetisch gezien niet economisch. Vijf verschillende substraten, alle secundaire alcoholen maar van verschillende grootte en samenstelling, zijn bestudeerd met het TTI model en met het hele enzymmodel. De richting van de enantioselectiviteit wordt door het TTI model correct voorspeld voor drie van de vijf substraten maar niet voor de substraten met bromidegroepen. Ter vergelijking: het hele enzym model maakt de juiste, en bovendien meer nauwkeurige voorspelling voor alle substraten. Ten slotte is gekeken naar een enzymmutant (Trp104Ala) voor een van de substraten (1-fenylethanol): de voorspelde enantioselectieve inversie van R naar S (-11.5 → 4.0 kJ/mol) is experimenteel waargenomen voor een vergelijkbaar substraat (-12.9 → 4.6 kJ/mol).

LIST OF PUBLICATIONS

Articles

1. Zhou, Y.; Oostenbrink, C.; Van Gunsteren, W. F.; Hagen, W. R.; De Leeuw, S. W.; Jongejan, J. A. (2005) “Relative stability of homochiral and heterochiral dialanine peptides. Effects of perturbation pathways and force-field parameters on free energy calculations” Molecular Physics 103 (14), 1961-1969
2. Zhou, Y.; Oostenbrink, C.; Jongejan, A.; Van Gunsteren, W. F.; Hagen, W. R.; De Leeuw, S. W.; Jongejan, J. A. (2006) “Computational study of ground-state chiral induction in small peptides: Comparison of the relative stability of selected amino acid dimers and oligomers in homochiral and heterochiral combinations” Journal of Computational Chemistry 27 (7), 857-867
3. Zhou, Y.; Oostenbrink, C.; Van Maanen, E.M.T.; Hagen, W. R.; De Leeuw, S. W.; Jongejan, J. A. Molecular “Modeling of the enantioselectivity of *Candida antarctica* lipase B – Free energy calculation” Journal of Computational Chemistry, accepted
4. Zhou, Y.; Jongejan, J. A. “Review: Computational prediction of enzyme enantioselectivity” Biocatalysis and Biotransformations, submitted

Oral presentations

1. Zhou, Y.; Oostenbrink, C.; Van Maanen, E.M.T.; Hagen, W. R.; De Leeuw, S. W.; Jongejan, J. A. “Computational study of the enantioselectivity of *CaLB* Lipase – Free Energy Calculation” International Conference on Biopartitioning and Purification 2005, the Netherlands

Posters and abstracts

1. Zhou, Y.; Oostenbrink, C.; Jongejan, A.; Van Gunsteren, W. F.; Hagen, W. R.; De Leeuw, S. W.; Jongejan, J. A. “Prediction of enzyme enantioselectivity through free energy calculation” 30th FEBS Congress: The Protein World, Budapest, Hungary, July 2005
2. Zhou, Y.; Oostenbrink, C.; Jongejan, A.; Van Gunsteren, W. F.; Hagen, W. R.; De Leeuw, S. W.; Jongejan, J. A. “Computational study of enzyme enantioselectivity” J&JPRD 4th Symposium on Drug Discovery, Antwerp, Belgium, April 2005
3. Zhou, Y.; Oostenbrink, C.; Jongejan, A.; Van Gunsteren, W. F.; Hagen, W. R.; De Leeuw, S. W.; Jongejan, J. A. “Computational prediction of enzyme enantioselectivity” NWO Computational Life Science (CLS) Group Meeting, Amsterdam, the Netherlands, Feb 2005

ABBREVIATIONS

BCL	<i>Burkholderia cepacia</i> lipase
CaLB	<i>Candida antarctica</i> lipase B
CRL	<i>Candida rugosa</i> lipase
DMW	Dynamical modified windows
ESP	Electrostatic potential
FEP	Free energy perturbation
Flops	Floating-point operations per second
HLL	<i>Humicola lanuginosa</i> lipase
<i>Hb</i> -HNL	<i>Hevea brasiliensis</i> hydroxynitrile lyase
HPC	High performance computing
MD	Molecular dynamics
MM	Molecular mechanics
PCL	<i>Pseudomonas cepacia</i> lipase
QM	Quantum mechanics
RHF	Restricted Hartree-Fock (<i>ab initio</i>)
RML	<i>Rhizomucor miehei</i> lipase
TI	Thermodynamic integration
TTI	Truncated tetrahedral intermediate

GENERAL INTRODUCTION

1. CHIRALITY AND LIFE

Chirality

Chirality (*handedness*) is derived from the Greek word $\chi\epsilon\rho\iota\varsigma$. It is a property of three-dimensional objects that are non-superimposable with their mirror image. In chemistry, this term applies to molecules that do not possess a plane, a center, or an alternating axis of symmetry. The mirror forms of chiral molecules are defined to display **enantiomeric** relationships. Molecules that possess a chiral center are commonly addressed as either *L* or *D* enantiomers (according to the Fischer convention) or *S* or *R* enantiomers (according to the Cahn-Ingold-Prelog, CIP, rules). When more than one chiral center is present in a molecule, isomers that display **diastereomeric** relationships arise as well. Enantiomers and diastereomers are jointly referred to as stereoisomers. Enantiomers are physically indistinguishable under isotropic conditions, except for a (minor) energetic difference resulting from the parity non-conserving character of the weak nuclear force (see below). Molecules or complexes that are diastereomerically related, on the other hand, are not mirror images. They may show considerable differences of physical and chemical properties. As an example, the enantiomeric and diastereomeric relationships between the stereoisomeric forms of a dipeptide are shown in Figure 1. Similarly, the complexes of a natural (containing all *L*-amino acids) enzyme with enantiomeric substrates are examples of combinations with diastereomeric character.

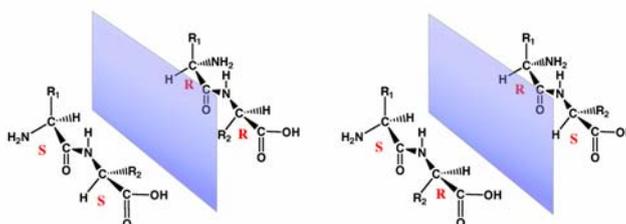


Figure 1. Stereoisomeric relationships between the isomers of dipeptides. The *S,S*- and *R,R*-peptides (left) and the *S,R*- and *R,S*-peptides (right) are enantiomerically related (mirror images). Other combinations, i.e. *S,S*- and *S,R*-peptides or *S,S*- and *R,S*-peptides, are diastereomerically related.

Mixtures containing one, or predominantly one, enantiomeric form are referred to as **enantiopure** or **enantiomerically enriched**, respectively. The **enantiomer-excess value**, *e.e.*-value, defined as the molar ratio of the excess enantiomer and the total ($e.e._s = \frac{S-R}{S+R}$) provides a quantitative measurement for the enantiopurity. When equal amounts of the two enantiomers are present, the mixture is a **racemate** (*e.e.*-value = 0). The process to separate the enantiomers of a mixture is called resolution. The (equilibrium) conversion of one enantiomer to the other is **racemization**. In diastereomers, the conversion affecting separate chiral centers is called **epimerization**.¹

The importance of chirality

Most biomolecules are chiral. Whereas laboratory chemical syntheses starting with achiral or prochiral reactants, routinely produced racemic product mixtures, the natural biopolymers such as proteins, nucleic acids, and

polysaccharides are (almost exclusively) found in nature to occur in *homochiral* form e.g. they are composed of only *L*-amino acids or *D*-sugars. This is surprising, because the *L*- and *D*-enantiomer building blocks are physically and chemically equal and might be chosen randomly with equal chance. Apparently, in biosynthesis a strong preference for single-enantiomer chemistry has evolved, to the extent that this is considered to be one of the prime characteristics of life on Earth. It is thus of great interest to understand this fundamental phenomenon.

While it is difficult to pinpoint the origin and propagation of homochirality, the advantages appear to be obvious: it enables the organism to collectively utilize single configurational series of molecules. This constitutes an efficient way of using resources since additional machinery to handle the enantiomeric forms is not required. Conversely, the occasional occurrence of “unnatural” configurations, e.g. *D*-amino acids and *L*-sugars, mainly in bacteria, insects and amphibians, can be interpreted as a defense mechanism. In particular, compounds containing *D*-amino acids are disfavored by the (*L*-specific) proteases of predators.

From a practical point of view, chirality has many interesting implications. First, some diseases appear to be related to the chiral composition of biomolecules in our body. As an example, abnormal levels of *D*-amino acids have been observed in Alzheimer’s disease patients.²⁻⁴ Knowledge of the cause and implications might provide additional insight into the mechanisms of these diseases and enable the development of appropriate strategies for therapeutic programmes.^{3;5;6} Secondly, chirality sets a principle for pharmacologic intervention. In fact, the stereochemistry of pharmaceutical compounds is now an important aspect in the screening, design, and development of drugs, supported by the current awareness of the role played by molecular recognition in many pharmacologically relevant events.⁷ When exogenous compounds are introduced into the body, physiological responses show a high degree of chiral discrimination, with the effects of stereoisomers often being markedly different as a consequence of their differential interaction with chiral targets, such as receptors, enzymes and ion channels, all of which are basically homochiral proteins (Table 1).⁸ The classical example is the tragic story of thalidomide in the 1960s. The *R* enantiomer of thalidomide has an anti-morning-sickness effect, whereas the *S* enantiomer is teratogenic for the developing fetus. A highly interesting account of this and related events is given by R. Hoffmann in his book “The same and not the same”.⁹ Since then, the close connection between chirality and biological interactions, and the implications for mankind, has started to be fully recognized. At the same time, the availability of technologies such as asymmetric synthesis, preferential crystallization, chiral chromatography, and chiral membrane made it possible to prepare and analyze pure enantiomers in quantity. From 1990 on, enantiopure drugs started to increase their market share dramatically at the expense of racemates (Figure 2). Worldwide sales of chiral drugs in single-enantiomer dosage forms continued growing at a more than 10% annual rate to \$133 billion in 2000. This figure is expected to reach \$200 billion in 2008.¹⁰ It must be emphasized that the interest of the pharmaceutical industry for chiral drugs or molecules has stimulated chirality research to a large extent.

Table 1. Drug action of enantiomeric compounds

Compounds	Effect of the R enantiomer	Effect of the S enantiomer	Reference
Ibuprofen	inactive	anti-inflammation	11
Chloramphenicol	antibacterial	inactive	12
Thalidomide*	sleep-inducing	teratogenic	13
2-Propyl-4-pentenoic acid	anticonvulsant	teratogenic	14
Propranolol	inactive or sterility-inducing	beta-blocker	15
Penicillamine	highly toxic	anti-chronic arthritis	16

*The recent discovery of its clinical effects in treating various cancers e.g. multiple myeloma adds more mysteries to this molecule.¹⁷

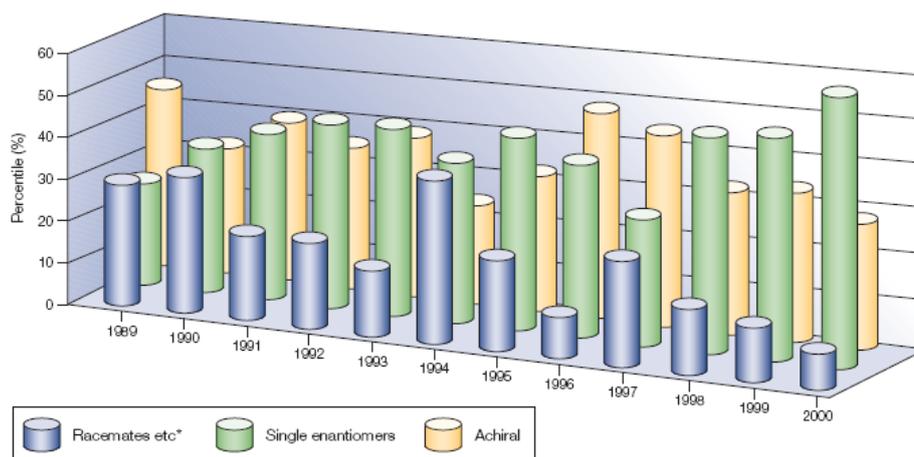


Figure 2. Annual distribution of worldwide approved drugs according to their chiral character (1989–2000).⁷ As shown, the single enantiomer drugs have increased their market share dramatically at the expense of racemic drugs. Figure is taken from ref. 7.

Thirdly, with a better understanding of the nature of chiral interactions, novel drug molecules can be designed. For instance, by intentional insertion of *D*-amino acids into peptides, one can obtain a special potency of a drug,^{18,19} or enhance the stability of the molecule since it will not be readily cleaved by proteases.²⁰⁻²⁴ One successful example is the design of a new type of HIV inhibitor (Figure 3).²² The inhibitor is a small peptide composed exclusively of *D*-amino acids, which binds to glycoprotein gp41 (one of the HIV envelope proteins) and is able to inhibit the gp41-mediated cell–cell fusion and HIV-1 infection. The use of *D*-amino acids for the synthesis of such HIV-inhibitors overcomes the problem of poor endogenous stability which is the case for the *L*-amino acid peptide. This strategy has been recently patented for the design of an oral HIV drug.²⁵ Yet, the potential of *D*-amino acid peptides and proteins has not been fully explored. According to a study of Smith and coworkers in 2003, only 56 distinct entries in the PDB protein database were found to contain *D*-amino acids. The majority (36 out of 56) is pharmaceutically related.²⁶

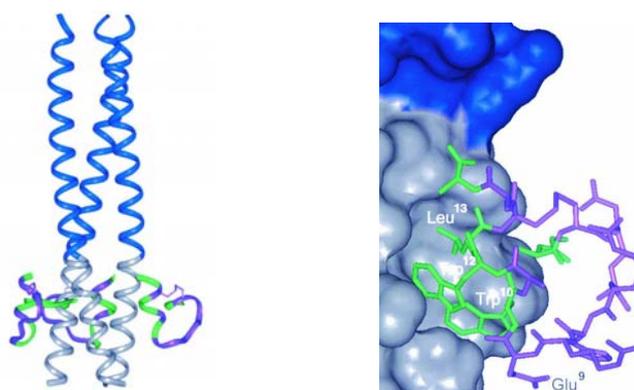


Figure 3. Crystal structure of a *D*-peptide inhibitor bound to the gp41 pocket of HIV envelope glycoprotein complex (adapted from ref.²²). Left: Overall structure of protein-peptide complex (vertical helix) with three peptide inhibitors (horizontal coil at lower part) bound to the hydrophobic pocket. Right: the *D*-amino acid residues (stick) making direct contacts with the protein (molecular surface). The *D*-peptide has enhanced stability compared to the *L*-peptide while retaining tight binding to the target pocket of the HIV protein.

Origin and propagation of homochirality

The evolutionary origin (symmetry-breaking) and accumulation (propagation) of homochirality presents one of the intriguing and possibly fundamental features of the biochemistry of life on Earth. Although it is not possible to

experimentally repeat these early events of life evolution, several propositions have been put forward to rationalize an initial symmetry-breaking event (for a review see Bonner²⁷). Random mechanisms such as spontaneous chiral symmetry breaking in enantiomeric crystallization, and selective adsorption on calcite, rely on chance. Determinate mechanisms, on the other hand, require identification of an intrinsic physical force that perturbs the racemic balance in a specific way. Of these, the electroweak force mechanism derived from the parity violating energy difference, initially proposed by Yamagata²⁸, has obtained much attention (for a review see Quack²⁹). Considering the very small energetic difference between e.g. *L*- and *D*-amino acid enantiomers induced by such forces, subsequent amplification is required. Clearly, the Accumulation Principle in the form originally proposed by Yamagata, where a (parity violating) energy difference is presumed to be operative at each step, fails to account for the formation of homochiral polymers, as has been argued by Bonner.³⁰ Additional selective pressure appears to be required. Several models have been proposed to explain the possible scenarios adapted in the prebiotic world which led to the propagation of homochirality.³¹⁻³⁴

Investigation of the relative stability of homochiral and heterochiral peptides may help to understand the propagation of homochirality in the biosynthesis of natural proteins. At fixed temperature and pressure, the relative stability is given by a (Gibbs) free energy difference that can (in principle) be determined experimentally or calculated using computational methods. The latter approach is described in some details in the first part of this thesis. Molecular simulations are quite suitable to study the relative stability of small peptides. However, the minor difference between the homochiral and heterochiral forms of these peptides, usually of the order of 1 kJ/mol, presents a big challenge since it requires high accuracy of the computational methods. Our study shows that inclusion of entropy into the calculation is of vital importance to obtain reliable results. Moreover, it also predicts that homochiral peptides are not always favored over their heterochiral counterparts under the conditions tested. The preference depends on the size, character, accessible conformational space of the peptide side chains, temperature, and solvent composition etc. It is concluded that more comprehensive investigations are needed to establish the role of dipeptide stability in the propagation of homochirality in detail.

In a wider context, the occurrence of related, yet opposite, manifestations of certain natural phenomena has intrigued mankind since the very beginning of civilization. Taoism, the ancient Chinese philosophy, considers nature and the universe to be organized in Tao (the Way) which generates Two opposite elements, namely Yin and Yang. These two basic elements, through Three (dynamic processes) further produce Four existing forms and finally build up All (the universe). So, following Taoism, one might think of *L* and *D* forms as two opposite extremes (the Yin and Yang of chirality), coexisting in one unity and interacting with each other via a dynamic process (Figure 4). While the protein world is dominated by *L*-amino acids, *D*-sugars of nucleic acids worlds can counteract the extremity and help to keep the whole natural system in balance. This philosophy, even though intuitive instead of quantitative, provides a systematic view helping us to understand nature and the universe.

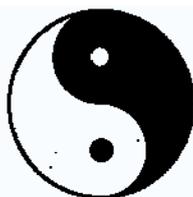


Figure 4. Taoism's systematic interpretation of the universe: Tao, the origin of the universe, generates Two (Yin and Yang), Two generates Three, and Three generates all things. (Tau Teh Ching, Lao Tzu, Chapter 42).

2. ENZYMES AND ENANTIOSELECTIVITY

Enzymes play an important role by catalyzing the biochemical reactions of an organism. The unique characteristics of enzymes are their high (enantio) selectivity and efficiency during catalysis. These properties are equally essential for their performance outside the organism in different industrial sectors. Although, applications of enzymes can be traced back to ancient time, only the past two decades have witnessed a rapidly growing interest in biocatalytic applications of enzymes, due to both scientific progresses and growing industrial recognition of the possibilities, as shown in Table 2.

Table 2. A timeline of scientific and application highlights during the development of biocatalysis (adapted from³⁹)

Timeline	Scientific and technology hallmark	Application Highlight
<i>Application of enzymes in early days</i>		
4000BC		Early recorded biocatalysis application: yoghurt and cheese making (China) and brewing (Mesopotamia)
800BC		The use of calf stomachs (the enzyme chymosin) in the production of cheese described in Homer's "Iliad" and "Odyssey"
1814	Kirchhoff discovered the conversion of starch into sugar by wheat extract	
1830s		Production of dextrin for French bakeries and breweries
<i>Science of enzyme established</i>		
1848	Pasteur discovered enantiomers	
1874		Standardized enzyme preparation by Chris Hansen's Laboratory for cheese making in Denmark
1878	Kühne coined the word "Enzyme"	
1890	"Lock and Key" theory by Fischer	
1900		First application of lipase in non-aqueous solvent ⁴⁰
1900s		Application of enzyme preparation in beer manufacturing at industrial scale
1907		Röhm patented the application of pancreatic extract for tanning
1908		First asymmetric synthesis with an enzyme: stereoselective addition of HCN to aldehyde ⁴¹
1926	Crystallization of urease by Sumner: proof of enzyme as protein ⁴²	
<i>Industrial application of enzymes</i>		
1930s		Preparation of enzyme via fermentation
1949	Invention of enzyme immobilization by Micheel	
1960		Development of penicillin acylase (PA) producing strains and process for industrial application (Bayer & Beecham)
1969		First industrial application of immobilized enzyme for amino acid production (Tanabe Seiyaku Co, Japan)
1970s		Immobilized biocatalyst for industrial application: PA (Lilly's group and Beecham); glucose isomerase (NOVO, DSM)
(1997)		First commercial enzymatic route for the synthesis of beta-lactam antibiotics (DSM)
(1999)		Production of Vitamin C via whole cell biocatalysts
<i>Use of toolbox to understand and manipulate enzymes</i>		
1973	Recombinant plasmid constructed by Cohen & Boyer	
1976	QM/MM computational model by Warshel for theoretical study of enzyme reactions ⁴³	
1980		Production of amylase through genetic engineering (NOVO)
1982		Production of recombinant enzymes at industrial scale (Boehringer Mannheim/Roche)
1986		Biocatalysis in organic solvent well established (Klibanov) ⁴⁴
1988	PCR invented by Mullis	
1990	HTS assay developed for enzyme activity and enantioselectivity ⁴⁵	
1991		Durazyme [®] , a bleach-stable detergent protease developed using computer simulation (NOVO)
1995		Over 200 recombinant enzyme made available (Roche)
1997		Directed evolution of biocatalyst by DNA-shuffling ⁴⁶ ; Kannase [®] , a low-temperature protease produced using directed evolution (NOVO)
2000s		Application of high performance computing in computational biocatalysis
2006	Blue Gene/L (IBM) ready for use for protein folding with peak performance of 360TFlops ⁴⁷	

The enantioselective enzymes often display not only high enantioselectivity in the catalytic conversion of their natural (chiral) substrates, but also show a preference for specific chiral forms, e.g. enantiomers, of non-natural substrates. Therefore, the enantioselectivity of enzymes is of great industrial interest and can be exploited to prepare enantiomerically pure compounds through asymmetric synthesis or kinetic resolution (Figure 5). The latter is more dominantly practiced, especially when other approaches such as resolution by crystallization and chromatographic separation are not feasible. In addition to its enantioselectivity, the ideal industrial enzyme should accept a wide range of substrates, be stable in organic solvents, and don't require cofactors. A good example of a class of enzymes with such features are lipases, which have found wide applications in food, detergent, cosmetics, and pharmaceutical industry at a scale up to hundreds of tons per year (Table 3)³⁵⁻³⁸.

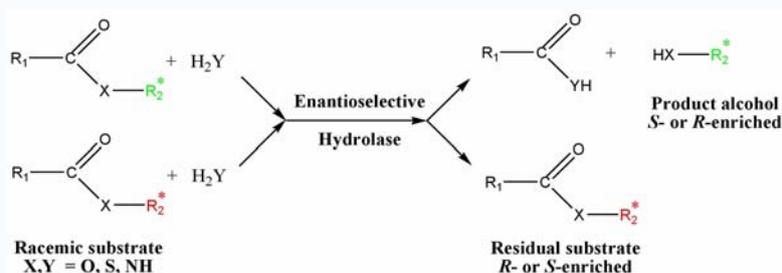


Figure 5. Kinetic resolution of racemic carboxylic acid derivatives with an enantioselective enzyme

Table 3. Examples of industrial applications of lipases

Industry	Application example	Lipase source	Source
Detergent	Detergent additives for fat-containing stains removal	<i>Thermomyces lanuginosa</i>	Novozyme ⁴⁸
Food	Acceleration of ripening cheese and flavor enhancement	<i>Mucor miehei</i>	DSM ⁴⁹
Pharmaceutical	Production of S-methoxyisopropylamine (building block for herbicide)	<i>Burkholderia plantarii</i>	BASF ⁵⁰
Pharmaceutical	Synthesis of Flesinoxan	<i>Candida rugosa</i>	Solvay
Cosmetics	Production of isopropyl myristate (cosmetic component)	<i>Candida antarctica</i> (lipase B)	Novozyme ³⁵
Bioenergy	Production of biodiesel	<i>Pseudomonas fluorescens</i>	Tokyo Univ Agr & Technol ⁵¹
Oleochemical	Production of soaps	<i>Candida cylindracea</i>	Miyoshi Oil & Fat Co
Paper	Control of pitch	<i>Candida rugosa</i>	Novozyme ⁴⁸
Textile	Biopolishing the cotton fabric	<i>Candida antarctica</i> (lipase A)	Novozyme ^{48;52}

In practice, finding an enantioselective, which meets the enantioselectivity requirements for industrial applications, is not straightforward. Early efforts have aimed at experimental approaches, e.g. medium engineering (solvent, water activity), enzyme immobilization and reaction condition optimization (temperature, pH and pressure).^{53;54} In the past several decades, the increasing knowledge of the protein structure and reaction mechanism has stimulated the use of strategies to modify the enzyme using protein engineering (site-directed and random mutagenesis, DNA shuffling) for the improvement of the enzyme enantioselectivity. This process is still rather time-consuming, labor-intensive and random. This prompted us to explore a more rational approach by computer modeling to predict the enantioselectivity of an enzyme (mutant) before the experiment is performed.⁵⁵⁻⁶²

3. COMPUTATIONAL PREDICTION OF ENZYME ENANTIOSELECTIVITY

An introduction to the computational prediction of enzyme enantioselectivity is discussed in some detail in Chapter II. Here, a brief introduction is given of the computational methods and the computational protocol that has been developed.

Enantioselectivity can be quantitatively expressed by the Enantiomeric Ratio, the ratio of the specificity constants for the conversion of *R*- *viz.* *S*-substrate. The Enantiomeric Ratio is also addressed as the *E*-value, or $E_{R/S}$ (Eqn 2).⁶³ In general, the *E*-value is correlated with the free energy difference between the (diastereomeric!) enzyme/*R*-substrate, and enzyme/*S*-substrate combinations at the transition state, $\Delta\Delta G^\ddagger$ (Figure 6).⁶⁴ Using molecular modeling, $\Delta\Delta G^\ddagger$ -values may be calculated, affording an estimate of $E_{R/S}$ for the enzyme-substrate combination (Eqn 4).

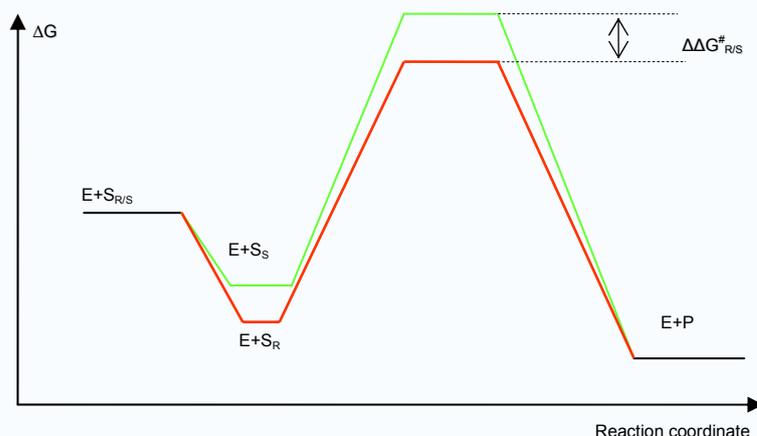


Figure 6. Generalized Gibbs free energy profiles of an enzyme-catalyzed conversion of enantiomeric substrates (S_S , S_R).

$$k_{sp} = k_{cat} / K_m \quad (1)$$

$$E_{R/S} = \frac{k_{sp}^R}{k_{sp}^S} \quad (2)$$

$$k_{sp} = \kappa \cdot \frac{k_B T}{h} \cdot e^{-\frac{\Delta G^\ddagger}{RT}} \quad (3)$$

$$\ln E_{R/S} = \frac{\Delta G_S^\ddagger - \Delta G_R^\ddagger}{RT} = \frac{-\Delta\Delta G_{R-S}^\ddagger}{RT} \quad (4)$$

In these equations k_{sp} is the specificity constant, k_{cat} and K_m are turnover number and Michaelis constant respectively, h is Planck's constant, k_B is the Boltzmann constant, T is the absolute temperature and R is the gas constant. The transmission factor, κ , is included in the Eyring Transition State Theory, TST, to account for non-productive events. Due to its elusive character, it is routinely ignored.

Most of the methods for the computational prediction of enzyme enantioselectivity are based on quantum mechanical, QM, molecular mechanics, MM, or mixed, QM/MM, calculations. In theory, QM calculations are required to describe enzyme-catalyzed reactions where bond breaking and formation occurs. In practice, however, because of the severe computational demands of *ab initio* and semi-empirical methods, simplified models are used. These approximations,

i.e. neglecting atoms outside a defined boundary, neglecting explicit solvent effects, and approximations to the exact solution of the wave function, contribute to the errors involved in the calculation.

MM calculations use an empirical force field obtained by fitting mechanical parameters to a suitable set of quantum-mechanical, statistical-mechanical and/or experimental data. MM calculations are generally much faster than QM calculations. The results, however, can be off when the molecule being computed is not similar enough to the molecules in the database used to parameterize the force field. In addition, the assumption has to be made that the structure of the transition state can be approximated by that of a TS analogue, which can be handled with MM. This approach has been used extensively for molecular simulations of serine-type hydrolases, such as lipases, esterases and proteases, where the transition state can be approximated by the tetrahedral intermediate.^{57;65-68} However, accurate prediction of the enantioselectivity still remains difficult and the predicted *E*-value can be several orders different from the value that is determined by experimental methods. One of the reasons is that in the majority of reports the numerical equivalence of the force field potential energy difference between the diastereomeric transition state models and the Gibbs free energy difference is taken for granted, thus neglecting the entropic contribution. Since the effects of entropy on enzyme enantioselectivity can be substantial, this neglect is unlikely to give an accurate estimate of the free energy difference which determines the enantioselectivity. In the second part of this thesis an approach is developed to improve this situation by calculating the free energy difference in a more direct manner.

QM/MM hybrid approaches use *ab initio* or semi-empirical quantum mechanical techniques to model the active site where bond breaking and formation take place, while classical molecular mechanics is used to model the rest of the molecule in order to improve the computation efficiency. The biggest challenges of a QM/MM calculation are the accuracy of the calculation and the link atom problem. The current *ab initio* QM calculation can bring the accuracy of calculation for free energy difference up to 5 kcal/mol,⁶⁹ but this is still not enough for accurate enantioselectivity prediction since the ΔG involved can easily fall below 5 kcal/mol. When the development of the computational power allows us to include more atoms in the QM calculation, accuracy of the calculation is expected to improve. The link atoms problem refers to the difficulties in choosing the boundary between the QM and MM calculation. At this moment, it is not clear to which extent the choice of link atoms affects the accuracy of the ΔG calculation.⁷⁰

Our first attempt to set up the free energy difference protocol turned out to be unsuccessful. The simulation was performed with Discover (Accelrys, San Diego) software on an SGI O₂ workstation (developed by Silicon Graphics Inc., California). We found that the free energy calculation protocol used in Insight/Discover FDTI (finite difference thermodynamic integration) was not robust and accurate enough to meet our expectations. The hard core problem (see Chapter 3) affected a proper sampling of configuration space. The all-atom CVFF (Consistent Valence Force Field) implemented in Discover module also limited the choice of pathways for the energy perturbation. In addition, the calculating speed of the O₂ processor was too low for efficient testing of the free energy calculation protocol. This prompted us to switch to Gromos96, a biomolecular simulation software package developed by the group of Prof. van Gunsteren at ETH, Zürich. In Gromos96, a soft-core methodology is available which avoids computational singularity, and enhances the sampling accuracy.⁷¹ Besides, we developed an additional protocol for thermodynamic integration (TI) using Gromos96. During the energy perturbation of the TI from *R* to *S*, the united atom force field allowed a pathway by which an improper dihedral at the C atom of one of the residues could be changed from -35.26° (*R*) to 35.26° (*S*). To reduce the strain in the bond angles at intermediate values of the improper dihedral, it turned out to be advantageous to insert an intermediate state in which the minimum energy value of the improper dihedral is set to 0° and the minimum energy values of the three bond angles around the C atom are set to 120°. This pathway, which was

tested first in our study on peptides and proved to be able to give accurate estimates of the (small) free energy differences,⁷² has been subsequently applied for enzyme modeling studies.

4. HIGH PERFORMANCE COMPUTING

Although, the successful prediction of enzyme enantioselectivity reported in this thesis gives an appealing result, a statistically meaningful validation of the method cannot yet be given, due to the limitations set by computational resources at our disposal. Because of their size, biomolecules like enzymes are computationally demanding. For instance, it takes about two months to simulate 1ms of protein folding by molecular dynamics on a machine with 20 TFlops (Tetra Floating Point Operations Per Second, a measurement of the speed of computer), a speed of the 13th fastest computer on the Top500 in June 2006, not to mention the high level calculations, i.e. quantum mechanics calculations. As a result, compromises are needed between the accuracy of the modeling and the computational expenses. Fortunately, with the fast development of high performance computing (HPC) and algorithms for molecular simulations, the computer simulation of enzymes is getting faster and faster. HPC is a branch of computer science that concentrates on developing supercomputers and software to run on supercomputers. An important area of this discipline is the development of parallel processing algorithms and software: programs that can divide the tasks into little pieces so that each piece can be executed simultaneously by separate processors. This parallelization methodology seems to be very suitable in dealing with complex system such as biomolecules. Indeed, recent years have seen many examples of HPC in life sciences, some of which are listed in Table 4.

Table 4: Some examples of HPC applications in biology

Research theme	Institute	Computers	Peak Performance (TFlops)
Sequencing and development of molecular models of proteins in biological agents (2002)	University at Buffalo, USA	2,000-node cluster composed of Dell PowerEdge servers	2
Docking in drug design (2003)	Novartis AG, Basel, Switzerland	Grid-computing, 2,700 P-4 processor PCs within its R&D organization	5
Molecular modelling of proteins (2003)	Virginia Tech, USA	1,100 Apple G5, each node dual 64-bit PowerPC 2GHz processors.	12
Binding site study in smallpox vaccine research (2003)	Edward Jenner Inst for Vaccine Research and Oxford University, UK	Grid-computing, 2.5 million PC in >200 countries	100
Protein folding	IBM Life Science, USA	Blue Gene/L	360
Protein folding	Riken, Japan	MDGRAPE-3 (Protein Explorer), upgraded and ready for use in 2006	1000

In the past, HPC was primarily the domain of supercomputers — dedicated, specialized number-crunching machines housed in a special environment, which makes it expensive to maintain. The supercomputers had limited applications and were mainly used by engineers for production technical computing and by scientists for research. In the past five years, Linux-based PC clusters emerged and started to play an increasingly important role in the arena of HPC because of their high performance-cost ratio (significant cost reduction up to 80% compared to supercomputers), scalability and low cost of maintenance, which makes these computational tools accessible to a wide range of researchers.⁷³

Great efforts have been made to improve both the hardware and software in HPC. The hardware improvement focuses on more advanced microprocessor CPU design, faster memory and interconnects. The Blue Gene built by IBM has a simplified RISC (Reduced Instruction Set Computing) with an “instruction set” of only 57 compared to 256 for most RISC machines. It also places memory for storing data on the same chip as the microprocessor. This enables the 1 million-CPU machine to operate at about 200 TFlops. Blue Gene’s speed target is one Pflops (1,000 TFlops) which will allow the molecular mechanics simulation of 1 millisecond of protein folding within one day of computer time. A similar design philosophy is now being used for desktop computers arranged in PC clusters. Dozens of parallel processors are packed onto each chip to achieve a massive data parallelism. This can improve the performance of the chip up to 25 GFlops resulting in simulations within Gromacs that are much faster than on conventional chips.⁷⁴

By comparison, methodology for the development of parallel programs is far from mature, with automatic parallelization compilers and languages still quite limited. This is especially true for CPU-intensive applications such as molecular modeling which needs sequential calculations. One of the problems is that not all the programs available are able to recognize flawed processors during the simulation run in parallel mode and subsequently re-route the data. An experiment showed that when the clusters were scaled up from 300 to over 1,000 machines, the job-scheduling program sometimes objected to deal with any dead or dysfunctional machines and refused to schedule jobs at all.⁷³ Ideal algorithms should be able to split the computational jobs to the utmost so that these jobs can be executed in parallel modes efficiently. This, however, remains difficult. Even one of the fastest biomolecular simulation programs, Gromacs, reaches its limits at thirty-two nodes and cannot be further scaled up efficiently, partly due to communication problems of the hardware and partly due to the inefficiency of the software to handle these problems. New algorithms e.g. neutral territory methods as well as programs like NAMD for biomolecular simulations may improve the efficiency of HPC, especially on large parallel computers.^{74,75}

5. OUTLINE OF THIS THESIS

The goal of this thesis research work has been to establish a computational platform for the computational prediction of enzyme enantioselectivity. Emphasis has been placed on (Gibbs) free energy calculations employing MM/MD methods to achieve a good tradeoff between computational efficiency and accuracy. The thesis consists of three parts: the first part (Chapter 2) describes an overview of the computational tools currently used in the prediction of enzyme enantioselectivity; the second part (Chapters 3 and 4) focuses on the validation of the free energy calculation protocol applied to small peptides. At the same time, the propagation of homochirality, as a topic of great fundamental research interest, is addressed; the third part (Chapter 5) describes in details the application of the free energy calculation methods to calculate free energy differences that can be compared with those derived from experimental data. In addition, the *E*-value of a mutant enzyme is predicted.

Chapter 2 starts with a historical overview of computational predictions of enzyme enantioselectivity. The concept and theory of enzyme enantioselectivity and its dependence on several parameters is introduced and discussed. Different strategies and methods for enzyme enantioselectivity calculations are compared. Finally several free energy calculation methods are discussed.

In **Chapter 3**, dialanine peptide is used as a (small) model molecule to investigate the relative stability of homochiral and heterochiral dialanine. The free energy calculation method and the thermodynamic integration protocol are compared with experimental data obtained from peptide syntheses.

In **Chapter 4**, the methods and protocols for free energy calculation described in Chapter 3 are tested further on additional dipeptides, tripeptides and a pentapeptide. The influences of size, character, accessible conformational space of the peptides, temperature, and solvent composition on the free energy difference between the homochiral and heterochiral peptides are investigated.

In **Chapter 5**, the free energy calculation method, validated by the results reported in Chapters 3 and 4, is applied to the lipase B from *Candida antarctica*, CaLB, for the prediction of enantioselectivity. Two model systems, the truncated tetrahedral intermediate (TTI) model, and the whole enzyme model, are tested in terms of accuracy and efficiency. The TTI model, which is simple and computationally less demanding, serves as an efficient test system to probe the enantioselectivity in the enzymatic catalysis. Five different substrates, all esters of (chiral) secondary alcohols, are studied both with the TTI model and with the whole enzyme model. In addition, a blind test is performed on a Trp104Ala mutant of CaLB for which the enantioselectivity was not known at the time of the experiment.

Reference

- (1) Commission on the Nomenclature of Organic Chemistry; Blackwell Scientific Publications: 1993, pp 149-154.
- (2) Fisher, G. H.; Daniello, A.; Vetere, A.; Padula, L.; Cusano, G. P.; Man, E. H. *Brain Res Bull* 1991, 26, 983-985.
- (3) Fuchs, S. A.; Berger, R.; Klomp, L. W. J.; de Koning, T. J. *Mol Genet Metab* 2005, 85, 168-180.
- (4) Fisher, G.; Lorenzo, N.; Abe, H.; Fujita, E.; Frey, W. H.; Emory, C.; Di Fiore, M. M.; D' Aniello, A. *Amino Acids* 1998, 15, 263-269.
- (5) Blanch, E. W.; Morozova-Roche, L. A.; Cochran, D. A. E.; Doig, A. J.; Hecht, L.; Barron, L. D. *J Mol Biol* 2000, 301, 553-563.
- (6) Stott Kelvin UK Patent WO 0107474 2001.
- (7) Agranat, I.; Caner, H.; Caldwell, A. *Nature Rev Drug Dis* 2002, 1, 753-768.
- (8) Triggle, D. J. *Drug Dis Tod* 1997, 2, 138-147.
- (9) Roald Hoffmann. *The Same and Not the Same*, Columbia University Press: 1997.
- (10) Stinson, S. C. *Chem Eng News* 2001, 79, 79-+.
- (11) Adams, S. S.; Bresloff, P.; Mason, C. G. *J Pharm Pharmacol* 1976, 28, 256-257.
- (12) Drayer, D. E. *Clin Pharmacol & Ther* 1986, 40, 125-133.
- (13) Fabro, S.; Smith, R. L.; Williams, R. T. *Nature* 1967, 215, 296-.
- (14) Hauck, R. S.; Nau, H.; Elmazar, M. M. A. *Naturwissenschaften* 1991, 78, 272-274.
- (15) Patil, P. N.; Miller, D. D.; Trendelenburg, U. *Pharmacol Rev* 1974, 26, 323-392.
- (16) Kean, W. F.; Dwosh, I. L.; Anastassiades, T. P.; Ford, P. M.; Kelly, H. G. *Arthritis Rheum* 1980, 23, 158-164.
- (17) Garcia-Sanz, R. *Expert Opin Pharmacother* 2006, 7, 195-213.
- (18) Shai, Y.; Oren, Z. *J Biol Chem* 1996, 271, 7305-7308.
- (19) Yanaihara, Noboru, Pponda, Seiji, and Kuzuha, Noboru Patent JP1992-331327 1994.
- (20) Kuwada, M.; Teramoto, T.; Kumagaye, K. Y.; Nakajima, K.; Watanabe, T.; Kawai, T.; Kawakami, Y.; Niidome, T.; Sawada, K.; Nishizawa, Y.; Katayama, K. *Mol Pharmacol* 1994, 46, 587-593.

- (21) Castillo Gerardo M. US Patent WO 2005060683, 2005.
- (22) Eckert, D. M.; Malashkevich, V. N.; Hong, L. H.; Carr, P. A.; Kim, P. S. *Cell* 1999, 99, 103-115.
- (23) Soto, C.; Kindy, M. S.; Baumann, M.; Frangione, B. *Biochem Biophys Res Commun* 1996, 226, 672-680.
- (24) Chalifour, Robert, Hebert, Lise, Kong, Xianqi, and Gervais, Francine. Patent US2002094335 2001.
- (25) Eckert D. M., Chan D.C., Malashkevich V., Carr P. A., and Kim P. S. Patent US 6841657 2005.
- (26) Mitchell, J. B. O.; Smith, J. *Proteins* 2003, 50, 563-571.
- (27) Bonner, W. A. *D-amino acids in sequences of secreted peptides of multicellular organisms*; Birkhäuser: Basel, 1998, pp 159-188.
- (28) Yamagata, Y. *J Theor Biol* 1966, 11, 495-498.
- (29) Quack, M. *Angew Chem-Int Edit* 2002, 41, 4618-4630.
- (30) Bonner, W. A. *Origins Life Evol B* 1999, 29, 615-623.
- (31) Popa, R. *J Mol Evol* 1997, 44, 121-127.
- (32) Thiemann, W. H. P.; Rosenbauer, H.; Meierhenrich, U. J. *Space Life Sciences: Life in the Solar System: Prebiotic Chemistry, Chirality Space Biol* 2001, 27, 323-328.
- (33) Munegumi, T.; Shimoyama, A. *Chirality* 2003, 15, S108-S115.
- (34) Mikami, K.; Yamanaka, M. *Chem Rev* 2003, 103, 3369-3400.
- (35) Houde, A.; Kademi, A.; Leblanc, D. *Appl Biochem Biotechnol* 2004, 118, 155-170.
- (36) Ghanem, A.; Boul-Enein, H. Y. *Chirality* 2005, 17, 1-15.
- (37) Anderson, E. M.; Karin, M.; Kirk, O. *Biocataly Biotransfor* 1998, 16, 181-204.
- (38) van Rantwijk, F.; Sheldon, R. A. *Tetrahedron* 2004, 60, 501-519.
- (39) Bornscheuer, U. T.; Buchholz, K. *Eng Life Sci* 2005, 5, 309-323.
- (40) Kastle J.H.; Loevenhart A.S. *Am Chem J* 1900, 24, 491-525.
- (41) Van Duin, A. T.; Larter, S. R. *Organic Geochem* 1998, 29, 1043-1050.
- (42) Sumner, J. B. *J Biol Chem* 1926, 69, 435-441.
- (43) Warshel, A.; Levitt, M. *J Mol Biol* 1976, 103, 227-249.
- (44) Klibanov, A. M. *Chemtech* 1986, 16, 354-359.
- (45) Reetz, M. T.; Becker, M. H.; Klein, H. W.; Stockigt, D. *Angew Chemie-Int Edit* 999, 38, 1758-1761.
- (46) Reetz, M. T.; Zonta, A.; Schimossek, K.; Liebeton, K.; Jaeger, K. E. *Angew Chem-Int Edit* 1997, 36, 2830-2832.
- (47) Suits, F.; Pitman, M. C.; Pitera, J. W.; Swope, W. C.; Germain, R. S. *IBM J Res Develop* 2005, 49, 475-487.
- (48) Faber, K.; Kroutil, W. *Curr Opin Chem Biol* 2005, 9, 181-187.
- (49) Dairy Ingredients Products: Piccantase. 2006. (http://www.dsm.com/en_US/html/dfs/dairy-products-enzymes-piccantase.htm)
- (50) Schmid, A.; Dordick, J. S.; Hauer, B.; Kiener, A.; Wubbolts, M.; Witholt, B. *Nature* 2001, 409, 258-268.
- (51) Iso, M.; Chen, B. X.; Eguchi, M.; Kudo, T.; Shrestha, S. *J Mol Cataly B-Enzym* 2001, 16, 53-58.
- (52) Khoddami, A.; Morshed, M.; Tavanai, H. *Iranian Poly J* 2001, 10, 363-370.

- (53) Lutz, S. *Tetrahedron-Asymmetr* 2004, 15, 2743-2748.
- (54) Berglund, P. *Biomol Eng* 2001, 18, 13-22.
- (55) Uppenberg, J.; Ohrner, N.; Norin, M.; Hult, K.; Kleywegt, G. J.; Patkar, S.; Waagen, V.; Anthonsen, T.; Jones, T. A. *Biochemistry* 1995, 34, 16838-16851.
- (56) Orrenius, C.; Haeffner, F.; Rotticci, D.; Ohrner, N.; Norin, T.; Hult, K. *Biocataly Biotransfor* 1998, 16, 1-15.
- (57) Haffner, F.; Norin, T.; Hult, K. *Biophys J* 1998, 74, 1251-1262.
- (58) Magnusson, A. O.; Takwa, M.; Harnberg, A.; Hult, K. *Angew Chemie-Int Edit* 2005, 44, 4582-4585.
- (59) Magnusson, A. O.; Rotticci-Mulder, J. C.; Santagostino, A.; Hult, K. *Chembiochem* 2005, 6, 1051-1056.
- (60) Zhang, N. Y.; Suen, W. C.; Windsor, W.; Xiao, L.; Madison, V.; Zaks, A. *Protein Eng* 2003, 16, 599-605.
- (61) Patkar, S.; Vind, J.; Kelstrup, E.; Christensen, M. W.; Svendsen, A.; Borch, K.; Kirk, O. *Chem Phys Lipids* 1998, 93, 95-101.
- (62) Suen, W. C.; Zhang, N. Y.; Xiao, L.; Madison, V.; Zaks, A. *Protein Eng Design & Select* 2004, 17, 133-140.
- (63) Chen, C. S.; Fujimoto, Y.; Girdaukas, G.; Sih, C. J. *J Am Chem Soc* 1982, 104, 7294-7299.
- (64) Overbeeke, P. A.; Orrenius, S. C.; Jongejan, J. A.; Duine, J. A. *Chem Phys Lipids* 1998, 93, 81-93.
- (65) Wipff, G.; Dearing, A.; Weiner, P. K.; Blaney, J. M.; Kollman, P. A. *J Am Chem Soc* 1983, 105, 997-1005.
- (66) Ohagan, D.; Rzepa, H. S. *J Chem Soc-Perk T2* 1994, 3-4.
- (67) Colombo, G.; Toba, S.; Merz, K. M. *J Am Chem Soc* 1999, 121, 3486-3493.
- (68) Micaelo, N. M.; Teixeira, V. H.; Baptista, A. M.; Soares, C. M. *Biophys J* 2005, 89, 999-1008.
- (69) Kollman, P. A.; Kuhn, B.; Perakyla, M. *J Phys Chem B* 2002, 106, 1537-1542.
- (70) Reuter, N.; Dejaegere, A.; Maignret, B.; Karplus, M. *J Phys Chem A* 2000, 104, 1720-1735.
- (71) Beutler, T. C.; Mark, A. E.; Van Schaik, R. C.; Gerber, P. R.; Van Gunsteren, W. F. *Chem Phys Lett* 1994, 222, 529-539.
- (72) Zhou, Y.; Oostenbrink, C.; Van Gunsteren, W. F.; Hagen, W. R.; De Leeuw, S. W.; Jongejan, J. A. *Mol Phys* 2005, 103, 1961-1969.
- (73) Okur, A.; Strockbine, B.; Hornak, V.; Simmerling, C. *J Comput Chem* 2003, 24, 21-31.
- (74) Wilson, E. K. *Chem Eng News* 2004, 82(39), 35-+.
- (75) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *J Comput Chem* 2005, 26, 1781-1802.

COMPUTATIONAL PREDICTION OF ENZYME ENANTIOSELECTIVITY

SUMMARY

Enantioselectivity is one of the characteristic features of enzymatic catalysis. As a result of the chiral architecture of the protein scaffold, virtually all reactions of (pro)chiral substrates, whether of physiological importance or of non-natural origin, are catalyzed with a (high) degree of preference for one of the enantiomeric configurations. This preference is interesting not only from a fundamental point of view (since it is strongly linked to the catalytic propensity of enzymes) but also very appealing to many industrial applications e.g. in enantiomer resolution and asymmetric synthesis. Consequently, a substantial amount of research has been devoted to the experimental determination of enzyme enantioselectivity¹. With the currently available knowledge of enzyme structures and the improved computing power, molecular modeling has also become a useful tool to study enzyme enantioselectivity. Especially in the ongoing search for optimized enzyme enantioselectivity, i.e. via protein engineering, the results of computational methods can be helpful for the rational design of experimental strategies. It should be emphasized, however, that useful predictions of enzyme enantioselectivity require the accuracy of a calculation to be of the order of 1-5 kJ/mol, which is a challenging target for computational methods. In this chapter, an overview is presented of historical developments and of the current status of computational predictions of enzyme enantioselectivity.

1. HISTORICAL HIGHLIGHTS

Initial attempts to gain insight into the differential binding of substrates to enzymes by computational methods were described by the group of Scheraga in the early 1970ies.^{2,3} They studied the stability of noncovalent Michaelis complexes in order to estimate the selectivity of α -chymotrypsin for small peptides using molecular mechanics methods. Less than ten years later, publications by DeTar and the group of Kollman already reported results based on a computational strategy that set the stage for most of the work that has been carried out since. In particular, they showed that computation of the (steric) energy difference of the covalently-bound tetrahedral intermediates, generated from the substrate enantiomers and the active site serine of the enzyme (α -chymotrypsin), was more suitable to describe the enantioselectivity of the enzyme than the noncovalent models.^{4,5} Another ten years later, this approach was further improved by Norin *et al.* who included both steric and electrostatic effects, using molecular dynamics (MD) in combination with simulated annealing techniques in order to locate the minimum energy conformations of the respective tetrahedral intermediates.⁶ Meanwhile, the importance of electrostatic interaction in enzymatic catalysis in general has been well documented.⁷ Curiously, using semi-empirical quantum mechanical calculations, O'Hagan proposed that electrostatic effects by themselves, i.e. in the absence of the catalytic site cavity, might be sufficient to explain preferences in enantioselective enzymatic catalysis.⁸ A possible rationale will be discussed in Chapter 5. Later developments comprised several approaches (see Orrenius *et al.* for an overview⁹). One strategy has been to focus on the calculation of the potential energy of the respective TIs, using MD to generate starting structures for subsequent minimization, counting contributions of selected residue subsets within the modeled system.¹⁰ In the case of *Candida*

This chapter is based on the review article to be published in Biocatalysis and Biotransformation

antarctica lipase (CalB), a so-called ‘function-based’ subset appeared to give a better prediction of the experimentally determined enantioselectivity than structure-based or energy-based subsets.¹¹ Other approaches have attempted to include entropic effects and focused on free energy calculation. Different free energy calculation methods have been applied, ranging from free energy perturbation to thermodynamic integration. Already in 1994, attempts were described to model the binding of the enantiomeric substrates to three lipases in implicit solvent, by calculating the free energy difference using a thermodynamic integration protocol.¹² However, the calculation was reported to fail due to convergence problems.¹³ In a recent study by Micaelo and coworkers on the enantioselectivity of cutinase, on the other hand, a thermodynamic integration protocol could be successfully applied for the calculation of the free energy differences. The simulations were run with the enzyme placed in explicit solvent. A soft-core interaction was included in the calculation, in order to avoid singularity problems.¹⁴ The enantioselectivity of subtilisin has been investigated by the group of Merz. They applied a (semi-empirical) QM/MM calculation to account for the developing charge distribution during the free energy calculation.¹⁵ The free energy was perturbed in the slow growth method and the charge calculated by QM was periodically updated during the energy perturbation in MD. Encouragingly, it was found that the calculated results were qualitatively in agreement with the experimental values. However, they had to exclude explicit solvent effects from the QM region of the computational cost.¹¹ In summary, it appears that qualitative predictions of the enantiopreference are now quite feasible for those enzymes that allow the TS to be approximated by a covalently-bound tetrahedral intermediate. Quantitative predictions of the enantioselectivity, however, can still be off by one or more order(s) of magnitude. As it has been well established that the entropic contribution to the enantioselectivity needs to be taken into account,¹⁶⁻¹⁹ future developments will almost certainly have to rely on more sophisticated and efficient methods for free energy calculations.

2. DEFINITION AND ORIGIN OF ENANTIOSELECTIVITY

For an enzyme that follows the Minimal Michaelis-Menten, 3M, kinetic model the traditional formulation of the rate equation (Eqn 1) for the conversion of substrate, S , by enzyme, E , can be rewritten in the format of a second order rate equation (Eqn 2) without loss of generality. Clearly, in practical cases, where the experimenter controls only the total concentration of enzyme species, the concentration of free enzyme will be a function of the substrate (and product) concentration.

$$r = \frac{k_{cat} \cdot S}{K_M + S} \cdot E_{total} \quad (1)$$

$$r = \frac{k_{cat}}{K_M} \cdot S \cdot E_{free} \quad (2)$$

The apparent (or: pseudo) second-order rate constant, k_{cat} / K_M , is addressed as the *specificity constant*, k_{sp} , of that substrate in the enzymatic reaction.²⁰ For 3M-enzymes that catalyse the conversion of two or more substrates in a competitive way, the rates can still be formulated as in Eqn 2, albeit that the free enzyme concentration will now be a function of all substrates and products. Clearly, this will be of no concern when one describes the ratio of rates for the enzymatic conversion of two enantiomers, R and S , in a single reaction mix (Eqn 3).

$$\frac{r_R}{r_S} = \frac{(k_{cat} / K_M)_R}{(k_{cat} / K_M)_S} \cdot \frac{R}{S} = \frac{k_{sp}^R}{k_{sp}^S} \cdot \frac{R}{S} = E \cdot \frac{R}{S} \quad (3)$$

Accordingly, the ratio of specificity constants is the parameter of choice to express the relative rates of competing enzymatic reactions²¹. The first application of specificity constants to characterize the enantioselectivity of enzymes as the ratio $(k_{cat}/K_M)_R/(k_{cat}/K_M)_S$ was reported for the conversion of *L*- and *D*-amino acid derivatives by α -chymotrypsin.²¹ Since the work of Chen *et al*²² ratios of specificity constants for enantiomers are commonly expressed as the enantiomeric ratio *E* (or *E*-value, to avoid confusion with the symbol for enzyme). By definition, the *E*-value is an intrinsic property of the enzyme for a specified couple of substrates.

E-values can be determined experimentally in number of ways.¹ In the context of this thesis, the accuracy of these methods is a matter of concern.²³ It would appear that explicit determination of the catalytic constant and the Michaelis constant for each enantiomer would provide unbiased results. This, however is certainly not the case, since the chiral purity of available enantiomer preparations does not normally exceed 99.5% (which may still leave up to 0.5% of the highly reactive substrate). In this respect, an implicit method turns out to (slightly) more robust. Integration of Eqn 3 leads to:

$$\frac{-dR/dt}{-dS/dt} = E \cdot \frac{R}{S} \Rightarrow E = \frac{\ln\{(1-\xi)(1-ee_s)\}}{\ln\{(1-\xi)(1+ee_s)\}} \quad (4)$$

with $ee_s = \frac{S-R}{S+R}$, the enantiomer-excess value of the remaining substrate, and $\xi = 1 - \frac{S+R}{S_0+R_0}$, the extent of conversion.

Provided that the assumptions underlying Eqn 2 hold, and sophisticated analytical tools are available, plots of ee_s vs ξ allow determination of the *E*-value with an accuracy that is roughly $\pm 20\%$. As assumed, Eqn 2 holds for a 3M kinetic scheme. For the enzymes discussed in this thesis, i.e. lipases, a more complicated bi bi ping pong kinetic scheme applies, leading to specificity constants comprising a number of microscopic kinetic constants.²⁴

The enantiomeric ratio *E* can be written in exponential format by invoking the relationship between the kinetic constants and the (Gibbs) free energy difference between the ground state and the activated complex, according to the Eyring Transition State Theory, TST.²⁵ Since the specificity constants are lumped parameters of the rate equations that involve microscopic constants representing chemical (bond breaking/formation) and physical (diffusion, adsorption) processes, their identification with a single “rate-determining” step is not straightforward. Proper analysis shows that they can be related to the exponentially weighted average of the free energy barriers measured from the ground-state level^{26,27}. With this restriction in mind the *E*-value can be correlated to the free energy difference between the *R* and *S* form of enzyme-substrate complex at the transition state $\Delta\Delta G^\ddagger$ (Eqns. 5-8).

$$E_{R/S} = \frac{k_{sp}^R}{k_{sp}^S} \quad (5)$$

$$k_{sp} = \kappa \cdot \frac{k_B T}{h} \cdot e^{-\frac{\Delta G^\ddagger}{RT}} \quad (6)$$

$$\ln E_{R/S} = \frac{\Delta G_R^\ddagger - \Delta G_S^\ddagger}{RT} = \frac{-\Delta\Delta G_{R/S}^\ddagger}{RT} \quad (7)$$

$$k_{sp} = k_{cat}/K_m \quad (8)$$

In the Eyring TST equation (thermodynamic format) k_{sp} is the specificity constant as defined in Eqn. 8, k_{cat} and K_m are turnover number and apparent dissociation constant in enzymatic reaction respectively, h is Planck’s constant, k_B is Boltzmann’s constant, T is the absolute temperature and R is the gas constant.

For enzymes following bi bi ping pong kinetics, with the first substrate harbouring chirality in a single part of the molecule, i.e. esters of a chiral alcohol, and the second substrate non-chiral, i.e. water (in hydrolysis) or a neutral ester (in transesterification) only the free energy barriers occurring in one stage of the catalytic scheme have to be accounted for. An example of such a situation is given in Figure 1 for an enzyme (i.e. a serine-type hydrolase) involving (a) tetrahedral intermediate state(s).

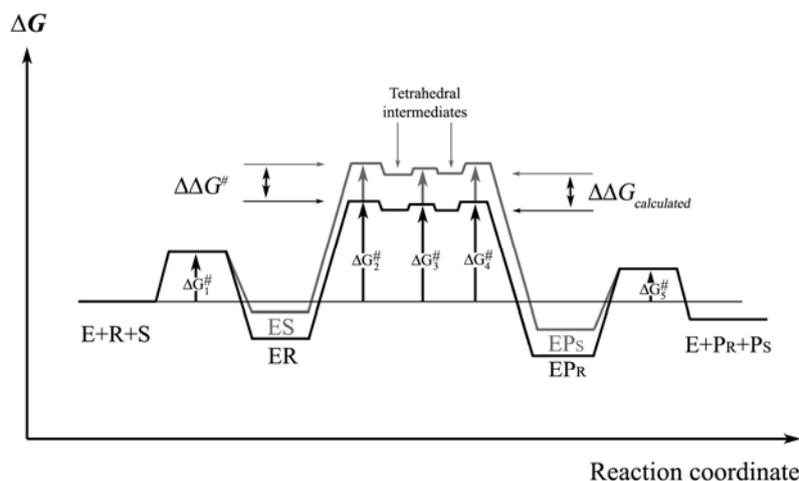


Figure 1. The schematic illustration of free energy profile of enzyme-catalyzed reaction for two enantiomers

Knowledge about the enzyme structure and reaction mechanism helps us to understand the origin of enzyme enantioselectivity. The bi bi ping pong scheme for a lipase-catalyzed transesterification is shown in Figure 2. One of the substrates ester R_1COOR_2 comes in as the acyl donor generating the acyl enzyme, which further reacts with the second substrate alcohol R_3OH . There are supposedly five transition states in each stage of the scheme. Considering the relative barrier heights, it is commonly assumed that only the three transition states flanking the tetrahedral intermediates contribute significantly to the Gibbs free energy difference involved in the chiral discrimination.²⁸ This assumption, together with the neglect of the transmission factors (believed to cancel out in the relation for the ratio), adds another factor to the overall (in)accuracy. By way of the electrostatic preorganization in the active site cavity, the enzyme is able to stabilize the tetrahedral intermediates and the transition states by hydrogen bonding to the negative charge on the oxygen atom by accomodating this in the so-called oxyanion hole of the enzyme.²⁹ When the stabilization of the *R* and *S* substrate is of a different magnitude, enantioselectivity results (Figure 3).

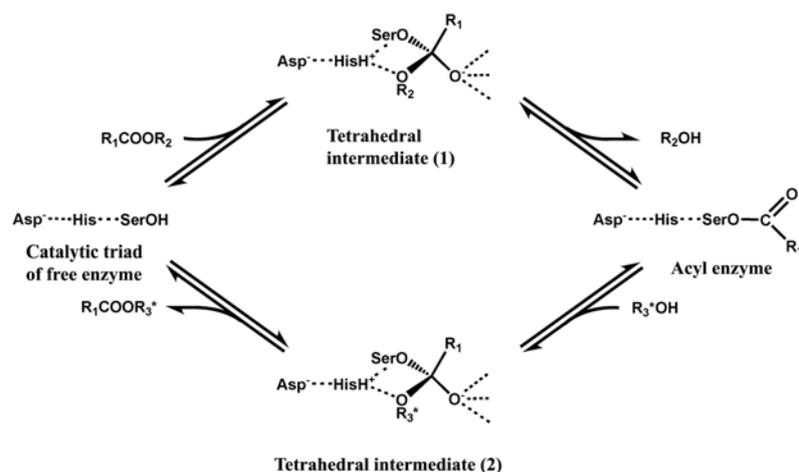


Figure 2. The catalytic mechanism of lipase-catalyzed transesterification (R_3 =alkyl) or hydrolysis (R_3 =hydrogen) of an ester R_1COOR_2 , using a chiral alcohol R_3OH .

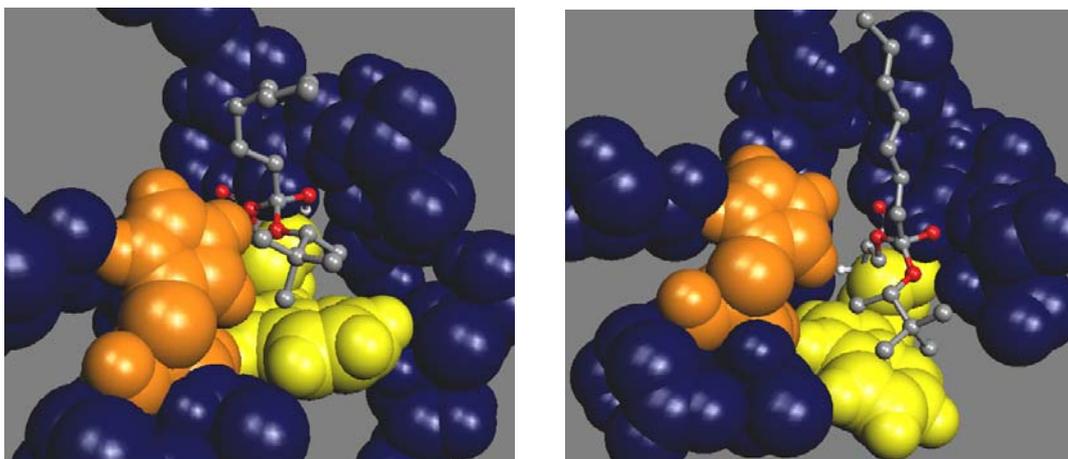


Figure 3: The substrate-enzyme complex at the active site of CaLB. The *R* (left) and *S* (right) form of the substrate 3, 3-dimethyl-2-butanol connected to the acyl chain and Ser105 (ball and stick) fit differently into the enantioselectivity pocket (light yellow, space and filling) of the enzyme. The *S* conformer adapts itself to avoid steric clashes on the expense of impaired hydrogen bond formation with His224 which results in the enantiopreference of the enzyme towards the *R* enantiomer.

3. PARAMETERS INFLUENCING ENANTIOSELECTIVITY

The interactions between enzyme, substrate and solvent in the reaction system determine the enantioselectivity. The influence of other parameters like temperature and pressure etc is realized through these interactions. It is important to understand their effects when experiments or computer simulations are designed to modify or improve the enzymatic enantioselectivity. Here the main contributing parameters for the enantioselectivity are discussed with the focus on lipases. More details can be found in the review literature.³⁰

- Enzyme

There are two groups of residues of the enzyme which influence the enantioselectivity: 1). The catalytic triad in the active site which is responsible for the catalysis and those near the active site which accommodate the substrates or help to stabilize the substrates-enzyme binding 2). Those far away from the active site which influence the conformation of the enzyme and thus change the enantioselectivity indirectly. For instance, some of these residues can control the degree of exposure of the enzyme active site to the outside solvent or substrates. Genetic engineering can help to elucidate the effects of these residues on the enzyme enantioselectivity.³¹⁻³³ Jaeger et al. has shown by the directed evolution of lipase that mutating several residues on the surface-exposed loops into glycine improved the enantioselectivity from $E=1.1$ to $E=25.8$, due to the enhanced flexibility of the loop which changed the conformation of the enzyme from a closed one into an open one.³³ However, if the active site is deeply buried inside the enzyme, the enantioselectivity is then less sensitive to the change of these residues.

Several approaches to improve the lipase enantioselectivity were reviewed by Haas et al.³² In case the enzyme structure and mechanism are not clearly known, chemical modification, enzyme immobilization and random mutagenesis might be employed. The directed evolution of lipase is a powerful and practical strategy which is gaining popularity nowadays especially in industry. It can bring the enantioselectivity of the enzyme from low value ($E = 1$) to moderate value ($E = 10-30$).^{33,34} If we have sufficient information about the enzyme structure and mechanism, a rational approach can be applied which focuses on the residues around the active site of enzyme. In this case, the computer modeling is often used to give a prediction or explanation before and after the site-directed mutagenesis is carried out.^{35,36}

- Substrate

The structure of the substrates has substantial effects on the enantioselectivity of the enzyme. In the case of ester substrate, the influence comes from the alcohol moiety (containing the large and medium group) and the acyl moiety of the substrates. Their steric and electrostatic interactions with the enzyme contribute to differential enthalpy while the differences in the degrees of freedom of the enzyme-substrate complexes give rise to differential entropy.

The enantioselectivity of lipase has different sensitivity to the change of alcohol moiety and acyl chain. In the case of CalB, the enzyme seems more sensitive to the change of alcohol group. When changing 2-propanol to 2-pentanol, the *E* value increases from *E* = 9 to *E* = 390 (Table 1).

The influence of the alcohol moiety on enantioselectivity sometimes complies with the Kazlauskas rule (explained in Section 4). The enantioselectivity increases as the difference between the middle-sized and large-sized group of alcohol becomes bigger, as can be seen from the left and middle panel of Table 1 where the comparisons are made with varying the size of large-sized group and middle-sized group respectively. The presence of halogen atoms in the alcohol moiety complicates the situation (right panel of Table 1). In the case of short aliphatic chain in the alcohol moiety, the *E* value increases dramatically with the halogenated alcohols. This is out of the expectation of Kazlauskas rule since the size of halogen atoms is not much bigger than the corresponding methyl group in other substrate. In the case of a long aliphatic chain, not only the *E* value decreases, but also there is inversion of enantioselectivity for halogenated alcohols, which can not be predicted by Kazlauskas rule at all. Repulsive forces due to the non-steric interactions between enzyme residues in the enantioselectivity pocket and the halogen atom pointing into the pocket have been proposed to explain the uncommon enantioselectivity for halogenated alcohols.³⁷

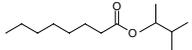
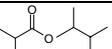
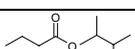
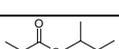
Compared to the alcohol moiety, the acyl moiety does not have a strong effect on the enantioselectivity. The change of acyl chain from propyl to octyl does not result in a dramatic *E*-value change (*E* = 450 to *E* = 760, Table 2) and is mainly due to the entropic component.³⁸

Table 1. Influence of *sec*-alcohol moiety on the CalB enantioselectivity.³⁹

Substrate	<i>E</i>	f.e.	$\Delta\Delta G$ (kJ/mol)	Substrate	<i>E</i>	f.e.	$\Delta\Delta G$ (kJ/mol)	Substrate	<i>E</i>	f.e.	$\Delta\Delta G$ (kJ/mol)
	1	n.d.	0.8		390	R	15.5		9	R	5.8
	7	R	5.0		97	R	11.8		164	R	12.6
	9	R	5.8						371	R	14.7
	390	R	15.5		705	R	17.2		340	R	14.3
	705	R	17.2		109	R	12.2		14	S	6.3
	350	R	15.1						7	S	5.8

Left panel (varying large group on fixed medium group), middle panel (varying medium group on fixed large group), right panel, upper three (varying halogen containing large group on fixed medium group), right panel, lower three (varying halogen containing medium group on fixed large group). Left and middle panel: 39°C, *S*-ethyl thiooctanoate as acyl donor; right panel: 18-23°C, vinyl alkananoate as acyl donor.

Table2. Influence of acyl moiety on the CalB enantioselectivity¹¹

Substrate	E	f.e.	$\Delta\Delta G$ (kJ/mol)
	760	R	16.5
	450	R	15.2
	360	R	14.7
	450	R	15.2

varying acyl length on fixed alcohol moiety

- Solvent

Several arguments can be raised to explain the effect of the solvent on the enantioselectivity of an enzyme. Here we discuss the subject in topics: water and organic solvent.

Water activity

It is generally thought that a certain amount of water is needed for the lipase catalysis in low water organic medium, since water molecules are supposed to influence the enzyme charge distribution, conformational flexibility, proton transfer thus influencing the structure and geometry of the enzyme and the catalytic center.⁴⁰ Some people use 50 essential water molecules in the simulation with other organic solvent.^{15;41} However, the effect of water activity on the enantioselectivity is still unclear.³⁰ It has been reported from experiments that both decreased,⁴² increased^{14;43;44} and unaffected.^{14;44;45} *E*-values are obtained upon lowering the water activity (or water content).

Organic solvent

Several hypotheses have been proposed to explain the mechanism by which organic solvent influences enzymatic enantioselectivity including:

- 1). Modification of interactions between the enzyme and its substrate when the solvent molecules enter the active site and strip the essential water in the enzyme⁴⁶
- 2). Modification of the enzyme conformation and rigidity thus changing the molecular recognition process⁴⁷
- 3). Orientation of substrates in the catalytic center of enzymes

However, so far all these hypotheses are not generally applicable since there is no consistent correlation of the enzyme enantioselectivity with the physicochemical properties of the solvent (hydrophobicity, dielectric constant, dipole moment etc.). Sometimes, the best solvent for a certain substrate and enzyme can be the worst for another substrate and enzyme.⁴⁸ Two things are clear: the enantioselectivity of enzyme can be manipulated by the solvent which offers flexibility in controlling the reaction; catalysis in organic solvent is important for industrial application since many substrates and products of industrial interests are only soluble in organic solvent. Besides, many reactions which are not favoured in aqueous media can be realized in organic solvents.

- Temperature

The temperature dependence of the E -value can be appreciated from the separate enthalpic and entropic contributions to the free energy difference (Eqn 9).

$$RT \ln E = \Delta\Delta G^\ddagger = \Delta\Delta H^\ddagger - T\Delta\Delta S^\ddagger \quad (9)$$

The temperature that leads to $E = 1$ or $\Delta\Delta G_{RS}^\ddagger = 0$ is defined as racemic temperature T_r . Since most of the known enzymatic reactions appear to have $T_r < 300$ K, with the enthalpy difference as the leading term, enantioselectivity usually seen to decrease with the temperature. In some exceptional cases the enantioselectivity is seen to increase with temperature.⁴⁹ The simultaneous improvement of enantioselectivity and activity can then be achieved, which is of great practical interest (as long as the enzyme stays active!).

The influence of temperature on enantioselectivity is dependent on the enzyme, solvent and substrate, thus it should be treated with caution.⁵⁰ While the enantioselectivity is not so sensitive to temperature for subtilisin Carlsberg, *R. miehei* lipase has reduced enantioselectivity from $E = 170$ (7°C) to $E = 19$ (45°C) in dioxane. On the other hand, the enantioselectivity of *R. miehei* shows completely different temperature dependence in dioxane and triethylamine ($E = 150$ at 7°C to $E = 170$ at 45°C). The enantioselectivity is more likely to be influenced by temperature when there are non-polar groups in the substrate. If the substrate has polar groups (e.g. halogenated atoms) that interact with the enzyme by non-steric effects, the enthalpy contribution can be quite large and will dominate the free energy of activation, resulting in little temperature effects.⁴⁹

4. ENZYME ENANTIOSELECTIVITY PREDICTION

To model the enantioselectivity of the enzyme, empirical, semi-empirical and ab initio methods can be applied. Among the empirical models, Kazlauskas rule is the most popular one.⁵¹ The rule predicts that the size difference between the medium and large group in the alcohol is the key factor for the enantioselectivity. The bigger the difference, the higher enantioselectivity is. The rule gives a correct prediction for most of the 130 secondary alcohols substrates included in the study for different enzymes CRL, PCL etc. but it doesn't work so well for primary alcohols and carboxylic acid esters,⁵¹ nor for halogenated alcohols. The reason lies in the fact that Kazlauskas rule doesn't take the entropy contribution and non-steric interaction between substrate and the enzyme into account thus it only works well when the two mentioned effects are not obvious. However, the rule is simple and helpful especially when computational tools are not available.

The empirical model cannot account for the subtleties in the selectivity of the different enzymes; neither can it produce quantitative result. A higher level of computational models is needed to obtain more accurate solution.⁵² By now, most of computational models for enantioselectivity are built on QM, MM or QM/MM hybrid methods. The enantioselectivity is correlated by Eq 5 with the free energy difference between the different conformers at transition state. QM method is a desirable method to describe the transition state structure. Although this method aims to give more realistic charge distributions than other methods, the point charge is dependent on the conformation of the substrate and needs to be updated frequently.⁵³ And this still doesn't guarantee the correct description of the molecules. This is especially the case when the system is treated in gas phase without the solvent and other surrounding residues. In addition, QM calculation, especially the *ab initio* one, is computationally very expensive.⁵⁴

By comparison, the MM calculation uses force fields and is easier to implement. Even though it can not describe the electronic transition state, it can be used for tetrahedral intermediate which closely resembles the structure of transition state as supported by Hammond postulate and the calculation from the *ab initio* studies.⁵⁵⁻⁵⁷ This concept was successfully used for the modeling of the enantioselectivity of different lipase.^{10;15}

Many previous MM studies for enzyme enantioselectivity are based on the potential energy calculation without taking entropy contribution into account. This is a more straightforward and a less time-consuming method compared to the free energy calculation. Potential energy calculation can be done via molecular mechanics and molecular dynamics using classical force fields such as AMBER,^{12;15;58-65} CHARMM,^{10;54;63} CVFF,^{9;66} and GROMOS¹⁴ etc. The R and S configurations are minimized (hopefully we find the structure in global minimum even though it is almost impossible) and then the potential energy difference between R and S is calculated. To overcome the sampling problem, several strategies have been used to search conformation space for the evaluation of the enantioselectivity, ranging from statistical conformational search with the substrate docking in the active site⁶⁵ and simulated annealing⁹ mainly through molecular dynamics (MD). Norin et al calculated the potential energy on function-based subset which includes the residues most relevant to the catalysis. This approach was proved to improve the calculation accuracy in the case of *Candida antarctica* lipase B (CaLB) compared to the calculation on structure-based subset.^{10;11} However, all potential energy calculation approaches require the correct productive pre-orientation of both fast and slow-reacting substrate in the active site. For fast reacting substrate it is more straightforward, but for slow-reacting substrate it is difficult to define. If the conformation is not correct, then the predicted enantioselectivity is going to be wrong as well.

The accuracy of the result from potential energy calculation is so far not satisfactory because in most of the cases it is a semi-quantitative result, i.e. the direction of the enantioselectivity is correctly predicted but the degree is not accurate with commonly off several orders of difference between experimental and calculated data.

One of the drawbacks of the potential energy calculation approach is the neglect of the entropy contribution, which is only valid when the enzyme is considered to be rigid and react similarly with the substrate fragments during the catalysis, and the solvent does not interact with the enzyme-substrate complex in different way. But it has been shown that entropy can make a significant contribution to the enantioselectivity to the relatively small $\Delta\Delta G_{R/S}^\ddagger$ and thus cannot be neglected in many cases.¹⁶⁻¹⁹ The global minimum of potential energy and free energy has been proved to be different for some peptide and protein systems.⁶⁷ Neglecting the entropy calculation easily introduces some errors in the prediction of enzyme enantioselectivity. Thus in order to get a more comprehensive description of the system, the free energy difference calculation is more desirable. Several attempts have been made to predict the enzyme enantioselectivity via free energy calculation, which are listed in Table 3.

QM/MM approach circumvents the computational costs encountered by treating the active site with QM and the rest of the system with MM. The connection between the QM and MM regions are accomplished via e.g. linked-atom treatment. The reliability of QM/MM approaches is strongly dependent on the quality of QM calculation, which is still limited with the computing power nowadays. Using empirical valence bond (EVB) potential provides another way to reduce the sampling problem in QM calculation.⁶⁸ However, we have not seen its application in computation for enzyme enantioselectivity prediction so far.

Table 3. Examples of computational study of enzyme enantioselectivity with different methods

Who	Enzyme	Solute	Solvent	Calculation methods and protocol	FF	ΔE_{cal} (kJ/mol)	ΔE_{exp} (kJ/mol)
Hult et al. 1994 ¹²	CRL, RML and HLL	1-phenyl ethanol	water	MM; TI; 2.4 ps MD simulation for data collection; partial charges derived by ESP fitting in gas phase;	Amber; all-atoms	failed	9.2
Menge et al. 1997 ⁶²	CRL, PCL and RML	1-phenyl ethanol	Water	QM/MM; FEP-DMW for free energy calculation; 34 windows in DMW, 1ps MD per window.	Amber; all-atoms	20.5	11.2
Hult et al. 1998 ¹⁰	CaLB	3,3-dimethyl-2-butanol	Water (279)	MM; Energy-based subsets for potential energy calculation; partial charges derived by RHF in gas phase; 1.5-ns MD simulation for data collection	Charmm all22_prot set	137 ± 11.8	15.1
Colombo et al. 1999 ¹⁵	Subtilisin	1-phenyl ethanol	DMF/water (1515/50)	QM/MM; FEP (slow growth) for free energy calculation; 750ps MD simulation; partial charges updated periodically by QM calculations	Amber; all-atoms	5.4 ± 2.0	1.8
Gruber 2001 ⁶¹	<i>Hb</i> -HNL	2,3-dimethyl-2-hydroxyl-butyrionitrile	Water	MM; TI for free energy calculation; 40 steps with 8ps data collection/step in MD simulation in both forward and backward	Amber; all-atoms	7.1 ± 5.5	11.8
Raza 2001 ¹¹	CaLB	3,3-dimethyl-2-butanol	Water	MM/MD; Function-based subsets for potential energy calculation; partial charges derived by Pullman method in gas phase; 50ps MD simulation for data collection	Kollman; all-atoms	3.2	15.1
Tomić et al. 2006 ⁶⁹	BCL	1-phenoxy-2-hydroxybutane	Water (291, MM methods)	MM/MD: Amber Semi-empirical: MOPAC/GAMESS ab initio: RHF	Amber	9.2 79.8 100.8	12.9
Zhou et al. 2006	CaLB	3,3-dimethyl-2-butanol	Water (~8000)	MM/MD; TI for free energy calculation; 21 steps with >1ns data collection/step in MD simulation; soft-core approach used; improper dihedral changed from R to S; intermediate inserted to reduce the strains	Gromos96 53A6 set; united atoms	14.3 ± 0.8 WZ model; 4.7 ± 0.6 TTI model	15.1

TI: thermodynamic integration

DMW: dynamical modified windows

OPC: 5-oxo-2-pentylpyrrolidine-3-carboxylate

BCL: *Burkholderia cepacia* lipase*Hb*-HNL: *Hevea brasiliensis* hydroxynitrile lyase

TTI: truncated tetrahedral intermediate

FEP: free energy perturbation

RHF: Restricted Hartree-Fock (*ab initio*)CRL: *Candida rugosa* lipasePCL: *Pseudomonas cepacia* lipase

WZ: Whole enzyme

ESP: electrostatic potential

CaLB: *Candida antarctica* lipase BHLL: *Humicola lanuginosa* lipaseRML: *Rhizomucor miehei* lipase

5. FREE ENERGY CALCULATION

Evidently the Gibbs free energy is the relevant thermodynamic quantity, which determines enantioselectivity. According to Statistical Mechanics the Gibbs free energy of a (biochemical) system composed of N_α particles of species α , with $\alpha=1, \dots, n$, is given by:

$$G(N_1, \dots, N_n, P, T) = -k_B T \ln \Delta(N_1, \dots, N_n, P, T) \quad (10)$$

where k_B is Boltzmann's constant, T is the absolute temperature, P the pressure and Δ is the isothermal-isobaric partition function, which can be written as:

$$\Delta(N_1, \dots, N_n, P, T) = C_{N_1, \dots, N_n, P, T} \int dV \int d\mathbf{r}_1 \dots \int d\mathbf{r}_N \exp(-PV/k_B T - U(\mathbf{r}_1, \dots, \mathbf{r}_N)/k_B T) \quad (11)$$

In Eqn. 11 $C_{N_1, \dots, N_n, P, T}$ is a constant, V refers to volume and U is the potential energy of particles positioned at $\mathbf{r}_1, \mathbf{r}_2, \dots$ respectively. Direct evaluation of the free energy is difficult, because Eqn. 11 shows that it is related to the partition function, which cannot be expressed in terms of a statistical average, but depends on the extent of the configuration space accessible to the system. In fact, direct computation of absolute free energy is not feasible. However, many interesting thermodynamic properties, including the enzyme enantioselectivity, are related to free energy differences, which can be computed. To generate representative configurations for the calculation of the free energy either Monte Carlo (MC) or molecular dynamics (MD) simulation can be applied. However, for complex systems involving many covalently bound atoms like proteins, MD seems more efficient than MC.⁷⁰ To calculate free energy, two methods: free energy perturbation (FEP) and thermodynamic integration (TI) are generally applied. We will briefly discuss them together with other derived methods. More details about free energy calculation can be found in the literature.⁷¹⁻⁸¹

Central to the calculation of free energy difference between two states A and B is the concept of a coupling parameter λ , which transforms the system from state A into state B. The concept of a coupling parameter was originally introduced by Kirkwood.⁸⁵ The potential energy is considered as a function of the parameter λ i.e. $U = U(\mathbf{r}_1, \dots, \mathbf{r}_N; \lambda)$, such that, when $\lambda = \lambda_A$ the potential energy corresponds to system A, and when $\lambda = \lambda_B$ it corresponds to system B. The simplest choice is:

$$U(\mathbf{r}_1, \dots, \mathbf{r}_N; \lambda) = (1 - \lambda)U_A(\mathbf{r}_1, \dots, \mathbf{r}_N) + \lambda U_B(\mathbf{r}_1, \dots, \mathbf{r}_N) \quad (12)$$

with $\lambda_A = 0$ and $\lambda_B = 1$. Clearly the Gibbs free energy now depends on the value of the coupling parameter:

$$G(\lambda) = -k_B T \ln C \int dV \int d\mathbf{r}_1 \dots \int d\mathbf{r}_N \exp(-PV/k_B T - U(\mathbf{r}_1, \dots, \mathbf{r}_N; \lambda)/k_B T) \quad (13)$$

where, for brevity, we have dropped the dependence of the free energy G and the normalization constant C on the thermodynamic variables N_α , P and T .

If we use Eqn. 12 for the potential energy and evaluate equation 13 for $\lambda=0$ and $\lambda=1$ and insert $1 = \exp(-U_A/k_B T) \exp(U_A/k_B T)$ we obtain:

$$\Delta G = G_B - G_A = -k_B T \ln \langle \exp(-(U_B - U_A)/k_B T) \rangle_A \quad (14)$$

where $\langle \dots \rangle_A$ denotes an isothermal-isobaric ensemble average in system A. Thus by sampling the potential energy of system B for configurations generated by a distribution for system A the free energy difference can be calculated by averaging the Boltzmann factor of the difference in energy in the two systems. This is only practicable when states A and B are very similar, since otherwise the exponential in Eqn. 14 is either vanishingly small or very large, a recipe for bad statistics. When the difference is large intermediate states can be inserted, so that changes take place gradually.

When we are dealing with semi-flexible molecules, in which some degrees are constrained, e.g. by fixing bond lengths, an additional contribution arising from the constraints can be expected to free energy differences. This point is discussed more fully by van Gunsteren. When only bond lengths are constrained, as is the case in the calculations described in this thesis, these contributions can generally be neglected. Therefore we do not consider these contributions in this thesis.

- Free Energy Perturbation (FEP)

In FEP, a number of intermediate states are inserted between states A and B. The total free energy change can then be obtained by summing the free energy differences between the intermediate states using Eqn. 14.

$$\Delta G_{BA} = -k_B T \sum_{i=0}^{N_i-1} \ln \left\langle \exp \left(- \left(U(\lambda_i) - U(\lambda_{i+1}) \right) / k_B T \right) \right\rangle_{\lambda_i} \quad (15)$$

Although this technique is generally referred to as free energy perturbation Eqn. 15 is exact. Usually, the intermediate states are defined with a λ -coordinate that differs successively by less than $2k_B T$ or ~ 1.5 kcal/mol.⁸² Reducing the spacing between the adjacent λ steps always means a significant additional cost. There might be also convergence problem for FEP in sampling the λ_i and λ_{i+1} state.

- One-step Perturbation (OSP)

The method reduces the required computational time by avoiding the simulation of uninteresting intermediate states. It estimates relative free energy by extrapolating in one single step from a well chosen reference state, which is not necessarily corresponding to a physically meaningful state.⁸³ Eqn. 16 is used to compute the free energy difference between the reference R and the states A and B. The free energy difference between states A and B is then simply

$$\Delta G_{BA} = \Delta G_{BR} - \Delta G_{AR} \quad (16)$$

The advantage of OSP method is its efficiency: compared to traditional MD, it can save simulation by 10-100 times. But choosing the right reference state is not very straightforward and requires experience.⁸⁴

- Thermodynamic Integration (TI)

The method, originated from Kirkwood,⁸⁵ has been extensively used for calculation of free energy differences. Differentiating equation 13 with respect to λ , followed by integration at constant temperature, pressure and particle numbers, yields:

$$\Delta G_{BA} = \int_{\lambda_A}^{\lambda_B} \left\langle \frac{\partial U(r, p, \lambda)}{\partial \lambda} \right\rangle_{\lambda} d\lambda \quad (17)$$

Monte Carlo or molecular dynamics simulation allows efficient evaluation of the integrand at different values of λ . The free energy difference can therefore be computed through numerical integration of Eqn. 17. The convergence of the simulations at each λ point can be checked independently.⁸⁶ The advantage of TI over FEP is that it avoids the problem of sampling the λ_{i+1} –state with the Hamiltonian corresponding to λ_i . It only requires averages of the energy or rather its derivative with respect to λ , which converge relatively fast. It has been demonstrated that TI can be more robust and efficient than FEP.⁸⁷ Moreover, the derivative can be evaluated analytically rather than as a finite difference, which leads to a considerable reduction of the statistical error. TI is susceptible to errors when a simulation is not run long enough to obtain convergence. Several enzyme enantioselectivity studies have applied TI to compute the free energy difference between enantiomers.^{12;14;61}

- Finite difference thermodynamic integration (FDTI)

FDTI is a combination of FEP and TI proposed by Mezei.⁸⁸ The integral in Eqn. 17 is evaluated using a suitable numerical quadrature, e.g. Gaussian integration. The integrand is approximated through Eqn. 15 in the form:

$$\Delta G_{BA} = -k_B T \sum_{i=1}^k w_i \frac{\ln \langle \exp^{-\frac{U(\lambda_i) - U(\lambda_{i+1})}{k_B T}} \rangle_{\lambda_i}}{\delta \lambda} \quad (18)$$

Here k is the number quadrature points in the numerical integration and w_i is the weight at the point i . In addition the average in Eqn. 17 is evaluated using a reference state with energy $U = (U(\lambda_i) + U(\lambda_{i+1})) / 2k_B T$. The advantages of FDTI are: 1). Unlike FEP, large changes in free energy can be calculated in fewer steps 2). Unlike TI, analytical derivatives of the Hamiltonian with respect to the coupling parameter are not needed. 3). It may converge faster than TI or PM, suggested by Mezei.⁸⁸ So far, we haven't seen any example to apply FDTI in free energy calculation for enzyme enantioselectivity study.

- Dynamical modified windows method

In this method, the $\delta\lambda$ between consecutive windows in FEP or TI is continuously changing in order to achieve a relatively constant free energy change per window. This allows us to focus on those ranges of λ where the free energy changes more rapidly. This method can improve the efficiency of sampling but since we don't know the free energy integration curve before starting the simulation, we can't predict the simulation time needed. Dynamical modified windows method was proposed by Kollman⁸⁹ and has been used by Menge to study enantioselectivity of lipases.⁶²

- Slow Growth method

This method is a modification of FEP or TI methods. The integration can be performed continuously while slowly changing the coupling parameter λ from λ_A to λ_B during the simulation. However this approach is problematic because the system lags behind the changing Hamiltonian and never equilibrates appropriately. So fundamentally it is a non-equilibrium estimate.⁹⁰ This method was used by Colombo to study the enantioselectivity of Subtilisin.¹⁵ For this reason, this method is abandoned or developed into other methods like fast growth approach.

- Fast Growth method

This method is based on slow growth method. It can be shown that the exponential average of the non-equilibrium slow-growth free energy estimates yields an equilibrium value of the free energy. Several slow growth calculations are

run and the required sampling of conformational space is obtained from the difference trajectories of the slow growth runs. The method is well suited for parallel computer architectures, requiring only the simplest parallelism with repeated runs for different starting conditions.^{91;92} The application of fast growth method in the enantioselectivity study is to be explored in the future.

In theory, the free energy is a state function, which means that the free energy difference is only dependent on the initial and final state, no matter what path is taken to go from one to the other. As a consequence, any non-physical pathway can be chosen to perform the calculations. But in practice, the choice of the pathway is crucial for the quality and accuracy of the result of free energy calculation. Singularities (or in general large changes) in the energy along the path will introduce large statistical and numerical errors, which render the technique unfeasible. The best pathway should be the one which brings the least disturbance to the system so that the system can be perturbed smoothly from initial state to final state.^{71;93}

Reference

- (1) Straathof, A. J. J.; Jongejan, J. A. *Enzym Microb Techno* 1997, 21, 559-571.
- (2) Platzer, K. E.; Momany, F. A.; Scheraga, H. A. *Int J Pept Prot Res.* 1972, 4, 187-200.
- (3) Platzer KEB; Momany, F. A. *Int J Pept Prot Res* 1972, 4, 209.
- (4) Detar, D. F. *Biochemistry* 1981, 20, 1730-1743.
- (5) Wipff, G.; Dearing, A.; Weiner, P. K.; Blaney, J. M.; Kollman, P. A. *J Am Chem Soc* 1983, 105, 997-1005.
- (6) Norin, M.; Norin, T.; Hult, K. *Biocataly* 1993, 7, 131-147.
- (7) Warshel, A. *J Biol Chem* 1998, 273, 27035-27038.
- (8) Ohagan, D.; Rzepa, H. S. *J Chem Soc-Perk T2* 1994, 3-4.
- (9) Orrenius, C.; van Heusden, C.; van Ruiten, J.; Overbeeke, P. A.; Kierkels, H.; Duine, J. A.; Jongejan, J. A. *Protein Eng* 1998, 11, 1147-1153.
- (10) Haffner, F.; Norin, T.; Hult, K. *Biophys J* 1998, 74, 1251-1262.
- (11) Raza, S.; Fransson, L.; Hult, K. *Protein Sci* 2001, 10, 329-338.
- (12) Norin, M.; Haeffner, F.; Achour, A.; Norin, T.; Hult, K. *Protein Sci* 1994, 3, 1493-1503.
- (13) Edholm, O.; Ghosh, I. *Mol Simulat* 1993, 10, 241-253.
- (14) Micaelo, N. M.; Teixeira, V. H.; Baptista, A. M.; Soares, C. M. *Biophys J* 2005, 89, 999-1008.
- (15) Colombo, G.; Toba, S.; Merz, K. M. *J Am Chem Soc* 1999, 121, 3486-3493.
- (16) Aqvist, J.; Warshel, A. *Chem Rev* 1993, 93, 2523-2544.
- (17) Ottosson, J.; Rotticci-Mulder, J. C.; Rotticci, D.; Hult, K. *Protein Sci* 2001, 10, 1769-1774.
- (18) Ottosson, J.; Fransson, L.; Hult, K. *Protein Sci* 2002, 11, 1462-1471.
- (19) Kollman, P. A.; Kuhn, B.; Perakyla, M. *J Phys Chem B* 2002, 106, 1537-1542.
- (20) NC-IUB. *Eur J Biochemistry*, 1982, 128, 281-291.
- (21) Fersht, A. W. H. Freeman and Company: New York, 1985, p 105.

- (22) Chen, C. S.; Fujimoto, Y.; Girdaukas, G.; Sih, C. J. *J Am Chem Soc* 1982, 104, 7294-7299.
- (23) Van Tol, J. B. A.; Jongejan, A.; Geerlof, A.; Duine, *Recl Trav Chim Pay-B* 1991, 110, 255-262.
- (24) Cleland, W. W. *BBA* 1963, 67, 104-137.
- (25) Truhlar, D. G.; Garrett, B. C.; Klippenstein, S. J. *J Chem Phys* 1996, 12771-12800.
- (26) Overbeeke, P. A.; Orrenius, S. C.; Jongejan, J. A.; Duine, J. A. *Chem Phys Lipids* 1998, 93, 81-93.
- (27) Anthonsen, T.; Jongejan, A. *Methods Enzymol.* 1997, 286, 473-495.
- (28) Martinelle, M.; Hult, K. *BBA-Prot Struct M* 1995, 1251, 191-197.
- (29) Kraut, J. *Annu Rev Biochemistry* 1977, 46, 331-358.
- (30) Berglund, P. *Biomol Eng* 2001, 18, 13-22.
- (31) Qian, Z.; Lutz, S. *J Am Chem Soc* 2005, 127, 13466-13467.
- (32) Villeneuve, P.; Muderhwa, J. M.; Graille, J.; Haas, M. J. *J Mol Cataly B-Enzym* 2000, 9, 113-148.
- (33) Liebeton, K.; Zonta, A.; Schimossek, K.; Nardini, M.; Lang, D.; Dijkstra, B. W.; Reetz, M. T.; Jaeger, K. E. *Chem Biol* 2000, 7, 709-718.
- (34) Reetz, M. T. *P Natl Acad Sci USA* 2004, 101, 5716-5722.
- (35) Rotticci, D.; Rotticci-Mulder, J. C.; Denman, S.; Norin, T.; Hult, K. *Chembiochem* 2001, 2, 766-770.
- (36) Magnusson, A. O.; Takwa, M.; Harnberg, A.; Hult, K. *Angew Chem-Intl Edit* 2005, 44, 4582-4585.
- (37) Rotticci, D.; Orrenius, C.; Hult, K.; Norin, T. *Tetrahedron-Asymmetr* 1997, 8, 359-362.
- (38) Ottosson, J.; Hult, K. *J Mol Cataly B: Enzym* 2001, 11, 1025-1028.
- (39) Rotticci, D.; Haeffner, F.; Orrenius, C.; Norin, T.; Hult, K. *J Mol Cataly B: Enzym* 1998, 5, 267-272.
- (40) Klibanov, A. M. *Trends Biochem Sci* 1989, 14, 141-144.
- (41) Toba, S.; Hartsough, D. S.; Merz, K. M. *J Am Chem Soc* 1996, 118, 6490-6498.
- (42) Yasufuku, Y.; Ueji, S. *Bioorg Chem* 1997, 25, 88-99.
- (43) Orrenius, C.; Norin, T.; Hult, K.; Carrea, G. *Tetrahedron-Asymmetr* 1995, 6, 3023-3030.
- (44) Rariy, R. V.; Klibanov, A. M. *Biocataly Biotransfor* 2000, 18, 401-407.
- (45) Wehtje, E.; Costes, D.; Adlercreutz, P. *J Mol Cataly B-Enzym* 1997, 3, 221-230.
- (46) Nakamura, K.; Takebe, Y.; Kitayama, T.; Ohno, A. *Tetrahedron Lett* 1991, 32, 4941-4944.
- (47) Fitzpatrick, P. A.; Klibanov, A. M. *J Am Chem Soc* 1991, 113, 3166-3171.
- (48) Carrea, G.; Ottolina, G.; Riva, S. *Trends Biotechnol* 1995, 13, 63-70.
- (49) Phillips, R. S. *Trends Biotechnol* 1996, 14, 13-16.
- (50) Noritomi, H.; Almarsson, O.; Barletta, G. L.; Klibanov, A. M. *Biotechnol Bioeng* 1996, 51, 95-99.
- (51) Kazlauskas, R. J.; Weissfloch, A. N. E.; Rappaport, A. T.; Cuccia, L. A. *J Organ Chem* 1991, 56, 2656-2665.
- (52) Warshel, A. *Annu Rev Biophys Biomol Struct* 2003, 32, 425-443.
- (53) Singh, U. C.; Kollman, P. A. *J Comput Chem* 1984, 5, 129-145.

- (54) Topf, M.; Varnai, P.; Richards, W. G. *J Am Chem Soc* 2002, 124, 14780-14788.
- (55) Hammond, G. S. *J Am Chem Soc* 1955, 77, 334-338.
- (56) Hu, C. H.; Brinck, T.; Hult, K. *Int J Quantum Chem* 1998, 69, 89-103.
- (57) Warshel, A.; Narayszabo, G.; Sussman, F.; Hwang, J. K. *Biochemistry* 1989, 28, 3629-3637.
- (58) Orrenius, C.; Haeffner, F.; Rotticci, D.; Ohrner, N.; Norin, T.; Hult, K. *Biocataly Biotransfor* 1998, 16, 1-15.
- (59) Tafi, A.; Manetti, F.; Botta, M.; Casati, S.; Santaniello, E. *Tetrahedron-Asymmetr* 2004, 15, 2345-2350.
- (60) Felluga, F.; Pitacco, G.; Valentin, E.; Coslanich, A.; Fermeglia, M.; Ferrone, M.; Pricl, S. *Tetrahedron-Asymmetr* 2003, 14, 3385-3399.
- (61) Gruber, K. *Proteins* 2001, 44, 26-31.
- (62) SainzDiaz, C. I.; Wohlfahrt, G.; Nogoceke, E.; HernandezLaguna, A.; Smeyers, Y. G.; Menge, U. *Theochem-J Mol Struct* 1997, 390, 225-237.
- (63) Colombo, G.; Ottolina, G.; Carrea, G.; Merz, K. M. *Chem Commun* 2000, 559-560.
- (64) Tomic, S.; Bertosa, B.; Kojic-Prodic, B.; Kolosvary, I. *Tetrahedron-Asymmetr* 2004, 15, 1163-1172.
- (65) Manetti, F.; Mileto, D.; Corelli, F.; Soro, S.; Palocci, C.; Cernia, E.; D'Acquarica, I.; Lotti, M.; Alberghina, L.; Botta, M. *BBA-Prot Struct Mol Enzym* 2000, 1543, 146-158.
- (66) Brady, L.; Brzozowski, A. M.; Derewenda, Z. S.; Dodson, E.; Dodson, G.; Tolley, S.; Turkenburg, J. P.; Christiansen, L.; Hugejensen, B.; Norskov, L.; Thim, L.; Menge, U. *Nature* 1990, 343, 767-770.
- (67) Evans, D. A.; Wales, D. J. *J Chem Phys* 2003, 118, 3891-3897.
- (68) Aqvist, J.; Warshel, A. *Chem Rev* 1993, 93, 2523-2544.
- (69) Tomic, S.; Ramek, M. *J Mol Cataly B-Enzym* 2006, 38, 139-147.
- (70) Van Gunsteren, W. F.; P.K.Weiner *Computer simulation of biomolecular systems, theoretical and experimental applications*; Excom Science Publishers: Leiden, The Netherlands, 1989, pp 27-59.
- (71) Straatsma, T.P. *Reviews in computational chemistry*; VCH: 1996; Chapter 2, pp 81-121.
- (72) Van Gunsteren, W. F.; Daura, X.; Mark, A. E. *Helvetica Chimica Acta* 2002, 85, 3113-3129.
- (73) Warshel, A.; Sussman, F.; Hwang, J. K. *J Mol Biol* 1988, 201, 139-159.
- (74) Mark, A. E.; Van Gunsteren, W. F.; Berendsen, H. J. C. *J Chem Phys* 1991, 94, 3808-3816.
- (75) Michael R.Shirts; Jed W.Pitera; William C.Swope. *J Chem Phys* 2004, 119, 5740-5761.
- (76) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; Debolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. *Comput Phys Commun* 1995, 91, 1-41.
- (77) Radmer, R. J.; Kollman, P. A. *J Comput Chem* 1997, 18, 902-919.
- (78) Rodinger, T.; Pomes, R. *Curr Opin Structl Biol* 2005, 15, 164-170.
- (79) Burgi, R.; Lang, F.; Van Gunsteren, W. F. *Mol Simulat* 2001, 27, 215-236.
- (80) Pearlman D. A. Comparison of free energy calculation methods
www.cmpharm.ucsf.edu/~troyer/troff2html/amber/gibbs.html 2005
- (81) Oostenbrink C. *Free energies from biomolecular simulation: Force fields, methodology and applications*; Shaker Verlag: Aachen, 2005.
- (82) Straatsma, T. P.; Berendsen, H. J. C.; Postma, J. P. M. *J Chem Phys* 1986, 85, 6720-6727.

- (83) Liu, H. Y.; Mark, A. E.; Van Gunsteren, W. F. *J Phys Chem* 1996, 100, 9485-9494.
- (84) Oostenbrink, C.; Van Gunsteren, W. F. *J Comput Chem* 2003, 24, 1730-1739.
- (85) Kirkwood, J. G. *Theory of Liquids*; Gordon and Breach: New York, 1968.
- (86) Straatsma, T. P.; Berendsen, H. J. C. *J Chem Phys* 1988, 89, 5876-5886.
- (87) Berendsen, H. J. C.; Van Gunsteren, W. F. *Molecular dynamics and protein structure*; Polycrystal Book Service: 1985, pp 43-66.
- (88) Mezei, M. *J Chem Phys* 1987, 86, 7084-7088.
- (89) Pearlman, D. A.; Kollman, P. A. *J Chem Phys* 1989, 90, 2460-2470.
- (90) Pearlman, D. A.; Kollman, P. A. *J Chem Phys* 1989, 91, 7831-7839.
- (91) Hummer, G. *J Chem Phys* 2001, 114, 7330-7337.
- (92) Hendrix, D. A.; Jarzynski, C. *J Chem Phys* 2001, 114, 5974-5981.
- (93) Zhou, Y.; Oostenbrink, C.; Van Gunsteren, W. F.; Hagen, W. R.; De Leeuw, S. W.; Jongejan, J. A. *Mol Phys* 2005, 103, 1961-1969.

THE RELATIVE STABILITY OF HOMOCHIRAL AND HETEROCHIRAL DIALANINE PEPTIDES. EFFECTS OF PERTURBATION PATHWAYS AND FORCE-FIELD PARAMETERS ON FREE ENERGY CALCULATIONS

SUMMARY

The relative stability of homochiral (*D,D* or *L,L*) and heterochiral (*D,L* or *L,D*) dipeptides may have been a decisive factor in the evolutionary propagation of a symmetry-breaking event leading to the present-day predominance of *L*-amino acids in natural proteins. Kinetic resolution in the solid-phase peptide synthesis of blocked dialanine suggests the activation free energy difference of formation of (*D,D* or *L,L*)- and (*D,L* or *L,D*)-dialanine to be 0.22 kJ/mol in favour of the formation of the homochiral dipeptide. Computer simulation studies were performed on water-solvated dialanine, applying a thermodynamic integration protocol using the GROMOS force field. Five different pathways and three force-field parameter sets have been used to assess the possibility of a computational prediction of the chiral preference. Inversion of the configuration around either one of the C_α-atoms by changing the improper dihedral angle with concomitant relaxation of the bond angles, leads to an excellent reproduction of the experimental result.

Keywords:

chirality, evolution, dipeptide, dialanine, kinetic resolution, free energy, thermodynamic integration, force field, GROMOS, simulation

INTRODUCTION

The predominance of *L*-amino acids and *D*-sugars in natural biopolymers is well-established. Compared to the (almost) exclusive formation of racemic mixtures in the chemical synthesis of chiral compounds from non-chiral precursors, the evolutionary origin (symmetry-breaking) and accumulation (propagation) of this chiral preference presents one of the intriguing and possibly fundamental features of the biochemistry of life on Earth. Several propositions have been put forward to rationalize an initial symmetry-breaking event (for a review see Bonner [1]). Random mechanisms such as spontaneous chiral symmetry breaking in enantiomeric crystallization, and selective adsorption on calcite, rely on chance. Determinate mechanisms, on the other hand, require identification of an intrinsic physical force that perturbs the racemic balance in a specific way. Of these, the electroweak force mechanism derived from the parity violating energy difference, initially proposed by Yamagata [2], has obtained much attention (for a review see Quack [3]). Considering the very small energetic difference between e.g. *L*- and *D*-amino acid enantiomers induced by such forces, subsequent amplification is required. Clearly, the Accumulation Principle in the form originally proposed by Yamagata, where a (parity violating) energy difference is presumed to be operative at each step, fails to account for the formation of homochiral polymers, as has been argued by Bonner [4]. Additional selective pressure appears to be required. Indeed,

This chapter has been published in Molecular Physics 2005, 103(14), 1961-1969.

by invoking a catalytic (pseudo-) steady state feed back model, it can be shown that an initial disturbance of the racemic balance of an amino acid may lead to its eventual incorporation as a single enantiomer [5]. Formation of homochiral proteins, however, requires additional selection. As early as 1957, Wald suggested that the oligopeptide secondary structure would dictate the selection of one amino acid enantiomer rather than its antipode during polymer growth [6]. This hypothesis has since been validated both by experiment [7] and by simulation [8,9]. In particular, configurational randomness lowers the stability of an elongated α -helix structure. So far, data on the relative stability of homo- and heterochiral amino acid dimers and trimers formed in the initial stages of protein polymerization are lacking. In the present study we investigate the free energy difference between a homochiral dipeptide and its heterochiral counterpart. Dialanine peptide is chosen as a model molecule since alanine is the simplest chiral amino acid. Its presumed presence in the “primordial soup” is largely undisputed [10]. Considering the lack of experimental data and the difficulties encountered during racemization in vitro, we investigated and evaluated computational protocols based on thermodynamic integration (TI) methods in order to calculate the free energy difference involved.

MATERIALS AND METHODS

L-Alanine, *D*-Alanine, *L*-Ala-*L*-Ala, *D*-Ala-*D*-Ala and alanine racemase from *Bacillus stearothermophilus* were purchased from Sigma-Aldrich Chemie, B.V., Zwijndrecht, The Netherlands. *L*-Ala-*D*-Ala and *D*-Ala-*L*-Ala were synthesized at Neosystem, Strasbourg, France. The amino acids and peptides were more than 99% pure as confirmed by chiral HPLC analysis.

Dipeptide synthesis

Solid phase synthesis was employed to investigate the kinetic preference of formation of the different forms of the dipeptides. Syntheses were conducted at Neosystem (Strasbourg, France) using Fmoc strategy. Benzotriazol-1-yl-oxytritypyrrolidinophosphonium hexafluorophosphate (781 mg), 1-hydroxybenzotriazole (230 mg), diisopropyl ethyl amine (513 μ l), and 1.5 mmoles of amino acid were dissolved in 5 ml of dimethylformamide (DMF) at pH 8.5. The amino acids were coupled to an equimolar mixture of Fmoc-*L*-Ala-OH and Fmoc-*D*-Ala-OH resin that was in three-fold excess of free amino acids in the mixture described above. The synthesis was conducted in 5 steps: 1) Coupling, in DMF (1 h); 2) Washing in DMF (3 x 1 min); 3) Deprotection in 25% piperidine in DMF (1 min); 4) Deprotection as 3) (2 x 15 min); 5) Washing, 7 x 1 min, in DMF. The peptide was cleaved from the resin by treatment with trifluoroacetic acid (1%) in dichloromethane for 1 h. The synthesized peptide was purified on a Silica C18 HPLC column, using gradient elution (Solvent A: water, 0.1% TFA; Solvent B: acetonitrile/A, 60/40 (v/v), linear gradient from 0 to 10% B in 20 min) at 20 ml/min. The experiment was carried out at room temperature.

HPLC analysis

Crownpak CR (+) HPLC chiral column (0.4x15 cm) was from Chiral Technologies, Europe (Illkirch, France). Elution was carried out with 5%-HClO₄/water (w/w), pH 1.3 as the mobile phase, at T = 27 °C, flow rate 0.4 ml/min, on a Waters Alliance HPLC (Milford, MA, USA) system consisting of a 2690 separation module and a 996 PDA detector. Data were stored and processed using Waters Millennium software.

Computer simulation

All simulations were performed with the GROMOS96 simulation package[11,12]. Using the thermodynamic integration (TI) formula [13] the free energy difference, ΔG_{BA} , is calculated between the diastereomeric dipeptides, which are defined as state A and state B

$$\Delta G_{BA} = \int_{\lambda_A}^{\lambda_B} \left\langle \frac{\partial H(r, p, \lambda)}{\partial \lambda} \right\rangle_{\lambda} d\lambda$$

$H(r, p, \lambda)$ is the Hamiltonian describing the system as a function of the molecular positions, r , conjugate momenta, p , and the coupling parameter λ . The Hamiltonian describes the homochiral peptide if $\lambda = \lambda_A = 0$ and the heterochiral peptide if $\lambda = \lambda_B = 1$. The angular brackets indicate an ensemble average at a specific value of λ . The integration was performed by changing the value of λ from λ_A to λ_B in a number of discrete steps at which the ensemble average is collected. Different pathways connecting states A and B can be chosen, which should all lead to the same value of ΔG_{BA} in the limit of infinite sampling. The accuracy and efficiency of five pathways are studied here. Some pathways described below require the addition of non-interacting so-called dummy atoms as indicated in Figure 1.

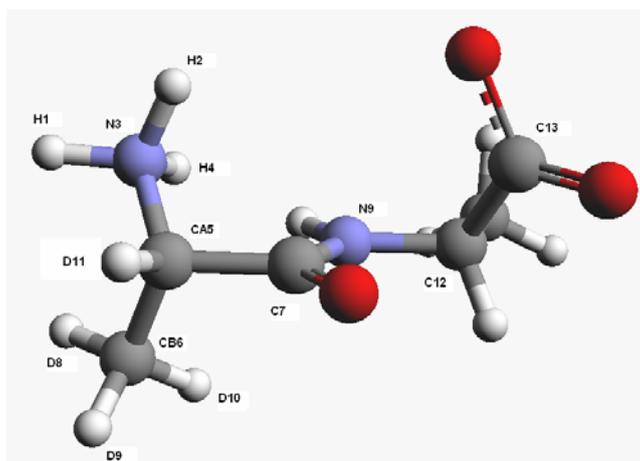


Figure 1: Structure of the (*LL*) alanine dipeptide. The GROMOS force fields treat aliphatic carbons (CH_n) as united atoms, as a single interaction site. D denotes dummy atoms for which (non-bonded) interactions can be switched on and off as function of λ .

L-Ala-*L*-ala (zwitterionic, COO^- and NH_3^+ at the termini, see Figure 1) in an extended conformation ($\varphi = 180^\circ$, $\psi = 180^\circ$) was solvated in (periodic) truncated octahedral boxes containing 1242 to 1394 simple point charge (SPC) water molecules [14]. The systems were energy minimized using steepest descent minimisation, and then gradually heated to 300 K in three 100 K intervals of 10 ps at constant volume using a weakly coupled temperature bath with relaxation time 0.1 ps [15]. Another 20 ps of equilibration was performed at a constant pressure of 1 atm using weak coupling to a pressure bath with a relaxation time of 0.5 ps and an estimated isothermal compressibility of $4.575 \cdot 10^{-4} \text{ (kJ mol}^{-1} \text{ nm}^{-3})^{-1}$ [15]. At every λ point 160 ps equilibration were followed by 240 ps of data collection. Bond lengths were constrained (with relative precision of 10^{-4}) at the minimum energy values using the SHAKE algorithm [16] allowing for a time step of 2 fs. Long-range interactions were calculated using a triple range cut-off scheme. All interactions within 0.8 nm were calculated every time step from a pair list that was updated every fifth step. Interactions between pairs that were between 0.8 and 1.4 nm were calculated at pair list (re)construction and kept constant between pair list updates. A reaction-field contribution [17] was added to the electrostatic interactions to account for a homogeneous medium outside the long range cutoff with a relative permittivity of 62 [18].

The five pathways that were used in the simulations are listed in Table 1. Pathways 1, 2 and 3 involve mutating atoms of different groups between state A and state B. Of these, pathway 3 is the only one that can be used straightforwardly for other, non-alanine, peptides. A soft core interaction [19] was used for all atoms for which the perturbation or mutation involves a change in atom character. Pathways 4 and 5 are two other generally applicable pathways. They involve perturbation of an improper dihedral change at the C_α atom of one of the residues from 35.26° (*L*) to -35.26°

(D). In pathway 4, the molecule is expected to experience much strain in the bond angles at intermediate values of the improper dihedral. For that reason, we inserted an intermediate state I in pathway 4, leading to pathway 5, in which the minimum energy value of the improper dihedral is set to 0° and the minimum energy values of the three bond angles around the C_α atom are set to 120° . All integrations were initially split up into ten steps ($\Delta\lambda = 0.1$) from 0 to 1. For pathway 5, the two half pathways were each performed in 5 steps with $\Delta\lambda = 0.2$. Additional λ -values and longer simulation times were used if necessary to obtain smooth $\langle\partial H/\partial\lambda\rangle$ curves. The perturbation was also carried out in a reverse way from *D*-ala-*L*-ala to *L*-ala-*L*-ala to check the reliability of the TI protocol.

The force-field parameters were taken from the GROMOS96 force field [12]. The 45A3 parameter set [20] was used for all the five pathways. Parameter sets 43A1 [21] and 53A6 [22] were additionally used for pathway 5 only.

Table 1: Five pathways to change the *N*-terminal alanine in the dipeptide from *L* to *D*. For pathways 1-3 the atoms for which the force-field parameters are exchanged are listed. D12, D13, D14 are dummy atoms connected to dummy atom D11. These are not shown in Figure 1. For pathways 4 and 5 the force field terms that are changed are listed, improper dihedral: CA5-N3-C7-CB6, bond angles: N3-CA5-C7, CB6-CA5-C7 and N3-CA5-CB6. Pathway 5 is defined through an intermediate state I.

Pathway	State A(LL)		State B(DL)	
(1) $H \Leftrightarrow CH_3$	CB6		D11	
	D11		CB6	
(2) $NH_3^+ \Leftrightarrow CH_3$	N3		CB6	
	H1		D8	
	H2		D9	
	H4		D10	
	CB6		N3	
	D8		H1	
	D9		H2	
	D10		H4	
(3) $H \Leftrightarrow NH_3^+$	N3		D11	
	H2		D12	
	H1		D13	
	H4		D14	
	D11		N3	
	D12		H2	
	D13		H1	
	D14		H4	
(4) Improper dihedral change	35°		-35°	
Pathway	State A (LL)	State I	State B (DL)	
(5)	Improper dihedral change	35°	0°	-35°
	bond angle change	109.5°	120°	109.5°

RESULTS AND DISCUSSION

To the best of our knowledge, explicit experimental data for the (relative) stability of homo- and heterochiral dialanine are not available in the literature. A single report by Saetia *et al.* [24], describes the formation of dialanine after several

cycles of evaporation and dilution in a simulation of salt-induced peptide synthesis under possibly prebiotic conditions. A slight excess of homochiral (52%) over heterochiral (48%) dialanine is reported by the authors. If we accept these numbers to reflect the equilibrium composition at 300 K, the homochiral dipeptide would be more stable by a Gibbs free energy difference of 0.22 kJ/mol. Since the diastereomeric dipeptides represent only a very small portion of the products obtained and the chromatography materials used by these authors to separate the predominantly formed diketopiperazines were not available to us, other possibilities to estimate the free energy difference were explored. Despite the fact that alanine racemase from *Bacillus stearothermophilus* efficiently racemizes alanine under neutral conditions at room temperature, no activity was observed with the dipeptide, neither at room temperature nor at elevated temperatures. Epimerization of dialanine at high pH (0.1 M NaOH) produced substantial amounts of intermediate cyclic compounds, such as diketopiperazine, which strongly interfered with the analysis [23].

Applying solid phase dipeptide synthesis by reacting either resin-bound *L*-alanine or *D*-alanine with racemic (*D,L*)-alanine in the condensation step, we found the relative rate of formation of the homochiral and the heterochiral alanine dipeptide with respect to total dipeptide production to be 52.2% and 47.8%, respectively, as deduced from the peak areas in the HPLC chromatograms. From these data, we could derive the activation free energy difference ΔG^\ddagger (from the thermodynamic format of the Eyring TS theory) in the formation of (*DL* or *LD*)-dialanine and (*DD* or *LL*)-dialanine to be 0.22 kJ/mol in favour of the formation the latter dipeptide. This result agrees well with difference that could be deduced from the experiments of Saetia *et al.* [24]. Although, strictly speaking, a comparison of these values is not completely warranted since they refer to the ΔG_R° (free energy difference of the epimerization reaction products) and the ΔG^\ddagger (difference of the free energy of activation), respectively, the Hammond-Leffler relationship suggests that these two free energy differences are generally correlated [25]. Awaiting forthcoming experimental data, we explored the possibilities of current computational tools to estimate the free energy difference, accepting a slightly (~ 0.2 kJ/mol) higher stability of the homochiral dipeptide as a calibration mark.

Free energy is a state function and thus independent of the chosen pathways for the calculation. In practice, however, the choice of pathway will determine to a considerable extent the accuracy and efficiency of the calculations. The best sampling will arise from a pathway that changes the system from initial to final state with the least additional disturbance to the system, *i.e.* a path for which the relaxation of the remaining degrees of freedom of the system is fast [26,27]. For such a path, a smooth, gradual change in free energy is expected. This is indeed illustrated by the examples in our study. Most of the simulations yield results that are very close to the experimental value (Table 2). However, the resulting ΔG values do show sensitivity to the different pathways and force fields used. Pathways 2 and 3 involve mutations of a charged (NH_3^+) group and bring a large disturbance to the system, which results in large (10^3 kJ mol⁻¹) values of $\partial H/\partial \lambda$ (Figure 2) and tends to yield slightly larger absolute free energy differences than expected. Also the error bars are larger. Pathways 1 and 4 show smaller error bars and smaller (10^2 kJ mol⁻¹) values of $\partial H/\partial \lambda$. The larger variation in $\partial H/\partial \lambda$ for pathway 4 (Figure 2a) is probably due to the bond-angular strain that is building up when the improper dihedral approaches 0°. In pathway 5 this strain is reduced, leading to yet smaller error bars and $\partial H/\partial \lambda$ values (10 kJ mol⁻¹). Together with the good convergence behaviour and its general applicability, this makes pathway 5 the method of choice to further investigate the homo- / heterochiral free energy difference and its dependence on the force field used.

Table 2: Free energy differences between heterochiral and homochiral dialanine. Statistical errors are calculated according to [28].

Pathway	Force-field parameter set	ΔG (forward) (kJ/mol)	ΔG (backward) (kJ/mol)	Number of λ points	Total simulation time (ns)
1	45A3	-0.3 ± 1.2		13	3.7
2	45A3	1.0 ± 2.2		13	3.7
3	45A3	-2.5 ± 2.1		13	3.7
4	45A3	-0.7 ± 1.5		13	3.7
5	45A3	0.2 ± 1.1	0.2 ± 1.1	12	3.6
5	45A3	1.0 ± 1.2	-0.1 ± 1.1	22	10.6
5	43A1	0.0 ± 0.9	0.0 ± 1.0	12	3.6
5	43A1	0.2 ± 1.1	-0.1 ± 1.1	22	10.6
5	53A6	0.9 ± 1.1	0.7 ± 0.9	12	3.6
5	53A6	0.2 ± 1.0	0.8 ± 0.9	22	10.6
5*	53A6	0.12 ± 0.03		22*	106*
Experiment		0.22			

* Average (standard deviation) of ten separate runs.

The free energy profiles for different force fields along pathway 5 show similar trends (Figure 2c). All three force fields yield results that are in close agreement with the very small experimental value, with a statistical error of about 1.0 kJ/mol (Table 2). A one-way analysis of variance on the free energy profiles resulted in a p-value of 0.94, indicating no statistical difference between the force fields. The GROMOS force field 43A1 was developed with a derivation of the van der Waals parameters for the aliphatic CH_n united atoms based on thermodynamic data for short n -alkanes [21]. As the density for longer alkanes appeared too high, a reparameterization of the aliphatic united atoms was done introducing two additional atoms types for branched and cyclic alkanes. This resulted in the 45A3 set of parameters [20]. The parameter set 53A6 was developed to overcome the underestimated free energy of hydration of polar compounds in conjunction with the SPC model for liquid water [22].

Although, our main objective has been to establish a suitable low-energy integration path, the accuracy of the results is a point of concern. In order to obtain a more appropriate estimate, a number of runs, using different seed-values for the initial velocity distribution, were combined. As the final entry of Table 2 shows, a major reduction of the standard deviation can be obtained at an acceptable computational cost.

Analysis of the trajectories at $\lambda = 0$ and $\lambda = 1$ discloses the difference between the *LL* and *DL* dialanines. From the dihedral angle distributions (Figure 3), we derived the most-populated conformations for the *LL* and *DL* dialanines (Figure 4). Slightly different configurations, $[(\varphi_2, \psi_1) \sim (-121^\circ, 144^\circ)]$, for *LL*, and $[(\varphi_2, \psi_1) \sim (-126^\circ, -137^\circ)]$ for *DL* are favored. In *LL*, the amino group and carboxyl group point in opposite directions, making the molecule dipolar (dipole moment is 31 D) In *DL*, the two terminal groups are positioned on the same side, attracting each other, but making the molecule slightly less dipolar (dipole moment is 29 D), see Table 3. The two molecules show no significant differences in peptide-water hydrogen bonding (Table 3).

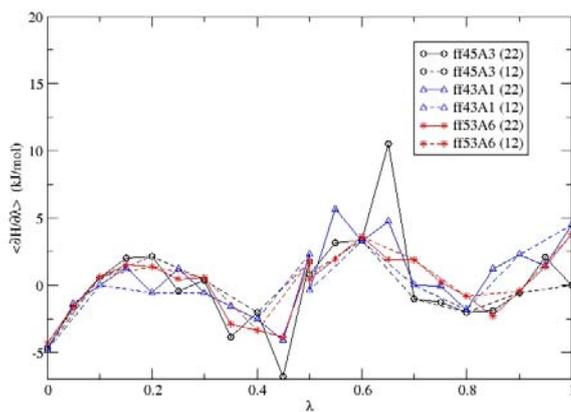
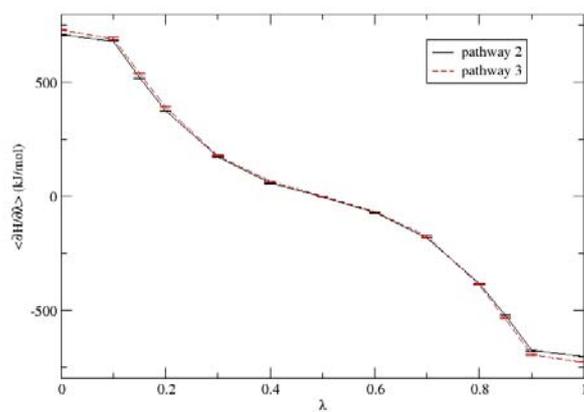
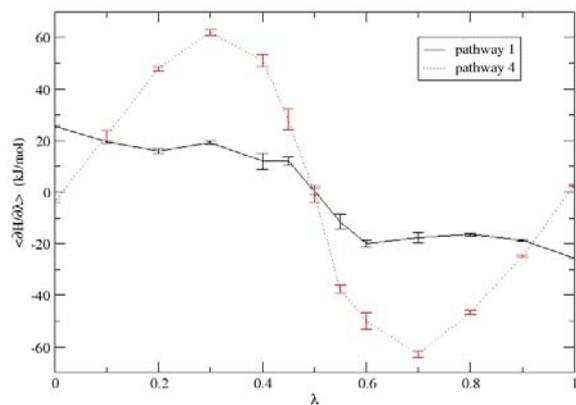


Figure 2. Free energy profile for the five pathways with different GROMOS force fields. (a) (b) Pathways 1- 4, with the 45A3 force field. (c) Pathway 5, with three different GROMOS force fields: 43A1, 45A3 and 53A6. (22) and (12) denote the number of λ points used in the integration. Different scales are used for (a), (b) and (c) to show the details of the free energy profiles.

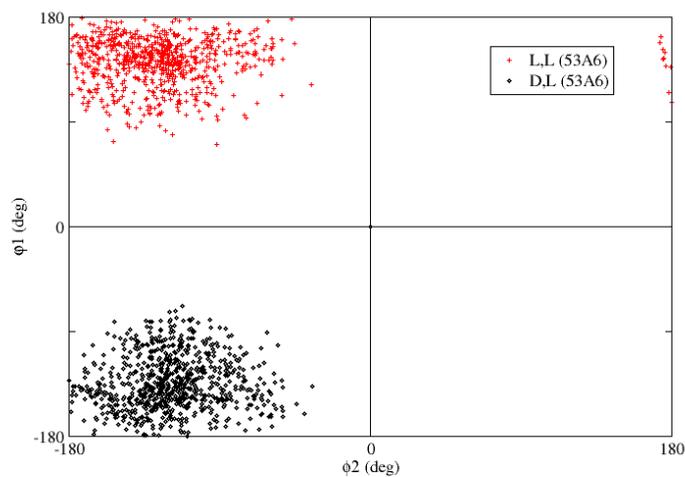


Figure 3. Ramachandran plots for the (ϕ_2, ψ_1) distribution of *LL* (+) and *DL* (◆) dipeptides in aqueous solution using the GROMOS force field 53A6.



Figure 4. The most populated conformation of *LL* (left) and *DL* (right) dipeptides extracted from the MD trajectories based on the 53A6 force-field parameter set.

Table 3: Comparison of average values of the torsional angles ψ_1 and ϕ_2 , of the molecular dipole moment and of the number of solute-water hydrogen bonds formed for *LL* and *DL* dialanines in the simulations using different force fields.

Force field parameter set	Configur- ation	Dihedral angle		Average dipole moment of dipeptide (D)	Average number of peptide-water hydrogen bonds
		N3-CA5-C7-N9	C7-N9-CA11-C13		
		$\langle\psi_1\rangle$ (deg)	$\langle\phi_2\rangle$ (deg)		
43A1	LL	148	-107	30.7	10.6
	DL	-152	-121	28.8	10.6
45A3	LL	138	-128	31.2	10.2
	DL	-123	-138	28.8	10.1
53A6	LL	145	-128	31.2	10.1
	DL	-135	-119	28.8	10.3

CONCLUSION

The free energy difference between heterochiral and homochiral dialanine was obtained by experiment as well as by computer simulation. Experimentally, the formation of the homochiral dipeptide seems to be preferred by up to 0.22 kJ/mol over the heterochiral peptide. Even though this is a small value, it might indicate a possible thermodynamic explanation for the propagation of chirality during the evolution of life. Alanine is the smallest chiral amino acid and future investigations will include larger amino acids in order to determine whether the homochiral preference is a universal phenomenon for all peptides.

Computationally, five different pathways and three different force-field parameter sets have been used to determine the free energy difference. All but one pathway do reproduce the experimental result within the statistical error. Considering the statistical error and the smoothness of the free energy profiles, a pathway involving a planar intermediate seems to perform best. Using this pathway all three different force-field parameter sets reproduce the experimental value within the statistical accuracy. This shows that molecular dynamics simulations can be used to estimate the free energy difference between homo- and heterochiral dipeptides. Further investigations on more complex dipeptides are underway and will shed more light on the propagation of chirality.

Acknowledgement

M.C.Heeren is thanked for her contribution to the experimental work. This research work is supported by Delft University of Technology in the framework of DIOC (Delft Interdisciplinary Research Centre) Life Tech.

References

1. W.A. Bonner, in: P. Jolles (Ed.), *D-amino Acids in Sequences of Secreted Peptides of Multicellular Organisms*, Birkauer, Basel, 1998, p.159.
2. Y. Yamagata, *J. Theor. Biol.* 11 (1966) 495.
3. M. Quack, *Angew Chem Int Edit.* 41 (2002) 4618.
4. W.A. Bonner, *Origins Life Evol B.* 29 (1999) 615.
5. J.A. Jongejan and J.A. Duine, in: Y. Murakami (Ed), *Comprehensive Supramolecular Chemistry*, Elsevier Science Ltd., Oxford, 1996, p.473.
6. G. Wald, *Ann NY Acad Sci.* 69 (1957) 352.
7. W.A. Bonner, in: D.B. Cline (Ed.), *Physical Origin of Homochirality in Life*, American Institute of Physics, Woodbury and New York, 1996, p. 17.
8. R.D. Lins, T.A. Soares, and R. Ferreira, *Z Naturforsch C.* 51 (1996) 70.
9. T.A. Soares, R.D. Lins, R. Longo, R. Garratt, R. Ferreira, *Z Naturforsch C.* 52 (1997) 89.
10. S.L. Miller, *J Am Chem Soc.* 77 (1955) 2351.
11. 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, W.F. van Gunsteren, *J. Phys. Chem. A.* 103 (1999) 3596

12. W.F. van Gunsteren, S.R. Billeter, A.A. Eising, P.H. Hünenberger, P. Krüger, A.E. Mark, W.R.P. Scott, I.G. Tironi, in *Biomolecular Simulation: The GROMOS96 Manual and User Guide*, vdf, Zurich, 1996.
13. J.G. Kirkwood, *J Chem Phys.* 3 (1935) 300.
14. H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, J. Hermans, in: *Intermolecular Forces*, D. Reidel, Dordrecht, 1981, p.331.
15. H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. Dinola, J.R. Haak, *J Chem Phys.* 81 (1984) 3684.
16. J.P. Ryckaert, G. Ciccotti, H.J.C. Berendsen, *J Comput Phys.* 23 (1977) 327.
17. I.G. Tironi, R. Sperb, P.E. Smith, W.F. van Gunsteren, *J Chem Phys.* 102 (1995) 5451.
18. T.N. Heinz, W.F. van Gunsteren, P.H. Hunenberger, *J Chem Phys.* 115 (2001) 1125.
19. T.C. Beutler, A.E. Mark, R.C. van Schaik, P.R. Gerber, W.F. van Gunsteren, *Chem Phys Lett.* 222 (1994) 529.
20. L.D. Schuler, X. Daura, W.F. van Gunsteren, *J Comput Chem.* 22 (2001) 1205.
21. X. Daura, A.E. Mark, W.F. van Gunsteren, *J Comput Chem.* 19 (1998) 535.
22. C.Oostenbrink, W.F. van Gunsteren, A. Villa, A.E. Mark, *J Comput Chem.* 25 (2004) 1656
23. G.G. Smith, R. Baum, *J Org Chem.* 52 (1987) 2248.
24. S. Saetia, K.R. Liedl, A.H. Eder, B.M. Rode, *Origins Life Evol B.* 23 (1993) 167.
25. J.E.Leffler, E.Grunwald, *Rates and Equilibria of Organic Reactions*, Wiley, New York, 1963.
26. W.F. van Gunsteren, in: W.F. van Gunsteren, P.K. Weiner (Eds), *Computer simulation of biomolecular systems, theoretical and experimental applications*, Excom Science Publishers, Leiden, 1989, p.27.
27. A.E. Mark, W.F. van Gunsteren, H.J.C. Berendsen, *J Chem Phys.* 94 (1991) 3808.
28. M.P. Allen and D.J. Tildesly, *Computer Simulation of Liquids*, Clarendon Press, Oxford, 1987.

COMPUTATIONAL STUDY OF GROUND STATE CHIRAL INDUCTION IN SMALL PEPTIDES: COMPARISON OF THE RELATIVE STABILITY OF SELECTED AMINO ACID DIMERS AND OLIGOMERS IN HOMOCHIRAL AND HETEROCHIRAL COMBINATIONS

SUMMARY

The relative stabilities of homochiral and heterochiral forms of selected dipeptides, AA, AS, AC, AV, AF, AD, AK, tripeptides, AAA, AVA, and an acetylpentapeptide, AcGLSFA, have been calculated using thermodynamic integration protocols and the GROMOS 53A6 force field. Integration pathways have been designed that produce minimal disturbance to the system, including the use of soft-atoms, low-energy intermediates, and chiral inversion of the smaller amino acid in the peptide. Comparison of the results obtained by thermodynamic integration between the diastereomeric forms (in explicit water, at 300 K) and from exhaustive global minimum-energy searches for the individual dipeptides (implicit water, $\epsilon = 78$, 0 K) suggests that entropic contributions to the relative stability of the chiral forms are important. This conclusion is supported by the results of explicit calculation of the effect of temperature on the relative stability of alanylvalylalanine diastereomers. The Gibbs free energy calculations predict that at ambient temperature and pressure homochiral dipeptides with small side chains or polar groups in the vicinity of the peptide backbone, AA, AS, AD, are more stable than their heterochiral counterparts by fractions of a kJ/mol. For bigger side chains, AC, AV, AF, AK, the heterochiral diastereomers appear to be more stable. Predicted relative stabilities are in line with observations reported in the literature for AE, and YY. Excellent agreement is found for the calculated and experimentally determined relative stabilities of the diastereomers of the dipeptide AA and of all-*L* AcGLSFA and its diastereomer containing *D*-serine in the central position. Addition of counter-ions to the solvent box has no significant effects on charged and neutral forms. From the present findings it would appear unlikely that the intrinsic stability difference between homo- and heterochiral dipeptides has been a driving force in a primordial selection process leading to the incorporation of amino acids with a single enantiomeric configuration in natural proteins.

Keywords:

Free energy (FE) calculation, thermodynamic integration (TI), global energy minimum (GEM) search, chiral induction, (oligo-) peptides, propagation of chirality, evolution

1. INTRODUCTION

The predominance of *L*-amino acids and *D*-sugars in natural biopolymers such as proteins, polysaccharides, and nucleic acids appears to be a fundamental characteristic of the biochemistry of life on Earth. Unraveling the causes of this preference might well contribute to our understanding of the evolutionary process. A number of hypotheses concerning the possibilities of symmetry-breaking events have been put forward.¹ Both random mechanisms, such as the

This chapter has been published in Journal of Computational Chemistry 2006, 27, 857-867.

spontaneous formation of single-enantiomer crystals or chance contacts with abiogenetic chiral materials, and determinate mechanisms, drawing on an intrinsic physical property, have been considered. In particular, the parity-violating character of the electroweak force generating an energy difference between the enantiomers of amino acids and sugars under otherwise isotropic conditions, has received much attention.^{2,3,4} Whereas these propositions aim to identify possible causes for the initial disturbance of the racemic balance in favor of the presently observed preferences, their magnitude appears to be far too small to explain the occurrence of e.g. homochiral peptides and proteins solely by thermodynamic control.⁵ Hypotheses invoking kinetic control of the propagation of chirality leading to homochiral polymers appear to be more successful.^{6,7,8} Clearly, catalytic (pseudo-) steady state feed back models⁷ show that a tiny initial disturbance of the racemic balance of the monomers from which the catalyst is synthesized will be sufficient to drive the system to either end of the spectrum. However, while these models rationalize the evolution of enzymes (amino acid polymers) or ribozymes (nucleic acids) with a chiral composition that matches their chiral preference, they offer no explanation for the all-*L* (amino acids) viz. all-*D* (ribose) signature of the present-day biosynthetic system. In this respect, it would be quite conceivable even within the highly evolved framework of the protein biosynthetic machinery to envision its construction from say 19 *L* amino acids + *D*-alanine with matching specificity. Probably, the mechanistic details of the currently established polymerization catalyst (ribosome) would appear to interfere with a similar chirality inversion of amino acids that possess a more demanding side chain. However, a “chicken-or-egg” conundrum still remains! As early as 1957, Wald suggested an additional thermodynamic stability criterion to be involved alongside the kinetic control. Oligopeptide secondary structure would dictate the selection of one amino-acid enantiomer rather than its antipode during polymer growth.⁹ This hypothesis has since been validated both by experiment⁶ and by simulation^{10,11} for oligopeptides that show defined secondary (helical) structure. Again, this offers no explanation for oligo’s that do not adopt such discriminating structures. As an example we mention the acetylated all-*L* pentapeptide AcGLSFA, discussed below, in which case the diastereomer containing *D*-serine in the central position has been found to be the more stable species under physiological conditions. This prompted us to investigate the thermodynamic stabilities of even smaller fragments, mainly alanine-containing dipeptides.

Previously, we reported that the solid-phase synthesis of dialanine peptides from racemic mixtures of *L*- and *D*-alanine leads to a slight bias towards the formation of the homochiral diastereomers.¹² Combined with the results reported in the literature for the formation of dialanine after several cycles of evaporation and dilution in a simulation of salt-induced peptide synthesis under possibly prebiotic conditions¹³, we concluded that homochiral (*L,L* or *D,D*) dialanine is more stable than its heterochiral counterpart (*L,D* or *D,L*) dialanine by fractions of a kJ/mol. Efforts to reproduce these small differences by computational methods showed that reliable and efficient protocols could be designed.¹² Here, we describe the application of these methods to a selection of alanine-containing di-, and tripeptides.

2. METHODS

2.1 Molecular models

Interaction energies for all peptides under investigation were taken from the GROMOS 53A6 force field.¹⁴ This is a united atom force field in which aliphatic hydrogen atoms are treated as a single interaction site together with the carbon atom to which they are attached. Global minimum energy conformations were obtained in implicit water with an effective relative dielectric constant of 78. Free energy calculations were performed in explicit SPC water.¹⁵

2.2 Global minimum energy search

Overall potential energy minima were obtained for all investigated dipeptides from an exhaustive search method. Dihedral angles along all rotatable bonds were assigned 3 to 6 discreet values, after which peptide conformations for all

combinations of dihedral angle values were generated for the two diastereomers. All conformations were subsequently energy minimized using the steepest descent option of the GROMOS96 program^{16,17} with an energy cutoff of 0.01 kJ/mol. Lowest energy conformations were selected.

2.3 Free energy calculation

All simulations were performed with the GROMOS96 simulation package^{16,17}. Using the thermodynamic integration (TI) formula,¹⁸ the free energy difference, ΔG_{BA} , is calculated between state A and state B,

$$\Delta G_{BA} = \int_{\lambda_A}^{\lambda_B} \left\langle \frac{\partial H(r, p, \lambda)}{\partial \lambda} \right\rangle_{\lambda} d\lambda$$

$H(r, p, \lambda)$ is the Hamiltonian describing the system as a function of the molecular positions, r , conjugate momenta, p , and a coupling parameter λ . Initially, we used the Hamiltonian to describe the homochiral peptide if $\lambda = \lambda_A = 0$ and the heterochiral peptide if $\lambda = \lambda_B = 1$. The angular brackets indicate an ensemble average at a specific value of λ . The integration is then performed by changing the value of λ from λ_A to λ_B in a number of discrete steps at which the ensemble average is collected.

Different pathways connecting states A and B can be chosen, which will all lead to the same value of ΔG_{BA} in the limit of infinite sampling. Guided by the results of previous experiments we chose a pathway that involves the perturbation of the improper dihedral involving the C_α atom of one of the residues from 35.26° (*L*) to -35.26° (*D*)¹². To reduce the strain in the bond angles at intermediate values of this improper dihedral, we inserted an intermediate state in which the minimum energy value of the improper dihedral is set to 0° and the minimum energy values of the three bond angles around the C_α atom are set to 120° . Such a split pathway, in which state A and state B alternately describe the starting (homochiral) peptide and the intermediate structure *viz.* the intermediate and the final inverted (heterochiral) peptide, has been shown to improve the accuracy of the calculations considerably¹². In the dipeptides studied, the chirality of the alanine moiety was inverted during the free energy perturbation. In the case of the Ala-Val peptide, the effect of changing the chirality of the valine was also investigated in order to compare the two pathways.

2.3.1 MD settings

Peptides (COO^- and NH_3^+ at the termini) in a right-handed α helical conformation ($\varphi = -57^\circ$, $\psi = -47^\circ$) were solvated in (periodic) truncated octahedral boxes containing 1212 to 1830 simple point charge (SPC) water molecules.¹⁵ To check the effects of added counter-ions for Ala-Asp (deprotonated γ -carboxyl group) and Ala-Lys (protonated ϵ -amino group), 8 Na^+ and 7 Cl^- , and 7 Na^+ and 8 Cl^- , were added to the box during solvation of Ala-Asp and Ala-Lys, respectively. This is equivalent to a salt concentration of about 0.15 M. The systems were energy minimized using steepest descent minimization, and then gradually heated to 300 K in three 100 K intervals of 10 ps at constant volume using a weakly coupled temperature bath with relaxation time 0.1 ps.¹⁹ For Ala-Val-Ala, the system was heated up to final temperatures of 300 K, 325 K and 350 K, allowing us to deduce enthalpic and entropic contributions to the free energy difference. Another 20 ps of equilibration was performed at a constant pressure of 1 atm using weak coupling to a pressure bath with a relaxation time of 0.5 ps and an estimated isothermal compressibility of $4.575 \cdot 10^{-4} \text{ (kJ mol}^{-1} \text{ nm}^{-3})^{-1}$.¹⁹ At every λ -value of the TI simulations, 160 ps equilibration was followed by 480-1260 ps of data collection. Bond lengths were constrained (with relative precision of 10^{-4}) at the minimum energy values using the SHAKE algorithm²⁰ allowing for a time step of 2 fs. Long-range interactions were calculated using a triple range cut-off scheme. All interactions within

0.8 nm were calculated every time step from a pair list that was updated every fifth step. Interactions between pairs that were between 0.8 and 1.4 nm were calculated at pair list (re)construction and kept constant between pair list updates. A reaction-field contribution²¹ was added to the electrostatic interactions to account for a homogeneous medium outside the long range cutoff with a relative permittivity of 62.²² The error in the value of $\langle \partial H / \partial \lambda \rangle$ was calculated using a block averaging procedure.²³ All calculations were performed in parallel mode on a 14 nodes Linux-based cluster with 32-64 bit extension dual Intel CPU processors. A complete integration from homochiral to heterochiral (~10ns) for the peptides takes about 24-30 hours.

3. RESULTS AND DISCUSSION

3.1 Systems investigated

With few exceptions²⁴ free energy calculations on di- and tripeptides reported in the literature refer to molecules blocked at both termini. In this study we focus on the zwitterionic forms of some small peptides. We chose to investigate seven dipeptides, AA, AS, AC, AV, AF, AD, AK, all containing alanine at the *N*-terminus for reasons of computational economy. Of these, the dialanine served as a reference for which data have been reported earlier.¹² Alanylserine, AS, was chosen both because serine is the simplest amino acid with an additional functional group, and because serine has been implied as a chirality “starter” in primordial biochemistry.²⁵ Alanycysteine, AC, has been included because the cysteine moiety has a heavy atom distribution around the C_{α} , that is different from that of the other amino acids (as reflected by its different chirality label in the Fischer (F) and in the Cahn, Ingold, Prelog (CIP) nomenclature system). Apparently, evolution of homochirality in proteins followed chemical directives (amino, carboxyl, (substituted-) alkyl group) used in the Fischer convention, rather than the physical property (mass) that forms the basis of the CIP system. Alanylvaline, AV, and alanylphenylalanine, AF, represent combinations of amino acids with bulky side chains. In alanylaspargate, AD, and alanyllysine, AK, a side chain is introduced of which both the protonated and the deprotonated forms have been included in this study. Trialanine, AAA, the smallest peptide that has been reported to be able to adopt an α helical structure in its homochiral form based on physico-chemical evidence^{26,27,28,29} and on the results of molecular simulations,^{30,31} has been included to study the effect of inverting the chirality of the central alanine. Subsequently, we investigated alanylvalylalanine, AVA, to determine whether the effect of introducing valine as a bulky residue might magnify the effects of a *D*-amino acid in that position, while keeping the system as small as possible. Finally, the pentapeptide AcNH-Gly-Leu-Ser-Phe-Ala has been included as the single case where a free energy difference can be extracted from experimental data on the enzyme-catalyzed racemization of the central residue.³²

3.2 Free energy calculations

The calculated Gibbs free energy differences between the two diastereomeric forms of the peptides investigated are collected in Table 1. The values are quite small (of order 1 kJ/mol), challenging the accuracy of the force field and of the algorithms that are used in the calculation. To the best of our knowledge, comparable data from physico-chemical experiments exist only for two of the systems considered here. For those cases, alanylalanine and the pentapeptide, agreement between simulation and experiment is obtained. For two additional compounds, alanylglutamate, AE, and tyrosyltyrosine, YY, a slight preference can be deduced from reports on the composition of what appears to be an equilibrium mixture. Comparing the peak heights of the NMR signals in the spectrum of AE after treatment with the glutamate racemizing enzyme L-Ala-D/L-Glu epimerase³³ points at a slight excess (53:47) of the homochiral peptide over the heterochiral peptide corresponding to G(hetero) - G(homo) = + 0.3 kJ/mol, in line with the value calculated

here for the AD peptide. In contrast, the average composition of reaction mixture obtained after activation of D,L-tyrosine with CO on (Ni,Fe)S surfaces³⁴ shows the heterochiral dipeptides to be present in excess of the homochiral compounds, to an amount not unlike the value calculated for e.g. AF in this study ($G(\text{hetero}) - G(\text{homo}) = -0.6$ kJ/mol for YY and -0.9 kJ/mol for AF respectively). Since the limited size of the reference data set precludes proper calibration of the computational protocol, we rely on the quality of the integration results, as discussed previously.¹² The value of $\langle \partial H / \partial \lambda \rangle$ as a function of λ for selected peptides is given in Figure 1. The two panels indicate the free energy calculation from the initial state (e.g. homochiral) to the planar intermediate state (left panel) and onward to the final state (e.g. heterochiral, right panel). It should be noted that the value of $\langle \partial H / \partial \lambda \rangle$ does not need to be continuous between the two panels since on entering and leaving the intermediate state, the derivative points in different directions on the potential energy surface. As is shown in these graphs, profiles of reasonable roughness were obtained. In some cases we allowed considerable time (up to 1.3 ns) per λ -value for the peptides to sample between different conformations in order to reach sufficiently converged values of $\langle \partial H / \partial \lambda \rangle$. This is illustrated in Figure 2, in which the ϕ - ψ probability maps for the dipeptide AA are given at different points in the conversion. A gradual transition of the most occupied regions can be observed, and it is clear that in order to obtain sufficiently converged estimates of $\langle \partial H / \partial \lambda \rangle$ in the planar intermediate state (i.e. panel 2c), the peptide needs to sample both regions of the Ramachandran map. The effect of the pathway chosen on the statistical and empirical error (correspondence to experimental values) has been the subject of the previous study.¹² There, we showed that the insertion of a planar intermediate state during the TI simulations leads to more accurate results. Considering the fact that the value of the free energy difference is not affected by the simultaneous inversion of all centers (apart from minute parity violation effects), either the *N*-terminal, or the *C*-terminal residue may be chosen for the inversion, theoretically resulting in the same free energy difference in the mirror world. However, as we show for the dipeptide AV in Figure 1, the $\langle \partial H / \partial \lambda \rangle$ profile is smoother when inverting the Ala chiral center than when inverting the Val chiral center. This can be rationalized by considering the smaller disturbance introduced to the system when the Ala center is inverted, which enables the system to explore and sample phase space more efficiently. It has been shown that considerable inaccuracies may occur for high-energy conformations that are hardly sampled by the unbiased simulation.³⁰ As a rule of thumb, we aim to keep the variation in the $\langle \partial H / \partial \lambda \rangle$ values as low as possible in order to obtain as accurate results as possible. Even though the calculated free energy differences are small, the calculated error estimates indicate that the results seem conclusive about the preference for homochiral or heterochiral peptides. Free energy differences of ~ 0.5 kJ/mol would in the case of peptide bond formation in a completely achiral environment lead to homochiral / heterochiral compositions of approximately 55 % / 45 %, which may already be sufficient to explain a homochiral preference in the primordial soup. However, as becomes clear from Table 1, several dipeptides show preference for the heterochiral rather than the homochiral configuration.

Table 1. Global energy minima ($\Delta V = V(\text{hetero}) - V(\text{homo})$) and free energy ($\Delta G = G(\text{hetero}) - G(\text{homo})$) differences between heterochiral and homochiral peptides. V is the global energy minimum of the peptides in implicit solvent, ΔG values are obtained from explicit solvent simulations. Statistical errors are calculated using a block averaging according to ref 22. D-amino acids are indicated by underlining. Additions to the peptide name have the following meaning: (reverse) the simulations were performed in reverse direction, i.e. from heterochiral to homochiral; (deprot) corresponds to the deprotonated form; (prot) corresponds to the protonated form; (counterion) explicit counter ions were included at an ionic strength of 0.15 M; (300K, 325K, 350K) simulations performed at indicated temperature. Dipole moments and occurrence of hydrogen bonds (as % of simulation time) as averaged after the end-point simulations of the homochiral and heterochiral peptides are presented.

Peptide	Simulation		ΔV (kJ/mol)	ΔG (counter ions) (kJ/mol)	Experiment	Simulation		H-bond (counterion)		ref
	V (kJ/mol)				DG (kJ/mol)	Dipole (counterion) (D)		H-bond (counterion) (%)		
	Homo	Hetero				Homo	Hetero	Homo	Hetero	
<u>AA</u>	-11.1	-11.3	-0.2	0.2 ± 0.1	0.22	31	29	10.1	10.3	13,12
<u>AS</u>	-12.2	-12.4	-0.2	0.4 ± 0.4	n.a. ^{a)}	31	28	12.6	12.5	b)
<u>AV</u>	-9.0	-8.8	0.2	-0.4 ± 0.3	n.a.	31	29	10.2	10.4	b)
<u>AV</u>	n.a.	n.a.	n.a.	18 ± 1.2	n.a.	n.a.	n.a.	n.a.	n.a.	b)
<u>AC</u>	-12.0	-12.3	-0.3	-0.6 ± 0.4	n.a.	34	34	10.3	10.6	b)
<u>AF</u>	-18.3	-19.1	-0.8	-0.9 ± 0.4	n.a.	31	30	10.5	10.9	b)
<u>AD</u> (deprot)	-14.4	-14.2	0.2	0.5 ± 0.3 (0.8 ± 0.4)	n.a.	37 (38)	38 (39)	17.2 (16.8)	17.0 (17.0)	b)
<u>AD</u> (prot)	-15.7	-16.3	-0.6	0.3 ± 0.4 (0.2 ± 0.3)	n.a.	30 (30)	28 (28)	11.3 (12.7)	12.0 (12.9)	b)
<u>AK</u> (deprot)	-13.3	-15.0	-1.7	-0.7 ± 0.4 (-0.4 ± 0.4)	n.a.	29 (31)	29 (28)	12.3 (12.6)	12.8 (12.8)	b)
<u>AK</u> (prot)	-13.9	-15.1	-1.2	-0.4 ± 0.3 (-0.4 ± 0.4)	n.a.	29 (32)	32 (32)	12.8 (12.8)	12.7 (12.8)	b)
<u>AAA</u>	n.a.	n.a.	n.a.	-0.9 ± 0.5	n.a.	47	34	12.4	12.3	b)
<u>AVA</u> (300K)	n.a.	n.a.	n.a.	2.3 ± 0.5	n.a.	45	33	13.2	12.9	b)
<u>AVA</u> (325K)	n.a.	n.a.	n.a.	-0.3 ± 0.6	n.a.	43	34	11.7	11.8	
<u>AVA</u> (350K)	n.a.	n.a.	n.a.	-2.4 ± 0.6	n.a.	46	37	11.4	11.5	
<u>GLSFA</u>	n.a.	n.a.	n.a.	-0.5 ± 0.6	-0.6	43	32	18.3	16.9/ /0.2 ^{c)}	32

- a) n.a: not available or not accessible
b) this work
c) Intralolute hydrogen bond formation

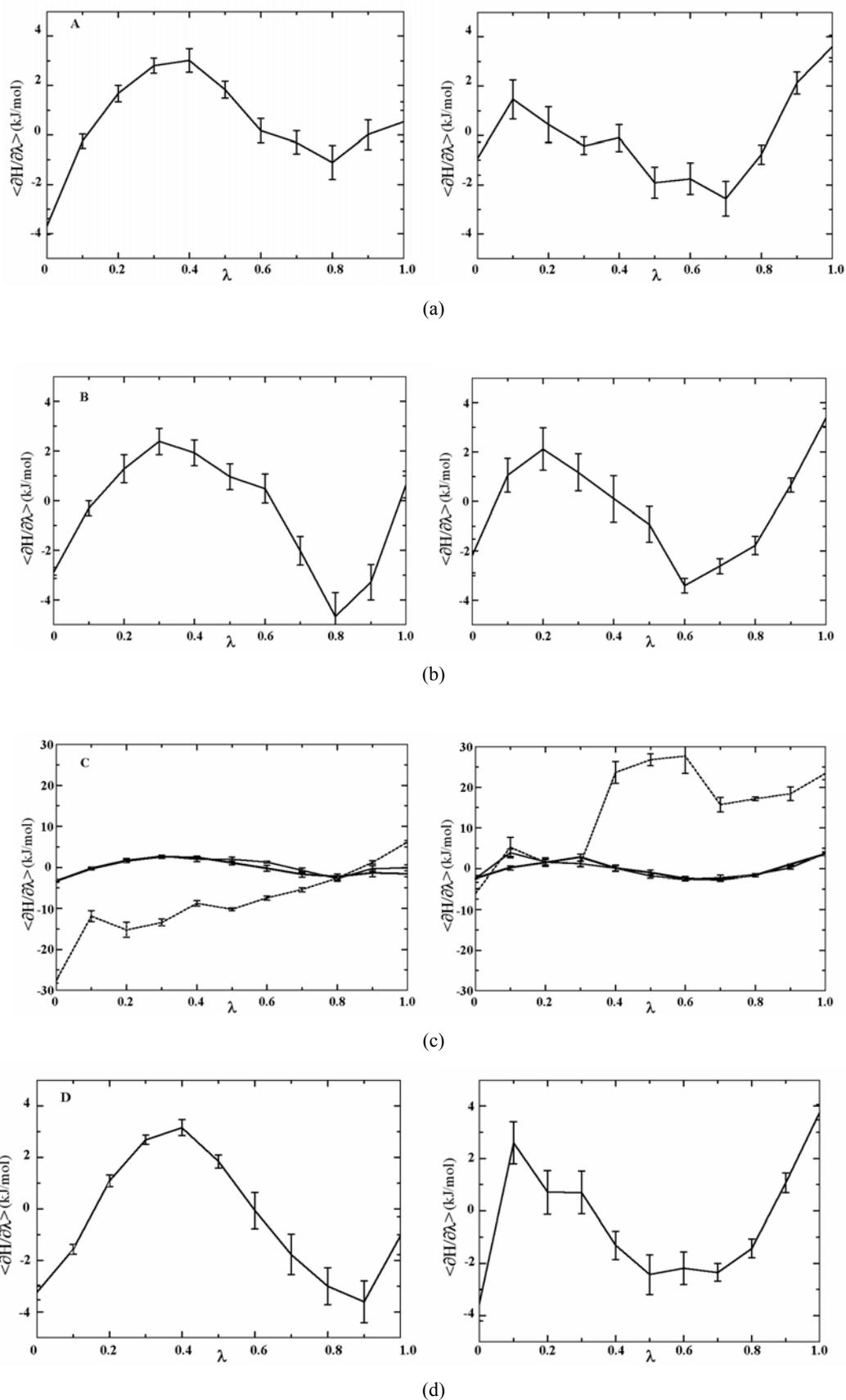
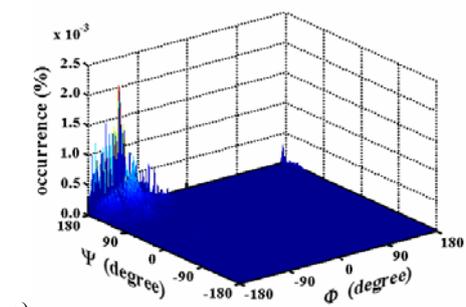
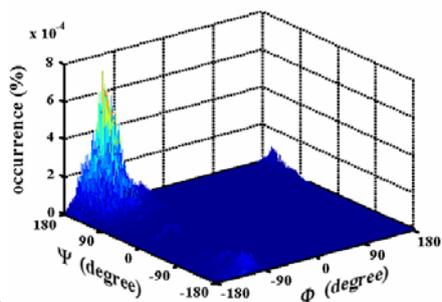


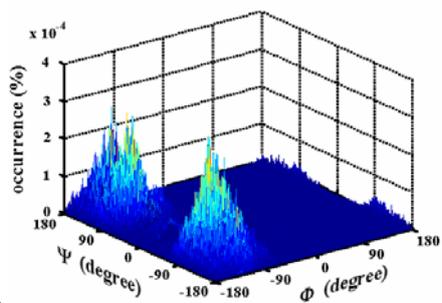
Figure 1: Free energy profiles for selected residues as a function of the scaling parameter λ . Free energy derivatives were calculated as the sum of two processes 1) thermodynamic integration from initial (homochiral) state to a planar intermediate state (left panels) and 2) thermodynamic integration from the planar intermediate state to the final (heterochiral) state (right panels). Profiles are shown for (a) AS, (b) AC, (c) AV mutating both the A-side forward (solid bold line), backward (solid thin line) and the V-side (dotted line) and d) AK.



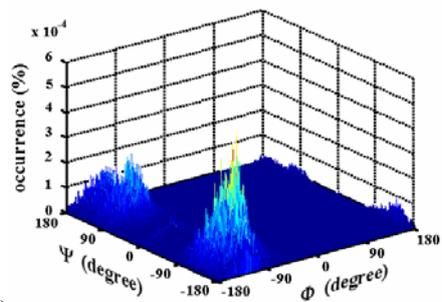
a)



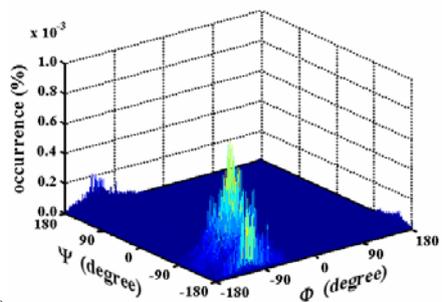
b)



c)



d)



e)

Figure 2: Ramachandran map of the ϕ, ψ distribution for the AA dipeptide at different λ -points of the simulation. ϕ, ψ -occurrences relate to 200 fs sampling of 160 ps simulation. a) initial homochiral state, b) halfway the thermodynamic integration towards the planar intermediate state, c) at the planar intermediate state, d) halfway the thermodynamic integration towards the final (heterochiral) state, e) the final, heterochiral state.

3.3 Counter-ions

The effect of counter-ions on peptide and protein conformation is not well understood. It has been argued by some authors that the effect is negligible, while others stress the importance of counter-ions in biophysical systems.^{35,36} It seems that depending on the system and conditions used in the study, the effect of adding counter-ions may vary considerably. We investigated the effect of addition of counter-ions (0.15 M) on the relative stability of homochiral vs heterochiral AD and AK in the case of both protonated and deprotonated side chains. The incorporation of counter-ions appears to have little effect in the case of the side chain protonated forms of AD (neutral) and AK (positively charged). In the deprotonated forms (AD, negatively charged, AK neutral) a small effect is observed (see Table 1). Donnini observed that the counter-ions had a significant effect on the mutation of 2-phosphoglycolic acid (PGA) to 3-phosphonopropanoic acid (3PP).³⁶ This effect is correlated with the short average distance of 0.22 nm between the ions and the PGA and 3PP molecules, which carry strongly negatively charged groups. We analyzed the distance between the peptide molecules and the closest ions and found that the latter are not attracted to the surroundings of the peptides. For the dipeptides, the positive and negative groups at the termini and the side chain appear to balance one another, reducing the net charge. As a result, the average distance between counter-ions and peptides is over 1 nm.

3.4 Global minimum energies vs. Free energy differences

The difference in minimum *potential intramolecular energy* as compared to the *free energy* difference between the diastereomeric forms of the peptides, is shown in Table 1. Due to the exponential increase of the number of conformations that need to be taken into account using exhaustive search methods, comparisons are made for dipeptides only. It appears that the energy difference observed for the global minimum-energy structures obtained with implicit solvation does not follow the free energy difference, either calculated using explicit solvent or experimentally determined, in all cases studied here. Agreement is found between the two approaches for AF, AC, AD (deprotonated) and AK but the different approaches yield different stabilities for AA, AS, AV and AD (protonated). The fact that results between the two approaches do not agree in 50 % of the cases indicates that entropy and details of solvation play an important role in the relative stability of the diastereomers. This is further substantiated by the investigation of the temperature dependence of the free energy difference for AVA. This peptide was selected for further entropy calculations because it shows the largest free energy difference observed in this study. A distinct effect of the simulation temperature on the free energy difference can, indeed, be observed for this peptide. At 300 K, the all-*L* form is more stable ($G(\text{hetero}) - G(\text{homo}) = +2.3 \text{ kJ/mol}$) than that of the molecule with the central valine in a *D*-configuration. At 350 K, the situation is reversed, with a higher stability for the heterochiral form (see Figure 3). From the temperature derivative of ΔG in this figure, the relative entropy difference between the two diastereomers is estimated to be $\Delta S = 93 \text{ J K}^{-1} \text{ mol}^{-1}$, which indicates that the entropic and enthalpic contributions to the difference in stability at 300 K are 27.9 kJ/mol and 30.5 kJ/mol respectively.

The relationship between potential energy and free energy landscapes has recently been studied by Wales and others.^{37,38,39} It was shown that the increased conformational flexibility (higher vibrational entropy) can make an important contribution to the free energy and sometimes outweighs the energetic benefit of forming hydrogen bonds.

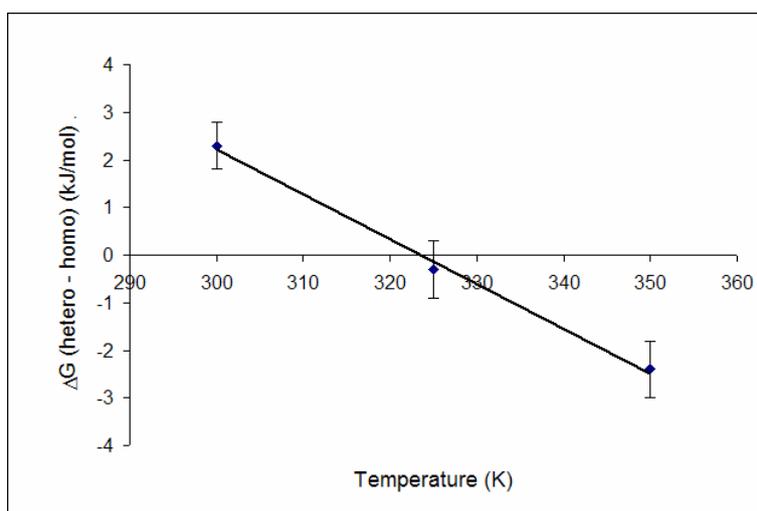
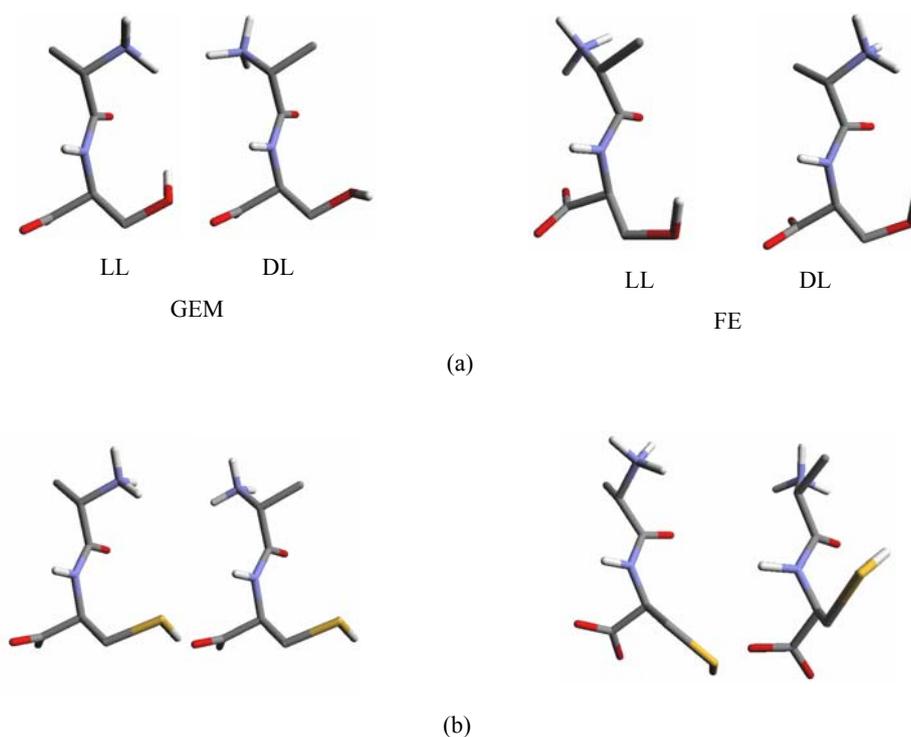


Figure 3: Temperature dependence of the free energy difference between heterochiral and homochiral AVA tripeptide. A least squares fit on these points results in a regression of $\Delta G = -0.093 T \text{ (K)} + 30.50 \text{ kJ/mol}$

The difference in the calculated result between the global energy minimum (GEM) and free energy (FE) approaches can be explained in part from the different solvent models used. In the GEM approach implicit water is used while in FE we used explicit water. For zwitterionic dipeptides that form strong hydrogen bonds with solvent water molecules, one may expect significant differences between the implicit and explicit inclusion of the solvent environment due to specific solvation cages that stabilize conformations other than the global minimum-energy one in an implicit solvent environment. However, inclusion of the solvent degrees of freedom in a global minimum energy search is impossible and makes no sense. In the explicit solvent calculations no signs of persistent solvent cages were observed in the course of the simulations. Figure 4 presents the global minimum conformations of the LL and DL forms together with the most observed conformations from the MD simulations for selected peptides (see below).



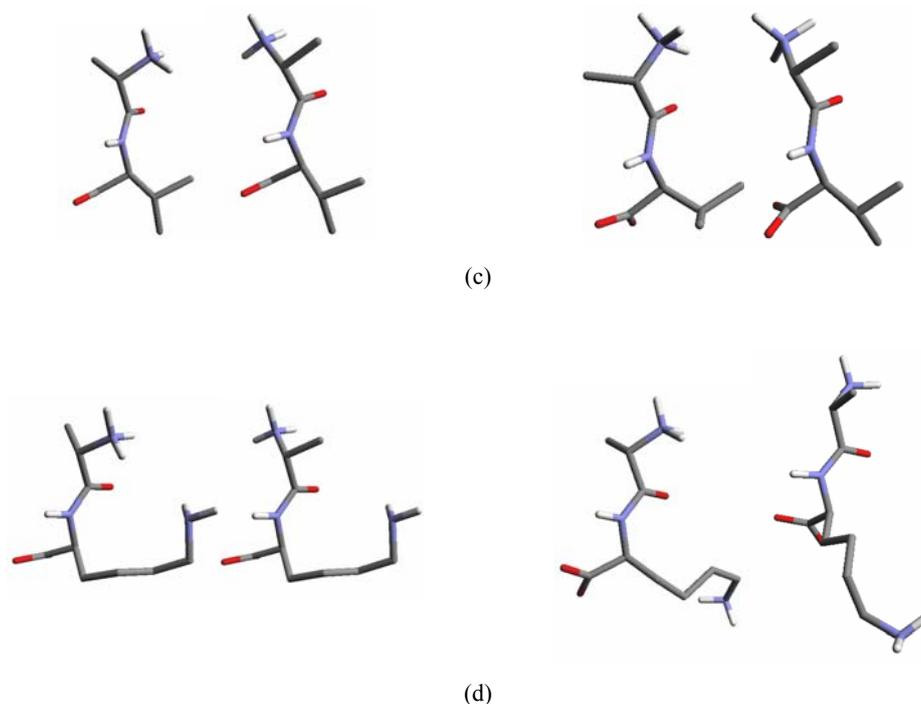


Figure 4: Global minimum (solute) potential energy structures (GEM) and most commonly observed conformations from the free energy (FE) simulations in homochiral and heterochiral configurations for selected dipeptides (a) AS (b) AC (c) AV (d) AK

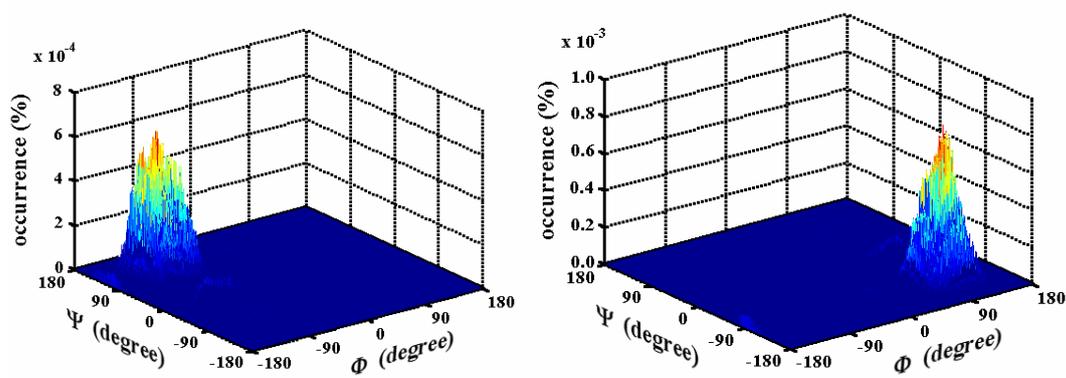
3.5 Free energy difference and structure

As argued above, both enthalpic and entropic effects contribute to the free energy of the system. Clearly, neither of these effects can be related in a straightforward way to the structural characteristics of the dissolved peptide alone. Thus, the enthalpy is dependent on the intrasolute and solute-solvent interactions (as well as on changes in the solvent-solvent enthalpy, which we estimate to be negligible in these cases) while the configurational freedom of the systems determines their entropy. Unfortunately, this means that it is not obvious to recognize general trends that make one peptide prefer the heterochiral form, while others prefer the homochiral form. The time-averaged number of solute-solvent hydrogen bonds for the homochiral and heterochiral forms, listed in Table 1, is within 0.7% except for AD (prot) and GLSFA. In addition, these small changes do not seem to correlate with the calculated free energies. Also the differences in average solute dipoles, which seemed indicative for the AA peptide in the previous study¹² do not explain the differences between the peptides observed here. From the simulations, we have extracted the most common structures in the homochiral and heterochiral states using a conformational clustering algorithm over the simulation trajectories.⁴⁰ Representative examples are shown in Figure 4. Interestingly, for almost all peptides the most sampled heterochiral conformation is characterized by the N- and C-termini pointing to the same side of the molecule, while in the homochiral case, the most sampled conformation shows the two termini pointing away from each other. In the case of bulky side chains of low polarity (AF, AV and AC), the enthalpic effect from the slightly reduced distance between the termini, resulting in a more favorable Coulombic intramolecular interaction, explains the preference for the heterochiral configurations. For smaller side chains (AA) or side chains that interact more strongly with water, the conformational variability of the peptide (and influence of solvent thereon) plays a more significant role. This is supported by the results of Monte Carlo simulations of the alanine dipeptide, reported by Drozdov, where at least sixteen conformations were found to be populated to some extent in water. The peptide-solvent interactions counterbalance the intrapeptide electrostatic interaction and thus obscure underlying conformational preferences that are a consequence of optimizing the intrapeptide energy only.⁴¹

Interestingly, the protonation state of the side chains does not seem to influence the diastereomeric preference. In all combinations of protonation state and ionic strength, AD seems to prefer the homochiral configuration, while AK prefers a heterochiral configuration. It is not unlikely that due to the relatively long side chain of lysine the strongly interacting group is on average relatively far removed from the peptide backbone. The bulky non-polar side chains of AV, AC and AF, on the other hand, show a similar preference for the heterochiral configuration. In addition, it appears that the experimental data reported for tyrosinylytyrosine, YY, lend support to this finding.³⁴

The most common conformations extracted from the MD runs of the homochiral alanylserine- and alanylcysteine peptides (see Figure 4) are rather similar with respect to the position of the side chain oxygen and sulfur atoms, respectively, relative to the cysteine C_{α} . They also compare well with the global minimum energy structures. For the heterochiral AC, however, cluster analysis shows a different most preferred conformation. It is not clear whether this has a bearing on the fact that our calculation suggests AC to favor “homochirality” in the Cahn, Ingold, Prelog sense (S-alanyl-S-cysteine being more stable than R-alanyl-S-cysteine) whereas it prefers heterochirality in the sense used throughout this report. In this respect, the results of relativistic parity nonconserving calculations reported by Wesendrup and coworkers⁴ on serine and cysteine add to the confusion. They estimate R-serine and S-cysteine to be slightly more stable than S-serine and R-cysteine, respectively. It should be noted, however, that these calculations were based on gas phase conformations of serine and cysteine that are substantially different.

The tripeptides AAA and AVA offer additional complexity, as shown in Table 1. Extending the peptide AA by one Ala, inverts the chiral preference. While AA shows a slight preference for a homochiral configuration, AAA rather exists in a heterochiral configuration. Upon extending AV by one Ala, the reciprocal situation holds: AV shows preference for the heterochiral configuration, while AVA prefers the homochiral form. The simulations at different temperatures for AVA provide more information on the energy and structure correlation. The expected increase in conformational flexibility with increasing temperature seems to be more pronounced for the heterochiral case as can be seen from the Ramachandran maps in Figure 5. At 350 K, the heterochiral form gains access to the PII region ($\phi \sim -70^{\circ}, \psi \sim 130^{\circ}$,) of the Ramachandran map and this increased flexibility (entropy) results in a decrease in the relative free energies as is shown in Figure 3. The linearity of this plot indicates that realistic effects are captured. Enthalpy-entropy compensation, EEC, as found here, is commonly observed in biological systems.⁴²



(a)

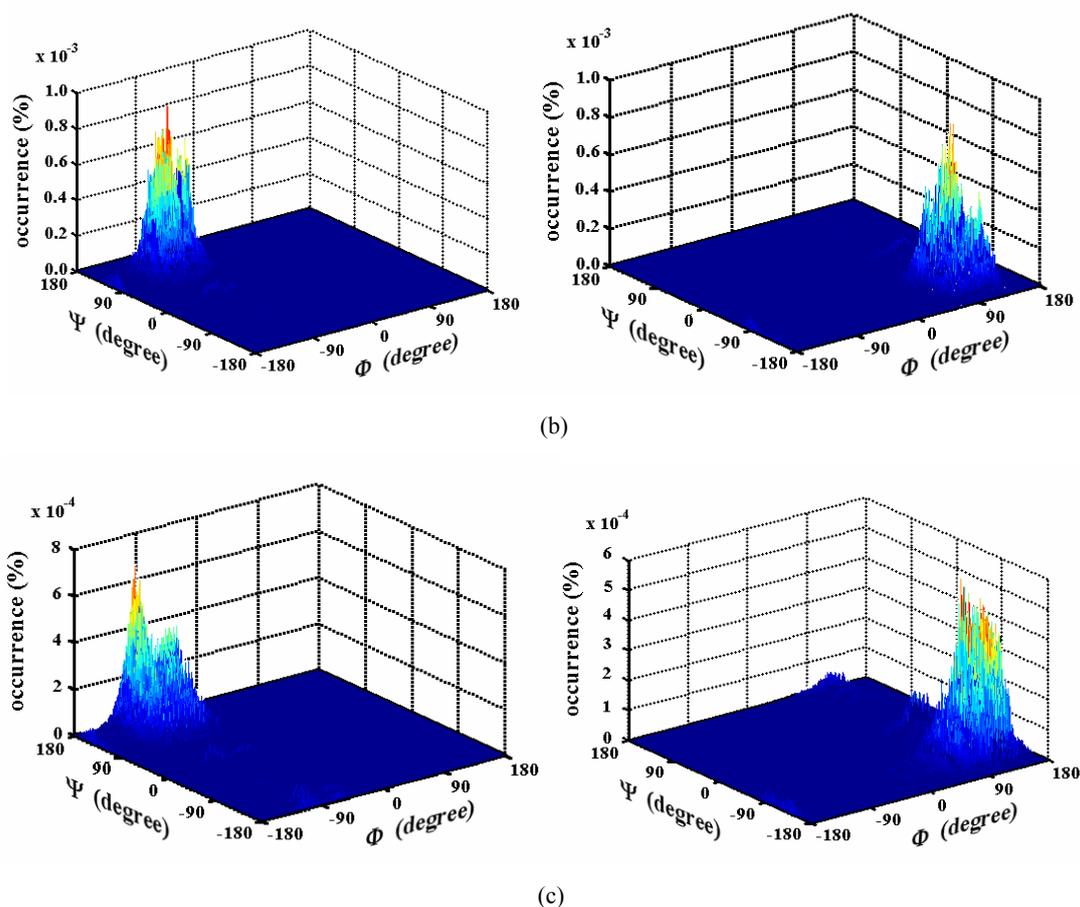


Figure 5: Ramachandran map for the ϕ, ψ distribution of the dihedral angles in the Valine residue of the tripeptide AVA at different temperatures. a) 300K, b) 325K, c) 350K (left: homochiral (valine) peptide; right: heterochiral (valine) peptide)

3.5 Pentapeptide

The natural presence of *D* amino acids in peptides and proteins is an interesting phenomenon both from the point of view of structure-function studies and for pharmaceutical applications. Investigation of the currently identified number of 492 cases of *D* amino acids distributed among 148 separate PDB entries shows that they are almost always found in fairly short-chain polypeptides, with 13 occurrences in chains of more than 20 residues, and only 1 in a chain longer than 42 residues.⁴³

While for the dipeptides and tripeptides studied here we did not detect the formation of intramolecular hydrogen bonds, we do observe such hydrogen bonds in the pentapeptide. The time-averaged total (inter- plus intramolecular) occurrences of hydrogen bonds in the homo- and heterochiral forms are 18.3% and 16.9%, respectively, while the intramolecular hydrogen bonds account for 0.2% and 0.9%, respectively. Overall, the heterochiral peptide has a lower free energy, in spite of the relative large loss of overall hydrogen bonds, but in agreement with a slightly increased occurrence of intramolecular hydrogen bonds. The calculated free energy agrees well with the value that can be deduced from the reported kinetic constants¹², using $\Delta\Delta G = -RT \ln K_{eq} = -RT \ln \frac{(k_{cat} / K_M)_{L \rightarrow D}}{(k_{cat} / K_M)_{D \rightarrow L}}$

CONCLUSION

We have presented calculations of the free energy differences between homochiral and heterochiral peptides for 10 molecules, of varying size and character. Good agreement with scarce experimental data was obtained, especially when taking into account the fact that free energy differences of the order of 1 kJ/mol are calculated. By comparing the results obtained from global energy minimization calculations and by evaluating the temperature dependence of the free energy, it has been shown that inclusion of entropy is of vital importance to obtain reliable results.

An attempt has been made to correlate the predicted chiral preference for specific diastereomers with molecular properties, such as the average occurrence of hydrogen bonds, the size of the peptide dipole and particular intrasolute interactions. However, in line with the observed importance of entropic contributions straightforward correlations appear to be elusive. One general observation to be made is that bulky, non-polar groups in the vicinity of the peptide backbone seem to result in a preference for heterochiral dipeptides.

Free energy calculations provide an appealing approach to investigate the relative stability of small peptides, and this research may give insight into the way in which homochirality was propagated in the evolution of biomolecules. This study predicts that homochiral peptides are not always favored over heterochiral peptides under the conditions tested. The preference depends on the size, character, accessible conformational space of the peptides and on temperature and solvent composition.

Reference

1. Bonner, W. A. D-amino acids in sequences of secreted peptides of multicellular organisms; Birkäuser: Basel, 1998, 159.
2. Yamagata, Y. J Theo Biol 1966, 11, 495.
3. Quack, M. Angew Chem Int Ed 2002, 41, 4618.
4. Wesendrup, R., Laerdahl, J. K., Compton, R. N., and Schwerdtfeger, P. J Phys Chem A 2003, 107, 6668.
5. Bonner, W. A. Origins Life Evol Biosphere 1999, 29, 615.
6. Bonner, W. A. American Institute of physics, Woodbury: New York 1996, 17.
7. Jongejan, J. A. and Duine, J. A. Comprehensive supramolecular chemistry; Elsevier Science: Oxford, 1996, 473.
8. Frank, F. C. Biochim Biophys Acta 1953, 11, 459.
9. Wald, G. Annals of the New York Academy of Sciences 1957, 69, 352.
10. Lins, R. D., Soares, T. A., and Ferreira, R. Z Naturforsch., C: Biosci. 1996, 51, 70.
11. Soares, T. A., Lins, R. D., Longo, R., Garratt, R., and Ferreira, R. Z Naturforsch., C: Biosci. 1997, 52, 89.
12. Zhou, Y., Oostenbrink, C., Van Gunsteren, W. F., Hagen, W. R., De Leeuw, S. W., and Jongejan, J. A. Mol Phy 2005, 103, 1961.
13. Saetia, S., Liedl, K. R., Eder, A. H., and Rode, B. M. Origins Life Evol Biosphere 1993, 23, 167.
14. Chris Oostenbrink, Alessandra Villa, Alan E. Mark, and Wilfred F. van Gunsteren J Comput Chem 2004, 25, 1656.
15. Berendsen, H. J. C., Postma, J. P. M., Van Gunsteren, W. F., and Hermans, J. Intermolecular Forces; Ed Pullman: Dordrecht, 1981, 331.

16. Van Gunsteren, W. F., Billeter, S. R., Eising, A. A., Hunenberger, P. H., Kruger, P., Mark, A. E., Scott, W. R. P, and Tironi, I. G. *Biomolecular Simulation: The GROMOS96 Manual and User Guide*; Hochschulverlag AG an der ETH Zurich: Zurich, 1996,
17. Scott, W. R. P., Hunenberger, P. H., Tironi, I. G., Mark, A. E., Billeter, S. R., Fennen, J., Torda, A. E., Huber, T., Kruger, P., and Van Gunsteren, W. F. *J Phys Chem A* 1999, 103, 3596.
18. Kirkwood, J. G *J Chem Phys* 1935, 3, 300.
19. Berendsen, H. J. C., Postma, J. P. M., vanGunsteren, W. F., Dinola, A., and Haak, J. R. *J Chem Phys* 1984, 81, 3684.
20. Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. C. *J Comput Phys* 1977, 23, 327.
21. Tironi, I. G., Sperb, R., Smith, P. E., and vanGunsteren, W. F. *J Chem Phys* 1995, 102, 5451.
22. Heinz, T. N., Van Gunsteren, W. F., and Hunenberger, P. H. *J Chem Phys* 2001, 115, 1125.
23. Allen, M. P and Tildesly D.J. *Computer simulation of liquids*; Clarendon Press: Oxford, 1987,
24. Knapp-Mohammady, M., Jalkanen, K. J., Nardi, F., Wade, R. C., and Suhai, S. *Chem Phys* 1999, 240, 63.
25. Takats, Z., Nanita, S. C., and Cooks, R. G. *Angew Chem Int Ed* 2003, 42, 3521.
26. Woutersen, S. and Hamm, P. *J Chem Phys* 2001, 114, 2727.
27. Woutersen, S. and Hamm, P. *J Phys Chem B* 2000, 104, 11316.
28. Schweitzer-Stenner, R., Eker, F., Huang, Q., and Griebenow, K. *J Am Chem Soc* 2001, 123, 9628.
29. Eker, F., Cao, X. L., Nafie, L., and Schweitzer-Stenner, R. *J Am Chem Soc* 2002, 124, 14330.
30. Mu, Y. G. and Stock, G. *J Phys Chem B* 2002, 106, 5294.
31. Mu, Y. G., Kosov, D. S., and Stock, G. *J Phys Chem B* 2003, 107, 5064.
32. Heck, Steven D., Faraci, W. Stephen, Kelbaugh, Paul R., Saccomano, Nicholas A., Thadeio, Peter F., and Volkmann, Robert A. *Proc Nat Acad Sci* 1996, 93, 4036.
33. Schmidt, DM Z., Hubbard, B. K., and Gerlt, J. A. *Biochem* 2001, 40, 15707.
34. Huber, C. and Wächtershäuser, G. *Science* 1998, 281, 670.
35. Omta, A. W., Kropman, M. F., Woutersen, S., and Bakker, H. J. *Science* 2003, 301, 347.
36. Donnini, S., Mark, A. E., Juffer, A. H., and Villa, A. *J Comput Chem* 2005, 26, 115.
37. Evans, D. A. and Wales, D. J. *J Chem Phys* 2003, 118, 3891.
38. Krivov, S. V. and Karplus, M. *J Chem Phys* 2002, 117, 10894.
39. Komatsuzaki, T., Hoshino, K., Matsunaga, Y., Rylance, G. J., Johnston, R. L., and Wales, D. J. *J Chem Phys* 2005, 122, 084714:1.
40. Daura, X., Van Gunsteren, W. F., and Mark, A. E. *Proteins: Struct. Funct. Genet* 1999, 34, 269.
41. Alexander N.Drozdo, Alan Grossfield, and Rohit V.Pappu *J Am Chem Soc* 2004, 126, 2574.
42. Dunitz, J. D. *Chem & Biol* 1995, 2, 709.
43. Mitchell, J. B. O. and Smith, J. *Proteins: Struct. Funct. Genet* 2003, 50, 563.

MOLECULAR MODELING OF THE ENANTIOSELECTIVITY OF *CANDIDA ANTARCTICA* LIPASE B – FREE ENERGY CALCULATION

SUMMARY

The relative stability of tetrahedral intermediates involved in the conversion of enantiomeric esters of chiral *sec*-alcohols by lipase B from *Candida antarctica*, CaLB, has been calculated from molecular dynamics trajectories using thermodynamic integration protocols and the GROMOS 53A6 force field. Integration pathways have been designed to produce minimal disturbance to the system, mainly by the use of soft-core atoms and the introduction of low-energy intermediates. Two model systems at different levels of complexity are introduced. Simulation studies of “truncated” tetrahedral intermediates with the active site residue, Ser105, representing the protein contribution, show that chiral induction by the successive chiral centers of Ser105 and the tetrahedral carbon alone may account for some 10-30% of the total free energy difference estimated from the experimentally determined enantioselectivity of CaLB. For the substrates that were studied, the results of the free energy calculations of a whole-enzyme model are fully consistent with experimentally determined enantiomeric ratio-values. The quantitative prediction of the inversion of the enantiopreference of a Trp104Ala mutant of CaLB has meanwhile been confirmed by experiment.

Keywords: *Candida antarctica* lipase B, transesterification, molecular modeling, free energy calculation, enantioselectivity, thermodynamic integration, tetrahedral intermediate

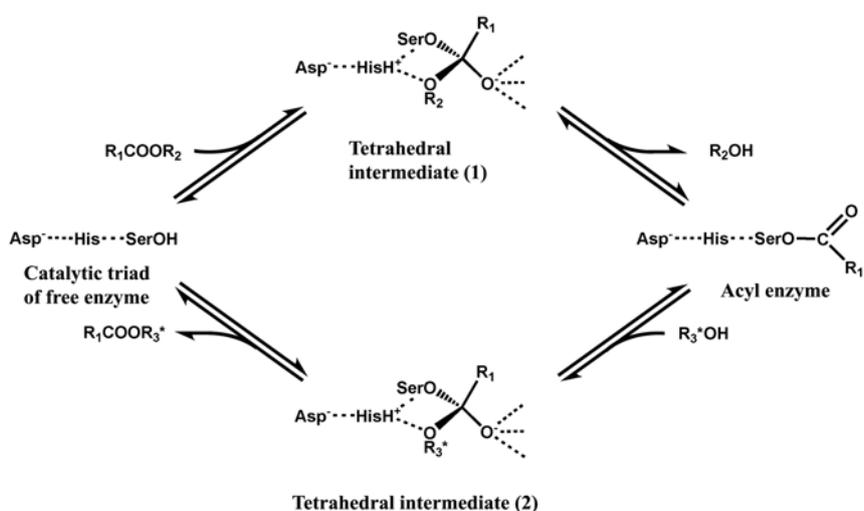
1. INTRODUCTION

Computational studies of enzyme enantioselectivity, the property of enzymes to discriminate between the mirror configurations of chiral substrates, were pioneered in the 1980s by the groups of DeTar¹ and Wipff² using α -chymotrypsin models. Their molecular mechanics (MM) based approach focused on a comparison of the minimum *steric* energy of the enzyme and the respective substrate enantiomers. The tetrahedral intermediate formed by the substrate and the serine residue at the active site of the enzyme was considered to mimic the structure of the transition state, which was, at the time, not accessible by quantum mechanical (QM) calculations. The model, though simple, gave results in fair agreement with experimental data. Several years later, the same enzyme was studied by Norin and coworkers.³ Both steric and electrostatic effects were taken into account when searching for structures of the tetrahedral intermediates of minimum energy using molecular dynamics (MD) combined with simulated annealing protocols. Despite the fact that only part of the enzyme was included in the model system and solvent molecules were not taken into account, this approach resulted in a qualitatively correct prediction of the enantiopreference in a statistically relevant number of cases. Next, attempts were made to model the enantioselective substrate binding in implicit solvent in a more quantitative fashion by calculating the *free energy* difference using a thermodynamic integration (TI) protocol.⁴ However, the calculations that were attempted for three different lipases, failed due to convergence problems. Subsequently, attention was shifted back to potential energy calculations using MD and simulated annealing protocols to locate the energy minima of the tetrahedral intermediates and monitoring the contributions to the total energy of defined subsets of the modeled system.^{5,6} Although certain choices of subsets were shown to give good agreement with the experimental data, this approach suffers from a need for calibration and a lack of portability to other enzymes and substrates. Meanwhile, it has been well recognized that the entropic contribution to the enantioselectivity needs to be taken into account as well.^{7,8} This has prompted renewed interest in methods to compute the free energy difference between the relevant transition states (or their mimics) that may provide a more accurate quantitative estimate of the enantioselectivity.⁹⁻¹² Colombo and coworkers chose a free energy perturbation (FEP) scheme to calculate the free energy difference of the tetrahedral intermediates relevant to the enantioselectivity of subtilisin.¹³ They used a combined quantum mechanics/molecular mechanics (QM/MM) approach, implementing slow-growth with periodical updating of the charge distribution. The free energy difference calculated for the tetrahedral intermediates formed between the enzyme and the *R* and *S* substrates is favorably comparable with the experimental data. Still, in the QM region the solvent effects had to be neglected because of the computational cost.⁵ Micaelo and coworkers, on the other hand, investigated the possibilities of thermodynamic integration (TI) schemes using molecular mechanics and full solvation to study the enantioselectivity of cutinase.^{14,15}

So far, qualitative predictions of the enantiopreference of this type of enzymes appear to be quite feasible, quantitative predictions of the enantioselectivity, on the other hand, may still be off by (an) order(s) of magnitude. This prompted us to investigate the possibilities for further improvement of the computational protocol in order to design a more accurate method for the prediction of enzyme enantioselectivity.

Recently, we developed a thermodynamic integration protocol to calculate the free energy differences between diastereomeric configurations of small molecules.^{16,17} This method affords reproducible results when small differences are involved. Thus, application of this protocol to the tetrahedral intermediates of an enzyme and its substrate enantiomers (diastereomeric combinations) is obviously straightforward. For a number of reasons, the lipase B from *Candida antarctica*, CaLB, serves as an attractive model molecule for this study. The enzyme is a medium-sized serine-type hydrolase belonging to α/β -hydrolase fold family. It does not involve any cofactor and there is no interfacial

activation. Besides, CaLB is active on a variety of substrates¹⁸ catalyzing hydrolysis,¹⁹ (inter/trans)esterification,^{20;21} acidolysis,²² aminolysis²³ and alcoholysis reactions.²² The enantioselective properties of CaLB have attracted much attention. A score of chiral substrates has been resolved by asymmetric transformation reactions in both laboratory and industrial applications, providing sufficient experimental reference data to benchmark a computational study. Fundamental aspects of CaLB (enantioselective) catalysis have been investigated in detail. Considering that the substrates tested for CaLB catalysis are, in general, not the compounds for which CaLB may have evolved to acquire high enantioselectivity, improvement of the enantioselectivity of CaLB for a broad range of substrates remains a challenging target. In this respect, strategies such as site directed mutagenesis might well benefit from the availability of reliable computational methods to predict the enantioselectivity of mutants. Similarly, computational prediction may be used to generate lead structures for the optimization of auxiliary moieties of the substrate structure, obviating the need to synthesize demanding substrate analogs.



Scheme 1: The catalytic mechanism of lipase-catalyzed transesterification (R_3 =alkyl) or hydrolysis (R_3 =hydrogen) of an ester, using a chiral alcohol R_3OH .

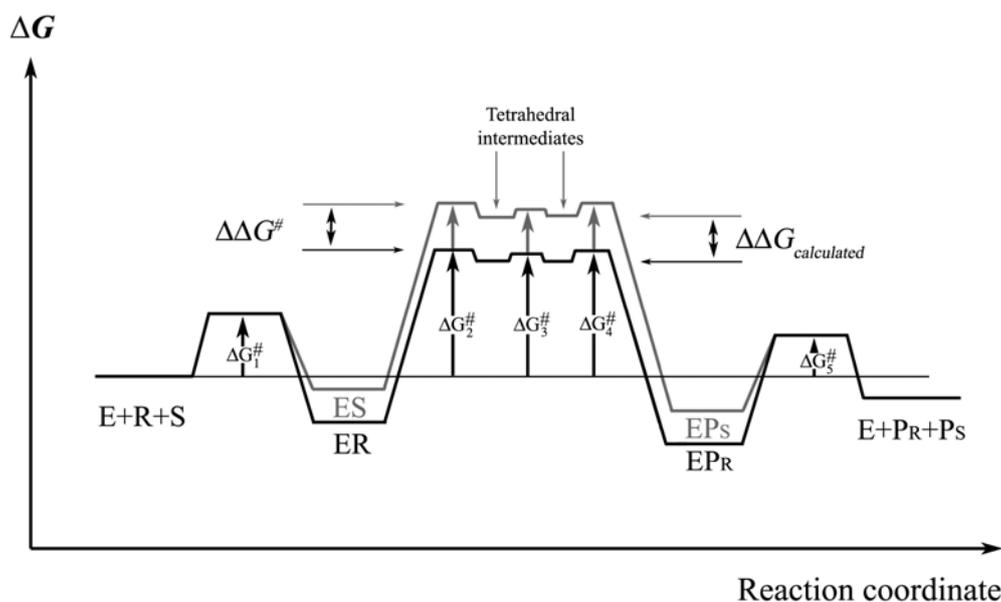


Figure 1. A schematic free energy profile for the enzymatic half-reaction of a serine-type hydrolase following ping-pong kinetics showing the relation between the transition state energies and the stability of the tetrahedral intermediates. $\Delta\Delta G^\ddagger$ (free energy difference at TS) may be approximated by $\Delta\Delta G^\ddagger_{TI}$ (free energy difference for the tetrahedral intermediates)

The enantiomeric ratio, or *E*-value,^{24,25} is defined in Eqn 1.

$$E_{R/S} = \frac{(k_{cat} / K_M)_R}{(k_{cat} / K_M)_S} \quad (1)$$

For enzymes following bi bi ping pong kinetics, it appears that the catalytic constants, $k_{cat}^R, k_{cat}^S, K_M^R, K_M^S$, are strictly dependent only on the microscopic kinetic constants appearing in the enzymatic half-reaction in which the enantiomeric species appear.²⁶ For the mechanism sketched in Scheme 1, this would be the lower part of the figure comprising the (reversible) conversion of free enzyme and chiral substrate (R_1COOR_3) to the chiral alcohol (released R_3OH) and the acylated enzyme species. A probable energetic profile of this part is given for the two enantiomers in Figure 1, following suggestions taken from the literature.²⁷ On the assumption that each kinetic constant can be related to an activation barrier according to the Eyring Transition State Theory (thermodynamic) formulation, straightforward analysis of the contributions of the individual kinetic constants to the over-all (lumped) parameters shows that the free energy barrier heights, as drawn in Figure 1, contribute to the *E*-value according to Eqn 2.

$$RT \ln E_{R/S} = \ln\left(\sum_{i=1}^5 \exp(\Delta G_i^\ddagger)^S\right) - \ln\left(\sum_{i=1}^5 \exp(\Delta G_i^\ddagger)^R\right) \quad (2)$$

Currently, most of the computational approaches to predict the enantioselectivity of serine-type hydrolases rely on the likely energetic equivalence of the tetrahedral intermediates (TI) and the actual transition state(s), in which case Eqn 3 applies.

$$RT \ln E_{R/S} \cong (\Delta G_{TI})^S - (\Delta G_{TI})^R = \Delta_{S-R}\Delta G = \Delta\Delta G_{R/S} \quad (3)$$

In deriving Eqn 3, it has been assumed that the contributions from barriers 1 and 5 (diffusion of substrates and products) are negligible (in the sum of exponentials) with respect to the other barriers, while the free energy of the tetrahedral intermediates is supposed to be sufficiently close to that of the (exponentially averaged) remaining barriers. The final equality in Eqn 3 serves to define our use of the difference of the Gibbs free energy values.²⁸ The main advantage of the use of Eqn 3 lies in the fact that simulation of the tetrahedral intermediates does not involve bond breaking or formation. This obviates the need for elaborate and time-consuming quantum mechanical calculations.

Here, we describe free energy calculations using two models: a truncated tetrahedral intermediate (TTI) model and a whole-enzyme model. In the truncated model, the protein part is represented solely by the active-site serine residue. This model does not only allow the separate assessment of the chiral induction component of the enantioselectivity, it may also provide a deep insight into the conditional aspects of equilibration, sampling, and integration. As a blind test, we investigated the performance of a Trp104Ala mutant, the construction of which has been reported in the literature.²⁹ While this work was in progress, the enantiomeric ratio of the mutant CaLB, on a substrate very similar to the one chosen for the simulation, has been reported.³⁰

2. METHODS

2.1 Substrates

Five substrates, all esters of an *n*-alkanoic acid and a secondary alcohol, were chosen for our computer simulation of the CaLB enantioselectivity (Table 1). For these substrates the tetrahedral intermediate with the side chain acyl moiety and

Ser105 from CaLB was constructed as shown in Figure 2. Substrate 1 contains a short alkyl chain in the secondary alcohol group; substrates 2 and 4 have bulky alcohol groups; substrates 3 and 5 contain bromide as an extra challenge to the computational modeling. CaLB catalyses the reaction involving the *R*-enantiomer most efficiently for all substrates, except for substrate 5, for which the *S*-enantiomer is preferred. A recent experimental study of Magnusson *et al.*²⁹ motivated us to also include the Trp104Ala mutant of CaLB in our modeling because an effect on the enantioselectivity could well be expected. We chose substrate 2 since this has the largest *E*-value for the wild-type enzyme.

Table 1. Structures of substrate with the different acyl chains and secondary alcohols

	Acyl chain	Alcohol moiety	Preference	E_{Exp} (kJ/mol)	$\Delta\Delta G_{R-S \text{ Exp}}$ (kJ/mol)
1	<i>n</i> -hexyl	2-butanol	R	9	-5.9
2	<i>n</i> -octyl	1-phenylethanol	R	>200 ^{a)}	> -12.9
3	<i>n</i> -hexyl	3-bromo-2-propanol	R	371	-14.7
4	<i>n</i> -octyl	3,3-dimethyl-2-butanol	R	350	-15.1
5	<i>n</i> -butyl	1-bromo-2-octanol	S	7	4.7

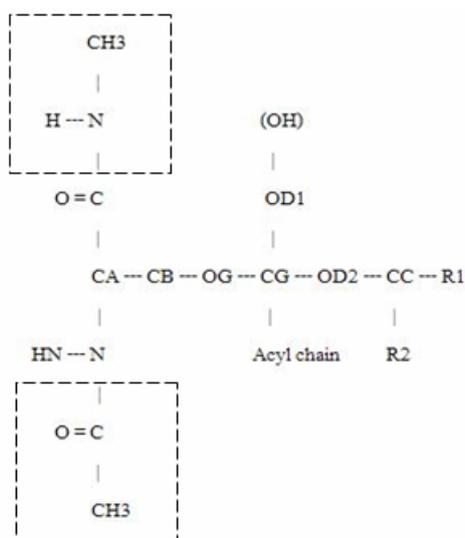


Figure 2: The substrate-Serine105 structure in the TTI model (including the blocks in dashed lines) and in the whole enzyme model (the blocks in dashed lines replaced by neighbor residues of the enzyme). CC is the chiral center to be inverted from *R* to *S*. CG is the other chiral center set to be in the *S* configuration.

2.2 Experiment

Experimental values of the enantioselectivities of CaLB for esters of various chiral *sec*-alcohol substrates were obtained from the literature (Table 1). Different acyl donors were used, hexanoic acid (substrate 1 and 3),³¹ octanoic acid (substrate 2 and 4),^{20,32} and butanoic acid (substrate 5)³³ in the reaction catalyzed by wild type CaLB. The reaction temperature lies between 30~39°C. Hexane is used as the solvent for all the substrates except for substrate 2 for which *S*-ethylthiooctanoate (wild type) and cyclohexane (mutant) was used.

2.3 Computer simulation

2.3.1 Preparation of molecules in the TTI model and in the whole enzyme model

TTI model

The TTI molecules were built based on part of the crystallographically determined structure of an inhibitor-enzyme complex (PDB entry 1LBS)³⁴. Ser105 and the covalently bound inhibitor *N*-hexylphosphonate ethyl ester were retained and the rest of the structure was removed. The amino and carboxyl moieties of Ser105 were capped with acetyl and methylamine groups, respectively. The inhibitor was converted into the tetrahedral structure illustrated in Figure 2. The OD1 atom with a formal charge of -1, was replaced by a hydroxyl group in the TTI model since this resembles the situation in the enzyme where the oxyanion is stabilized by Thr40 and Gln106 through hydrogen bonds. The chirality of carbon atom CG, connected to three oxygen atoms was changed from *R* to *S* which is justified by the fact that it enables the substrate to accommodate itself in the active site so that the acyl chain is pointing out to the surface of the protein and the alcohol moiety interacts favourably with the “enantioselectivity pocket” of the active site. The relevance of this mode has been demonstrated in previous studies^{5,16} and has been observed in other lipases as well.³⁵ The charges of the atoms were determined in analogy to the GROMOS force field parameter set 53A6. Selected atoms are listed in Table 2.

The TTI complexes were solvated in (periodic) truncated octahedral boxes containing 909 to 1009 methanol molecules, which was selected because the dielectric constant of methanol is closer to the average value of the protein interior ($\epsilon < 20$) than that of water ($\epsilon = 80$).^{36,37,38} The dielectric constant of the force field model used here was calculated to be 23.³⁹ The simulations were carried out with the GROMOS96 biomolecular simulation package using the 53A6 version of GROMOS96 force field.⁴⁰⁻⁴² The systems were minimized using steepest descent minimization, and then gradually heated to 300 K in three 100 K intervals of 10 ps at constant volume using a weakly coupled temperature bath with relaxation time 0.1 ps.⁴³ Another 20 ps of equilibration was performed at a constant pressure of 1 atm using weak coupling to a pressure bath with a relaxation time of 0.5 ps and an estimated isothermal compressibility of $4.575 \cdot 10^{-4} \text{ [(kJ mol}^{-1} \text{ nm}^{-3})^{-1}]$ ⁴³ Bond lengths were constrained (with relative precision of 10^{-4}) at the minimum energy values using the SHAKE algorithm⁴⁴ allowing for a time step of 2 fs. Long-range interactions were calculated using a triple range cut-off scheme. All interactions within 0.8 nm were calculated every time step from a pair list that was updated every fifth step. Interactions between pairs that were between 0.8 and 1.4 nm were calculated at pair list (re)construction and kept constant between pair list updates. A reaction-field contribution⁴⁵ was added to the electrostatic interactions to account for a homogeneous medium outside the long range cutoff with a relative permittivity of 62.⁴⁶ The serine residue was positionally restrained with a force constant of $2.5 \cdot 10^4 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ during the whole simulation while the substrate moiety was allowed to move. After equilibration, thermodynamic integration was performed for the free energy calculation as described in section 2.3.2.

Whole enzyme model

In the whole enzyme model, the tetrahedral intermediates were constructed as outline above. The total charge of the molecule is -1, mainly located on the oxyanion. The coordinates of CaLB were taken from the x-ray structure (PDB entry 1LBS).³⁴ The two sugar units (NAG) as well as crystallographically observed water molecules were removed. The free energy calculations were all started from the most commonly preferred *R*-configuration of atom CC, which can be oriented in the enzyme in two modes, as suggested by Haeffner.⁶ In mode I, the medium group of the sec-alcohol moiety (the smaller of R1 and R2 in Figure 2) is accommodated in the enantioselectivity pocket. In mode II, the larger group of the sec-alcohol moiety occupies this pocket. Mode I, also referred to as the productive mode, was used at the very beginning of the simulations for *R* conformer of substrates. The whole enzyme was solvated in a periodic box containing the enzyme tetrahedral intermediate and approximately 9000 SPC water molecules, with a total number of

roughly 30,000 atoms. The system was heated stepwise with 50 K interval in six steps up to 300 K at constant volume, during which position restraints on the enzyme were reduced ten-fold every step from $2.5 \cdot 10^4 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ in the first step to zero in the sixth step. Another period of equilibration under a constant pressure of 1 atm was performed, as described for the TTI model. The prepared molecules were subsequently used in the free energy calculations.

Table 2 Partial atomic charges of selected atoms in the TTI (see Figure 2)

Atom name	Charge _{TTI}	Charge _{Enz}	Atom name	Charge _{TTI}	Charge _{Enz}
N	-0.310	-0.310	CG	0.266	0.266
HN	0.310	0.310	OD1*	-0.674	-0.670
CA	0.000	0.000	(HO)	0.408	(Dummy)
CB	0.142	0.142	CD**	0.000	0.000
OG	-0.142	-0.440	OD2	-0.142	-0.440
C	0.450	0.450	CC	0.142	0.142
O	-0.450	-0.450			

* Atom type of OD1 is 3 (TTI) and 2 (Enz) corresponding to hydroxyl oxygen and carboxyl oxygen

** The first atom of the acyl chain connected to CG

2.3.2 Free energy calculation

Using the thermodynamic integration formula⁴⁷ the free energy difference, ΔG_{RS} , is calculated between the R and S configuration of the system:

$$\Delta G_{RS} = \int_{\lambda_R}^{\lambda_S} \left\langle \frac{\partial H(r, p, \lambda)}{\partial \lambda} \right\rangle_{\lambda} d\lambda \quad (4)$$

$H(r, p, \lambda)$ is the Hamiltonian describing the system as a function of the positions, r , of the constituent particles, their momenta, p , and the coupling parameter λ . The Hamiltonian describes the system with TI in which the CC atom is in the R -configuration if $\lambda = \lambda_R = 0$ and the S configuration if $\lambda = \lambda_S = 1$. The angular brackets indicate an ensemble average at a specific value of λ . The integration was performed by changing the value of λ from λ_R to λ_S in a number of discrete steps at which the ensemble average is collected. Different pathways connecting states R and S can be chosen, which should all lead to the same value of ΔG_{RS} in the limit of infinite sampling.

For the energy perturbation from R to S , we chose a pathway that involves the perturbation of an improper dihedral change at the CC atom from 35.26° (R) to -35.26° (S). To reduce the strain in the bond angles at intermediate values of the improper dihedral, we inserted an intermediate state in which the minimum energy value of the improper dihedral is set to 0° and the minimum energy values of the three bond angles around the CC atom are set to 120° . We have shown previously that this pathway was most efficient in calculations involving the chiral inversion of peptides.¹⁶ Both half pathways ($R \rightarrow I$; $I \rightarrow S$) were performed in 11 steps with $\Delta\lambda = 0.1$. In each step, 1 ns equilibration was followed by 500 ps data collection. Using a total of 22 integration steps and 33 ns of simulation, the system was changed from the R to the S -configuration of the CC atom.

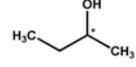
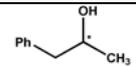
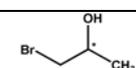
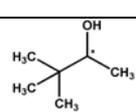
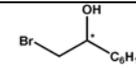
3. RESULTS AND DISCUSSION

The calculated $\Delta\Delta G_{RS}$ -values from computer simulations based on the TTI model and the whole enzyme model is listed in Table 3. The TTI model correctly predicts the preference for the R -configuration for substrates 1, 2, and 4 but yields the incorrect configuration for substrates 3 and 5. The whole enzyme model correctly predicts all preference and is quantitatively more accurate for all substrates. In general, the present method also gives more accurate predictions

compared to the results of previous studies where larger differences between the experimental and calculated values and larger errors were reported.^{6;31}

The protein structures do not appreciably change during the simulations, with heavy atom-positional root-mean-square deviations from the crystal structure of less than 2 Å, indicating stable simulations.

Table 3 Different substrates, enantiomer that is preferably catalyzed by CaLB, calculated $\Delta\Delta G_{R-S}^{Calc}$ compared to the experimentally determined $\Delta\Delta G_{R-S}^{Exp}$ as derived from the experimentally determined enantioselectivity, E_{exp}

	sec-alcohol	Pref.		E_{Exp}		$\Delta\Delta G_{R-S}^{Exp}$ (kJ/mol)		$\Delta\Delta G_{R-S}^{Calc}$ (kJ/mol)		
								Enzyme model		Truncated model
1		R		9		-5.9		-5.3		-0.9
2		R ^{a)}	S ^{b)}	>200 ^{c)}	6.6 ^{d)}	> -12.9	4.6	-11.5 ^{e)}	4.0 ^{e)}	-1.7
3		R		371		-14.7		-25.2		7.9
4		R		350		-15.1		-14.3		-4.7
5		S		7		4.7		11.3		-1.3

a) Wild type enzyme

b) Mutant Trp104Ala

c) Ethylthiooctanoate as acyl chain donor

d) Calculated from k_{cat}^{app} and K_M^{app} in the experiment with cyclohexane as solvent and vinyl butanoate as acyl chain donor

e) Water as solvent and octanoate as the acyl chain donor

Simulation time

The ensemble averages in Eqn 2 need to be properly converged in order to reach accurate free energy estimates. For this, sufficient sampling is required, which in the case of MD translates into sufficiently long simulations.

In the case of the TTI model, a long simulation time of up to 1ns is needed at the very beginning of the chiral inversion. In the example of substrate 2, the complex relaxes into a different conformation after about 800 ps (Figure 3). Starting the free energy calculation from the unrelaxed conformation leads to a different perturbation pathway (Figure 4). This phenomenon is observed in substrate 3 and 5 as well, which involve bulky groups. Once the proper relaxation is obtained and the subsequent free energy perturbation steps continue from the previously converged structure, the system can follow a smooth perturbation pathway and convergence seems to occur (Figure 5 a), which leads to a more reliable free energy difference result.

For the enzyme model, the situation is a somewhat different. Due to steric hindrance between the chiral alcohol moiety and the enzyme, the substrate does not have the same freedom to move as in the TTI model. This is especially true at the very beginning of the energy perturbation (i.e. $\lambda=0$), during which the substrate adapts itself well to the steric demands of the active site cavity, and convergence problems do not occur. As the system becomes more perturbed in the following steps, however, more time is needed for full equilibration (Figure 5 b). Micaelo *et al.* reported 5 ns

equilibration period at $\lambda = 0$, while during the following 11 integration step, only 10 ps of equilibration followed by 40 ps data collection are reported for each step.¹⁴ In the present case, long simulation period were needed at intermediate λ -values rather than at the beginning of the simulation. It must be emphasized, however, that the convergence and sampling are notorious problems in molecular simulation. Even 1 ns period do not guarantee that the system reaches equilibrium.

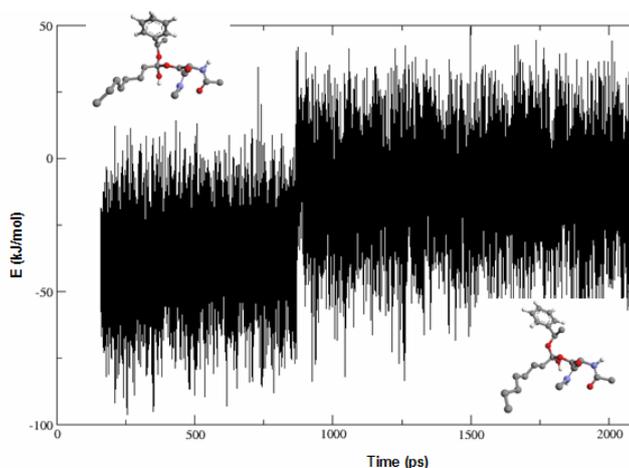


Figure 3. A long simulation time up to at least 1 ns is needed for the TTI model in order to obtain convergence of the system which assumes different conformation than otherwise.

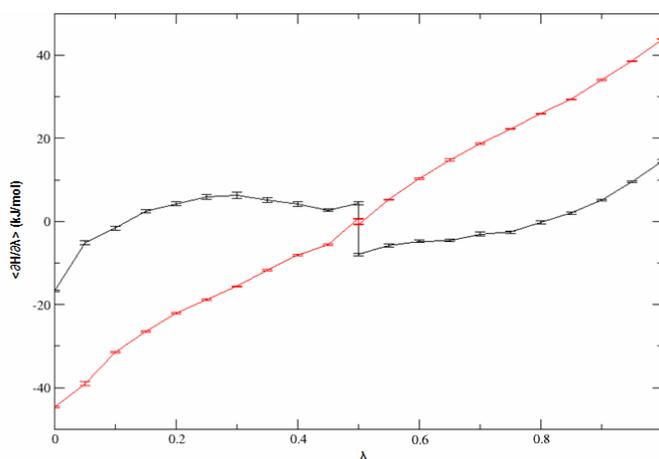


Figure 4. Insufficient simulation time at the very beginning of the TTI2 simulation leads to a different perturbation pathway. Light line: 500ps/step simulation time; dark line: 1 ns/step simulation time.

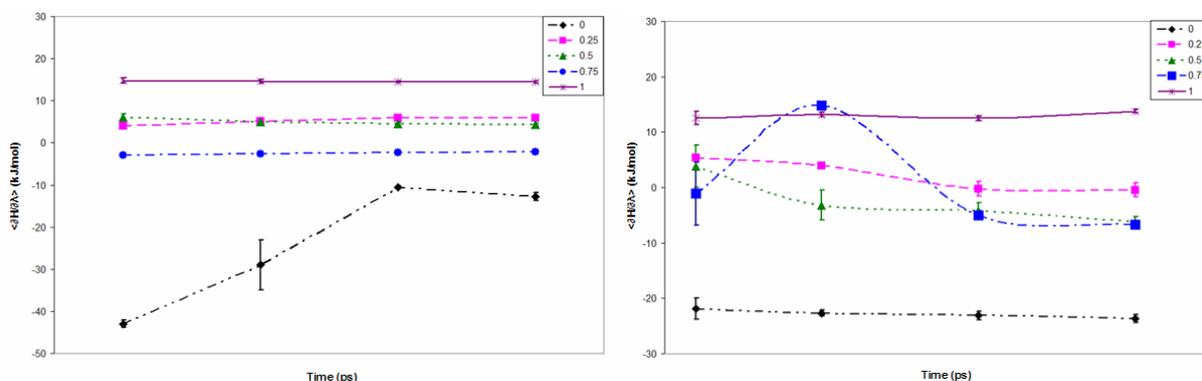


Figure 5. The fluctuation of $\langle \partial H / \partial \lambda \rangle$ during the simulation of a) TTI2 and b) Enz2_mut indicates that a long simulation time up to at least 1ns is needed to get equilibration of the system before the data collection. The different lines correspond to integration steps from initial state to end state during the energy perturbation. All lines start at 100ps and ends at 1500ps on the x-axis.

Solvent effect

There are many indications that the solvent affects the enantioselectivity of an enzyme.⁴⁸⁻⁵¹ It is generally thought that a minimum amount of water is needed for enzymatic catalysis to maintain the structure of the enzyme.⁵² For example 50 essential water molecules have been used in simulations with other organic solvents.^{13;53} However, the effect of these water molecules is unclear.⁵⁴ On lowering the water activity (or water content), either decreased,⁵⁵ increased^{14;56;57} and unaffected^{14;57;58} *E*-values have been reported. For CaLB, it seems that variations of the water content hardly affect the enantioselectivity.⁵⁸ On the other hand, the influence of the solvent on the enantioselectivity appears to vary with temperature. A study on CaLB dispersed in three solvents of different molecular size showed that the solvent effect becomes obvious only at relatively high temperatures (> 40°C).³⁰ Since, the simulations described here were run at 300 K, the effect of different solvents used (organic media in the transesterification experiments and water viz. methanol in our study) is not expected to add significantly to the discrepancy between measured and simulated *E*-values.

Crystallographic water is treated in different ways in reported simulations. The water molecules from the x-ray structure are either removed completely,⁵⁹ kept as the only solvent molecules used in the simulation,⁶ or extra solvent molecules are added.¹³ Using the water from the x-ray structure only, in our opinion, is not enough: They are in reality not the only waters that present, but are just those that are relatively immobile when the X-ray experiment is performed. All the cavities and the space in between the protein residues will in reality be filled with dynamically moving water, which shields the charge-charge interaction within the protein. The importance of including a sufficient amount of water in the enzyme is reflected by the fact that the enzyme enantioselectivity is influenced by the solvent's dielectric constant (ϵ).⁵⁰ In computer simulations the total or partial removal of aqueous solvent can kinetically trap the system at the starting configuration. Therefore, added enough water to ensure proper sampling of the accessible configurational space.

MM vs. QM

A fair comparison of the MM and QM approaches to the simulations of hydrolase enantioselectivity is hardly possible. Several groups have employed (semi-empirical) QM to calculate the partial charges for the starting structures of substrates in the gas phase for subsequent MM calculations of the enantioselectivity of CaLB^{6;31;60} and of other lipases.^{4;14;59;61-63} In these cases the influence of the enzyme environment and of solvent effects were not included in the QM calculation. This is a serious omission in the assignment of charges to the substrate progressing along different reaction pathways.⁶⁴ Attempts have been reported to extend the QM region to include the catalytic triad and use calculations at the *ab initio* level rather than semi-empirical QM methods, in order to improve the accuracy of the calculation.^{63;65} However, this approach is computationally demanding, even when explicit water is not included in the calculation, not to mention dynamic simulations. Recent modeling of the enantioselectivity of *Burkholderia cepacia* lipase by Tomic *et al.* showed that MM calculations produce results much closer to the experimental observations than QM calculations (both semi-empirical and *ab initio*), indicating the drawback of neglecting water and entropy effects in the QM approach.⁶³

Two studies have been reported using QM/MM in free energy calculations concerning the enantioselectivity of subtilisin and α -chymotrypsin, respectively.^{13;66} Although, the use of QM calculations is expected to give more realistic charge distributions, the distributions are dependent on the conformation of the substrate and need to be updated frequently.⁶⁷ From the *ab initio* studies, the structures of the transition states for the formation and the breakdown of the tetrahedral intermediate are found to closely resemble the structure of tetrahedral intermediate itself.^{68;69} Considering the successful use of the tetrahedral intermediate structures for the modeling of the enantioselectivity of a various

lipases^{6,13} this concept has been applied in the present investigation. By assigning the partial charges in accordance with the empirically validated charge distributions of the Gromos force field and by explicitly taking into account the protein environment and solvent effects we describe the system using MM. All degrees of freedom were treated dynamically, allowing for an accurate calculation of free energies, i.e. including entropic contributions.

Comparison with previous studies

So far, thermodynamic integration is the most widely used method for the free energy calculations on enantioselectivity.^{4,14,62,66} Norin first applied this method in 1994 to study the enantioselectivity of lipases CRL, RML and HLL. Due to the limited computing facilities at that time, only short time simulations (240 ps in total vs. 30 ns in this work) could be afforded. Furthermore, all the water molecules in the enzyme were removed and continuum implicit water was added afterwards, which describes the solvent effects in a mean-field manner but lacks the inclusion of direct solute-solvent interactions.⁴ The free energy calculation failed because of convergence problems during the simulation. Dreveny *et al.* used a similar methodology for hydroxynitrile lyases (HNL) catalyzing the enantioselective cleavage on hydroxynitriles. 41 integration steps with a total simulation time of 410 ps were used to transform the *R* configuration to the *S* configuration. The results of the simulation were in good agreement with the experimental results, but this study was not extended to include more substrates.⁷⁰ Pricl studied the enantioselectivity of α -chymotrypsin using combined ab initio and free energy calculations. The thermodynamic integration protocol was used for the QM region with 101 integration steps each consisting of only 0.2 ps equilibration and 0.3 ps data collection. However, during the subsequent MD simulation, implicit instead of explicit water was used. Recently, Soares used the thermodynamic integration protocol to calculate the free energy for cutinase in nonaqueous solvents. 11 integration steps were used, of 50 ps each.¹⁴ None of the reported simulations considered the use of nanoseconds per integration step, which we found to be required for accurate and reproducible results.

An alternative to the thermodynamic integration method is the slow growth method, in which the integration is performed continuously by slowly changing the coupling parameter λ from λ_A to λ_B during the simulation. Colombo *et al.* used this method to compute the enantioselectivity of subtilisin in DMF.¹³ A set of substrate atoms in the *S* configuration is gradually removed while the corresponding group in the *R* configuration appears. The free energy is calculated via this slow growth method in a total simulation time of 750 ps. However, in this approach the system always lags behind the changing Hamiltonian and never equilibrates properly. Therefore, it can be basically considered as a non-equilibrium estimate.¹⁵

TTI model vs. whole enzyme model

Calculations involving the TTI model are more time-efficient, by a factor of ten, than those using the whole enzyme model. However, the full details of the enzyme model are apparently required to obtain accurate results that match the experimental values, as shown in Table 3. It is of practical interest to consider the possible causes for the difference between the two models. First, it must be emphasized that the simulation conditions for the two models are not identical. In the TTI model the oxyanion is replaced by a hydroxyl group and methanol is used as a solvent in order to mimic the electrostatic dielectricum in the protein interior. In the whole enzyme model, the oxyanion is explicitly included and the enzyme is solvated in water. However, it is difficult to estimate the extent of the contribution that is made by each of these different conditions to the differences between the results of the two models. Of course, the most obvious factor is the steric hindrance that the substrate might encounter in the enzyme, which is not taken into account in the TTI model. This becomes clear from the different conformations, that the substrates assume in the two models (Figure 6-10).

Our attempts to simplify the system by trimming the enzyme down to the active site serine carrying the tetrahedral intermediate structure, the TTI model, are based on the results of free energy calculations of diastereomeric dipeptides. Since the observed and calculated energy differences for diastereomers with two chiral centers are of the order of 1 kcal/mol, even larger free energy differences may be expected for diastereomers containing three chiral centers. Clearly, the calculated energy differences for the TTI models reflect the chiral induction of the fixed chirality of the active site serine and the tetrahedral carbon of the ester carboxyl moiety on the stability of *R*- viz. *S*-configurations of the alcohol. Only in the case that the most populated conformation of the TTI complex of the preferred enantiomer is compatible with the steric demands of the enzyme active site, will this difference provide a lower estimate on the enantioselectivity of the enzyme. In the case of substrates 1, 2, and 4, the free energy seems to be determined to a large extent by this inherent free energy difference; for substrates 3 and 5, energetically less favorable conformations appear, leading to a different outcome. In fact, in the case of substrates 1, 2 and 4, the most commonly observed structures for the *R* substrates show more similarity between the TTI model and whole enzyme model than for the *S* substrates. The situation for substrates 3 and 5 seems more complex, probably due to the presence of the bromide group. When comparing the conformations of the TTI and whole enzyme models for these substrates, large differences are observed, especially for substrate 5. Analysis of the proper dihedral angle fluctuation shows that the large-sized group (containing bromine) in the alcohol moiety of the substrate experiences a different amount of freedom in the TTI model as compared to the whole enzyme model. In the TTI model the *S* enantiomer has more conformational freedom while it is the other way around in the whole enzyme model (Figure 11).

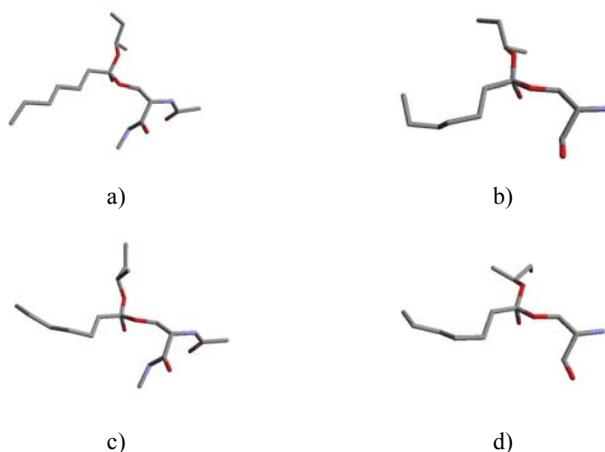


Figure 6. Comparison of most populated structure for substrate 1 extracted from the trajectory of TTI model and whole enzyme model. (a) TTI-*R* (b) Enz-*R* (c) TTI-*S* (d) Enz-*S*

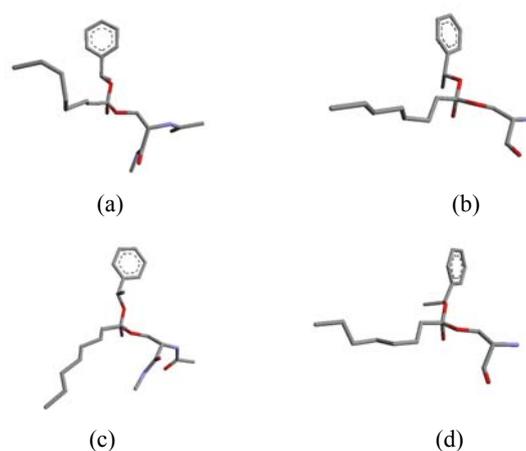


Figure 7. Comparison of most populated structures for substrate 2 extracted from the trajectory of the TTI model and the whole enzyme model. (a) TTI-*R* (b) Enz-*R* (c) TTI-*S* (d) Enz-*S*

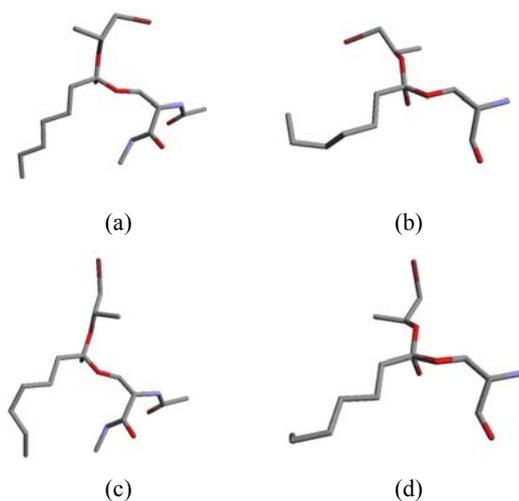


Figure 8. Comparison of most populated structures for substrate 3 extracted from the trajectory of the TTI model and the whole enzyme model. (a) TTI-R (b) Enz-R (c) TTI-S (d) Enz-S

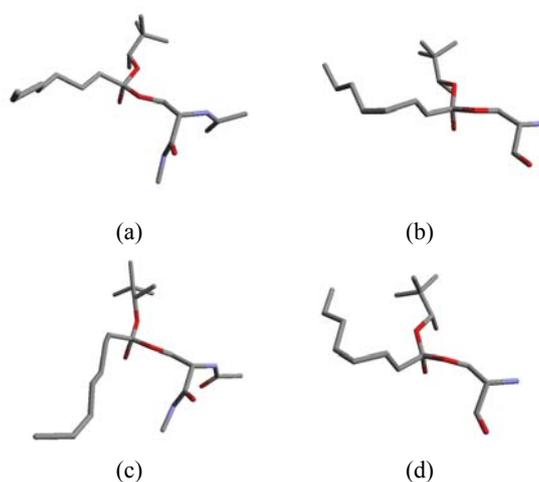


Figure 9. Comparison of most populated structures for substrate 4 extracted from the trajectory of the TTI model and the whole enzyme model. (a) TTI-R (b) Enz-R (c) TTI-S (d) Enz-S

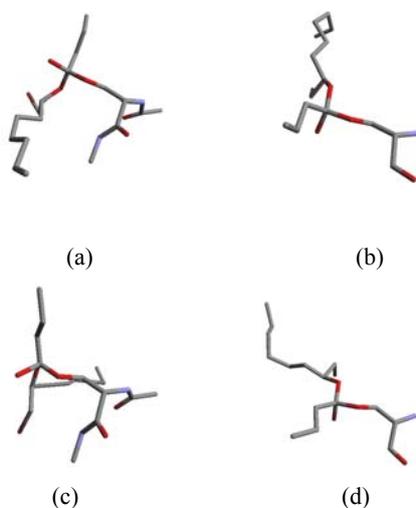


Figure 10 Comparison of most populated structures for substrate 5 extracted from the trajectory of the TTI model and the whole enzyme model. (a) TTI-R (b) Enz-R, (c) TTI-S, (d) Enz-S

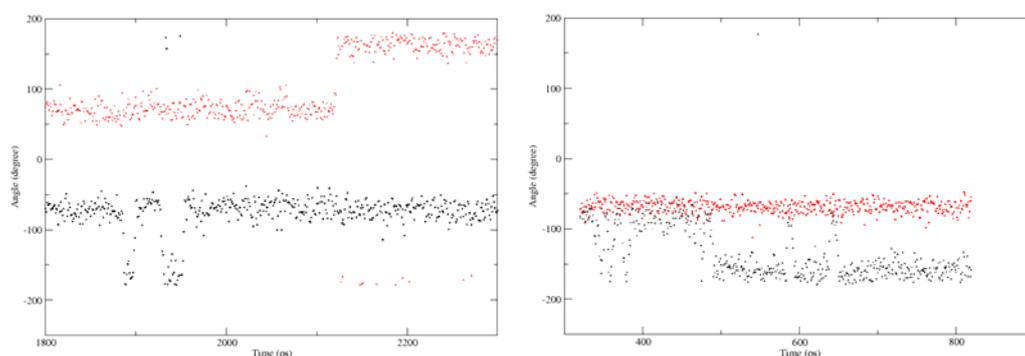


Figure 11. The proper dihedral angle fluctuation of OD₂-CC*-CK₁-Br in the *R* conformer (black) and the *S* conformer (red) of substrate 3 in the TTI model (left) and the enzyme model (right) simulation. The data have been extracted from the trajectory during the data collection stage when convergence has been reached.

The TTI model can provide us with a first estimate on the enantioselectivity of the enzyme in a quick but not very accurate way. Obviously, one of the disadvantages of the TTI model is the impossibility to study the effects of mutations. This can only be done with the enzyme model, and this is especially useful to support attempts to improve enantioselectivity of the enzyme by protein engineering.⁷¹

Different orientation mode

Previous results of CaLB enantioselectivity simulations have been interpreted in terms of two binding modes for both the fast and slow-reacting substrates, as described in Methods section 2.3.1.⁶ These two modes are applicable for substrates 1-4. The fast-reacting *R* substrate takes the productive mode, referred to as mode I, in which the medium group of the alcohol moiety is situated in the enantioselectivity pocket, and the hydrogen bonds between the substrate and His224, Thr40 and Gln106 are retained, and the large group is pointing outward from the active site. The results from our simulation show that for substrates 1-4 all the hydrogen bonds are indeed kept in the *R* configuration in mode I (Figure 12 a and Table 4). During thermodynamic integration the *R* configuration of the substrate is changed into the *S* configuration in 22 λ -steps ending in mode II where the large group is now occupying the enantioselectivity pocket. Due to steric hindrance between the large group and the enzyme, the conformations of the *S* configuration are subject to a change from mode II to mode I in which the small group sits in the pocket while the hydrogen bonding is disrupted (Figure 12 b and Table 4). Simulations were also done for the *R*-enantiomer of substrate 4 starting in mode II, but the structure of the complex in binding mode II seems to be unstable since conversion of mode II to mode I was observed during the early stage of perturbation. Substrate 5 is a “special” case since the large and medium groups are both quite big and it seems difficult to accommodate either the large group with a long chain or the medium group carrying the bromine into the enantioselectivity pocket. Finally, the small and medium groups fit in the pocket for both *R*- and *S*-configurations (Figure 13). In the *S*-configuration, the long chain of the large group is oriented far away from Trp104 and the bromine of the medium group is less protruding into the corner of the pocket (Fig 13b) avoiding steric hindrance existing in the *R*-configuration. However, the impact of the enantioselectivity pocket seems weaker, which explains why the enantioselectivity of CaLB towards this substrate is lower compared to the other substrates studied. Very interesting is the case of substrate 5, where the fast reacting and slow reacting substrates both choose mode I, which is the non-productive bindings for the slow-reacting substrate 1-4 (Scheme 2). The reason might lie in the structure of substrate 5, which is quite different from substrates 1-4.

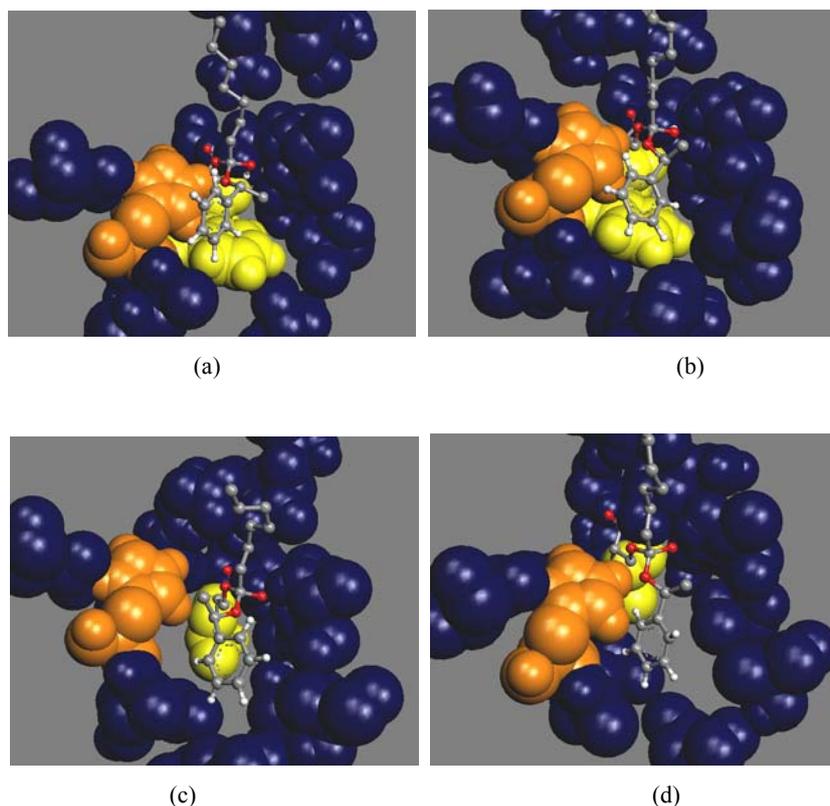


Figure 12: Active site of CaLB (space filling) with its Serine 105 connected to substrate 2 (ball cylinder): (a) *R*, wild type (b) *S*, wild type (c) *R*, mutant (d) *S*, mutant. His224 is displayed in grey, Trp(wt)/Ala(mutant) in white.

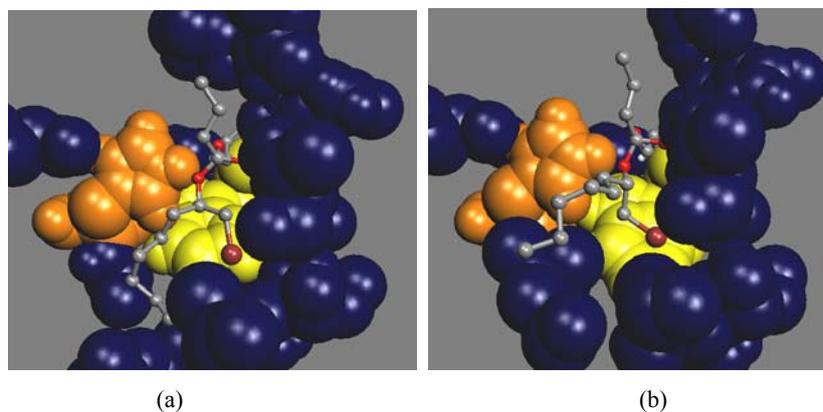
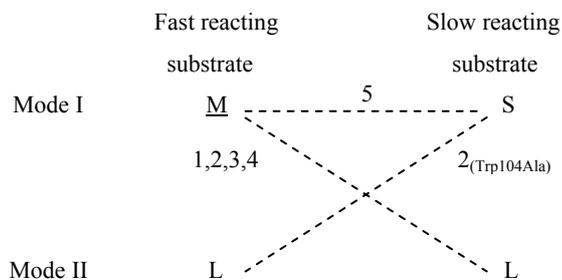


Figure 13: Active site of CaLB with its Serine 105 connected to substrate 5: (a) *R* (b) *S*



Scheme 2. Different binding mode preferences for substrates 1-5. L (large), M (medium) or S (small) groups of the alcohol moiety occupy the enantioselectivity pocket of CaLB in Mode I and II. Dashed lines indicate a conversion of binding mode from the starting configuration (*R* enantiomer, fast reacting substrate 1-4, slow reacting substrate 2_(Trp104Ala) and 5) to the end of energy perturbation (*S* configuration). Underlined groups have been suggested to be the productive binding types in an earlier study.⁶

A study on CaLB mutant W104A

Recently, Magnusson *et al.* published experimental results on a mutant of CaLB, in which Trp104 has been replaced by Ala.³⁰ This prompted us to test our computational method for the prediction of the enantioselectivity on this mutant. When experimental results of (the inversion of) the enantioselectivity of the mutant CaLB for certain substrates became available, it appeared that we happened to choose the same substrate 1-phenylethanol in our simulation, however, different acyl donors were used in the experiment (butanoate) and simulation (octanoate) because at the start of our simulations for the wild type enzyme, only experimental data for the reaction involving octanoate as an acyl donor were available. It would seem that the fact that our predictions of the enantioselectivity of the mutant CaLB were confirmed by the experimental results (wild type $\Delta\Delta G_{RS} = -11.5$ kJ/mol and mutant $\Delta\Delta G_{RS} = 4.0$ kJ/mol) is highly encouraging. Figure 12 shows how the *R* and *S* of substrate 2 orient differently in the wild type and mutant. Clearly, replacement of Trp104 with Ala removes the steric hindrance experienced mostly by the *S*-substrate in the wild type, allowing the bulky phenyl group of the *S*-substrate to fit comfortably in the pocket without disrupting the formation of hydrogen bonds between the substrate oxygen (OG and OD2) and the proton of His224. Such an orientation was also observed by Magnusson *et al.*³⁰ However, the orientation of the slow reacting *R*-substrate is different from the one reported in Magnusson's study. We observe that the group occupying the pocket changes from the medium one in the wild type enzyme to small one in the mutant (corresponding to mode I for slow-reacting substrate). Magnusson *et al.* report no changes, which seems to be inconsistent with the observation that k_{cat}^{app} of the *R*-enantiomer decreased about 60 times as compared to the wild type.³⁰ Moreover, the modes adapted by the substrates in the mutant are different from the ones adopted by the "normal" substrates 1-4: the fast- and slow-reacting substrates adopt modes II and I, exactly the non-productive modes for substrates 1-4. This implies again that when either substrate or enzyme experiences substantial structural changes, they are likely to adopt different binding.

Our calculation and prediction of the enantioselectivity of the mutant CaLB still needs to be validated with experimental data describing the enantioselectivity of the mutant enzyme for the substrate ester with the appropriate octanoate acyl chain donor. Previous studies have shown that the acyl moiety may have a noticeable but not a dramatic effect (<2 kJ/mol between butanoate and octanoate) on the enantioselectivity,⁵ so we expect that the *E*-value inversion will also be observed experimentally for the case of an octanoate acyl group.

Table 4: Number of intramolecular hydrogen bond formation for the substrate in the enzyme model (100ps/500frames).

1					
Molecules	OG ₁₀₅ -HE2 ₂₂₄ 967-2105	OD2 ₁₀₅ -HE2 ₂₂₄ 977-2105	OD1 ₁₀₅ -H ₄₀ 969-327	OD1 ₁₀₅ -HG1 ₄₀ 969-331	OD1 ₁₀₅ -H ₁₀₆ 969-985
$\lambda=0$ (R)	99%	18%	100%	100%	72%
$\lambda=1$ (S)	41%	93%	100%	100%	89%

2					
Molecules	OG ₁₀₅ -HE2 ₂₂₄ 967-2115	OD2 ₁₀₅ -HE2 ₂₂₄ 978-2115	OD1 ₁₀₅ -H ₄₀ 969-327	OD1 ₁₀₅ -HG1 ₄₀ 969-331	OD1 ₁₀₅ -H ₁₀₆ 969-995
$\lambda=0$ (R)	73%	84%	100%	100%	88%
$\lambda=1$ (S)	89%	56%	100%	100%	76%

2 Trp104Ala					
Molecules	OG ₁₀₅ -HE2 ₂₂₄ 952-2100	OD2 ₁₀₅ -HE2 ₂₂₄ 963-2100	OD1 ₁₀₅ -H ₄₀ 954-327	OD1 ₁₀₅ -HG1 ₄₀ 954-331	OD1 ₁₀₅ -H ₁₀₆ 954-980
$\lambda=0$ (R)	0%	0%	100%	100%	95%
$\lambda=1$ (S)	66%	70%	100%	94%	94%

3

Molecules	OG ₁₀₅ -HE2 ₂₂₄ 967-2105	OD2 ₁₀₅ -HE2 ₂₂₄ 977-2105	OD1 ₁₀₅ -H ₄₀ 969-327	OD1 ₁₀₅ -HG1 ₄₀ 969-331	OD1 ₁₀₅ -H ₁₀₆ 969-985
$\lambda=0$ (R)	72%	92%	100%	100%	96%
$\lambda=1$ (S)	67%	41%	100%	100%	99%

4

Molecules	OG ₁₀₅ -HE2 ₂₂₄ 967-2108	OD2 ₁₀₅ -HE2 ₂₂₄ 978-2108	OD1 ₁₀₅ -H ₄₀ 969-327	OD1 ₁₀₅ -HG1 ₄₀ 969-331	OD1 ₁₀₅ -H ₁₀₆ 969-988
$\lambda=0$ (R)	78%	9%	99%	85%	31%
$\lambda=1$ (S)	0	0	100%	99%	22%

5

Molecules	OG ₁₀₅ -HE2 ₂₂₄ 967-2107	OD2 ₁₀₅ -HE2 ₂₂₄ 974-2107	OD1 ₁₀₅ -H ₄₀ 969-327	OD1 ₁₀₅ -HG1 ₄₀ 969-331	OD1 ₁₀₅ -H ₁₀₆ 969-987
$\lambda=0$ (R)	83%	59%	100%	100%	12%
$\lambda=1$ (S)	88%	69%	100%	100%	87%

4. CONCLUSION

Free energy calculations based on molecular mechanics using a thermodynamic integration protocol appears to be a powerful computational tool to predict enzyme enantioselectivity considering the accuracy, efficiency and cost-effectiveness it can achieve. It is shown that the free energy difference resulting from the arrangement of chiral centers in the truncated tetrahedral intermediates plays an important, but not a decisive role in the enantioselectivity of the enzyme. As can be expected, the whole-enzyme model, which includes the influence of the enzyme environment, gives a more accurate estimate of the enzyme enantioselectivity value. In addition, this model allows prediction of the inversion of the enantioselectivity that has been observed experimentally for the CaLB mutant Trp104Ala. The productive and non-productive binding modes suggested previously might not apply when either the substrate or the enzyme experiences substantial structural changes. It appears that the computational protocol described here, offers a rational approach to identify targets for protein engineering to improve the enantioselectivity of this type of enzyme.

Acknowledgements

This research work has been supported by the Delft University of Technology in the framework of the Delft Interdisciplinary Research Centre "Life Tech".

References

- (1) Detar, D. F. *Biochemistry* 1981, 20, 1730-1743.
- (2) Wipff, G.; Dearing, A.; Weiner, P. K.; Blaney, J. M.; Kollman, P. A. *J Am Chem Soc* 1983, 105, 997-1005.
- (3) Norin, M.; Norin, T.; Hult, K. *Biocatalysis* 1993, 7, 131-147.
- (4) Norin, M.; Haefner, F.; Achour, A.; Norin, T.; Hult, K. *Protein Sci* 1994, 3, 1493-1503.
- (5) Raza, S.; Fransson, L.; Hult, K. *Protein Sci* 2001, 10, 329-338.
- (6) Haffner, F.; Norin, T.; Hult, K. *Biophys J* 1998, 74, 1251-1262.

- (7) Overbeeke, P. L. A.; Ottosson, J.; Hult, K.; Jongejan, J. A.; Duine, J. A. *Biocatal Biotransfor* 1999, 17, 61-79.
- (8) Overbeeke, P. A.; Orrenius, S. C.; Jongejan, J. A.; Duine, J. A. *Chem Phys Lipids* 1998, 93, 81-93.
- (9) Aqvist, J.; Warshel, A. *Chem Rev* 1993, 93, 2523-2544.
- (10) Ottosson, J.; Rotticci-Mulder, J. C.; Rotticci, D.; Hult, K. *Protein Sci* 2001, 10, 1769-1774.
- (11) Ottosson, J.; Fransson, L.; Hult, K. *Protein Sci* 2002, 11, 1462-1471.
- (12) Kollman, P. A.; Kuhn, B.; Perakyla, M. *J Phys Chem B* 2002, 106, 1537-1542.
- (13) Colombo, G.; Toba, S.; Merz, K. M. *J Am Chem Soc* 1999, 121, 3486-3493.
- (14) Micaelo, N. M.; Teixeira, V. H.; Baptista, A. M.; Soares, C. M. *Biophys J* 2005, 89, 999-1008.
- (15) Pearlman, D. A.; Kollman, P. A. *J Chem Phys* 1989, 91, 7831-7839.
- (16) Zhou, Y.; Oostenbrink, C.; Van Gunsteren, W. F.; Hagen, W. R.; De Leeuw, S. W.; Jongejan, J. A. *Mol Phys* 2005, 103, 1961-1969.
- (17) Zhou, Y.; Oostenbrink, C.; Jongejan, A.; Van Gunsteren, W. F.; Hagen, W. R.; De Leeuw, S. W.; Jongejan, J. A. *J Comput Chem* 2006, 27, 857-867.
- (18) Anderson, E. M.; Karin, M.; Kirk, O. *Biocatal Biotransfor* 1998, 16, 181-204.
- (19) Hoff, B. H.; Waagen, V.; Anthonsen, T. *Tetrahedron-Asymmetr* 1996, 7, 3181-3186.
- (20) Frykman, H.; Ohrner, N.; Norin, T.; Hult, K. *Tetrahedron Lett* 1993, 34, 1367-1370.
- (21) Ohrner, N.; Martinelle, M.; Mattson, A.; Norin, T.; Hult, K. *Biotechnol Lett* 1992, 14, 263-268.
- (22) Guibe-Jampel, E.; Chalecki, Z.; Bassir, M.; GeloPujic, M. *Tetrahedron* 1996, 52, 4397-4402.
- (23) Reetz, M. T.; Dreisbach, C. *Chimia* 1994, 48, 570.
- (24) Chen, C. S.; Fujimoto, Y.; Girdaukas, G.; Sih, C. J. *J Am Chem Soc* 1982, 104, 7294-7299.
- (25) Straathof, A. J. J.; Jongejan, J. A. *Enzyme Microb Tech* 1997, 21, 559-571.
- (26) Vantol, J. B. A.; Kraayveld, D. E.; Jongejan, J. A.; Duine, J. A. *Biocataly Biotransfor* 1995, 12, 119-136.
- (27) Hu, C. H.; Brinck, T.; Hult, K. *Int J Quantum Chem* 1998, 69, 89-103.
- (28) Martinelle, M.; Hult, K. *BBA-Protein Struct M* 1995, 1251, 191-197.
- (29) Magnusson, A. O.; Rotticci-Mulder, J. C.; Santagostino, A.; Hult, K. *Chembiochem* 2005, 6, 1051-1056.
- (30) Magnusson, A. O.; Takwa, M.; Harnberg, A.; Hult, K. *Angew Chem Int Edit* 2005, 44, 4582-4585.
- (31) Orrenius, C.; Haeffner, F.; Rotticci, D.; Ohrner, N.; Norin, T.; Hult, K. *Biocataly Biotransfor* 1998, 16, 1-15.
- (32) Rotticci, D.; Haeffner, F.; Orrenius, C.; Norin, T.; Hult, K. *J Mol Cataly B- Enzym* 1998, 5, 267-272.
- (33) Rotticci, D.; Rotticci-Mulder, J. C.; Denman, S.; Norin, T.; Hult, K. *Chembiochem* 2001, 2, 766-770.
- (34) Uppenberg, J.; Hansen, M. T.; Patkar, S.; Jones, T. A. *Structure* 1994, 2, 293-308.
- (35) Grochulski, P.; Bouthillier, F.; Kazlauskas, R. J.; Serreqi, A. N.; Schrag, J. D.; Ziomek, E.; Cygler, M. *Biochemistry* 1994, 33, 3494-3500.
- (36) Norin, M.; Haeffner, F.; Hult, K.; Edholm, O. *Biophys J* 1994, 67, 548-559.
- (37) CRC Handbook of Chemistry and Physics; CRC Press, LLC: 2005.

- (38) Antosiewicz, J.; McCammon, J. A.; Gilson, M. K. *J Mol Biol* 1994, 238, 415-436.
- (39) Walser, R.; Mark, A. E.; Van Gunsteren, W. F.; Lauterbach, M.; Wipff, G. *J Chem Phys* 2000, 112, 10450-10459.
- (40) Van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hunenberger, P. H.; Kruger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G. *Biomolecular Simulation: The GROMOS96 Manual and User Guide*; Hochschulverlag AG an der ETH Zurich: Zurich, 1996.
- (41) Oostenbrink, C.; Villa, A.; Mark, A.E.; van Gunsteren, W.F. *J Comput Chem* 2004, 25, 1656-1676.
- (42) Scott, W. R. P.; Hunenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennel, J.; Torda, A. E.; Huber, T.; Kruger, P.; Van Gunsteren, W. F. *J Phys Chem A* 1999, 103, 3596-3607.
- (43) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Dinola, A.; Haak, J. R. *J Chem Phys* 1984, 81, 3684-3690.
- (44) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J Comput Phys* 1977, 23, 327-341.
- (45) Tironi, I. G.; Sperb, R.; Smith, P. E.; Van Gunsteren, W. F. *J Chem Phys* 1995, 102, 5451-5459.
- (46) Heinz, T. N.; Van Gunsteren, W. F.; Hunenberger, P. H. *J Chem Phys* 2001, 115, 1125-1136.
- (47) Kirkwood, J. G. *J Chem Phys* 1935.
- (48) Colombo, G.; Ottolina, G.; Carrea, G.; Merz, K. M. *Chem Commun* 2000, 559-560.
- (49) Colombo, G.; Carrea, G. *J Biotech* 2002, 96, 23-33.
- (50) Carrea, G.; Ottolina, G.; Riva, S. *Trends Biotechnol* 1995, 13, 63-70.
- (51) Topf, M.; Varnai, P.; Schofield, C. J.; Richards, W. G. *Proteins* 2002, 47, 357-369.
- (52) Klibanov, A. M. *Trends Biochem Sci* 1989, 14, 141-144.
- (53) Toba, S.; Hartsough, D. S.; Merz, K. M. *J Am Chem Soc* 1996, 118, 6490-6498.
- (54) Berglund, P. *Biomol Eng* 2001, 18, 13-22.
- (55) Yasufuku, Y.; Ueji, S. *Bioorg Chem* 1997, 25, 88-99.
- (56) Orrenius, C.; Norin, T.; Hult, K.; Carrea, G. *Tetrahedron-Asymmetr* 1995, 6, 3023-3030.
- (57) Rariy, R. V.; Klibanov, A. M. *Biocataly and Biotransfor* 2000, 18, 401-407.
- (58) Wehtje, E.; Costes, D.; Adlercreutz, P. *J Mol Cataly B-Enzym* 1997, 3, 221-230.
- (59) Manetti, F.; Mileto, D.; Corelli, F.; Soro, S.; Palocci, C.; Cernia, E.; D'Acquarica, I.; Lotti, M.; Alberghina, L.; Botta, M. *BBA-Protein Struct M* 2000, 1543, 146-158.
- (60) Uppenberg, J.; Ohrner, N.; Norin, M.; Hult, K.; Kleywegt, G. J.; Patkar, S.; Waagen, V.; Anthonsen, T.; Jones, T. A. *Biochemistry* 1995, 34, 16838-16851.
- (61) Sainz-Diaz, C. I.; Wohlfahrt, G.; Nogoceke, E.; Smeyers, Y. G.; Menge, U. *Theochem-J Mol Struct* 1997, 390, 225-237.
- (62) Gruber, K. *Proteins* 2001, 44, 26-31.
- (63) Tomic, S.; Ramek, M. *J Mol Cataly B-Enzym* 2006, 38, 139-147.
- (64) Stanton, R. V.; Perakyla, M.; Bakowies, D.; Kollman, P. A. *J Am Chem Soc* 1998, 120, 3448-3457.
- (65) Topf, M.; Varnai, P.; Richards, W. G. *J Am Chem Soc* 2002, 124, 14780-14788.
- (66) Felluga, F.; Pitacco, G.; Valentin, E.; Coslanich, A.; Fermeglia, M.; Ferrone, M.; Pricl, S. *Tetrahedron-Asymmetr* 2003, 14, 3385-3399.
- (67) Singh, U. C.; Kollman, P. A. *J Comput Chem* 1984, 5, 129-145.

- (68) Hu, C. H.; Brinck, T.; Hult, K. *Int J Quantum Chem* 1998, 69, 89-103.
- (69) Warshel, A.; Narayshabo, G.; Sussman, F.; Hwang, J. K. *Biochemistry* 1989, 28, 3629-3637.
- (70) Dreveny, I.; Kratky, C.; Gruber, K. *Protein Sci* 2002, 11, 292-300.
- (71) Bocola, M.; Otte, N.; Jaeger, K. E.; Reetz, M. T.; Thiel, W. *Chembiochem* 2004, 5, 214-223.

CONCLUSIONS AND OUTLOOK

CONCLUSIONS

Computational prediction of enzyme enantioselectivity is of great practical interest due to the number of industrial applications of enzymes in the preparation of enantiomerically pure compounds. Much effort has already been put into the design of computational protocols to predict enzyme enantioselectivity efficiently. However, this remains difficult and unless the researchers continue until an acceptable value is reached, the predicted results are commonly off by one or more orders of magnitude from experimental data.¹ It has been the goal of this thesis work to design a computational protocol for the prediction of enzyme enantioselectivity with improved efficiency and accuracy.

Much like the work reported earlier, our strategy has been to employ the tetrahedral intermediate structure to approximate the transition state in the simulation. By free energy calculation based on MM/MD methods we have, for the first time, brought the calculated *E*-value very close to the experimental observations. Several substrates, all secondary alcohols but with difference size and composition, have been used to test two models (truncated tetrahedral intermediate (TTI) model and whole enzyme model) that were built to calculate the free energy difference of diastereomeric substrate-enzyme tetrahedral intermediate complex during the catalysis by the lipase CaLB. The TTI model, though not as accurate as the enzyme model, allowed a rough and quick estimate of the enzyme enantioselectivity. The whole enzyme models, though more time consuming than the TTI model, provided a substantially more accurate prediction. The result from a blind test on the mutant Trp104Ala with the whole enzyme model is very encouraging. As the protocol is robust and easy to implement in high performance computing based on PC-cluster technology, it is considered to be a very promising computational tool for the future study of enzyme enantioselectivity.

Several factors have contributed to the success achieved so far. The force field Gromos96 was tested intensively for many biomolecular systems as well as the peptides in our study, increasing its reliability in the application for the enzyme system in our study. The soft-core methodology embedded avoids the singularity problem in the sampling. The united atom option in the Gromos96 force field allows us to invert the chirality via the change of improper dihedral angle, a more universal approach than the “vanishing-and-growing atoms” approach applied in other methods. The pathway during the free energy perturbation is carefully chosen: by inserting an intermediate state with a relaxed structure in between, a smoother perturbation pathway is found, improving the quality of free energy calculation. The thermodynamic integration (TI) protocol employed seems robust and efficient for the calculation. All these factors enable us to apply the MM/MD method to study enzyme enantioselectivity without using the computationally expensive QM method.

The testing of the force field and free energy calculation on peptides led to a side line of this thesis: a study of homochirality propagation in nature during evolution. From the present findings it would appear unlikely that the intrinsic stability difference between homo- and heterochiral dipeptides has been a driving force in a primordial selection process leading to the incorporation of amino acids with a single enantiomeric configuration in natural

proteins. The preference depends on the size, character, accessible conformational space of the peptides and on temperature and solvent composition.

OUTLOOK

Perspective

By employing molecular modeling, we have been able to simulate an important aspect of an enzymatic reaction process by computer in a very moderate time frame. This would seem to be an ambitious project since nature has had billions of years to evolve while we are attempting to unencrypt one of its secrets within our short life time!

Many models, from semi-empirical ones based on a classical force field to high level *ab initio* QM, have been built to study enzymatic reactions, but their applications have been more or less restricted due to the limitation of computing power. This is especially true for QM calculations. Even with the fast development of high performance computing, our quest for computing power for biomolecules simulation is unlikely to end in the coming decade. On the other hand, computer simulation will play an increasingly important role in the study of properties of biomolecules including enzyme enantioselectivity. More efficient and accurate models based on more advanced algorithms will appear which brings us closer to an accurate description and prediction of biomolecules properties.

It can be expected that further developments of force field, QM/MM, and free energy calculations will take place. As the classical force field has reached its limit, inclusion of explicit polarization of the atomic group is needed for the next generation of force fields.²⁻⁶ QM calculations like CPMD are gaining momentum, but they are unlikely to replace MM calculations. The simplicity of MM calculations enables one to invest more time in the sampling of the conformational space, instead of describing the electronic structure in QM which is not always necessary. This has been illustrated by the study on enzyme enantioselectivity in Chapter 5. However, it can be foreseen that there could well be a better integration of QM and MM approaches in the future.

As one of the seven DIOC LifeTech projects launched by Delft University of Technology in 2001 (among the others 1.Molecular recognition of single Biomolecules. 2.Single biomacromolecule electronic properties 3.Single biomacromolecule arraying and catalysis 4.Nanostructured biomacromolecules. 5.Simulation of transport processes in protein crystals and gels 6. Nanoarrays-based multi-parameter diagnostics at the nucleic acid and protein level),⁷ this thesis focuses more on the computational aspects to study the biomolecules. It is complimentary to approaches in other projects relying more on engineering and scientific experimental aspects in exploring the “Molecular Recognition of Biomacromolecules in Life Science and Technology” in order to “Make the Biorecognition-Tools for Tomorrow”, the collective goal of the DIOC LifeTech program. This reflects the fact that life science becomes more multi-disciplinary and the trend that the integration of biology, engineering and computation will be further strengthened, accelerating the process of our understanding and exploring nature.

Recommendations

The accuracy of MM/MD calculation is dependent on the quality of the force field used in the simulation. The Gromos96 force field used in our current work needs to be further evaluated in simulations with other enzymes and substrates for the enantioselectivity computational study. The thermodynamic integration protocol needs to be further tested. Both forward and backward integration are advised. Some efficient sampling methods e.g. local elevation to tackle the convergence problem encountered in our free energy calculation are worthwhile trying.^{8;9}

The current work mainly focuses on the productive orientation of fast-reacting substrate. A more detailed investigation into productive orientation of slow-reacting substrate is needed in order to get a more systematic overview of the stereo selective enzyme system.

A model with a complexity between TTI and whole enzyme could be tested. For example, a catalytic triad model consisting of tetrahedral intermediate, catalytic triad of the enzyme might lead to a better trade-off between the efficiency and accuracy than previous models.

A comparative study between the current MM/MD methods for free energy calculation and QM/MM methods will be useful when sufficient computing power and suitable protocols are available.

The benefits and costs of HPC need to be carefully assessed before it is set up and implemented in a project. This is especially true in industry, where timing and return on investment are of high priority. The desirable time period to develop a new biocatalyst should be as short and possible, usually in 3 months. HPC only becomes interesting when it can help accelerate this development process at lower cost. With the free energy calculation protocol developed for enzyme enantioselectivity prediction in this thesis, the simulation can be completed on a twenty-node PC cluster (64-bits, Intel dual Xeon CPU) within one week for a total simulation time of 45ns for a system consisting enzyme and aqueous solvent (~30,000 atoms). Despite the reasonably short simulation time needed, the total cost of €50,000 ~ €60,000 on the HPC facilities can only be justified if the simulation can deliver rather accurate results which give experimentalists a good guideline, and the simulation protocol is robust enough and easy to implement.

Reference

- (1) Kazlauskas, R. J. *Current Opinion in Chemical Biology* 2000, 4, 81-88.
- (2) Oostenbrink, C.; Villa, A.; Mark, A.E.; Van Gunsteren, W.F. *Journal Of Computational Chemistry* 2004, 25, 1656-1676.
- (3) Yu, H. B.; Hansson, T.; Van Gunsteren, W. F. *Journal of Chemical Physics* 2003, 118, 221-234.
- (4) Van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *Journal of Computational Chemistry* 2005, 26, 1701-1718.
- (5) Christen, T.; Hunenberger, P. H.; Bakowies, D.; Baron, R.; Burgi, R.; Geerke, D. P.; Heinz, T. N.; Kastenholz, M. A.; Krautler, V.; Oostenbrink, C.; Peter, C.; Trzesniak, D.; Van Gunsteren, W. F. *Journal of Computational Chemistry* 2005, 26, 1719-1751.
- (6) Chen, B.; Xing, J. H.; Siepmann, J. I. *J Phys Chem B* 2000, 104, 2391-2401.
- (7) *Molecular Recognition of Biomacromolecules in Life Science and Technology*.
<http://www.tnw.tudelft.nl/live/pagina.jsp?id=26800cf6-0f3b-4148-ac99-af61829d8ae6&lang=en> . 2001.
- (8) Huber, T.; Torda, A. E.; Van Gunsteren, W. F. *J.Computer Aided Molecular Design* 1994, 8, 695-708.
- (9) Laio, A.; Parrinello, M. *Proceedings of the National Academy of Sciences of USA* 2000, 99, 12562-12566.

CURRICULUM VITAE

Yu Zhou was born in Chongqing, China in January 12, 1971. After completing secondary school, he did his undergraduate study at the Sichuan University, where he majored in biochemistry and graduated cum laude with a BSc degree in 1992. From 1992, he worked for three years in the group of Prof. Jing-xing Wang and Prof. Xiao-pu Xiao at the Peking Union Medical College, Chinese Academy of Medical Science obtaining his MSc degree in biochemistry. His thesis was titled “Preparation of monoclonal antibody against human blood factor VIII/vWF and its application in immunoassay and preparative chromatography. From 1995 to 2000, Yu worked as a researcher at the Chinese Academy of Medical Sciences and led a group in the R&D work for plasma derived products and therapeutic recombinant proteins. In 2000, he got a scholarship from Delft University of Technology for an advanced study in biotechnology. He did his diploma thesis “Application of simulated moving bed in purification of recombinant human insulin” in the Bioseparation Group, Department of Biotechnology, Delft University of Technology and in the Protein Separation group, Novo Nordisk AS, Denmark under the supervision of Prof. dr. L.A.M. van der Wielen. He graduated in 2003 as a Master in Bioprocess Technology at the Delft University of Technology. From 2002 to 2006 he carried out PhD research work involving the “Computational study of enzyme enantioselectivity” in the Enzymology Section, Department of Biotechnology of the Delft University of Technology under the supervision of Dr. ir. Jaap A. Jongejan, Prof. dr. Wilfred R. Hagen and Prof. dr. Simon W. de Leeuw. From 2006, he has been working as a consultant for Inspirees International in the Netherlands.