Entwicklung und Anwendung einer flexiblen Dockingmethode für Enzyme und Substrate: *Substrate-imprinted Docking*

Von der Fakultät Energie-, Verfahrens- und Biotechnik der Universität Stuttgart zur Erlangung der Würde eines Doktors der Naturwissenschaften (Dr. rer. nat.) genehmigte Abhandlung

Vorgelegt von

Benjamin Juhl
aus Kiel

Hauptberichter:	Prof. Dr. Jürgen Pleiss
Mitberichter:	Prof. Dr. Bernhard Hauer
Tag der mündlichen Prüfung:	20.Januar 2011

Institut für Technische Biochemie
2011

Teile dieser Arbeit wurden bereits veröffentlicht / Parts of this thesis have been published previously:

1. P Benjamin Juhl, Peter Trodler, Sadhna Tyagi, Jürgen Pleiss, Modelling substrate specificity and enantioselectivity for lipases and esterases by substrate-imprinted docking. 2009, *BMC Structural Biology* 9:39.

2. P Benjamin Juhl, Kai Doderer, Frank Hollmann, Oliver Thum, Jürgen Pleiss, Engineering of *Candida antarctica* lipase B for hydrolysis of bulky carboxylic acid esters. 2010, *Journal of Biotechnology* 150:4

3. David I. Habeych, P Benjamin Juhl, Jürgen Pleiss, Gerrit Eggink, Diana Vanegas, Carmen G. Boeriu, Biocatalytic Synthesis of Polyesters from Sugar-Based building blocks using immobilized Candida antarctica lipase B. 2011, *Journal of Molecular Catalysis B* Accepted Manuscript, Article in Press

4. Widmann, M., Juhl, P., B., Pleiss, J., Structural classification by the Lipase Engineering Database: a case study of Candida antarctica lipase A. 2010, *BMC Genomics* 11:123

Danksagung

Für die wissenschaftliche Betreuung und die Überlassung der verschiedenen Themen dieser Arbeit möchte ich mich bei Herrn Prof. Dr. Jürgen Pleiss bedanken. Die enthusiastische, aber auch kritische Diskussion neuer Ideen und Ergebnisse trug maßgeblich zum Erfolg dieser Arbeit bei.

Bei Prof. Dr. Schmid und Prof. Dr. Hauer möchte ich mich für die Möglichkeit bedanken meine Arbeit um Institut für Technische Biochemie durchzuführen. Die Arbeitsbedingungen waren immer hervorragend, und die enge Zusammenarbeit zwischen Labor und Bioinformatik sorgte für ein kreatives und diskussionsfreudiges Umfeld.

Von den Mitgliedern der Bioinformatik Arbeitsgruppe möchte ich im Besonderen Dr. Fabian Bös, Dr. Michael Widmann, Dr. Peter Trodler und Dr. Alexander Steudle danken. Ihr Rat und ihre Hilfe waren eine große Unterstützung bei der Bewältigung der technischen und wissenschaftlichen Herausforderungen während dieser Arbeit.

Abschließend möchte ich meinen Eltern Marlene Werfl und Prof. Dr. Paulgeorg Juhl für die Unterstützung während dem Studium und der Promotion danken.

Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis

Abkürzungsverzeichnis

Proteine

BCL	*Burkholderia cepacia* Lipase
CALA	*Candida antarctica* Lipase A
CALB	*Candida antarctica* Lipase B
CRL	*Candida rugosa* Lipase
huBuChE	humane Butyrylcholinesterase
PPL	*Porcine pancreatic* Lipase
TcAChE	*Torpedo californica* Acetylcholinesterase

Aminosäuren

A, Ala	Alanin
C, Cys	Cystein
D, Asp	Asparaginsäure
E, Glu	Glutaminsäure
F, Phe	Phenylalanin
G, Gly	Glycin
H, His	Histidin
I, Ile	Isoleucin
K, Lys	Lysin
L, Leu	Leucin
M, Met	Methionin
N, Asn	Asparagin
P, Pro	Prolin
Q, Gln	Glutamin
R, Arg	Arginin
S, Ser	Serin
T, Thr	Threonin
V, Val	Valin
W, Trp	Tryptophan
Y, Tyr	Tyrosin

Weitere Abkürzungen

Acetylcholin	ACh
AS	Aminosäure
BLAST	basic local alignment search tool
Butyrylcholin	BuCh
CEO	cyclic ester oligomer
DAH	Di-anhydro hexitols
DBVS	Datenbankverwaltungssystem
EHA	2-Ethylhexansäure
EHA-EE	2-Ethylhexansäureethylester
E. coli	*Escherichia coli*
HOB	Hydroxyoctansäurebutylester
INA	Isononansäure
INA-EE	Isononansäureethylester
LED	*Lipase Engineering Database*
LEO	linear ester oligomer
MDB	Methyldecansäurebutylester
MPP	Methylvaleriansäurepentylester
PDB	Protein Data Bank
PEB	1-Phenylbutyrat
RMSD	*root mean square deviation*
SQL	*Structured Query Language*

Kapitel 1

Zusammenfassung

Digitale Ressourcen in Form von Sequenz- und Strukturdatenbanken sind für Proteine mittlerweile in großem Umfang verfügbar. So finden sich in der UniProt mehr als 500 000 Proteineinträge und in der Protein Data Bank mehr als 60 000 Proteinstrukturen (Stand Juli 2010). Mit diesen frei verfügbaren Ressourcen eröffnen sich viele Möglichkeiten zur Analyse, zur Modellierung und zur Vorhersage von Proteineigenschaften und Proteinvarianten. Die Bioinformatik und computergestützte Strukturbiologie befassen sich mit dieser Thematik und entwickeln Werkzeuge zur Bereitstellung und Auswertung dieser Ressourcen. Insbesondere die Analyse von Proteinstrukturen und Strukturmodellen kann dazu dienen, Proteine gezielt zu verändern.
Die hier vorgestellte Arbeit befasst sich mit der Entwicklung und Anwendung einer flexiblen Dockingmethode für Enzyme und Substrate: *Substrate-imprinted Docking*. Die Methode wurde ausgehend von dem Programm FlexX entwickelt, mittels Literaturdaten evaluiert und dazu verwendet, *Candida antarctica* Lipase B Varianten mit veränderter Substratspezifität zu entwickeln, sowie experimentell

beobachtete Substratspezifitäten auf molekularer Ebene zu erklären. Um die Anwendung dieser neuen Technik auf die Familie der α/β-Hydrolasen zu unterstützen wurde die *Lipase Engineering Database* aktualisiert und erweitert.

Substrate-imprinted Docking ist besonders geeignet, um die Substratspezifität und Stereoselektivität von Enzymen zu erklären und vorherzusagen, da das Substrat in einer zum Übergangszustand analogen Form gedockt wird und so die Funktion des Enzyms in der Stabilisierung des Übergangszustands abbildet. Zusätzlich wird mittels eines geometrischen Filters erkannt, ob das Substrat in der Bindetasche eine Konformation annimmt, die für eine Umsetzung günstig ist (produktiv), oder ob die Konformation des Substrats in der Bindetasche ungünstig für eine Umsetzung ist (nicht-produktiv), und es wird über eine Energieminimierung der Flexibilität des Enzyms Rechnung getragen. *Substrate-imprinted Docking* ist ein mehrstufiger Prozess. Zunächst wird das Substrat in einem tetrahedralen Übergangszustand konstruiert und kovalent an das katalytische Serin des Enzyms gedockt. Die Substratplatzierungen mit dem besten Dockingscore wird verwendet um einen Enzym-Substrat-Komplex zu bilden, der anschließend einer Energieminimierung unterzogen wird, um Überlappungen zwischen Enzym und Substrat zu korrigieren. Durch die Entfernung des Substrats aus dem minimierten Komplex erhält man eine optimierte Proteinstruktur, die für ein erneutes Docken des Substrats verwendet wird. Die aus dem zweiten Docken resultierenden Substratplatzierungen in der optimierten Struktur werden mit einem geometrischen Filter in produktive und nicht-produktive Lösungen eingeteilt und anhand ihres Dockingscores geordnet. Zur Validierung der Methode wurde *Substrate-imprinted Docking* dazu verwendet, um die Substratspezifität von *Candida antarctica* Lipase B, der W104A Variante von *Candida antarctica* Lipase B, *Candida rugosa*

Lipase, *Burkholderia cepacia* Lipase, *Torpedo californica* Acetylcholinesterase und der humanen Butyrylcholinesterase hinsichtlich einiger in der Literatur beschriebener Substrate zu modellieren. Die Modellierung mit *Substrate-imprinted Docking* stimmte zu 81% mit der Substratspezifität und Stereoselektivität aus den Literatur überein, während eine vergleichbare Dockingmethode ohne Berücksichtigung der Proteinflexibilität nur eine Zuverlässigkeit von 50% erzielte.
Nach der retrospektiven Validierung von *Substrate-imprinted Docking* anhand von Literaturdaten wurde dieses neue Dockingverfahren angewendet, um *Candida antarctica* Lipase B Varianten mit einer erhöhten Aktivität für Substrate mit einem sperrigen Acylanteil zu finden. Die beiden Acyldonoren, die für diese Studie verwendet wurden, waren Isononansäure und 2-Ethylhexansäure. Zunächst wurden Aminosäurepositionen identifiziert, die bei der Stabilisierung des Acylanteils in der Bindetasche eine wichtige Rolle spielen. Diese sogenannten *hotspots* wurden durch Docken der beiden Substrate und die Analyse der auftretenden Überlappungen zwischen Substrat und Enzym identifiziert. Fünf *hotspots* wurden identifiziert: D134, T138, Q157, I189 und V190. Um die Aktivität gegenüber sperrigen Substraten zu erhöhen, wurden für jede dieser Aminosäurepositionen mögliche Aminosäureaustausche bestimmt, um den Platz in der Bindetasche zu vergrößern, und ein für die Substrate günstiges biochemisches Umfeld zu schaffen. Mit den fünf *hotspots* und den alternativen Aminosäuren wurde eine kombinatorische *in silico* Bibliothek von Enzymvarianten angelegt. Diese Bibliothek wurde anschließend mittels *Substrate-imprinted Docking* und den beiden Substraten durchmustert. Mit der Durchmusterung wurden die Aminosäureaustausche identifiziert, die besonders häufig in Enzymvarianten mit guten Dockingergebnissen vorkamen. Insgesamt wurden auf diese Weise neun Aminosäureaustausche gefunden, von denen allerdings nur fünf

erfolgreich exprimiert werden konnten. Von diesen fünf Enzymvarianten besaßen alle eine erhöhte Biofeuchtmasseaktivität gegenüber Isononansäureestern, und drei besaßen zusätzlich eine erhöhte Aktivität gegenüber 2-Ethylhexansäureestern. Die *Candida antarctica* Lipase B T138S Variante wurde fermentativ hergestellt und aufgereinigt. Sie besitzt im Vergleich zum Wildtyp eine 5-fach höhere spezifische und 8,6-fach höhere relative Aktivität gegenüber Isononansäureestern.
Ein zweiter Anwendungsfall von *Substrate-imprinted Docking* war die Untersuchung der Spezifität von *Candida antarctica* Lipase B bezüglich der Veresterung von Isosorbid, Isomannid und Isoidid mit Succinat. Experimentell wird bei der Veresterung mit CALB eine Präferenz für Isomannid vor Isosorbid vor Isoidid beobachtet. Diese Präferenz ist entgegengesetzt zu der chemischen Reaktivität der Substrate (Isoidid > Isosorbid > Isomannid). *Substrate-imprinted Docking* zeigte, dass die Konfiguration der Hydroxylgruppen entscheidend für die Spezifität der Veresterung mit *Candida antarctica* Lipase B ist. Alle drei Substrate bestehen aus zwei Ringen, die "V"-förmig angeordnet sind, und besitzen zwei Hydroxylgruppen, die in die "V"-Form hinein (*endo*) oder aus ihr heraus (*exo*) zeigen können. Ester an *endo*-Hydroxylgruppen konnten in einer produktiven Orientierung in die Bindetasche gedockt werden, während Ester an *exo*-Hydroxylgruppen zwar in die Bindetasche gedockt werden konnten, allerdings nicht in einer produktiven Orientierung. *Endo*-Hydroxylgruppen sind demzufolge gute Alkohol-Donoren für *Candida antarctica* Lipase B, während *exo*-Hydroxylgruppen nicht oder nur mit sehr geringen Aktivitäten umgesetzt werden. Dies erklärt die experimentell beobachtete Präferenz von *Candida antarctica* Lipase B für Isomannid vor Isosorbid vor Isoidid, denn Isomannid hat beide Hydroxylgruppen in *endo*-Konfiguration, Isosorbid hat eine Hydroxylgruppe in *endo*- und eine in *exo*-Konfiguration und Isoidid hat

beide Hydroxylgruppen in *exo*-Konfiguration.
Die Entwicklung neuer Biokatalysatoren beginnt oft mit einer Datenbanksuche. Hierfür sind spezialisierte und gut annotierte Datenbanken am besten geeignet. Aus diesem Grund wurde die *Lipase Engineering Database* als Teil dieser Arbeit aktualisiert und erweitert. Die aktualisierte Version 3.0 der Datenbank ist unter http://www.led.uni-stuttgart.de/ öffentlich zugänglich. Im Vergleich zur vorherigen Version wurde die Anzahl der in der Datenbank enthaltenen Protein- und Struktureinträge etwa vervierfacht. Fast alle neuen Einträge konnten bestehenden Familien in der *Lipase Engineering Database* zugeordnet werden. Allerdings war es wegen geringer Sequenzähnlichkeit zu bestehenden Familien sinnvoll sechs neue homologe Familien und eine neue Superfamilie einzuführen. Die neue Superfamilie und vier der neuen homologen Familien beinhalten *Candida antarctica* Lipase A und verwandte Proteine, da diese Gruppe von Proteinen nur sehr geringe Seuqnzähnlichkeiten zu anderen Lipasen besitzt.
Im Rahmen dieser Arbeit wurde die Methode des *Substrate-imprinted Docking* erfolgreich etabliert und validiert. Die Anwendung dieser Dockingmethode bei der Durchmusterung einer *in silico* Bibliothek von 2400 *Candida antarctica* Lipase B Varianten führte Entwicklung von fünf *Candida antarctica* Lipase B Varianten mit verbesserter Aktivität hinsichtlich der Zielsubstrate. Des Weiteren konnte *Substrate-imprinted Docking* dazu verwendet werden, die Molekulare Grundlage für die Präferenz von *Candida antarctica* Lipase B für Isomannid vor Isosorbid vor Isoidid aufzuklären. Zur Unterstützung der Entwicklung neuer Biokatalysatoren aus der Familie der α/β-Hydrolasen wurde die *Lipase Engineering Database* aktualisiert und erweitert, so dass die aktuelle Version etwa viermal so viele Proteinsequenzen und Strukturen enthält wie die vorherige Version.

Kapitel 2

Summary

Digital ressources in the forum of sequence and structure databases on Proteins have become available in recent years to a large extent (2010). UniProt and the Protein Data Bank for instance now contain more than 500 000 protein entries and more than 60 000 structure entries respectively. These freely available online ressources have the potential to allow new ways to analyse, model, and predict protein properties and variants. The disciplnes of bioinformatics, computational biology, and structural biology are aiming at this complex and develope new tools to exploit these digital ressources. Especially the analysis of protein structures and structure models can be used for the design of protein variants with changed properties.

The aim of this work was to develope, update, and apply new digital tools and ressources with a special focus on the family of α/β-hydrolases. This includes the developement of substrate-imprinted docking, an evaluation of the method with published experimental data, the application of substrate-imprinted docking in the design of *Candida antarctica* lipase B variants with altered substrate specifi-

city and the rationalizing of *Candida antarctica* lipase B substrate preference in regards to isosorbide, isomannid, and isoidide, the update of the Lipase Engineering Database, and the integration and classification of *Candida antarctica* lipase A into a new superfamily in the Lipase Engineering Database.

Substrate-imprinted docking was developed to model enzyme-substrate interaction. therefore, substrates are docked as reaction intermediates analogous to the transition state that is stabilized during the enzyme catalysed reaction. Furthermore, substrate-imprinted docking classifies substrate poses in the binding pocket as productive or non-productive by use of geometric filter criteria and accounts for protein flexibility by means of energy minimization. Substrate-imprinted docking is a multistep process. First, a substrate is covalently docked to the active serine as a tetrahedral intermediate form. The best scoring substrate pose is then used to build an enzyme-substrate-complex, which is then energy minimized in order to correct for clashes between substrate and protein atoms. After energy minimization of the complex, the substrate is removed to yield an optimized protein structure, which is used for a seconf round of docking of the same substrate. Substrate poses from this second round of docking are classified into productive and non-productive poses by geometric filter criteria and ranked by their docking score. Substrate-imprinted Docking was validated by modeling systems that have been described in literature and comparing the results. The enzymes used in the validation were *Candida antarctica* lipase B, the W104A variant of *Candida antarctica* lipase B, textitCandida rugosa lipase, *Burkholderia cepacia* lipase, *Torpedo californica* acetylcholine esterase, and the human butyrylcholine esterase, each. Using substrate-omprinted docking to model substrate specificity correlated in 81% of the cases with the previously reported substrate specificity, while a similar

docking procedure without the energy minimization only achieved a correlation of 50% with the data from literature.
Subsequently, substrate-imprinted docking was applied to design *Candida antarctica* lipase B variants with the aim of increasing the activity for substrates with a bulky acyl moiety. The two substrates used in this study were isononaoic acid ethyl ester and 2-ethyl hexanoic acid ethyl ester.First, substrate-imprinted docking of the two substrates into the binding pocket of *Candida antarctica* lipase B was used to identify residues that play a role in the binding of the acyl moiety, so called hotspots. Following hotspots were found by analysing the clashes between substrate and enzyme atoms in the docked substrate poses: D134, T138, Q157, I189, and V190. for each of these five residues alternative amino acoids for selected with the aim of increasing the space in the binding pocket and providing a more beneficial biochemical environment for the bulky, aliphatic substrates. Then, a combinatorial *in silico* library of 2 400 enzyme variants was built from the five identified hotspots and the alternative and wild type amino acids. The liberary was screened by substrate-imprinted docking of isononaoic acid ethyl ester and 2-ethyl hexanoic acid ethyl ester. The screening identified nine amino acid exchanges the occur frequently in enzyme variants that lead to productive substrate poses with a good docking score. *Candida antarctica* lipase B variants with five of these amino acid exchanges could successfully be expressed in *Escherichia coli*. Alle five variants displayed an increased bio wet mass activity towards isononaoic acid ethyl ester, and three variants also displayed and increased activity towards 2-ethyl hexanoic acid ethyl ester. The most promising variant T138S was expressed in fusion with a his-tag, purified, and analysed. The variant displays a 8.6 fold increase in relative activity towards isononaoic acid ethyl ester and a 2-fold increase in relative activity towards 2-ethyl hexa-

noic acid ethyl ester in comparison to wild type *Candida antarctica* lipase B.

In a second application, substrate-imprinted docking was used to rationalize the selectivity of *Candida antarctica* lipase B in the synthesis of isosorbide, isomannide, and isoidide with succinic acid. The preference of *Candida antarctica* lipase B (isomannide > isosorbide > isoidide) is contrawise to the chemical reactivity of the substrates (isoidide > isosorbide > isomannide). The docking of the respective esters into the binding pocket showed, that the configuration of the acylated hydroxyl group is of essential importance for the specificity of the reaction. The three alcohols consist of two five member rings that form a "V"-shaped system. Hydroxyl groups can either point into this "V"-shape (*endo*) or out of it (*exo*). Esters that were formed at *endo* hydroxyl groups could be productively docked into the binding pocket of *Candida antarctica* lipase B, while esters formed at *exo* hydroxyl groups could be docked into the binding pocket, but in a productive orientation. Therefore, *endo* hydroxyl groups are good acyl aceptors in *Candida antarctica* lipase B catalyzed esterification, while *exo* hydroxyl groups are not accepted as acyl acceptors or only at very low activities. This finding explains the experimentally observed preference of *Candida antarctica* lipase B for isomannide over isosorbide over isoidide, as isomannide has two hydroxyl groups in *endo* orientation, isosorbide has one hydroxyl group in *endo* and one in *exo* orientation, and isoidide has two hydroxyl groups in *exo* orientation.

Another part of this work was the update and expansion of the Lipase Engineering Database. The resulting version 3.0 is available at http://www.led.uni-stuttgart.de/. This version has about four times as many protein and structure entries as rpevious versions. Almost all of the new entries could be assigned to already existing protein fami-

lies. However, six new homologous families and one new superfamily had to be introduced into the database. Four of the new homologous families are assigned to the new superfamily and contain *Candida antarctica* lipase A and closely related proteins, as these proteins only share a very low sequence identity with other lipases. The recently published structure of *Candida antarctica* lipase A allowed us to better classify the new superfamily by structure comparison, and to rationalize the mechanisms of substrate binding of the protein.
In conclusion, the method of substrate-imprinted docking has been successfully established and validated. The method was then applied to screen an *in silico* library of 2400 *Candida antarctica* lipase B variants and led to the deisgn of five enzyme variants with increased activity towards target substrates. In a second application, substrate-imprinted docking was successfully used to rationalize the preference of *Candida antarctica* lipase B for isomannid over isosorbid over isoidid in esterification on a molecular level. Furthermore, the Lipase Engineering Database was updated to version 3.0 and enlarged 4-fold, including the integration of *Candida antarctica* lipase A and closely related proteins into the database.

Kapitel 3

Einleitung

3.1 Lipasen

Lipasen (E.C. 3.1.1.3) sind Serin-Hydrolasen und katalysieren die hydrolytische Spaltung von Carboxylestern in Carbonäuren und Alkohole. Neben der Hydrolyse können in wasserfreiem Medium auch andere Reaktionen von Lipasen katalysiert werden. Bei der Veresterung werden neutrale Ester aus geladenen, freien Carbonsäuren und freien Alkoholen gebildet. Bei der Umesterung wird entweder der Acylrest eines Esters mit einem freien Alkohol verestert (Alkoholyse) oder der Alkoholrest eines Esters mit einer freien Carbonsäure (Acidolyse). Bei der Interesterifikation wird der Acylrest eines Esters mit dem Alkoholrest eines zweiten Esters verestert, während der Alkoholrest des ersten Esters mit dem Acylrest des zweiten Esters verestert wird (Abbildung 3.1).

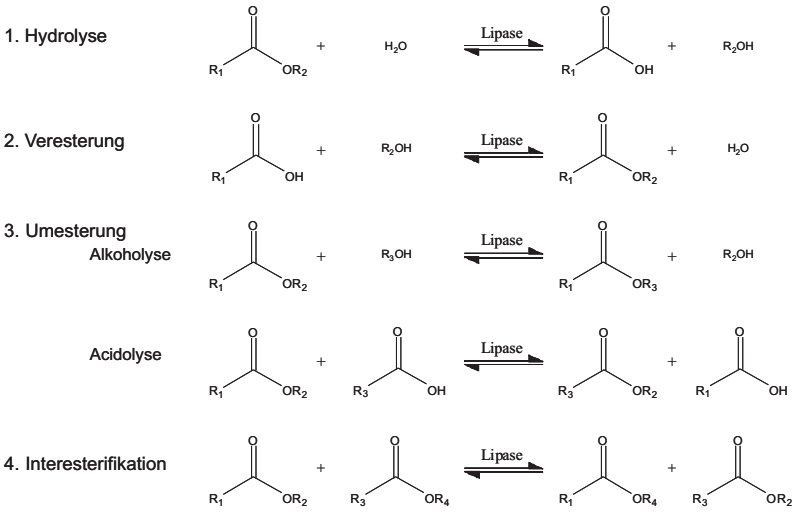

Abbildung 3.1: Von Lipasen katalysierte Reaktionen.

3.1.1 Vorkommen, Substrate und Anwendung von Lipasen

Lipasen kommen ubiquitär in der Natur vor. Bei höheren Tieren kommen Lipasen vor allem im Verdauungstrakt vor, wo sie der Verdauung von Fetten dienen. Im Pflanzenreich haben Lipasen hauptsächlich die Funktion gespeicherte Fette wieder verfügbar zu machen. Auch in Pilzen und Bakterien wurden Lipasen beschrieben (Hasan et al., 2009). Bakterielle Lipasen liegen oft extrazellulär vor, um Fette die nicht direkt aufgenommen werden können zu hydrolysieren. Die freien Carbonsäuren können dann von der Zelle aufgenommen werden (Jaeger et al., 1994). Die natürlichen Substrate von Lipasen sind typischerweise wasserunlösliche Triglyceride, die zu Monoglyceriden, Diglyceriden und freien Fettsäuren hydrolysiert werden. Diese Substrate bilden in wässrigem Medium eine eigene Phase und es entsteht

eine Fett/Öl-Wasser Grenzfläche an welcher die Lipasen aktiviert werden (Grenzflächenaktivierung) (Verger, 1997).
Lipasen besitzen interessante Eigenschaften für Forschung und biotechnologische Anwendungen. Viele Lipasen besitzen ein breites Substratspektrum und sind neben Triglyceriden auch gegenüber einfachen Estern, Lactonen, Thioestern und Amiden aktiv. Die hohe Stereo- und Regioselektivität der Lipasen ermöglicht es, enantiomerenreine Verbindungen herzustellen. Viele Lipasen sind in organischen Lösungsmitteln stabil und können so neben der Hydrolyse auch für organische Synthesen eingesetzt werden, die in wasserfreiem Medium stattfinden müssen. Es wurden auch mehrere Lipasen mit einer hohen Hitzestabilität beschrieben (Haki and Rakshit, 2003), die sich dadurch besonders als Waschmittelzusatz eignen.

3.1.2 Struktur von Lipasen

Lipasen sind Mitglied der Proteinfamilie der α/β-Hydrolasen und benötigen keinen Kofaktor. Die Enzyme aus dieser Familie weisen alle eine bestimmte Tertiärstruktur auf: den α/β-Hydrolase Fold (Ollis et al., 1992). Der kanonische α/β-Hydrolase Fold besteht aus drei Schichten, einer mittleren Schicht die von einem überwiegend parallelen β-Faltblatt gebildet wird, und zwei äußeren Schichten, die aus α-Helices bestehen (Abbildung 3.2). Dabei besteht das β-Faltblatt aus acht β-Strängen (1 bis 8) und die äußeren Schichten aus zwei und vier α-Helics (A bis F). Das zentrale β-Faltblatt ist leicht gebogen und linkshändig verdrillt, so dass die äußersten β-Stränge in einem Winkel von 90° zueinander stehen. Die Helix Schicht auf der konvexen Seite des β-Faltblatts wird von den Helices B, C, D und E gebildet, und die Schicht auf der konkaven Seite von den Helices A

und F. In der Sequenz der Lipasen folgen diese Sekundärstrukturen in der Reihenfolge β1, β2, β4, αA, β3, αB, β5, αC, β6, αD, β7, αE, β8 und αF aufeinander.
Die Abfolge der katalytisch aktiven Aminosäuren in der Sequenz ist immer Nukleophil (Serin), Säure (Aspartat oder Glutamat) und Histidin. Das Histidin befindet sich zwischen β8 und αF auf einem Loop, die Säure direkt hinter β7 und das Serin in einem hoch konservierten Motive. Dieses für α/β-Hydrolasen typische Motiv (G-X-S-X-G) ist in vielen Lipasefamilien konserviert, allerdings können auch Variationen vorkommen (Brenner, 1988). So können die Glycine des Motivs durch andere, kleine Aminosäuren ausgetauscht sein. Das G-X-S-X-G-Motiv befindet sich auf einem kurzen *Turn* zwischen β5 und αC. α-Helix und β-Strang bilden einen Winkel von etwa 20°, weshalb dieses Strukturmotiv als *nucleophilic elbow* bezeichnet wird.
Im Bereich der Helix αD befindet sich bei einigen Lipasen das sogenannte *cap*. Dieser Bereich stellt einen Aufsatz auf den kanonischen α/β-Hydrolase Fold dar, ist sehr variabel und kann verschiedene Funktionen wahrnehmen (Wei et al., 1999). Dies kann die bloße Abschirmung des aktiven Zentrums vom Lösungsmittel sein (Vincent et al., 2003), aber auch die Bildung und Formung der Bindetasche sein, die die Substratespezifität des Enzyms wesentlich beeinflusst (Ericsson et al., 2008).

3.1.3 Candida antarctica Lipase B

*Candida antarctic*a Lipase B (CALB, Swiss-Prot: P41365), auch *Pseudozyma antarctica* Lipase B, wurde zuerst 1993 beschrieben (Patkar et al., 1993) und ist mittlerweile eine der bekanntesten und weit verbreitetsten Lipasen in Forschung und biotechnologischer Anwendung.

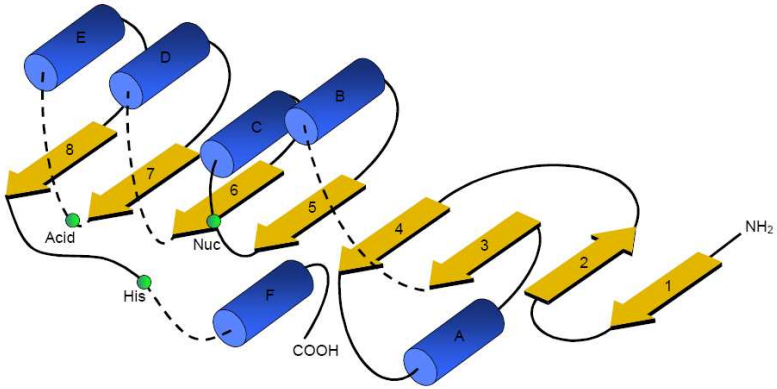

Abbildung 3.2: α/β-Hydrolase Fold von M.Fischer (Fischer, 2004) nucleophilic elbow histidine serine nach Ollis *et al* (Ollis et al., 1992).

CALB wird zusammen mit *Candida antarctica* Lipase A (CALA) von der Hefe *Candida antarctica* produziert (Hoegh et al., 1995). Das Protein besteht aus 342 Aminosäuren und hat ein Molekulargewicht von 35 518 Da. Die ersten 25 Aminosäuren des Proteins bestehen aus einem Signalpeptid (18 AS) und einem Propeptid (7 AS), so dass das prozessierte Protein aus 317 AS besteht (Aminosäurepositionen beziehen sich auf das prozessierte Protein). Trotz der Herkunft aus einer sehr kalten Umgebung sind CALA und CALB zwei der hitzestabilsten bekannten lipolytischen Enzyme. CALB zeichnet sich durch eine hohe Stereoselektivität gegenüber einer Vielzahl sekundärer Alkohole (Anderson et al., 1998) und ein breites Substratspektrum aus. CALB ist in organischen Lösungsmitteln über lange Zeiten stabil und kann deshalb gut für organische Synthesen verwendet werden.

Anders als die meisten anderen Lipasen besitzt CALB kein Lid und dadurch auch keine Grenzflächenaktivierung (Martinelle et al., 1995). CALB besitzt an der Position N74 eine N-Glycosylierungsstelle. Die

katalytische Triade setzt sich aus dem Nukleophil S105, dem Histidin H224 und der Säure D187 zusammen (Abbildung: 3.3). Das *oxyanion hole* wird von T40 und Q106 gebildet. Alle Cysteine des prozessierten Proteins sind an der Bildung von drei Disulfidbrücken beteiligt: C22-C64, C216-C258 und C293-C311. Die Sequenz um das katalytische Serin (TWSQG) weicht von dem bei Lipasen weit verbreiteten G-X-S-X-G Motiv deutlich ab. Die Tertiärstruktur von CALB besteht, in einer geringfügigen Abweichung vom kanonischen α/β-Hydrolase Fold, aus einem zentralen siebensträngigen β-Faltblatt, dessen erster β-Strang antiparallel zu den letzten sechs verläuft. Ein zusätzliches zweisträngiges antiparalleles β-Faltblatt befindet sich am C-Terminus des Proteins. Neben den β-Strängen besitzt CALB 10 α-Helices. α1 befindet sich unmittelbar vor dem ersten β-Strang, α2, α3, α4 und α7 verbinden benachbarte β-Stränge des zentralen Faltblatts und α5, α6 und α10 formen große Teile der Substratbindetasche (Uppenberg et al., 1994).

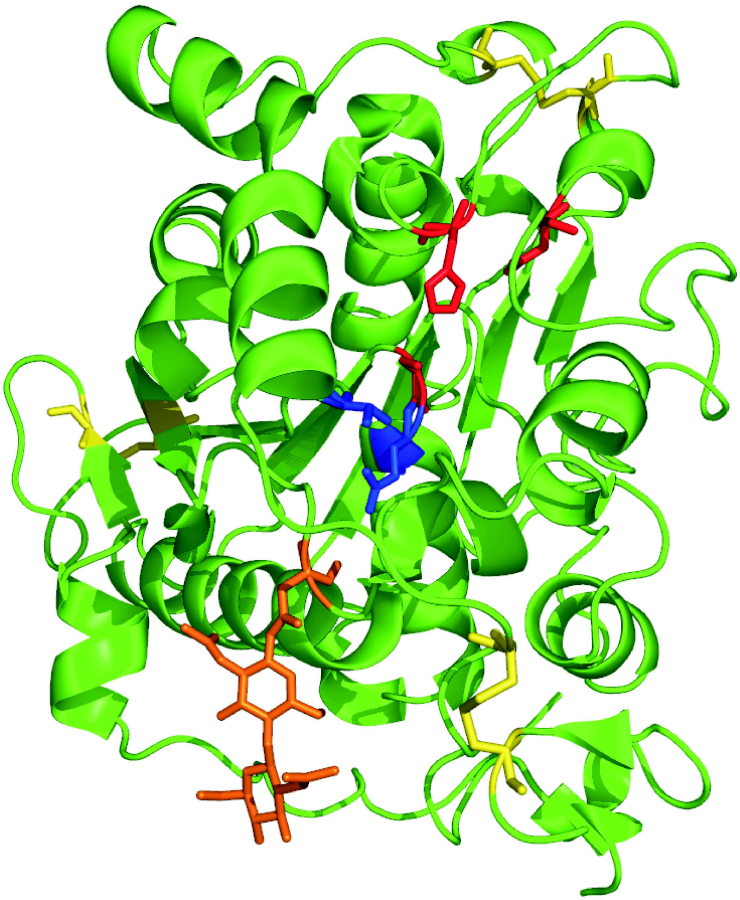

Abbildung 3.3: *Candida antarctica* Lipase B (PDB-ID: 1TCA) in *cartoon* Darstellung (grün). Funktionelle Aminosäuren und organische Moleküle sind in *sticks* dargestellt: *oxyanion hole* (blau), katalytische Triade (rot), Disulfidbrücken (gelb), Glycosilierungsstelle und gebundenes N-Acetyl-D-Glucosamin (orange).

3.1.4 Weitere Lipasen

Neben CALB wurden in dieser Arbeit weitere industriell relevante Lipasen verwendet. *Candida antarctica* Lipase A (CALA) wurde zusammen mit CALB entdeckt (Hoegh et al., 1995, Patkar et al., 1993), allerdings wurde die Struktur von CALA, im Gegensatz zu CALB, erst kürzlich aufgeklärt (Ericsson et al., 2008). Im Gegensatz zu CALB zeigt CALA Grenzflächenaktivierung und Aktivität gegenüber tertiären Alkoholen (Henke et al., 2002, 2003). CALA zeigt die höchste Aktivität in der Hydrolyse von kurzkettigen (C4-C8) Triglyceriden. In der Hydrolyse dieser Substrate zeigt sich eine Besonderheit von CALA, die Präferenz der *sn*2 Position in der Hydrolyse von Triglyceriden (Rogalska et al., 1993). Das pH-Optimum des Enzyms liegt bei 7, und ist noch bei Temperaturen über 90°C aktiv (Domínguez de María et al., 2005). Die Hydrolyseaktivität nimmt gegenüber längerkettigen Substraten rapide ab (Pfeffer et al., 2006), und in Veresterungsreaktionen konnten oft nur sehr geringe Aktivitäten beobachtet werden (Buthe et al., 2005).
Candida rugosa Lipase 1 (CRL) (früher auch *Candida cylindracea*) ist ein 57,2 kDa großes Protein. Neben der Lipase 1 gibt es in *Candida rugosa* noch 4 weitere Isoenzyme mit ähnlichem Molekulargewicht und einer Sequenzidentität zwischen 85% und 90% (Lotti et al., 1993). Die Strukturen von Lipase 1 (Grochulski et al., 1993) und den Isoenzymen Lipase 2 (Mancheño et al., 2003) und Lipase 3 (Ghosh et al., 1995) wurden aufgeklärt und zeigen fast keine strukturellen Unterschiede. Die Substratbindetasche von CRL besteht aus einem langen Tunnel für den Acylanteil des Esters, während es für der Alkoholanteil des Esters keine definierte Tasche gibt. Die Lipase zeigt Grenzflächenaktivierung. In geschlossenem Zustand ist die Bindetasche von einer α-Helix bedeckt (Grochulski et al., 1993, 1994a).

Burkholderia cepacia Lipase (BCL) (früher *Pseudomonas cepacia* Lipase) ist ein bekannter Biokatalysator für die selektive Umsetzung sekundärer Alkohole und ihrer Ester (Bornscheuer and Kazlauskas, 2006, Kazlauskas and Weber, 1998). BCL zeigt Grenzflächenaktivierung und besitzt ein *lid*, das von den Helices $\alpha 4$ und $\alpha 5$ gebildet wird (Schrag et al., 1997). Alle bekannten Strukturen von BCL zeigen die offene Konformation. Die Struktur von BCL wird durch ein Calcium-Ion stabilisiert (Tanaka et al., 2003), welches notwendig für ein aktives Enzym ist (El Khattabi et al., 2003).

3.2 Esterasen

Die Enzymklasse der Esterasen (E.C. 3.1.1) umfasst neben den Lipasen noch weitere Enzymfamilien, die dieselben Reaktionen katalysieren wie Lipasen, aber eine andere Substratspezifität haben. Zu diesen Enzymfamilien gehören unter anderem die Carboxylesterasen (E.C. 3.1.1.1), die Arylesterasen (E.C. 3.1.1.2), die Acetylesterasen (E.C. 3.1.1.5), die Acetylcholinesterasen (E.C. 3.1.1.7), die Cholinesterasen (E.C. 3.1.1.8) und die Cutinasen (E.C. 3.1.1.74). Wie Lipasen katalysieren sie die Bildung oder Hydrolyse von Esterbindeungen, allerdings haben sie zu den Lipasen deutliche Unterschiede in der Substratspezifität und sind eher gegenüber kurzkettigen Substraten aktiv. Allen Esterasen weisen den zuvor beschriebenen α/β-Hydrolase Fold auf, besitzen allerdings kein *lid* wie es für Lipasen beschrieben ist und zeigen deshalb auch keine Grenzflächenaktivierung.

3.2.1 Torpedo californica Acetylcholinesterase und humane Butyrylcholinesterase

Die Acetylcholinesterase aus *Torpedo californica* (TcAChE) besteht aus einem 12-strängigen β-Faltblatt, das von 14 α-Helices umgeben ist (Sussman et al., 1991). TcAChE hat ein Molekulargewicht von 68 kDa und bildet aktive Tetramere. Das aktive Zentrum befindet sich in einem 20 Å tiefen Tunnel. Die katalytische Triade wird von S200, H330 und E327 gebildet. Acetylcholinesterasen sind sehr spezifisch und setzten Ester mit Acyldonoren größer als Acetyl nicht um.
Die Struktur der humanen Butyrylcholinesterase (huBuChE) ist sehr ähnlich zur TcAChE Struktur (Nicolet et al., 2003). Kleine Unterschiede in der Acyltasche ermöglichen allerdings die Bindung und

Umsetzung von Estern mit deutlich größeren Acyldonoren. Die physiologische Funktion der huBuChE ist noch nicht vollständig geklärt. Die huBuChE ist gegenüber einem breiten Substratspektrum aktiv und spielt deshalb eine wichtige Rolle beim Abbau von Toxinen und pharmakologischen Wirkstoffen (Lockridge and Masson, 2000).

3.3 Substrate

Substrate von Lipasen und Esterasen sind Ester, wenn man die Hydrolyse in wässrigem Milieu betrachtet, und Alkohole und Carbonsäuren wenn man die Synthese unter Wasserausschluss betrachtet. Ester bestehen aus einem Alkoholteil und einem Säureteil, die über eine Esterbindung verbunden sind. Systematische Untersuchungen zur Substratspezifität von Lipasen bzw. Esterasen beschränken sich normalerweise auf Variationen eines Teils, entweder Alkohol oder Säure, während sich der andere Teil nicht ändert. Beim Docking wurde ebenfalls eine solche Strategie verfolgt.

3.3.1 Primäre und sekundäre Alkohole

Primäre Alkohole besitzen keine Stereozentrum in direkter Nachbarschaft der OH-Gruppe. Gibt es bei primären Alkoholen ein entferntliegendes (*remote*) Stereozentrum, so zeigen die meisten Lipasen nur eine geringe Enantiopräferenz für dieses Substrat. Selektive Umsetzungen razemischer primärer Alkohole sind fast ausschließlich für *Burkholderia cepacia* (BCL) Lipase und *Porcine pancreatic* Lipase (PPL) beschrieben (Bornscheuer and Kazlauskas, 2006).
Sekundäre Alkohole haben am Kohlenstoffatom, das auch die OH-

Gruppe trägt, nur einen Wasserstoffsubstituenden. In der Regel führt dies dazu, dass an diesem Kohlenstoff ein Stereozentrum vorliegt, der Alkohol also chiral ist. Ein nicht chiraler Alkohol liegt im Falle von sekundären Alkoholen nur vor, wenn die beiden Substituenden des Kohlenstoffs neben der H- und OH-Gruppe identisch sind, wie es zum Beispiel bei 2-Propanol der Fall ist.
Um zu bestimmen, welches Enantiomer eines sekundären Alkohols bevorzugt umgesetzt wird, wurde für Lipasen eine empirische Regel entwickelt, die auf einem Größenvergleich der beiden Substituenden am Stereozentrum beruht (Kazlauskas et al., 1991). Hat ein sekundärer Alkohol am Stereozentrum einen kleinen/mittleren Substituenden (M) und einen großen Substituenden (L), so wird das in Abbildung 3.4 gezeigte Enantiomer bevorzugt in der Hydrolyse umgesetzt. Zu dieser Regel gibt es allerdings einige Ausnahmen und auch Lipasen die eine entgegengesetzte Enantiopräferenz zeigen (Nagy et al., 2006, Szigeti et al., 2008, Xia et al., 2009).
Neben dieser auf der räumlichen Anordnung basierender Regel für die Enantiopräferenz von Lipasen haben elektrostatische Effekte ebenfalls einen wichtigen Einfluss auf die Enantiopräferenz. So ist die Enantiopräferenz bei Substraten mit einem stark polaren kleineren Substituenden deutlich geringer als bei Substraten mit einem unpolaren kleineren Substituenden mit vergleichbarer Größe (Orrenius et al., 1995a, Rotticci et al., 1997).

3.3.2 Tertiäre Alkohole

Tertiäre Alkohole besitzen an ihrem Stereozentrum eine OH-Gruppe und drei weitere, nicht Wasserstoff Substituenden. Deshalb sind sie sterisch anspruchsvoller als sekundäre Alkohole und werden nur von

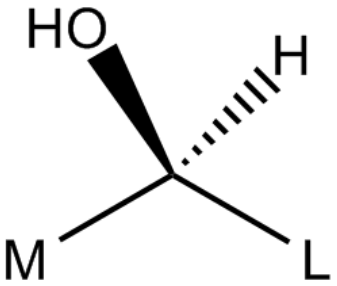

Abbildung 3.4: Das nach der Kazlauskas-Regel bevorzugte Stereoisomer eines sekundäreren Alkohols. Das chirale Kohlenstoffatome besitzt vier Substituenden: einem Wasserstoffsubstituenden (H), die OH-Gruppe und zwei weiteren Substituenden (M,L). M steht für einen kleinen oder mittelgroßen Substituenden wie eine Methyl- oder Ethylgruppe, während L für einen großen Substituenden steht.

wenigen Lipasen und Esterasen als Substrate akzeptiert (OHagan and Zaidi, 1994, Schlacher et al., 1998). Allen Lipasen und Esterasen, die tertiäre Alkohole umsetzten, ist auf Sequenzebene das sogenannte GGGX-Motiv gemein (Fischer and Pleiss, 2003, Henke et al., 2003). Hierbei steht G für Glycin und X für eine beliebige Aminosäure. Das dritte Glycin im Sequenzmotiv ist nicht vollständig konserviert. Bei einigen Enzymen kommt an dieser Position stattdessen ein Alanin vor. Lipasen und Esterasen mit dem weiter verbreiteten GX-Motiv (Pleiss et al., 2000) setzten keine Ester von tertiären Alkoholen um. Die Aminosäure vor dem X im Sequenzmotiv bildet mit der NH-Gruppe des Protein Rückgrats einen Teil des *oxyanion holes*. Durch die zwei zusätzlichen Glycine des GGGX-Motivs entsteht ausreichend Platz und Flexibilität, um anstatt des Wasserstoffs bei sekundären Alkoholen auch kleine Substituenden mit einem oder zwei Kohlenstoffatomen in der Substratbindetasche beherbergen zu

können.

3.3.3 Alpha- und Beta-substituierte Carbonsäuren

Die enzymatische Umsetzung von chiralen, α- und β-substituierten Fettsäuren und ihren Estern ist weit weniger untersucht als die Umsetzung von primären, sekundären und tertiären Alkoholen (Bornscheuer and Kazlauskas, 2006, Haraldsson, 1992). Chirale Carbonsäuren und ihre Ester können Ausgangsstoffe für Medikamente sein (Morrone et al., 1995), finden aber auch Anwendung als Antioxidantien (Majekodunmi et al., 2007), Emulgatoren oder Geschmacksstoffe (Su et al., 2010). Wegen der hohen sterischen Ansprüche von α- und β-substituierten Fettsäuren an die Bindetasche werden Ester dieser Säuren nur in wenigen Fällen effektiv umgesetzt. Die bekannteste Lipase für die Umsetzung solcher Ester ist die Lipase aus *Candida rugosa* (CRL). Für CRL wurde analog zur Kazlauskas Regel für sekundäre alkohole eine Regel für α-substituierte Fettsäuren entwickelt (Ahmed et al., 1994, Franssen et al., 1996).

3.4 Molekulares Docking

Molekulares Docking ist eine computergestützte Methode, um Bindungsaffinität und Bindungsgeometrie eines Liganden an einen Rezeptor vorherzusagen (Taylor et al., 2002). Die Strukturen von Protein und Ligand müssen hierfür bekannt sein. Molekulares Docking wurde für die Wirkstoffforschung entwickelt und wird auch bis heute fast ausschließlich in diesem Gebiet eingesetzt. In der Regel werden

dabei große Bibliotheken kleiner Moleküle in einen Rezeptor gedockt. Ein solches virtuelles Screening kann am Anfang der Entwicklung eines neuen Wirkstoffs Leitstrukturen identifizieren und so die Entwicklungszeit deutlich verkürzen. Da sehr viele kleine Moleküle gedockt werden um einen potentiellen, neuen Wirkstoff zu finden, sind gängige Dockingprogramme darauf ausgelegt, möglichst viele Moleküle in kurzer Zeit zu docken. Alle Dockingprogramme und Verfahren bestehen aus zwei Bestandteilen, einem Dockingalgorithmus, mit dem der Ligand am Rezeptor platziert wird, und einer Bewertungs oder Scoringfunktion, die gefundene Ligandenplatzierungen bewertet und ordnet.

3.4.1 Dockingalgorithmen

In bekannten Dockingprogrammen sind drei grundlegend verschiedene Ansätze vertreten, um für den gedockten Liganden eine Platzierung an der Bindestelle vorherzusagen (Brooijmans and Kuntz, 2003, Dias and de Azevedo, 2008). In einem ersten Ansatz werden mögliche Ligandenkonformationen und die Platzierung an der Bindestelle getrennt betrachtet (FLOG (Miller et al., 1994), EUDOCK (Pang et al., 2001), DOCK (Lang et al., 2009)). Zuerst werden mittels einer Konformationsanalyse Ligandenkonformatioenn mit niedriger Energie bestimmt. Die so gefundenen Ligandenkonformationen werden anschließend rigide, unter Berücksichtigung der verbliebenen sechs Freiheitsgrade von Rotation und Translation, an der Bindestelle platziert. Ein zweiter Ansatz ist es, die Ligandenkonformation und die Platzierung des Liganden an der Bindestelle gleichzeitig zu optimieren (AutoDock (Morris et al., 2009), GOLD (Jones et al., 1997), DOCK (Lang et al., 2009), MCDOCK (Liu and Wang, 1999)). Dieser

Ansatz ist allerdings sehr rechenintensive und es werden üblicherweise stochastische Methoden eingesetzt um zu einer Lösung zu gelangen. Ein dritter Ansatz zerlegt Liganden in rigide Fragmente und baut den Liganden dann ausgehend von einem *base* Fragment inkrementell an der Bindestelle auf. Jedes Fragment wird dabei einzeln optimiert (FlexX (Rarey et al., 1996), DOCK (Lang et al., 2009)).

3.4.2 Scoringfunktionen

Die mittels der Dockingalgorithmen gefundenen Platzierungen von Liganden an einer Bindestelle werden mittels Scoringfunktionen bewertet und so ein Anhaltspunkt für die Bindungsenergie angegeben. Dabei unterscheidet man drei Klassen von Scoringfunktionen (Ajay and Murcko, 1995): (i) empirische Scoringfunktionen, (ii) *knowledge-based* Scoringfunktionen und (iii) physikalische Scoringfunktionen. Empirische Scoringfunktionen finden spezifische Interaktionen zwischen Ligand und Protein, weisen diesen Interaktionen eine Energie zu und approximieren die freie Energie der Ligandenbindung als Summer dieser Energiewerte. Typische Interaktinen sind Wasserstoffbrücken, hydrophobe Interaktionen und Abstoßung bei großer Nähe. *Knowledge-based* Scoringfunktinoen basieren auf einer statistischen Analyyse experimentell bestimmter Kristallstrukturen von Protein-Ligand Komplexen. Hierbei werden Wechselwirkungen zwischen einem Ligand- und einem Rezeptoratom anhand ihrer Häufigkeit Werte zugewiesen. Der Dockingscore für eine Ligandenplatzierung ist die Summe der Werte aller gefundenen Wechselwirkungen. Physikalische Scoringfunktionen verwenden ein Kraftfeld um die Bindungsaffinität als Summe von elektrostatischen und van-der-Waals Wechselwirkungen abzuschätzen.

3.4.3 FlexX

Für alle Dockingverfahren in dieser Arbeit wurde das Programm FlexX (Claussen et al., 2001, Rarey et al., 1996) verwendet. FlexX verwendet einen fragmentbasierten Dockingalgorithmus. Zuerst wird der Ligand in Fragmente zerlegt. Die einzelnen Fragmente erhält man durch schneiden an jeder frei rotierbaren Bindung - ausgeschlossen sind Bindungen in Ringsystemen, Doppelbindungen und Dreifachbindungen. Ein Fragment wird als Basisfragment definiert und an der Bindestelle platziert. Um das Basisfragment an der Bindestelle zu platzieren werden Positionen gesucht an denen das Basisfragment drei Interaktionen mit dem Rezeptor ausbilden kann. Falls das Basisfragment mehr als drei Interaktionen mit dem Rezeptor ausbilden kann werden Platzierungen für alle Kombinationen von drei Interaktionen berechnet und geclustert (Rarey et al., 1997). Neben dieser Methode ein Basisfragment nichtkovalent an der Bindestelle zu platzieren gibt es in FlexX auch die Möglichkeit, ein Basisfragment kovalent an den Rezeptor zu binden.
Ausgehend vom Basisfragment werden alle weiteren Fragmente des Liganden nacheinander angefügt. In einem ersten Schritt des Aufbaus des Liganden an der Bindestelle wird das nächste Fragment an alle gefunden Platzierungen des Basisfragment in allen möglichen Orientierungen angefügt. Alle so erhaltenen Teilliganden aus dem Basisfragment und dem nächsten Fragment werden geclustert und über die Scoringfunktion bewertet, und die am besten bewerteten Teilliganden (Standard: Maximum 500) dienen als Ausgangspunkte für das Anfügen des nächsten Fragments. Dieser Vorgang wird wiederholt bis der Ligand vollständig an der Bindestelle aufgebaut ist.

Alle vollständig aufgebauten Liganden an der Bindestelle werden geclustert, über eine empirische Scoringfunktion bewertet und anhand des daraus resultierenden Dockingscore geordnet. Als Scoringfunktion verwendet FlexX eine leichte Abänderung der Böhm-Funktion (Böhm, 1994, Rarey et al., 1996). Die Scoringfunktion besteht aus drei Teilen (Gleichung 3.1): (1) Der erste Teil besteht aus einer Grundenergie ($\Delta G_0 = 5.4kJ/mol$) und einem Term der den Entropieverlust durch die Ligandenbindung repräsentiert. ΔG_{rot} hat dabei den Wert 1.4 kJ/mol und N_{rot} entspricht der Anzahl der frei rotierbaren Bindungen des Liganden. (2) Der zweite Teil ($\Delta G_{lipo} \sum_{lipo.cont.} f^*$ (ΔR)) ist ein Ausgleichsterm der aus den Überlappungen zwischen Ligand und Rezeptor berechnet wird. (3) Der dritte Teil setzt sich aus den Beiträgen der gefundenen Interaktionen zwischen Ligand und Rezeptor zusammen. Hierbei werden drei Interaktionsarten getrennt behandelt - Wasserstoffbrücken ($\Delta G_{hb} \sum_{neutral\ H-bonds} f(\Delta R, \Delta\alpha)$), Salzbrücken und Wasserstoffbrücken mit geladenen Gruppen (ΔG_{io} $\sum_{ionic\ int} f(\Delta R, \Delta\alpha)$), sowie aromatischen und Van-der-Waals Wechselwirkungen ($\Delta G_{aro} \sum_{aro\ int.} f(\Delta R, \Delta\alpha)$). ΔR steht hierbei für die Abweichung einer Interaktions- oder Bindungslänge von einem Idealwert, $\Delta\alpha$ steht für die Abweichung eines Interaktions- oder Bindungswinkels von einem Idealwert und f/f^* sind heuristische Funktionen für abstands- und winkelabhängige Strafen.

$$\begin{aligned}\Delta G &= \Delta G_0 + \Delta G_{rot} \times N_{rot} \\ &+ \Delta G_{lipo} \sum_{lipo.cont.} f^*(\Delta R) \\ &+ \Delta G_{hb} \sum_{neutral\ H-bonds} f(\Delta R, \Delta\alpha) \\ &+ \Delta G_{io} \sum_{ionic\ int} f(\Delta R, \Delta\alpha) \\ &+ \Delta G_{aro} \sum_{aro\ int.} f(\Delta R, \Delta\alpha)\end{aligned} \tag{3.1}$$

3.4.4 Flexibilität beim Docking

Die Flexibilität des Liganden wird mittlerweile von allen kommerziellen Dockingprogrammen berücksichtigt. Die Flexibilität des Rezeptors hingegen wird nur sehr begrenzt berücksichtigt. Diese fehlende Flexibilität des Rezeptors ist besonders dann von großem Einfluss, wenn nur Proteinstrukturen verfügbar sind die nicht in einer Konformation sind, die eine Ligandenbindung ermöglicht. Gründe hierfür können sein, dass nur die Struktur des apo Proteins bekannt ist, dass ein Ligand gebunden ist, der einen anderen Bindungsmodus besitzt, oder dass es während der Ligandenbindung zu großen Strukturveränderungen wie Domänenbewegungen kommt, die in der vorhandenen Struktur nicht repräsentiert sind. Da schon kleine strukturelle Unterschiede zu großen Änderungen beim Dockingergebnis führen können, ist es oft nicht möglich, in solchen Fällen ein zuverlässiges Dockingergebnis zu erhalten (Sandak et al., 1998). Die Berücksichtigung der Proteinflexibilität könnte helfen, in solchen Fällen zu einem verlässlicheren Dockingergebnis zu kommen. Aus diesem Grund ist die Berücksichtigung der Flexibilität beim molekularen Docking ei-

ner der Forschungsschwerpunkte bei der Weiterentwicklung des molekularen Dockings. Mehrere Gruppen haben Ansätze beschrieben, um die Flexibilität des Proteins teilweise zu berücksichtigen (Cozzini et al., 2008). Die Möglichkeit sich an vorher definierten Gelenkregionen während des Dockings zu bewegen ermöglichte es, einige Domänenbewegungen beim Docking zu berücksichtigen (Sandak et al., 1998). Ein Ansatz behandelt die Aminosäuren der Bindetasche als flexibel um das Dockingergebnis zu verbessern (Alberts et al., 2005, Leach, 1994). Weniger stringente Dockingparameter, die zum Beispiel eine räumliche Überlappung von Ligand und Rezeptor erlauben spiegeln ebenfalls die Proteinflexibilität teilweise wieder (Gabb et al., 1997). Andere Methoden berücksichtigen die Flexibilität des Proteins durch Verwendung mehrerer Proteinstrukturen und Docken den Liganden in ein Ensemble von Proteinstrukturen (Barril and Morley, 2005, Kamper et al., 2006) oder eine mittlere Struktur (Huang and Zou, 2007). Eine weitere Möglichkeit ist es, den Liganden in einen *Pharmacophore Grid* zu docken, der aus mehreren Proteinstrukturen berechnet wurde (Claussen et al., 2001). Alle diese Ansätze verbessern molekulares Docking, aber sie berücksichtigen die Flexibilität des Proteins nur in begrenztem Umfang und können nicht alle möglichen Konformationsänderungen des Rezeptors berücksichtigen, die während der Ligandenbindung auftreten (Murray et al., 1999). Um die Rezeptorflexibilität vollständig zu berücksichtigen ist es notwendig, eine Methode zu verwenden die auch alle Atome des Rezeptors umfasst, wie die Monte Carlo Methode oder molekularmechanische und molekulardynamische Simulationen. Ein Monte Carlo Dockingalgorithmus wurde erfolgreich für Ligand-DNA Bindung entwickelt (Rohs et al., 2005), und molekulardynamische Simulationen wurden mit der Bindetasche eines Rezeptors (Rydberg et al., 2008) oder dem gesamten Ligand-Rezeptor Komplex (Król et al., 2007)

durchgeführt, um Dockingergebnisse zu verbessern.

3.4.5 Docking und rationales Proteindesign

Der Ansatz, molekulares Docking von Substraten und Produkten in Enzyme für Vorhersagen bezüglich Aktivität und Enzymvarianten zu verwenden, ist eine recht neue Entwicklung. Im Bereich der Enzymkatalyse ist molekulares Docking eher etabliert, um die molekularen Grundlagen von experimentell beobachtete Änderungen von Enzymeigenschaften retrospektiv zu erklären (El-Hawari et al., 2009, Höst et al., 2006, Padhi et al., 2009). Der Einsatz von molekularem Docking für das Design von Enzymvarianten lässt sich grob in zwei Ansätze unterteilen: (1) Im ersten Ansatz wird molekulares Docking dazu verwendet, Aminosäurepositionen, so genannte *hotspots*, zu identifizieren, die relevant für die untersuche Enzymaktivität sind. An diesen Aminosäurepositionen wird dann eine Sättigungsmutagenese durchgeführt oder es werden bestimmte Aminosäureaustausche an diesen Positionen durchgeführt, ohne diese vorher über Dockingmethoden zu bewerten. Dieser Ansatz wurde erfolgreich verfolgt, um Enzymvarianten von Glutathiontransferasen (Kapoli et al., 2008), Fructosylamin Oxidasen (Miura et al., 2008) und Carbonylreduktasen (Zhu et al., 2008) mit verbesserter Aktivität zu finden. (2) Der zweite Ansatz geht über den ersten hinaus und identifiziert nicht nur relevante Aminosäurepositionen, sondern identifiziert auch spezifische Aminosäureaustausche an diesen Positionen, um die Enzymaktivität zu verbessern. So wurde ein zum Überganszustand eines Substrats analoges Phosphonat in den Wildtyp und zwei Varianten der humanen Anhydrase II gedockt, um eine Enzymvariante mit erhöhter Aktivität gegenüber mehreren para-nitrophenyl Substraten

zu identifizieren, und es wurde mittels Docking eine Hydantoinase Doppelmutante vorhergesagt, die eine 200-fache Aktivität gegenüber einem Zielsubstrat aufweist (Lee et al., 2009).

3.5 Modellierung von Proteinen

In der Molekularmechanik werden Proteine, Lösungsmittel und alle anderen Moleküle mittels eines Kraftfelds beschrieben. In einem Kraftfeld wird die Gesamtenergie des Systems als Summe mehrerer Einzelenergien beschrieben. In einem molekularmechanischem Kraftfeld setzt sich die Gesamtenergie E typischerweise aus bindenden und nicht-bindenden Einzelenergieen zusammen (Gleichung 3.2).

$$E = E_{Bindungen} + E_{Winkel} + E_{Torsion} + E_{nichtkovalent} \tag{3.2}$$

Die bindenden Energien sind wiederum aufgeteilt in die Energie aus dem Abstand zweier Atome ($E_{Bindeungen}$), die Energie aus dem Bindungswinkel (E_{Winkel}) und der Energie aus den Torsionswinkeln ($E_{Torsion}$). Die nichtbindende Energie ($E_{nichtkovalent}$) setzt sich aus den Beiträgen der Van-der-Waals Wechselwirkung und der elektrostatischen Wechselwirkung zusammen. Neben dem in dieser Arbeit verwendeten AMBER Kraftfeld (Duan et al., 2003) (Gleichung 3.3) sind in der Simulation von Biomolekülen auch das CHARMM (MacKerel Jr. et al., 1998), das GROMOS (van Gunsteren and Berendsen, 1990), das OPLS (Jorgensen and Tirado-Rives, 1988) und das Tripos Kraftfeld (Clark et al., 1989) geläufig.

$$E_{total} = \sum_{Bindungen} K_l(l - l_{eq})^2 + \sum_{\text{Winkel}} K_\theta(\theta - \theta_{eq})^2 + \sum_{\text{Torsion}} \frac{K_n}{2}[1 + \cos(n\phi - \gamma)] + \sum_{i<j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\epsilon R_{ij}} \right] \quad (3.3)$$

l : Bindungslänge

l_{eq} : Bindungslänge im Gleichgewicht

θ : Bindungswinkel

θ_{eq} : Bindungswinkel im Gleichgewicht

n : Anzahl der Minima

ϕ : Periodizität

γ : Phasenverschiebung

ϵ : Dielektrizitätskonstante

R : Abstand zwischen zwei Atomen

q : Partialladung

A, B, K : Konstanten

3.5.1 Energieminimierung

Die Beschreibung eines Moleküls durch ein Kraftfeld ermöglicht es, ausgehend von einer beliebigen Struktur ein lokales Energieminimum für das Molekül zu finden. Ein globales Energieminimum für ein Molekül erfordert es, ein das Minimum für ein Gleichungssystem mit

3N-6 Gleichungen zu finden, wobei N die Anzahl der Atome des Moleküls ist. Für Enzyme mit mehreren Tausend Atomen ist dies mathematisch sehr anspruchsvoll, weshalb sich Energieminimierungen in der Praxis auf ein lokales Minimum beschränken. Um dieses lokale Energieminimum zu finden wird, ausgehend von der Startstruktur, die Molekülstruktur in einem iterativen Prozess so verändert, dass die durch das Kraftfeld beschriebene Energie des Systems mit jedem Schritt abnimmt. Zwei verbreitete Methoden zur Veränderung der Molekülstruktur bei Energieminimierungen sind die *steepest descent* und die *conjugate gradient* Methode. Bei der *steepest descent* Methode wird die Struktur in Richtung des negativen Gradienten der Energiefunktion verändert, während bei der *conjugate gradient* Methode die Richtung der Strukturänderung zusätzlich von der Richtung der Strukturänderung des vorherigen Minimierungsschritts abhängt.

3.6 Die Lipase Engineering Database

Die *Lipase Engineering Database* (LED) (http://www.led.uni-stuttgart.de/) ist eine Datenbank für α/β-Hydrolasen. Die LED wurde im Jahr 2000 erstellt (Pleiss et al., 2000) und seitdem beständig weiterentwickelt. Neue Versionen mit erweiterten und aktualisierten Datensätzen, sowie neuen Funktionalitäten wurden 2003 (Fischer and Pleiss, 2003), 2006 (Fischer et al., 2006) und 2009 (Widmann et al., 2010) veröffentlicht. Wie andere computergestützte Datenbanksysteme besteht die LED aus zwei Komponenten, den Daten und dem Datenbankverwaltungssystem (DBVS). Eine Abfrage oder Manipulation der Daten erfolgt ausschließlich über das DBVS. Die Kommunikation von Anwendern und externen Programmen mit dem DBVS erfolgt mittels *Structured Query Language* (SQL) (Chamberlin and

Boyce, 1974), einer speziell für diesen Zweck entwickelten Sprache, die sich mittlerweile als Standard etabliert hat und von allen großen DBVS verstanden wird.
Die LED ist als relationale Datenbank (Codd, 1970) konzipiert. Bei einem relationalen Datenmodell werden nicht alle Daten in einem Eintrag gespeichert, sondern in Tabellen. Jedem Tabelleneintrag ist in der Tabelle ein Primärschlüssel und optional mehrere Fremdschlüssel zugeordnet. Diese Fremdschlüssel wiederum verweisen auf Primärschlüssel anderer Tabellen. Über diese Verknüpfungen können einem Protein alle in der Datenbank gespeicherten Informationen zu diesem Protein eindeutig zugewiesen werden. Dieses Datenbankmodell erlaubt es, Inhalte gezielt abzufragen und stellt sicher, dass sich Änderungen an einem Eintrag auf die gesamte Datenbank auswirken.
Die erste Version der LED wurde durch Homologiesuche in der GenBank (Benson et al., 2008) (http://www.ncbi.nlm.nih.gov/genbank/) mit annotierten α/β-Hydrolasesequenzen erstellt und enthielt 93 Proteine (Pleiss et al., 2000). Gefundene homologe Proteine mit einem E-Wert von 10^{-10} und kleiner wurden in die LED aufgenommen und über eine BLAST-Suche (Altschul et al., 1990) gegen die LED in Superfamilien und homologe Familien eingeteilt. Bei einer 98%igen Sequenzidentität wurde eine neue Proteinsequenz einem bereits existierenden Proteineintrag zugewiesen.
Die aktuelle Version der LED beinhaltet 18 585 Proteineinträge in 112 homologen und 38 Superfamilien (Widmann et al., 2010). Neben der Proteinsequenz und der Einteilung in Proteinfamilien stellt die LED auch Informationen über den zugehörigen Organismus, die isoelektrische Region, den isoelektrischen Punkt des Proteins, annotierte Alignments, phylogenetische Bäume und überlagerte Proteinstrukturen bereit.

3.7 Zielsetzung

Die Zielsetzung dieser Doktorarbeit war es, eine Methode basierend auf molekularem Docking zur Modellierung und Vorhersage von Substratspezifität und Stereoselektivität von Serinhydrolasen zu entwickeln und diese Methode einzusetzen, um Biokatalysatoren zur Umsetzung von Estern mit sperrigen Acyldonoren zu finden und zu verbessern. Bei der Entwicklung der Dockingmethode sollten Substrate nicht im Grundzustand, sondern in einem Zwischenzustand, der analog zum vom Enzym stabilisierten Übergangszustand ist, untersucht werden. Weitere wichtige Faktoren für die Entwicklung des Dockingverfahrens waren die Berücksichtigung der Proteinflexibilität und die eindeutige Klassifizierung von Substratplatzierungen in produktive und nicht-produktive Platzierungen. Die neu entwickelte Dockingmethode sollte zum Design von Enzymvarianten mit verbesserter Aktivität und zur Erklärung experimentell beobachteter Enzymaktivitäten eingesetzt werden. Zur Unterstützung der Entwicklung neuer Biokatalysatoren aus der Familie der α/β-Hydrolasen sollte CALA in die LED integriert und die LED aktualisiert werden.

Kapitel 4

Ergebnisse und Diskussion

4.1 Modellierung von Substratspezifität und enantioselektivität von Lipasen und Esterasen mittels Substrate-imprinted Docking

(siehe: Modelling substrate specificity and enantioselectivity for lipases and esterases by substrate-imprinted docking, Seite 87)

In den letzten zwanzig Jahren ist die Anzahl der Proteinstrukturen in öffentlichen Datenbanken exponentiell gewachsen. Die Protein Data Bank (PDB) (Berman et al., 2000) beinhaltet mittlerweile mehr als 57 000 Proteinstrukturen (Stand Dezember 2009). Zusätzlich zu diesen experimentell aufgeklärten Proteinstrukturen ist es

auch möglich, zuverlässige Strukturmodelle mittels Homologiemodellierung und verwandter Techniken zu entwickeln (Pieper et al., 2006). Die Entwicklung von neuen, strukturbasierten Methoden zur Vorhersage von Affinität, Aktivität, Spezifität und Stereoselektivität für Enzyme ist ein aktuelles Thema in der Forschung (Ortiz et al., 2006). Molekulares Docking wird regelmäßig in der medizinischen Forschung eingesetzt, um durch Screening großer Molekül datenbanken neue Wirkstoffe zu finden (Cavasotto and Orry, 2007). Neu entwickelte Dockingmethoden wurden mittlerweile erfolgreich zur Vorhersage von Enzymaktivitäten und zum Proteindesign eingesetzt. So war es möglich, anhand von Proteinstrukturen mit unbekannter Funktion, Dipeptid Epimerasen (Kalyanaraman et al., 2008), N-succinyl Aminosäure Racemasen (Song et al., 2007), Lacton Hydrolasen (Xiang et al., 2009) und Adenosin Deaminasen (Hermann et al., 2007) zu identifizieren. Durch Docking von vielen verschiedenen chemisch verwandten Substraten in eine Proteinstruktur wurde nicht nur die generelle Aktivität eines Enzyms, sondern auch neue, bisher unbekannte Substrate für dieses Enzyme identifiziert (Favia et al., 2008, Hermann et al., 2007, Irwin et al., 2005, Kalyanaraman et al., 2005, Macchiarulo et al., 2004). In einer speziellen Anpassung der Dockingmethode an Enzyme und Substrate wurden auch Reaktionszwischenprodukte (Hermann et al., 2006) und Übergangszustände (Rydberg et al., 2008) in Proteine gedockt um auf Funktion, Substrate und Aktivitäten zu schlussfolgern.
Alle diese Beispiele zeigen, dass die Anpassung von molekularem Docking an Enzyme und Substrate ein vielversprechender Ansatz ist, um Enzymaktivitäten vorherzusagen. Allerdings ist es wichtig, dass bei einer solchen Anpassung drei Faktoren berücksichtigt werden:

1. Das Substrat sollte im Übergangszustand oder einer ähnlichen

Form kovalent an das Enzym gedockt werden, denn die katalytische Funktion von Enzymen besteht in der Stabilisierung des Überganszustandes (Warshel and Florián, 1998). Die Betrachtung des Übergangszustands sollte deshalb für die Enzymkatalyse aussagekräftiger sein, als die Betrachtung des Grundzustands.

2. Die Konformation des Substrats in der Bindetasche Es muss zwischen Substratkonformationen die eine Umsetzung begünstigen (produktiv) und solchen die eine Umsetzung nicht begünstigen (nicht-produktiv) unterschieden werden. Bei α/β-Hydrolasen werden in einer produktiven Konformation vier Wasserstoffbrücken gebildet: (a) Zwei Wasserstoffbrücken zwischen dem *oxyanion* und zwei NH-Gruppen des Proteinrückgrats (*oxyanion hole*), (b) eine Wasserstoffbrücke zwischen dem katalytischem Histidin und dem O_γ des Substrats und (c) eine Wasserstoffbrücke zwischen dem katalytischem Histidin und dem Sauerstoff des Alkoholteils des Substratesters. Diese Eigenschaft wird in dieser Arbeit mittels eines geometrischen Filters überprüft (Schulz et al., 2000, Tyagi and Pleiss, 2006).

3. Die Flexibilität des Enzyms sollte neben der Flexibilität des Substrats ebenfalls berücksichtigt werden, da Kristallstrukturen und Homologiemodelle oft nicht in einer für die Substratbindung günstigen Konformation sind.

Um diesen Anforderungen gerecht zu werden haben wir *Substrate-imprinted Docking* entwickelt und an mehreren Lipasen und Esterasen in Kombination mit mehreren Substraten getestet. Die Methode baut auf dem Dockingprogramm FlexX (Rarey et al., 1996) auf, dockt die Substrate kovalent an das Enzym, erkennt produktive Substratkonformationen mittels eines geometrischen Filters und trägt

der Proteinflexibilität über eine Energieminimierung des Enzym-Substrat Komplexes mit dem Programm Amber (Case et al., 2006) Rechnung. Die getesteten Enzyme umfassten *Torpedo californica* Acetylcholinesterase (TcAChE), humane Buturylcholinesterase (huBuChE), *Candida rugosa* Lipase (CRL), *Burkholderia cepacia* Lipase (BCL), *Candida antarctica* Lipase B (CALB) und die W104A Variante von CALB. Um den Einfluss der Energieminimierung auf die Zuverlässigkeit der Dockingmethode zu evaluieren wurden alle Dockingexperimente zuerst ohne Energieminimierung durchgeführt (im folgenden als konventionelles Docking bezeichnet), und anschließend mit Energieminimierung (*Substrate-imprinted Docking*). Die Zuverlässigkeit der Ergebnisse wurde über einen Vergleich mit experimentellen Daten ermittelt, und konventionelles Docking mit *Substrate-imprinted Docking* verglichen.

Das chirale Substrat 1-Phenylethylbutyrat (PEB) wird von CALB stereoselektiv umgesetzt. CALB Wildtyp päferiert R-PEB mit einem E-Wert von 1 300 000, während die W104A CALB Variante keine Präferenz für eines der Enantiomere zeigt (Magnusson et al., 2005a). Beide Enantiomere wurden in fünf CALB Strukturen und fünf Homologiemodelle der W104A Variante gedockt. Ohne Berücksichtigung der Proteinflexibilität (konventionelles Docking) wird eine Übereinstimmung mit den experimentellen Befunden in 80% der Fälle bei CALB Wildtyp und 40% der Fälle bei der W104A Variante erzielt. *Substrate-imprinted Docking* erzielt bei den gleichen Beispielen eine Zuverlässigkeit von 90% für CALB Wildtyp und 100% für die W104A Variante.

Die chiralen Substrate 2- bis 8-Methyldecansäurebutylester (MDB) werden von CRL mit wechselnder Enantiopräferenz und E-Werten zwischen 2,8 und 91 umgesetzt (Hedenström et al., 2002). Enantiomere der sieben Substrate wurden in sieben CRL Strukturen ge-

dockt. Mit konventionellem Docking wird eine Übereinstimmung mit den experimentellen Befunden in 42% der Fälle erzielt. *Substrate-imprinted Docking* erzielt bei den gleichen Beispielen eine Zuverlässigkeit von 59%.

Das Substrat 3-Methylvaleriansäurepentylester (MPP) wird weder von CRL noch von BCL umgesetzt, 2-MPP wird von CRL, aber nicht BCL umgesetzt und 4-MPP wird von CRL und BCL umgesetzt (Kodera et al., 1986). 2-Hydroxyoctansäurebutylester (HOB) wird von CRL und BCL umgesetzt, jeweils mit einer Präferenz für R-HOB (E-Wert ≈ 20) (Sakaki et al., 2002). Enantiomere der Substrate wurden in sieben CRL und sieben BCL Strukturen gedockt. Mit konventionellem Docking wird eine Übereinstimmung mit den experimentellen Befunden für HOB in 64% und MPP in 44% der Fälle erzielt. *Substrate-imprinted Docking* erzielt bei den gleichen Beispielen eine Zuverlässigkeit von 75% für HOB und 66% für MPP.

Acetylcholin (ACh) und Butyrylcholin (BuCh) unterscheiden sich nur geringfügig. Die Acylgruppe in BuCh wird von Buttersäure gebildet, die um zwei CH_2-Gruppen länger ist als Acetat, welches bei Acetylcholin die Acylgruppe bildet. Dieser Unterschied zwischen ACh und BuCh hat zur Folge, dass TcAChE ACh etwa 850 mal besser umsetzt als BuCh (Selwood et al., 1993), während huBuChE beide Substrate etwa gleich gut umsetzt (Moralev and Rozengart, 2007, Nicolet et al., 2003). Beide Substrate wurden in vier TcAChE und vier huBuChE Strukturen gedockt. Mit konventionellem Docking wird eine Übereinstimmung mit den experimentellen Befunden in 50% der Fälle erzielt. *Substrate-imprinted Docking* erzielt bei den gleichen Beispielen eine Zuverlässigkeit von 80%.

Die Ergebnisse zeigen deutlich, dass konventionelles Docking ohne Berücksichtigung der Proteinflexibilität nicht in der Lage ist, die Enzym-Substrat Interaktion zuverlässig zu modellieren. Im Vergleich

zum konventionellen Docking erzielt *Substrate-imprinted Docking* eine deutlich höhere Übereinstimmung mit experimentellen Ergebnissen und ist prinzipiell geeignet, um die Substratspezifität von Enzymen zu modellieren. Wenn Proteinstrukturen mit einem, zum Beispiel durch Inhibitoren in der Kristallstruktur, stark verschobenem aktivem Zentrum nicht berücksichtigt werden steigt die Zuverlässigkeit in der Substratmodellierung von *Substrate-imprinted Docking* von 68% auf 81%. Allerdings konnte die Enantiopräferenz nur für CALB zuverlässig modelliert werden. Die deutlich niedrigeren E-Werte bei CRL und BCL erlaubten es nicht, die Enantiopräferenz dieser beiden Enzyme zu modellieren. *Substrate-imprinted Docking* erlaubt es also, wichtige Interaktionen zwischen Enzymen und Substraten zu modellieren und Substrate für Enzyme zu identifizieren, aber die exakte Vorhersage von Aktivitäten und Enantiopräferenzen ist in naher Zukunft wohl nicht zu erwarten.

Einige Faktoren, wie zum Beispiel Lösungsmittel, Temperatur und Entropie, werden in der hier vorgestellten Methode nicht oder nur sehr vereinfacht berücksichtigt. Neue Methoden, die diesen Faktoren besser Rechnung tragen haben also das Potential, die Zuverlässigkeit der Vorhersagen weiter zu erhöhen. Haben die noch nicht adäquat berücksichtigten Faktoren einen starken Einfluss auf eine Reaktion, kann dies dazu führen, dass die Modellierungen und Vorhersagen in solchen Fällen ungenau oder ungenügend ausfallen. Die freie Energie der Reaktion hat einen temperaturabhängigen Entropiebeitrag. Deshalb sind Aktivitätsraten von der Temperatur abhängig. Dies fällt besonders dann ins Gewicht, wenn die Entropie der bedeutendste Faktor für den E-Wert ist (Ottosson et al., 2002a,b). Neben der Temperatur hat auch das verwendete Lösungsmittel Einfluss auf Aktivität und Enantiopräferenz. Die Anzahl der Lösungsmittel die Anwendung finden ist groß. Neben wässrigen Umgebungen werden

viele verschiedene Kombinationen von organischen Lösungsmitteln verwendet, teilweise mit Zusätzen wie DMSO, um die Löslichkeit der Substrate zu verbessern. Des Weiteren gibt es noch die Möglichkeit in ionischen Flüssigkeiten oder superkritischem Kohlenstoffdioxid zu arbeiten. Eine enzymkatalysierte Reaktion ohne Lösungsmittel, nur in den flüssigen Substraten ist ebenfalls möglich. Ein besseres Verständnis der Entropie und die Entwicklung von computergestützen Methoden um die Entropie in verschiedenen Umgebungen (nicht nur in Wasser) besser und einfacher berechnen zu können, würde es ermöglichen den beiden oben genannten Problemen der Vorhersagbarkeit Rechnung zu tragen. Allerdings gibt es noch einen dritten Faktor, der nichts mit der Entropie zu tun hat, und der genaue Vorhersagen ebenfalls erschwert. Für Modellierungen und Vorhersagen werden üblicherweise freie Proteinstrukturen verwendet. Für industrielle Anwendungen werden aber fast nie gelöste Proteine verwendet. Kommerzielle Enzyme werden üblicherweise auf Trägermatieralien immobilisiert, um die Stabilität, Lagerungstoleranz und Wiederverwendbarkeit der Enzyme zu verbessern. Weder sind die Trägermaterialien homogene Strukturen, noch ist der Vorgang der Immobilisierung für jedes einzelne Makromolekül gleich. Die genauen Molekularen Vorgänge der Immobilisierung sind oft nicht bekannt. Allerdings hat die Immobilisierung nicht nur Einfluss auf Stabilität und Lagerungstoleranz, sondern kann auch Aktivität und Enantiopräferenz des Enzyms verändern.

4.2 Design von Candida antarctica Lipase B für die Umsetzung von Estern sperriger Carbonsäuren

(siehe: Engineering of Candida antarctica lipase B for hydrolysis of bulky carboxylic acid esters, Seite 149)

Candida antarctica Lipase B (CALB) ist ein weit verbreiteter Biokatalysator (Anderson et al., 1998, Gotor-Fernandez et al., 2006, Plou et al., 2002) mit einer hohen Aktivität gegenüber einer großen Bandbreite von primären und sekundären Alkoholen. Allerdings ist die Bandbreite der von CALB akzeptierten Carbonsäuren deutlich kleiner und beschränkt sich fast ausschließlich auf unverzweigte Fettsäuren (Kirk et al., 1992). Um das Anwendungsgebiet von CALB in der Biotechnologie zu erweitern ist es von Interesse, insbesondere die Bandbreite an akzeptierten Carbonsäuren zu vergrößern und um verzweigte Fettsäuren und Fettsäuren mit Substituenden in der Nähe der Säuregruppe zu erweitern.
Um CALB Varianten mit einem derart erweitertem Substratspektrum zu finden wurde CALB für zwei Modellsubstrate, Isononansäureethylester (INA-EE) und 2-Ethylhexansäureethylester (EHA-EE), optimiert. Zum Design von optimierten CALB Varianten wurden zwei Strategien verfolgt:

1. Die beiden Substrate INA und EHA wurden mittels *Substrate-imprinted Docking* (Juhl et al., 2009) in eine CALB Struktur gedockt, um für die Substratbindung relevante Aminosäuren zu identifizieren, sogenannte *hotspots*. Anschließend wurden an diesen Aminosäurepositionen Mutationen eingeführt, einzelne Strukturmodelle erstellt und die CALB Varianten mittels

Substrate-imprinted Docking bewertet. Die besten Varianten wurden in *E. coli* exprimiert und ihre Aktivität bestimmt.

2. Ähnlich der ersten Strategie wurden die beiden Substrate INA und EHA mittels *Substrate-imprinted Docking* in eine CALB Struktur gedockt um für die Substratbindung relevante Aminosäuren zu identifizieren, sogenannte *hotspots*. Für diese Aminosäurepositionen wurden mehrere alternative Aminosäuren ausgesucht, um den Platz in der Bindetasche zu vergrößern und ein für die Substrate günstiges biochemisches Umfeld zu schaffen. Anders als bei der ersten Strategie wurde mit diesen Aminosäurepositionen und identifizierten Aminosäureaustauschen dann eine kombinatorische *in silico* Bibliothek von 2 400 CALB Varianten angelegt, um über Docking der beiden Substrate Aminosäureaustausche zu identifizieren, die öfter als die ursprüngliche Aminosäure zu verbesserten Dockingscores führten. CALB Varianten mit solchen Aminosäureaustauschen wurden anschließend in *E. coli* exprimiert und im Labor analysiert.

Das Docking von INA-EE und EHA-EE in CALB Strukturen führte zur Identifikation von fünf *hotspots*: D134, T138, Q157, I189 und V190. Die Van-der-Waals Radien von Seitenkettenatomen dieser Aminosäuren überschneiden sich energetisch ungünstig mit Atomen der beiden Substrate. Eine solche Überschneidung wurde für diese Aminosäuren und ein gedocktes, lineare Substrate (Caprylsäuremethylester) nicht beobachtet. Im Rahmen der ersten Strategie wurden an den identifizierten Aminosäurepositionen kleinere und ungeladene Aminosäuren eingeführt. Die Enzymvarianten T138A, Q157L, Q157I, I189L, T138A/Q157L und Q157I/I189L resultierten in leicht verbesserten Dockingergebnissen im Vergleich zum Wildtyp, und wurden deshalb in *E. coli* exprimiert. Im Rahmen der zweiten Strategie wur-

de eine *in silico* Bibliothek mit 2 400 CALB Varianten erstellt. Die 2 400 Varianten erhält man durch die Kombination der fünf zuvor identifizierten *hotspots* mit den ursprünglichen und vorgeschlagenen, anderen Aminosäuren an diesen Positionen (zwischen zwei und acht, je nach Position). Die Bibliothek wurde mit INA-EE und EHA-EE und *Substrate-imprinted* Docking (Juhl et al., 2009) durchmustert, um Aminosäureaustausche zu identifizieren, die öfter als die ursprüngliche Aminosäure in Varianten vorkommen, die zu guten Dockingergebnissen führen. Da das Ziel eine möglichst vollständige Umsetzung der Substrate und keine spezifische Umsetzung eines Enantiomers, war, wurde zur Bewertung der Häufigkeit die Summe der Häufigkeit bei beiden Enantiomeren eines Substrats berücksichtigt. Für INA-EE waren die so identifizierten Aminosäureaustausche D134N, T138S, Q157T, Q157A, Q157V, Q157S und I189V. Für EHA-EE waren es die Aminosäureaustausche D134T, D134N, D134S, T138N, T138Q, Q157T, Q157N, Q157A, Q157V, Q157S, I189T, I189A und V190A. Die Aminosäureaustausche die bei beiden Substraten häufig oder gleich häufig wie die ursprüngliche Aminosäure vorkamen sind D134S, D134N, T138Q, T138S, Q157T, Q157N, Q157A, Q157V und Q157S. Diese CALB Varianten wurden in *E. coli* exprimiert und analysiert.

Nach Expression in *E. coli* wurde für alle exprimierten CALB Varianten die Aktivität der Biofeuchtmasse bestimmt. Die sechs Varianten aus der ersten Strategie zeigten gegenüber INA-EE (CALB Wildtyp: 1,7 U/g, T138A: 1,1 U/g, Q157L: 1,1 U/g, Q157I: 1,0 U/g, I189L: 1,7 U/g, T138A/Q157L: 1,5 U/g, Q157L/I189L: 0,9 U/g) und EHA-EE (CALB Wildtyp: 1,1 U/g, T138A: 0,9 U/g, Q157L: 0,9 U/g, Q157I: 1,0 U/g, I189L: 1,0 U/g, T138A/Q157L: 0,9 U/g, Q157L/I189L: 0,8 U/g) entweder keine Veränderung in der Aktivität, oder sogar eine Verschlechterung. Die fünf CALB Varianten, die im Rahmen

der zweiten Strategie exprimiert wurden, zeigten alle eine erhöhte Aktivität der Biofeuchtmasse für INA-EE (D134S: 5.8 U/g, T138S: 30.6 U/g, Q157N: 2.9 U/g, Q157S: 7.1 U/g, Q157T: 8.6 U/g), und drei ebenfalls eine erhöhte Aktivität für EHA-EE (D134S: 3.1 U/g, T138S: 2.1 U/g, Q157S: 1.5 U/g). Vier Varianten (D134N, T138Q, Q157A, Q157V) die ebenfalls im Rahmen der zweiten Strategie identifiziert worden waren konnten nicht exprimiert werden.
Die CALB Variante T138S zeigte in der Aktivitätsbestimmung der Biofeuchtmasse mit einer 18-fachen Aktivitätssteigerung gegenüber CALB den größten Effekt. Deshalb wurden diese Variante und CALB Wildtyp zur Kontrolle mit einem C-terminalen His-Tag exprimiert, aufgereinigt und die spezifische Aktivität bestimmt. CALB zeigte gegenüber dem Referenzsubstrat Methylcaprylat eine spezifische Aktivität von 741 U/mg, während die T138S Variante nur eine Aktivität von 427 U/mg aufwies. Gegenüber INA-EE zeigte CALB eine Aktivität von 3,1 U/mg, während die T138S Variante eine Aktivität von 15,3 U/mg aufwies. Gegenüber EHA-EE zeigt CALB eine Aktivität von 0,5 U/mg, und die T138S Variante zeigte eine Aktivität von 0,6 U/mg. Dies ist eine Erhöhung der spezifischen Aktivität für INA-EE um den Faktor 5, und eine Erhöhung der relativen Aktivität (normiert auf Methylcaprylat) um den Faktor 8,6 (Faktor 2 für EHA-EE). Diese Unterschiede in der spezifischen und relativen Aktivität zwischen CALB und der T138S Variante sind deutlich niedriger als die Unterschiede die bei der Biofeuchtmasse beobachtet wurden. Ein Vergleich der Expression beider Enzyme mittels SDS-PAGE zeigt in etwa vergleichbare Banden. Allerdings fand keine Unterscheidung zwischen löslichem und unlöslichem Protein statt. Eine Erklärung kann deshalb sein, dass bei der Expression der T138S Variante mehr Protein löslich und aktiv war als bei der Expression des Wildtyps.
In dieser Studie wurden zwei Arten von Vorhersagen getroffen. Ers-

tens wurden sogenannte *hotspots* identifiziert, und zweitens wurden spezifische Aminosäureaustausche bezüglich einer Aktivitätserhöhung für INA-EE und EHA-EE vorgeschlagen. Es wurden fünf *hotspots* in der Bindetasche von CALB identifiziert. Für vier dieser *hotspots* wurden Varianten mit verbesserter Aktivität gegenübee INA-EE und EHA-EE vorgeschlagen. Varianten mit Aminosäureaustauschen an drei (D134, T138, Q157) der vier *hotspots* zeigten eine deutliche Zunahme der Aktivität gegenüber INA-EE und EHA-EE. Daher ist die Identifizierung von *hotspots* mittels *Substrate-imprinted Docking* mit einer Erfolgsquote von 75% erfolgreich und zuverlässig. Dieses Ergebnis stimmt gut mit den Ergebnissen anderer Studien überein, die ähnliche Methoden nutzten, um aktivitätsrelevante Aminosäurepositionen für Carbonylreduktasen (Zhu et al., 2008), Fruktosylamineoxidasen (Miura et al., 2008) und Glutathiontransferasen (Kapoli et al., 2008) zu identifizieren. Die nicht verbesserte Aktivität der Enzymvarianten aus der ersten Strategie zeigt allerdings, dass *Substrate-imprinted Docking* noch nicht genau und zuverlässig genug ist, um durch Docking in einzelne, manuell erstellte Enzymvarianten diejenigen mit verbesserter Aktivität zu identifizieren. Im Gegensatz zur ersten Strategie war die zweite Strategie, die über eine kombinatorische *in silico* Bibliothek Aminosäureaustausche identifiziert hat, die häufig in Enzymvarianten vorkommen die gute Dockingergebnisse zeigen, sehr erfolgreich darin, Aminosäureaustausche vorherzusagen, die zu einer verbesserten Aktivität für INA-EE und EHA-EE führten. Fünf CALB Varianten wurde im Rahmen der zweiten Strategie erfolgreich exprimiert (D134S, T138S, Q157N, Q157S, Q157T)). In der Aktivität der Biofeuchtmasse zeigen alle fünf eine erhöhte Aktivität für INA-EE, und drei der fünf zeigen eine erhöhte Aktivität für EHA-EE.

4.3 Biokatalytische Synthese von Oligoestern aus Zuckerdiolen: Eine Studie der Enantioselektivität von Candida antarctica Lipase B

(siehe: Biocatalytic Synthesis of Polyesters from Sugar-Based building blocks using immobilized Candida antarctica lipase B, Seite 181)

Vor dem Hintergrund stark schwankender Ölpreise und begrenzter Ressourcen hat die ölunabhängige Produktion von Polymeren in den letzten Jahren an Bedeutung gewonnen. Mögliche Bausteine für solche neuen Polymere sind zum Beispiel Zucker wie 1,4:3,6-dianhydro-D-glucitol (Isosorbide), 1,4:3,6-dianhydro-D-mannitol (Isomannide) und 1,4:3,6-dianhydro-L-iditol (Isoidide), die aus natürlichen und erneuerbaren Rohstoffen gewonnen werden können. Durch Veresterung mit Succinat oder anderen Disäuren können aus diesen Zuckern lineare Oligomere (LEOs, *linear ester oligomers*) oder zyklische Oligomere (CEOs, *cyclic ester oligomers*) hergestellt werden. *Candida antarctica* lipase B (CALB) besitzt ein sehr breites Substratspektrum und zeigte in vorläufigen Untersuchungen Aktivität gegenüber diesen Substraten.
Substrate-imprinted Docking (Juhl et al., 2009) wurde verwendet, um zu untersuchen, wie die drei Substrate Isosorbid, Isomannid und Isoidid in die Bindetasche von CALB passen, welche Substrate bevorzugt umgesetzt werden und wo es in der Bindetasche Engpässe gibt. Die beiden Ringe der Zuckermoleküle sind "V"-förmig in einem Winkel von 120° zueinander angeordnet. Die beiden Hydroxylgruppen der Substrate zeigen entweder in diese "V"-Form hinein (*endo-*

Hydroxylgruppe) oder zeigen aus ihr heraus (*exo*-Hydroxylgruppe). Isosorbid besitzt eine *endo*- und eine *exo*-Hydroxylgruppe, Isomannid besitzt zwei *endo*-Hydroxylgruppen und Isoidid besitzt zwei *exo*-Hydroxylgruppen. Aufgrund der unterschiedlichen Orientierung der beiden Hydroxylgruppen bei Isosorbid war es notwendig, zwei Buttersäureester von Isosorbid (je einen an jeder Hydroxylgruppe) zu docken, während für Isomannid und Isoidid nur jeweils ein Buttersäureester gedockt wurde, da beide Hydroxylgruppen in den Substraten jeweils die gleiche Orientierung aufwiesen. Der Isomannidester und der Isosorbidester mit Esterbindung in *endo*-Konfiguration konnten in einer produktiven Orientierung in CALB gedockt werden, während der Isoididester und der Isosorbidester mit Esterbindung in *exo*-Konfiguration nicht produktiv in CALB gedockt werden konnten. Eine Analyse der gedockten Substrate zeigt, dass Veresterung an einer *endo*-Hydroxylgruppe dazu führt, dass das "V"-förmige Zuckermolekül gut in die Bindetasche eingepasst werden kann, während eine Veresterung an einer *exo*-Hydroxylgruppe dazu führt, dass das "V"förmige Zuckermolekül nur schlecht in die Bindetasche passt, und nicht in der Lage ist, dabei die notwendigen Wasserstoffbrücken für eine erfolgreiche Veresterung auszubilden. Wenn diese Wasserstoffbrücken nicht gebildet werden können findet entweder keine Katalyse oder nur eine sehr schwache Katalyse statt.
Die Ergebnisse der Modellierung durch Docking korrelieren mit experimentellen Befunden. Die drei Substrate Isosorbid, Isomannid und Isoidid wurden mit Succinat unter Verwendung des Biokatalysators CALB verestert. Es zeigte sich eine Präferenz von CALB für Isomannid, welches gut umgesetzt wurde. Isosorbid wurde deutlich schlechter umgesetzt, und Isoidid wurde fast nicht umgesetzt. Diese Präferenz von CALB ist entgegengesetzt zu der chemischen Reaktivität der drei Zucker (Isoidid > Isosorbid > Isomannid) (Fenouillot et al.,

2010) und der Ausbeute bei Kondensation mit 1,4-Benzenedicarbonylchlorid (Isoidid > Isosorbid > Isomannid) (Thiem and Lüders, 1984-04-01), und stellt so eine Möglichkeit zur Synthese neuer Polymere aus diesen Ausgangsstoffen dar.

4.4 Strukturelle Einordnung mittels der Lipase Engineering Database: Eine Fallstudie anhand der Lipase A aus Candida antarctica

(siehe: Structural classification by the Lipase Engineering Database: a case study of Candida antarctica lipase A, Seite 215)

Die *Lipase Engineering Database* (LED) (http://www.led.uni-stuttgart.de/) ist eine relationale Datenbank für Sequenz- und Strukturinformationen von α/β-Hydrolasen (Pleiss et al., 2000). Um die Anwendung bioinformatischer Techniken zur Identifizierung neuer Biokatalysatoren zu unterstützen, wurde die LED auf die Version 3.0 aktualisiert und um sechs homologe Familien erweitert, die Lipasen, wie Lipase A aus *Candida antarctica* (CALA), enthalten, die bisher nicht in der LED enthaltenen waren.
Im Vergleich zu früheren Versionen wurde die Anzahl der in der LED enthaltenen Proteine mehr als vervierfacht. Durch eine BLAST-Suche in der GenBank mit den bereits in der Datenbank enthaltenen 4 322 Proteinsequenzen mit einem *E-Value* Schwellenwert von 10^{-50} wurde die Anzahl der Proteine auf 18 587 erweitert. Im Rahmen dieser Aktualisierung erhöhte sich ebenfalls die Zahl der in der LED enthaltenen Strukturen von 167 auf 656.
Die neuen Proteineinträge in der Datenbank wurden größtenteils in bereits bestehende homologe Familien eingeordnet. Einige bekannte α/β-Hydrolasen hatten allerdings nur geringe Sequenzähnlichkeiten zu bereits in der Datenbank enthaltenen Proteinen, so dass es notwendig war, sechs neue homologe Familien und eine neue Superfa-

milie einzuführen. Die neue Superfamilie *Candida antarctica lipase A like* (LED Nummer: abH38) umfasst vier der sechs neuen homologen Familien. Während diese neue Superfamilie klein ist, und nur sehr nah verwandte Enzyme enthält (32 Proteineinträge), sind die meisten der in ihr enthaltenen Enzyme in anderen Datenbanken in Proteinfamilien eingeordnet die deutlich diverser sind, und ein breites Spektrum an Lipasen enthalten. So findet sich CALA in der Pfam Datenbank (Finn et al., 2010) in der LIP Familie für sekretierte Lipasen (PF03583) (465 Proteine), in der InterPro Datenbank (Hunter et al., 2009) in der Familie IPR005152 (585 Proteine), und in der ESTHER Datenbank (Hotelier et al., 2004) in der FungalBact_LIP Familie für sekretierte Lipasen (85 Proteine). Allerdings beinhaltet die Familie abH38 nur Proteine mit einer sehr hohen Sequenzidentität, insbesondere an den katalytisch wichtigen Positionen, so dass sie deutlich kleiner ist als die eher allgemeinen Lipasefamilien in den drei genannten anderen Datenbanken.

50% alle Proteineinträge in der neuen Version der LED entfallen auf nur vier der 38 Superfamilien. Die größte dieser dominierenden Superfamilien ist die *Cytosolic Hydrolases* Familie (abH08 mit 15 homologen Familien und 3 188 Sequenzen, die hauptsächlich Epoxidhydrolasen und Haloalkandehalogenasen beinhaltet. Die zweitgrößte ist die *Carboxylesterases* Familie (abH01) mit 12 homologen Familien und 2 998 Proteinen, die eine großen Auswahl an Carboxylesterasen wie zum Beispiel Acetylcholinesterasen und *bile salt activated* Lipasen beinhaltet. Die drittgrößte ist die *Moraxella lipase 2 like* Familie (abH04) mit vier homologen Familien und 1 781 Proteineinträgen, die hauptsächlich Lipasen und andere Carboxylesterasen beinhaltet. Die viertgrößte ist die *Microsomal Hydrolases* Familie (abH09) mit drei homologen Familien und 1 336 Proteineinträgen, die hauptsächlich mikrosomale Peptidasen und Epoxidhydrolasen beinhaltet. Die *Cy-*

tosolic Hydrolases Familie und die *Microsomal Hydrolases* Familie gehören zzur GX-Klasse, während die *Carboxylesterases* Familie und die *Moraxella lipase 2 like* Familie zur GGGX-Klasse der α/β-Hydrolasen gehören.

Die systematische Klassifizierung von α/β-Hydrolasen in Kombination mit einer durchgängigen Annotation und weiteren Zusatzinformationen hat die LED zu einem nützlichen Werkzeug zur Analyse von Proteinfamilien und zur Entwicklung neuer Biokatalysatoren gemacht. So führte eine BLAST-Suche in der LED zur Identifikation neuer Enzyme mit α/β-Hydrolasefaltung (Kim et al., 2009, Lämmle et al., 2007), und ein Modell zur Vorhersage der Löslichkeit von Proteinen wurde mittels einer Analyse der Proteinfamilien in der LED entwickelt und verbessert (Koschorreck et al., 2005).

Publikationen und Publikationsmanuskripte

Die Ergebnisse der vorliegenden Arbeit wurden in folgenden Publikationen und Publikationsmanuskripten veröffentlicht:

1. P Benjamin Juhl, Peter Trodler, Sadhna Tyagi, Jürgen Pleiss, Modelling substrate specificity and enantioselectivity for lipases and esterases by substrate-imprinted docking. 2009, *BMC Structural Biology* 9:39.

2. P Benjamin Juhl, Kai Doderer, Frank Hollmann, Oliver Thum, Jürgen Pleiss, Engineering of *Candida antarctica* lipase B for hydrolysis of bulky carboxylic acid esters. 2010, *Journal of Biotechnology* 150:4

3. David I. Habeych, P Benjamin Juhl, Jürgen Pleiss, Gerrit Eggink, Diana Vanegas, Carmen G. Boeriu, Biocatalytic Synthesis of Polyesters from Sugar-Based building blocks using immobilized Candida antarctica lipase B. 2011, *Journal of Molecular*

Catalysis B Accepted Manuscript, Article in Press

4. Widmann, M., Juhl, P., B., Pleiss, J., Structural classification by the Lipase Engineering Database: a case study of Candida antarctica lipase A. 2010, *BMC Genomics* 11:123

Kapitel 5

Modelling substrate specificity and enantioselectivity for lipases and esterases by substrate-imprinted docking

P Benjamin Juhl[1], **Peter Trodler**[1], **Sadhna Tyagi**[1], **Jürgen Pleiss**[1*]

[1] Institute of Technical Biochemistry, University of Stuttgart, Allmandring 31, 70569 Stuttgart, Germany

5.1 Abstract

Background: Previously, ways to adapt docking programs that were developed for modelling inhibitor-receptor interaction have been explored. Two main issues were discussed. First, when trying to model catalysis a reaction intermediate of the substrate is expected to provide more valid information than the ground state of the substrate. Second, the incorporation of protein flexibility is essential for reliable predictions.

Results: Here we present a predictive and robust method to model substrate specificity and enantioselectivity of lipases and esterases that uses reaction intermediates and incorporates protein flexibility. Substrate-imprinted docking starts with covalent docking of reaction intermediates, followed by geometry optimisation of the resulting enzyme-substrate complex. After a second round of docking the same substrate into the geometry-optimised structures, productive poses are identified by geometric filter criteria and ranked by their docking scores. Substrate-imprinted docking was applied in order to model (i) enantioselectivity of *Candida antarctica* lipase B and a W104A mutant, (ii) enantioselectivity and substrate specificity of *Candida rugosa* lipase and *Burkholderia cepacia* lipase, and (iii) substrate specificity of an acetyl- and a butyrylcholine esterase toward the substrates acetyl- and butyrylcholine.

Conclusions: The experimentally observed differences in selectivity and specificity of the enzymes were reproduced with an accuracy of 81%. The method was robust toward small differences in initial structures (different crystallisation conditions or a co-crystallised ligand), although large displacements of catalytic residues often resulted in

substrate poses that did not pass the geometric filter criteria.

5.2 Background

The number of protein structures available to researchers has grown exponentially over the last two decades and more than 50 000 experimentally determined structure entries are now held in the Protein Data Bank (Berman et al., 2000). Furthermore, comparative structure prediction allows to derive reliable structure models from sequence information (Pieper et al., 2006). In silico methods are being developed to predict affinity, activity, specificity, and selectivity of newly discovered proteins based on structure information (Ortiz et al., 2006). In drug development, molecular docking is routinely used to identify new lead compounds by virtual screening of libraries of small compounds (Cavasotto and Orry, 2007). Recently, docking methods have also been successfully applied to predict the most probable substrates of enzymes with unknown function, but known structure (Hermann et al., 2007, Song et al., 2007). Previously, the specificity of enzymes was investigated by non-covalent docking of putative metabolites into the substrate binding site (Macchiarulo et al., 2004) and substrates for short chain dehydrogenases/reductases were identified by molecular docking (Favia et al., 2008). A similar method was used to identify eight new substrates for *Pseudomonas diminuta* phosphotriesterase (Irwin et al., 2005). Use of an improved scoring function made it possible to predict relative binding free energies for α-β barrel proteins and their metabolites (Kalyanaraman et al., 2005). The docking results were further improved for protein structures which had been resolved without a ligand by a restricted energy minimisation of the binding pocket around the docked metabolite. While all these methods considered the ground

state of the substrate, reaction intermediates of putative substrates were successfully used to predict substrates of amidohydrolases (Hermann et al., 2006), and docking of transition-states of flunitrazepam and progesterone have been docked into cytochrome P450 monooxygenases to predict hydroxylation patterns (Rydberg et al., 2008). Especially these two later findings support our approach of focusing on reaction intermediates when docking substrates into enzymes.

Carboxylic ester hydrolases (EC 3.1.1) are a large family of industrially relevant biocatalysts because they have been shown to catalyse hydrolysis of ester substrates with high regio- and enantioselectivity as well as the reverse reaction, the acylation of alcohols (Bornscheuer and Kazlauskas, 2006, Panda and Gowrishankar, 2005, Schlacher et al., 1998). Their reaction mechanism is well understood (Brady et al., 1990, Winkler et al., 1990): Upon nucleophilic attack of the catalytic serine, a tetrahedral intermediate is formed which is considered the rate limiting step. The binding pockets of esterases provide a pre-organised environment to specifically stabilise this intermediate by hydrogen-bonding. Therefore, a predictive model for esterase substrates has to take into account the following points:

1. The substrate has to be covalently docked to the enzyme in its tetrahedral intermediate state. While docking of molecules in their ground state allows predictions of the binding of that molecule to an enzyme, it does not allow to draw direct conclusions whether the molecule is converted by the enzyme or not. A docking method that aims to model enzymatic catalysis should reflect the molecular role of the enzyme in stabilising the transition-state (Warshel and Florián, 1998). A tetrahedral intermediate that is covalently bound to the catalytic serine is very close to the transition state which is formed during the enzyme-catalysed ester hydrolysis (Smith et al., 2008). Since in both states the interactions of the enzyme with the

acid moiety as well as with the alcohol moiety are identical, the tetrahedral intermediate is considered to be appropriate to predict the relative catalytic activity towards different substrates.

2. In addition, the docking pose of a putative substrate is essential. In order to be converted, the hydrogen bond network stabilising the intermediate has to be fully formed. Therefore, a simple geometric filter allows to distinguish between productive and non-productive substrate poses (Schulz et al., 2000, Tyagi and Pleiss, 2006).

3. X-ray structures and structure models based on homology are often not in a conformation to accommodate putative substrates, because even small differences in structures can have a strong effect on molecular docking results (Sandak et al., 1998). To overcome this problem, it is necessary to introduce protein flexibility into the docking procedure, allowing the enzyme to adjust its conformation to the substrate. Current docking programs treat the ligand as a flexible molecule, but consider the protein to be rigid. Ways to account for protein flexibility are a point of focus in current molecular docking research and a variety of methods have been suggested (Cozzini et al., 2008). Methods that incorporate limited flexibility for the proteins allow the receptor to bend in hinge regions (Sandak et al., 1998), introduce a limited flexibility of amino acid side chains in the active site (Alberts et al., 2005, Leach, 1994), or change the allowed overlap between ligand and protein (Gabb et al., 1997). Other docking methods represent protein flexibility by different protein structures or a rotamer library of substrate-interacting residues. The ligand is docked either into an ensemble of protein structures (Barril and Morley, 2005, Kamper et al., 2006), into an averaged structure (Huang and Zou, 2007), or into a pharmacophore grid (Claussen et al., 2001). However, this limited flexibility is not able to account for all possible conformational changes that occur in proteins upon ligand bin-

ding (Murray et al., 1999). A fully flexible protein can be simulated by molecular mechanics/molecular dynamics and Monte Carlo methods. Molecular dynamics simulations of a defined binding site (Rydberg et al., 2008) or the whole ligand-protein complex (Król et al., 2007) have been applied to improve docking results from rigid protein docking. Similarly, all-atom Monte Carlo docking algorithms have been successfully used to model drug-DNA binding (Rohs et al., 2005).
Here we introduce a strategy of substrate-imprinted docking, which uses the docking program FlexX (Kramer et al., 1999, Rarey et al., 1996), geometric filter criteria, and structure optimisation by molecular mechanics to account for full protein flexibility. The capability of this strategy was assessed in a case study on several lipases and two esterases to model enantioselectivity and substrate specificity:

- The wild type of *Candida antarctica* lipase B (CALB) was compared to a mutant (W104A) with altered enantioselectivity (Magnusson et al., 2005) by docking the two enantiomers of 1-phenylethyl butyrate ((*R*)-PEB and (*S*)-PEB) to model enantioselectivity.

- The enantiomers of 2- to 8-methyldecanoic acid butyl esters ((*R/S*)-2- to (*R/S*)-8-MDB) were docked into *Candida rugosa* lipase (CRL) to assess the capabilities of modelling lower enantioselectivities.

- CRL and *Burkholderia cepacia* lipase (BCL) were compared by docking the enantiomers of 2-hydroxyoctanoic acid butyl ester

((*R/S*)-2-HOB) and 2- to 4-methylpentanoic acid pentyl esters ((*R/S*)-2-MPP, (*R/S*)-3-MPP, 4-MPP) in order to model enantioselectivity and substrate specificity.

- *Torpedo californica* acetylcholine esterase (TcAChE) was compared to the human butyrylcholine esterase (huBuChE) by docking of acetylcholine (ACh) and butyrylcholine (BuCh) to model substrate specificity.

5.3 Results

5.3.1 Docking esters of chiral secondary alcohols into C. antarctica lipase B structures

Conventional docking

Tetrahedral reaction intermediates were covalently docked into CALB and its W104A mutant. During docking, the protein structure was assumed to be rigid, while the docked substrate was treated flexible. The docking procedure consists of three steps: (i) the construction of the putative substrates in their tetrahedral intermediate forms, (ii) the covalent docking into the active site, and (iii) the application of the geometric filter criteria for docked substrate poses. (*R*)-PEB and (*S*)-PEB were docked into five X-ray structures of CALB and the five models of its W104A mutant. Experimentally, CALB shows a enantiopreference in transesterification toward the (*R*)-enantiomer of PEB with a very high E-value of 1 300 000 (Magnusson et al., 2005), while the W104A mutant is non-selective. While all the structures were highly similar (all-atom RMSDs between the structures were

less than 0.5 Å), the docking scores differed considerably (Table 5.1 ans see Supplementary material Table 5.17). For four structures productive poses for a reaction intermediate of (*R*)-PEB were found. For the structure 1TCB no productive pose could be found by docking, which corresponds to a false negative result. For four structures no productive pose was found for the reaction intermediate of (*S*)-PEB, while a productive pose was found for 1LBT (false positive). Thus, the accuracy for the wild type without optimising the geometry is 80% - eight correct predictions, one false negative and one false positive.

The same docking procedure was performed with the five models of the W104A mutant. In four models (*R*)-PEB could be docked in a productive pose (Table 5.1), while no productive pose could be found for 1LBTW104A (false negative). For the (*S*)-enantiomer of PEB no productive pose could be found for any of the five mutant structures. This corresponds to five false negative results, because experimentally the (*S*)-enantiomer of PEB is converted as efficiently as the (*R*)-enantiomer. Thus, the accuracy for the mutant without optimising the geometry is 40% - four correct predictions and six false negatives.

In previous studies (McGovern and Shoichet, 2003), protein structures that were resolved with a particular ligand tended to give good docking results for similar ligands or ligands that have a similar mode of binding, while protein structures without inhibitor or in complex with a structurally different inhibitor failed more often. For docking of PEB into CALB and its mutant, structures with and without inhibitor have similar predictive accuracies. As expected, structures without a bound inhibitor have a tendency to lead to false negatives, such as for docking of (*R*)-PEB into 1TCB, while structures with inhibitor have a tendency to lead to false positives, such as docking of

(*S*)-PEB into 1LBT. This is caused by small differences in the structures, which lead to large differences in docking scores, as previously observed for trypsin, thrombin, and HIV-1-protease (Erickson et al., 2004). To overcome these limitations of protein rigidity (Davis and Teague, 1999) and to increase the accuracy, the docking procedure has to take into account protein flexibility.

Substrate-imprinted docking

To account for protein flexibility, protein-substrate complexes obtained by docking were subsequently optimised by energy minimisation. The resulting geometry-optimised structures of the protein are referred to as substrate-imprinted structures and were then used in a second round of covalent docking of the same substrate. The resulting poses were then analysed for the geometric filter criteria, the docking score, and the overlap volume (Table 5.1). Docking of (*R*)-PEB into CALB wild type resulted in productive poses for all five CALB structures. In contrast, docking of (*S*)-PEB led only for one structure (1LBS) to a productive pose (false positive). Thus, the accuracy of substrate-imprinted docking increased to 90% (nine correct predictions and one false positive) as compared to 80% for conventional docking, and the deviation between the docking scores was slightly reduced from 2.0 kJ/mol to 1.7 kJ/mol [see Supplementary material Table 5.17]. In contrast to docking into the X-ray structures, no false negative result was found. While docking of (*R*)-PEB into the X-ray structure 1TCB led to a false negative result, substrate-imprinted docking based on 1TCB led to a productive pose. Similarly, the productive pose upon docking of (*S*)-PEB into the X-ray structure 1LBT (false positive result) was not found upon substrate-

imprinted docking, but a new false positive result was found (1LBS). The largest impact of substrate-imprinted docking was observed for the mutant W104A. Here, docking into rigid model structures failed in six out of ten cases. However, docking of (*R*)-PEB into substrate-imprinted mutant structures resulted in productive poses for all five structures. Similarly, substrate-imprinted docking of (*S*)-PEB also led to productive poses for all structures. This result for the mutant is in agreement with experimental observations and corresponds to an accuracy of 100% - ten correct predictions.

The structural changes upon geometry optimisation are generally small. This also applies to the optimisation of the structure 1TCB (Fig. 5.1), which led to a false negative result upon docking of (*R*)-PEB into the X-ray structure, while substrate-imprinted docking found a productive pose. However, these small conformational changes in the alcohol binding pocket (all-atom RMSD of 0.46 Å, Table 5.2) are sufficient to remove clashes between the docked substrate and the enzyme, especially in complexes where the substrate moieties fit tightly into buried protein pockets, and thus allow to dock (*R*)-PEB in a productive pose. These changes in the alcohol binding pocket are in the same range as the overall conformational changes upon geometry optimisation (between 0.36 Å and 0.45 Å) for the CALB structures (Table 5.2). Previously it has been shown that a side chain optimisation was sufficient to successfully dock inhibitors into kinase structures (Rockey and Elcock, 2006). This method needs a X-ray structure of the inhibitor under investigation with a homologous protein as a starting point and assumes a rigid backbone. In contrast, substrate-imprinted docking can be applied to docking of new substrates and is able to improve binding pockets which are partially formed by backbone atoms, such as the oxyanion hole of lipases and esterases. For a typical substrate-enzyme complex, such

a full geometry optimisation takes less than 15 minutes on a dual core 2.0 GHz Opteron CPU.

5.3.2 Docking esters of chiral and non-chiral carboxylic acids into CRL and BCL structures

Conventional docking

Tetrahedral reaction intermediates of 2- to 8-MDB were docked into seven CRL X-ray structures (two of those structures had a displaced histidine) in order to model enantiopreference. Similar intermediates of 2-HOB and 2- to 4-MPP were docked into the same CRL structures and seven BCL X-ray structures in order to model substrate specificity. It has been shown experimentally that 2- to 8-MDBs can be synthesised by CRL with E-values between 2.8 an 91, alternately preferring the (*R*)- or the (*S*)-enantiomer (Hedenström et al., 2002). 2-HOB is synthesised by CRL and BCL, with a preference for the (*R*)-enantiomer (E-values of about 20) (Sakaki et al., 2002), 4-MPP is synthesised by CRL and BCL, 3-MPP is synthesised by neither CRL nor BCL, and 2-MPP is synthesised by CRL, but not BCL (Kodera et al., 1986).

Docking both enantiomers of 2- to 8-MDBs into CRL did most often result in predictions, that were either positive or negative for both enantiomers (Table 5.3). Thus, no stereoselectivity could be seen in the docking results. In particular, docking into the two structures 1LPN and 1LPP never resulted in a productive pose, due to the displacement of the catalytic histidine in these structures. For (*R*)-2-MDB productive poses could only be found for two structures, while for the (*S*)-enantiomer, a productive pose could only be found for one structure. Thus, productive poses for MDBs were only found in 42%

of the cases (41 correct predictions, 57 false negatives) and no enantiopreference could be observed in the docking results. The E-values CRL and 2- to 8-MDB are much lower than those observed in the case of CALB and PEB (E-value = 1 300 000), and the synthesis of the less prefered enantiomer did still occur. Therefore, both enantiomers were considered to be experimentally validated substrates for CRL and BCL.

Docking 2-HOB into CRL and BCL resulted in productive poses in most cases, but no distinction between the two enantiomers could be made. The experimentally observed E-value (Sakaki et al., 2002) was in the range of the E-values observed for CRL and 2- to 8- MDB, and both enantiomers were therefore considered to be experimentally converted substrates, too. For four CRL structures productive poses for the (R)-enantiomer and the (S)-enantiomer could be found (Table 5.4). No productive poses for any enantiomer could be found when docking into the other three CRL structures (1LPN, 1LPP, 1LPS). Productive poses for both enantiomers were also found for five BCL structures, while for two structures no productive poses could be found. 2-HOB was correctly identified as a substrate with an accuracy of 64% - 18 correct predictions, and 10 false negatives, but no enantiopreference could be observed in the docking results.

Docking 2- to 4-MPP into CRL X-ray structures resulted in only 17 correct predictions [see Additional file 2], where neither the substrates 2-MPP and 4-MPP nor the non-substrate 3-MPP were correctly predicted. When docking into the seven BCL X-ray structures, the substrate 4-MPP resulted in productive poses, and the non-substrates 2-MPP and 3-MPP also resulted in productive poses in many cases, leading to 21 false predictions. For eight structures productive poses were always found, no matter whether a substrate or a non-substrate was docked. For four structures (1LPN, 1LPP,

1LPS, 2LIP) no productive pose was found, regardless of the docked ligand. Docking into crystal structures of CRL and BCL is therefore not able to differentiate between substrates and non-substrates in the case of MPPs. Thus the experimentally described substrates 4-MPP for CRL and BCL and 2-MPP for CRL were correctly modelled with an accuracy of 67%, while the non-substrates 3-MPP for CRL and BCL and 2-MPP for BCL were correctly modelled with an accuracy of 33%. The overall accuracy for docking MPP was 44% - 31 correct predictions, 11 false negatives, and 28 false positives,

Substrate-imprinted docking

The capabilities of molecular docking to identify substrates and non-substrates were improved by using the method of substrate-imprinted docking. Docking 2- to 8-MDBs into substrate-imprinted CRL structures led to 58 productive poses (Table 5.3). The two structures with the displaced histidine (1LPN, 1LPP) did not provide any productive poses, as was already observed for the conventional docking. Thus, the identification of these esters as substrates was improved by substrate-imprinted docking to an accuracy of 59%, compared to the accuracy of 42% that was achieved with conventional docking. In contrast, substrate-imprinted docking was not able to identify enantioselectivities in the case of CRL and MDBs.

When 2-HOB was docked into substrate-imprinted CRL structures, four productive poses could be found for the (*R*)-enantiomer and five for the (*S*)-enantiomer (Table 5.4). When using substrate-imprinted BCL structures, six productive poses were found for (*R*)-2-HOB and six productive poses were found for the (*S*)-enantiomer. Thus, substrate-imprinted docking improved the identification of 2-HOB as

a substrate for CRL and BCL from 64% to 75%, but did not result predictions that reflected the experimentally determined enantioselectivity (E≈20).

Docking 2-MPP into substrate-imprinted CRL structures resulted in two productive poses for the (*S*)-enantiomer and none for the (*R*)-enantiomer [see Supplementary material 5.19]. When docking into substrate-imprinted BCL structures, four productive poses were found for the (*R*)-enantiomer, and none for the (*S*)-enantiomer. No productive poses could be found for docking 3-MPP into substrate-imprinted CRL structures, three productive poses could be found for each enantiomer when docking 3-MPP into substrate-imprinted BCL structures. When docking 4-MPP into substrate-imprinted CRL structures, five productive poses were found. For the structures 1LPN and 1LPP, no productive poses were found. When docking 4-MPP into substrate-imprinted BCL structures, productive poses were found for all seven structures. Substrate-imprinted docking was therefore able to identify the substrates 4-MPP for CRL and BCL, and 2-MPP for CRL with an accuracy of 50%. However, while the recognition of 4-MPP as a substrate was improved by substrate-imprinted docking, the recognition of 2-MPP as a substrate was better by conventional docking. The non-substrates 3-MPP for CRL and BCL and 2-MPP for BCL were correctly modelled with an accurracy of 76%. Thus, substrate-imprinted docking can, in the case of MPPs and CRL/BCL, differentiate between substrates and non-substrates with an accurracy of 66%, while conventional docking only achieved an accurracy of 44%.

When docking into the CRL structure 1LPP and its substrate-imprinted forms, no productive pose could be found for any of the docked molecules, and when using the structure 1LPN, the only productive pose found was for 2-HOB. A closer examination of these two

X-ray structures reveals that in both of them a inhibitor is bound to the catalytic serine, and a second inhibitor molecule is bound to the catalytic histidine (Grochulski et al., 1994). Because of this, the catalytic histidine (H449) in both structures is displaced by 3.1 Å when compared to the X-ray structure 1CRL. Such a large displacement was not corrected during the geometry optimisation.

5.3.3 Docking acetylcholine and butyrylcholine into AChE and BuChE structures

Conventional docking

In order to evaluate the capabilities of this method to correctly model substrate specificity with X-ray structures, tetrahedral reaction intermediates of ACh and BuCh were covalently docked into six TcAChE X-ray structures and four huBuChE X-ray structures. TcAChE only converts esters with a small acetyl moiety, because the acyl pocket of the protein is small. Therefore, TcAChE activity toward butyrylthiocholine is 850-fold lower than toward acetylthiocholine (Selwood et al., 1993). In contrast, huBuChE has a similar activity towards ACh and BuCh, because of its larger acyl pocket (Moralev and Rozengart, 2007, Nicolet et al., 2003).
Conventional docking into TcAChE and huBuChE did not differentiate between the two substrates. No docking solution could be found with two TcAChE structures and two huBuChE structures, while all other structures provided productive poses for both substrates (Table 5.5). The accuracy of conventional docking was 50% - 10 correct predictions, six false negatives, and four false positives.
While the docking results differ considerably, the differences between the structures of each enzyme are small. The RMSD of the backbo-

ne atoms between the six TcAChE or between four huBuChE X-ray structures is below 0.5 Å and 0.4 Å respectively. Co-crystallised inhibitors had no influence on the ability to find productive substrate poses. While the two TcAChE structures that did not lead to a productive pose (1E3Q, 1VXR) had been resolved with inhibitors, the TcAChE structure that had been resolved without inhibitor led to productive poses. Similarly, the huBuChE structure that had been resolved in complex with a choline ligand (1P0M) did not lead to productive poses, as well as one of the structures that was resolved with an inhibitor.

Substrate-imprinted docking

To improve predictability of substrate specificity, substrate-imprinted docking was applied. Docking ACh into substrate-imprinted TcAChE structures led to five productive poses (Table 5.5). It was not possible to dock ACh into the substrate-imprinted structure 1VXR (false negative). When docking BuCh into substrate-imprinted TcAChE structures, five of the six structures did not bind BuCh in a productive pose, while the substrate-imprinted structure 1DX6 led to a productive pose for BuCh (false positive). Substrate-imprinted huBuChE structures led to productive poses for ACh and BuCh in three out of four cases. The substrate-imprinted structure 1P0M did not lead to a productive pose for any of the substrates.
Thus, substrate-imprinted docking into TcAChE and huBuChE achieved an overall accuracy of 80% (16 correct predictions, three false negatives, and one false positive), while docking into structures that had not been optimised to fit the docked substrates only achieved an accuracy of 50%. In addition to the higher accuracy, substrate-

imprinted docking resulted in lower docking scores and a smaller spread of docking scores of true positive results [see Supplementary material 5.18].

5.4 Discussion

5.4.1 Accuracy of the method

It has been shown that substrate specificity and enantioselectivity of lipases and esterases are a consequence of a delicate balance between enthalpic and entropic contributions (Ottosson et al., 2002a). While shape fitting and enthalpic terms are well represented by substrate-imprinted docking, entropic contributions are only partially accounted for in the scoring function of FlexX. Previously, improved scoring functions have been proposed (Kalyanaraman et al., 2005). In addition, it has been observed for lipases that different organic solvents can mediate the experimentally determined enantioselectivity (Ottosson et al., 2002b, Overbeeke et al., 2000). However, none of the docking methods used today accounts for the molecular effects of organic solvents.

Beside the energy minimisation used in substrate-imprinted docking in order to optimise the structure of the substrate-enzyme complex, there are other more computational intensive methods like molecular dynamics or simulated annealing available that could be employed for the optimisation. However, clashes between atoms can easily be relaxed by a simple energy minimisation. In fact, such a minimisation is performed in many molecular dynamic protocols prior to the simulation itself for the purpose of relaxing such clashes. Furthermore, observed structural changes upon ligand binding are dominated

by small motions (Heringa and Argos, 1999, Zavodszky and Kuhn, 2005), which can be modelled well by energy minimisation (Cozzini et al., 2008).

Despite these limitations, substrate-imprinted docking can achieve a high predictive accuracy. As for other docking methods, the choice of the protein structure used for docking is crucial. Lipase structures which are adequate for substrate-imprinted docking must have an accessible substrate binding site and a functional orientation of the side chains in the active site. In the AChE X-ray structure 1VXR and the two CRL X-ray structures 1LPN and 1LPP, the catalytic histidine is considerably displaced by the bound inhibitors. Therefore, for all substrates these structures led to non-productive poses due to a failure of the geometric filter criteria. In contrast to other docking methods, substrate-imprinted docking is robust for other differences in protein structures: X-ray structures of free proteins and inhibitor complexes showed the same predictive accuracy. If the three X-ray structures with a displaced histidine are removed from the dataset, the accuracy of the method is 81%. Thus, substrate-imprinted docking allows to model substrate specificity and in some cases enantioselectivity of lipases and esterases with a good accuracy and with moderate computational and manual effort. The stereoselectivity could be accurately modelled for CALB, where the E-value was very high, while it was not possible to accureately model the stereoselectivity for CRL and BCL, where E-values were lower.

Docking reaction intermediates covalently into enzymes without accounting for flexibility did yield poor results, as can be seen in the results of the conventional docking. Likewise, it has been demonstrated by others, that performing an energy minimisation of ligand-protein complexes without applying filter criteria increased the number of false positives (Rockey and Elcock, 2005). Thus, all three steps of

the substrate-imprinted docking procedure are essential to achieve high accuracy.

5.4.2 False positive predictions

The conformational changes upon geometry optimisation of the substrate-protein complex often result in a widening of the binding pocket and can lead to false positive docking results in the substrate-imprinted docking approach. It can be argued, that the structures are optimised in a way that would fit any putative substrate used for imprinting whether a substrate or not, resulting in an increase of false positive predictions. This risk of false positives could reduces the ability of substrate-imprinted docking to discriminate between substrates and non-substrates. Previously, it has indeed been shown that energy minimisation of kinase-inhibitor complexes followed by scoring with Autodock resulted in an increase of false positives (Rockey and Elcock, 2005), thus a decreased ability to discriminate between substrates and non-substrates. This shortcoming of flexible protein structures can be counteracted by using more stringent parameters during docking, as we do by using smaller *maximum overlap volumes* in the second round of docking as compared to the first round of docking (1st round: 2.5 Å^3 to 7.5 Å^3, 2nd round: 2.0 Å^3 to 3.5 Å^3), and by applying geometric filter criteria that will discard all non-productive poses, even if they have a good score.
For CALB and its W104A mutant, the accuracy of docking into the substrate-imprinted structures increased from 60% to 95% when docking into substrate-imprinted structures, and only one false positive result occurred ((*S*)-PEB with 1LBS). This false positive could be identified by analysing the RMSD in the alcohol binding pocket

(T40, G41, T42, W104, A281). In this complex, the side chains of W104 and T42 have been displaced by more than 1 Å, and the backbone of T40 and G41 is twisted by almost 90°, thereby displacing the backbone oxygen by 2.1 Å (Fig. 5.2). This led to a high RMSD in the alcohol pocket (0.68 Å), which considerably exceeded the overall changes in protein structure (0.45 Å). In contrast, for 17 complexes (CALB and its mutant with (*R*)-PEB and (*S*)-PEB) the RMSD of the alcohol binding pocket was in the range of 65% and 121% of their total all-atom RMSD (Table 5.2). The RMSD of the alcohol pocket exceeded the overall RMSD considerably (170% and 149%) for only one further wild type complex (1LBT/(*S*)-PEB) and one mutant complex (1LBSW104A/(*R*)-PEB), although they were true negatives or positives. Thus, a RMSD exceeding 130% of the overall RMSD can indicate an unreliable optimised structure, which often leads to false predictions. However, this additional analysis also rejects some correct predictions.

Additionally, the increased total accuracy for docking 2- to 4-MPPs into substrate-imprinted CRL and BCL structures was due to a much improved identification of the non-substrates (76%) as compared to docking into the X-ray structures (33%). Therefore we think that the applied docking parameters and filter criteria are suitable to prevent false positives.

5.4.3 False negative predictions

One major effect of substrate-imprinted docking is the reduction of false negatives. When docking into TcAChE and huBuChE, the number of false negatives is reduced from ten to four by substrate-imprinted docking. In X-ray structures and homology models, the

orientation of side chains is not optimised, thus resulting in clashes with docked molecules (Jacobson and Sali, 2004). Therefore, docking into non-optimised structures resulted in ten false negatives. During geometry optimisation with the covalently bound substrate, the binding pocket adjusted to the substrate. As a result, seven of the ten false negatives did not occur when docking into substrate-imprinted structures. However, one additional false negative occurred when using the substrate-imprinted structures, that did not occur when using the non-optimised structures.

False negative results happen for two reasons. Either no pose for the substrate is found or none of the poses pass the geometric filter criteria. Two false negative results (ACh and BuCh docked into 1P0M) that occurred with both, the substrate-imprinted and the non-optimised structures, are examples for the first case and occurred due to clashes between substrate and protein in the binding pocket. The false negative that occurred with the substrate-imprinted and the conventional docking (ACh docked into 1VXR) is an example for poses that did not pass the geometric filter criteria. In these structures, the binding pocket has adopted a conformation that allows substrate binding, but not in a productive orientation, due to the orientation of the catalytic histidine. In 1VXR, the catalytic histidine has been displaced by the co-crystallised inhibitor (Millard et al., 1999), which was also the case for the two CRL structures 1LPN and 1LPP. In this conformation catalytic histidine the N_ϵ can not interact with the catalytic serine. With the histidine being unable to form a hydrogen bond to the serine O_γ, the docking pose did not pass the geometric filter criteria and was considered to be non-productive.

The false negative predictions for the huBuChE can be identified by analysing the RMSD of the choline pocket. A comparison of the overall RMSD and the RMSD of the choline pocket after the geo-

metry optimisation revealed that the choline pocket formed by W82, G115, G116, E197, H438, and G439 showed a considerably higher or lower RMSD than the rest of the protein. The all-atom RMSD of the whole protein after geometry optimisation ranged from 0.48 Å to 0.52 Å for huBuChE X-ray structures (Table 5.6). The RMSD of the choline pocket was 0.29 Å and 0.33 Å for the structure 1P0M, imprinted with ACh and BuCh, which is only 59% and 66% of the total RMSD. The three other substrate-imprinted structures that led to correct docking results had a RMSD for their choline binding pocket between 107% and 113% of the RMSD of the whole structure.
Thus, all false negative predictions of the huBuChE could be identified by a similar method that also identified the false positive docking results for CALB. A RMSD of the relevant binding pocket of the substrate-imprinted structure, that deviates more than 30% from the all-atom RMSD of the whole structure can be used as an indicator for an aberration in the geometry optimisation, resulting in a less reliable docking result.

5.5 Conclusions

Substrate-imprinted enzyme docking combines covalent docking, geometry optimisation, and geometric filter criteria to identify productive substrate poses. For the enzymes examined here, substrate specificity and enantioselectivity of wild type enzymes and mutants were modelled with an accuracy of 81% if the three structures with distorted active site were excluded (68% if the three structures are included). The process consists of five steps:

1. As protein structure, X-ray structures of free enzymes or inhibi-

tor complexes are suitable, as well as reliable homology models. However, it is crucial that the side chains of the catalytic serine and histidine are in a functional orientation.

2. Substrates are covalently docked in a tetrahedral intermediate form at an elevated *maximum overlap volume*. Productive poses are selected by geometric filter criteria and the docking score.

3. The geometry of the selected complexes is optimised by unconstrained energy minimisation.

4. In order to assess the reliability of the optimised structures, the deviation of the structure of the substrate binding site in respect to the overall deviation of the protein during energy minimisation of the complex can be evaluated. Structures where the difference between these deviations is larger than 30%, often led to false positive or false negative predictions.

5. The relaxed protein structure is used for a second round of substrate docking using more stringent docking parameters. Productive poses are again selected by geometric filter criteria and the docking score.

The method seems to be most accurate for modelling substrate specificity and less accurate for modelling enantioselectivity. Substrate-imprinted docking was able to model the differences in substrate specificity of CRL and BCL, and TcAChE and huBuChE, and differences between the enantioselectivity of CALB wild type and its W104A mutant. For CRL and BCL, enantioselectivity could not be reliably modelled. Substrate-imprinted docking was reproducible and robust toward different X-ray structures of the same protein. Because it combines good accuracy with a moderate computational and

manual effort, it is most suited to screen enzyme and mutant libraries with selected substrates.

5.6 Methods

5.6.1 Preparation of protein structures and substrates

X-ray structures of CALB (1LBS, 1LBT, 1TCA, 1TCB, 1TCC), CRL (1CLE, 1CRL, 1LPM, 1LPN, 1LPO, 1LPP, 1LPS), BCL (2LIP, 3LIP, 4LIP, 5LIP, 1OIL, 1YS1, 1YS2), TcAChE (1CFJ, 1DX6, 1E3Q, 1EVE, 1VXR, 1QIM), and huBuChE (1P0M, 1XLU, 1XLV, 1XLW) were retrieved from the Protein Data Bank (Berman et al., 2000) (http://www.pdb.org/). Two CALB structures (1LBS, 1LBT), six CRL structures (1CLE, 1LPM, 1LPN, 1LPO, 1LPP, 1LPS) and four BCL structures (4LIP, 5LIP, 1YS1, 1YS2) had been resolved with a bound inhibitor. From the six selected TcAChE structures, three had been resolved in complex with a large inhibitor (1DX6, 1E3Q, 1EVE), two with a small inhibitor (1CFJ, 1VXR), and one without any inhibitor (1QIM). From the four huBuChE X-ray structures, one had been resolved with a non-covalently bound product molecule (1P0M) and three had been resolved in a covalent complex with a small substrate analogous inhibitor (1XLU, 1XLV, 1XLW). Experimentally, the structures 1VXR, 1LPP, and 1LPN contain inhibitors that caused a very large displacement of the catalytic histidine. These three structures can therefore be considered to be not suited for modelling of catalytic activity, despite having a bound inhibitor, but were included in this study to better assess whether substrate-imprinted docking can correct these structural artefacts or

not. Models for the W104A mutant of CALB (1LBSW104A, 1LBTW104A, 1TCAW104A, 1TCBW104A, 1TCCW104A) were created by changing W104 to alanine in the X-ray structures of the wild type using the Swiss-Pdb viewer (Guex and Peitsch, 1997) and selecting the rotamer with the lowest score. W104 is located in the binding pocket for the medium-sized moiety of a secondary alcohol. For the huBuChE structures, the missing residues D378 and D379 were supplemented by MODELLER (Sali and Blundell, 1993), while keeping all other atoms fixed. The two residues are located on a loop far away from the substrate binding site. These models are referred to as "X-ray structures". The RMSD between two structures was calculated after fitting with the McLachlan algorithm (Mclachlan, 1982), implemented in the program ProFit (A C R Martin, http://www.bioinf.org.uk/software/profit/).

Protonation states of titrateable residues at pH 7 [see Supplementary material 5.20] were calculated by TITRA (Petersen et al., 1997), using the Tanford-Kirkwood sphere model, and MEAD (Miteva et al., 2005), using finite-difference methods to solve the Poisson-Boltzmann equation. Both methods predicted the same protonation states for the large majority of all titrateable residues. In few cases (≈2%) where the two methods predicted a different protonation state for the same residue, we relied on the predictions made by TITRA. The catalytic histidine in the protein structure was protonated, because we model a substrate-protein complex where the substrate is covalently bound to the catalytic serine. The formation of the covalent bond between the serine and the substrate ester is the result of a nucleophilic attack of the serine O_γ at the ester carbon. while the proton of the serine is transfered to the histidine.

Substrate esters were constructed as tetrahedral reaction intermediates of the lipase-catalysed ester hydrolysis (Fig. 5.3), including two

atoms of the catalytic serine, which forms a covalent bond to the intermediate. This tetrahedral carbon atom has four substituents - the alkyl moiety, the alcohol moiety, a negatively charged oxygen (oxyanion) and the O_γ-C_β-fragment of the catalytic serine.

5.6.2 Conventional docking

The conventional docking procedure consists of covalent docking of a reaction intermediate into the X-ray structure of an enzyme with a subsequent scoring and classification of the poses into productive and non-productive ones (Fig. 5.4). FlexX covalent docking superimposes a fragment (base fragment) of the ligand on a part of the X-ray structure. The base fragment serves as root for the incremental build-up of the whole ligand in the binding pocket. The substrate O_γ and C_β atoms form the base fragment and are superimposed on the O_γ and C_β atoms of the catalytic serine. Up to 50 different conformations of the base fragment are allowed during this superimposition, and the torsion angle of the bond between O_γ and C_β is sampled in a 10° range, according to the default settings of FlexX. The *maximum overlap volume* parameter in FlexX sets a limit for the overlap between the protein and a ligand atom. The allowed average overlap from every ligand atom is 0.4 * *maximum overlap volume*. Poses that exceed any of these values are automatically discarded. During every single conventional docking, the *maximum overlap volume* was gradually increased in 0.5 $Å^3$ steps from 2.5 $Å^3$ to 7.5 $Å^3$. Docking with gradually increasing *maximum overlap volumes* is necessary, because the incremental construction algorithm of the ligand used by FlexX (Rarey et al., 1996) can result in some substrate poses that are found at a small *maximum overlap volume*, but not at a larger *maximum*

overlap volume, and vice versa. The superimposed atoms of the base fragment and hydrogen atoms are not taken into account in overlap tests, nor is the base fragment considered when clashes between ligand and protein are calculated. The generated substrate poses are classified into productive and non-productive poses by the geometric filter criteria and ranked by score. The geometric filter checks for:(a) the existence of hydrogen bonds between the backbone N-H-groups of the two oxyanion hole residues and the oxyanion of the substrate, (b) a hydrogen bond between a side chain N-H-group of the catalytic histidine and the O_γ of the substrate, and (c) a hydrogen bond between a side chain N-H-group of the catalytic histidine and the oxygen of the alcohol moiety of the substrate (Fig. 5.3). A substrate pose with those four hydrogen bonds formed is considered to be productive. Hydrogen bonds were identified by FlexX (Kramer et al., 1999, Rarey et al., 1996) and defined according to the pairwise interaction scheme of FlexX. For each group able to act as a hydrogen acceptor or donor, a special interaction surface is defined as part of a sphere centred on the interacting atom. If two interaction centres lie near to each others interaction surface, they form an interaction. The docking scores given by FlexX are calculated by an empirical scoring function that estimates the free energy of binding (Böhm, 1994, Kramer et al., 1999). The function contains contributions for the loss of entropy, for hydrogen bonds, for ionic and hydrophobic interactions between the protein and the ligand, and for unfavourably close contacts between ligand and protein atoms. A productive pose with a negative score was considered to model a substrate that is converted by the enzyme, while the absence of such a pose was considered to correspond to a non-substrate.

5.6.3 Substrate-imprinted docking

Substrate-imprinted docking consists of a first round of docking by FlexX, a geometry optimisation, a second round of docking, and a final classification and scoring of the resulting poses (Fig. 5.5).The procedure starts with a X-ray structure and a putative substrate. Stereoisomers of one compound are treated as separate substrates. The putative substrate is covalently docked into the X-ray structure of the enzyme. During this first docking, the *maximum overlap volume* is gradually increased in 0.5 Å^3 steps from 2.5 Å^3 to 7.5 Å^3, as described for the conventional docking. A substrate-protein complex is built from the X-ray structure and the pose with the best score by removing the O_γ and C_β atoms of the catalytic serine in the X-ray structure and defining a bond between the C_β atom of the substrate and the C_α atom of the catalytic serine, as described above. If no substrate pose was found during the first round of docking, the method stopped here and the result was considered to be negative. However this occurred only twice in the 236 substrate-imprinted docking runs (2LIP with (*R*)- and (*S*)-MPP). This complex is optimised by energy minimisation (200 steps steepest decent followed by 800 steps conjugate gradient). A new, substrate-imprinted protein structure is extracted from the optimised complex by removing all substrate atoms except for the O_γ and C_β atoms that form the side chain of the catalytic serine. A second round of docking follows, where the same substrate that was used in the first round of docking is covalently docked into the optimised structure. The *maximum overlap volume* parameter is set more stringent in this second docking than in the first docking, and is gradually increased in 0.1 Å^3 steps from 2 Å^3 to 3.5 Å^3. All generated substrate poses are scored and classified into productive and non-productive poses as described for

the conventional docking. A productive pose with a negative score was considered to model a substrate that is converted by the enzyme, while the absence of such a pose was considered to correspond to a fake substrate, that is not converted by the enzyme.

5.6.4 Geometry optimisation

In its docked pose, the substrate partially overlaps with the catalytic serine. A substrate-protein complex with the substrate covalently bound to the catalytic serine was created by removing the O_γ and C_β of the catalytic serine and defining a bond between the C_β of the substrate and the C_α of the catalytic serine. Atom types and parameters of the AMBER ff99 force field (Cornell et al., 1995) were used. Parameters and atom types for the new serine-substrate residue were derived by analogy. The partial charges for the serine-substrate residue were assigned with the RESP fit methodology (Bayly et al., 1993) after ab initio geometry optimisation in the gas phase at the Hartree-Fock level of theory with the 6-31G* basis set and calculation of the electrostatic potential in gridpoints according to the Merz-Singh-Kollman scheme (Besler et al., 1990, Singh and Kollman, 1984). Protonation states of titratable residues were used as calculated for the docking steps. Hydrogens were added by LEaP (Case et al., 2006). The system was solvated by placing it in a truncated octahedral water box using the TIP3P water model (Jorgensen et al., 1983) with a minimal distance of 1 Å between protein and water molecules and a minimal distance of 12 Å between protein and the wall of the box. Counter ions were added in LEaP to neutralise the system. LEaP places the counter ions in a shell around the protein using a Coulombic potential. The protein-ligand complexes were mi-

nimised using the AMBER program package (Case et al., 2006) and the all-atom AMBER force field ff99. The Sander tool of AMBER was used to perform a 200 step steepest descent minimisation, followed by 800 steps conjugate gradient minimisation in order to relax clashes in the system. Except for the O_γ and C_β atoms that form the serine side chain, all atoms that belong to the substrate were removed from the optimised complex. These structures were referred to as substrate-imprinted structures.

5.7 Abbreviations

ACh - acetylcholine, BCL - *Burkholderia cepacia* lipase, BuCh - butyrylcholine, CALB - *Candida antarctica* lipase B, CRL - *Candida rugosa* lipase, HOB - hydroxyoctanoic acid butyl ester, huBuChE - human butyrylcholine esterase, MDB - methyldecanoic acid butyl ester, MPP - methylpentanoic acid pentyl ester, PEB - 1-phenylethyl butyrate, RMSD - root mean square deviation, TcAChE - *Torpedo californica* acetylcholine esterase

5.8 Authors contributions

PT and ST carried out docking and geometry optimisation for TcAChE. PBJ conducted the homology modelling and carried out docking and geometry optimisation for all other enzymes. JP was involved in designing and overseeing the study. All authors have read and approved the final manuscript.

5.9 Acknowledgements

This work was supported by the “Fachagentur Nachwachsende Rohstoffe e. V.“ and the “Deutsche Bundesstiftung Umwelt“. Furthermore, the authors wish to thank Fabian Bös, Alexander Steudle, and Michael Widmann for technical support.

References

I. L. Alberts, N. P. Todorov, and P. M. Dean. Receptor flexibility in de novo ligand design and docking. *J. Med. Chem.*, 48:6585–6596, 2005.

X. Barril and S. D. Morley. Unveiling the full potential of flexible receptor docking using multiple crystallographic structures. *Journal of Medicinal Chemistry*, 48(13):4432–4443, 2005.

C.I. Bayly, P. Cieplak, W.D. Cornell, and P.A. Kollman. A well-behaved electrostatic potential based method using charge restraints for determining atom-centered charges: The resp model. *J.Phys.Chem*, 97:10269–10280, 1993.

H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. The protein data bank. *Nucleic Acids Res*, 28(1):235–42, 2000.

Brent H. Besler, Kenneth M. Merz Jr., and Peter A. Kollman. Atomic charges derived from semiempirical methods. *J Comput Chem*, 11 (4):431–439, 1990.

H. J. Böhm. The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of

known three-dimensional structure. *J Comput Aided Mol Des*, 8: 243–256, Jun 1994.

Uwe T. Bornscheuer and Romas J. Kazlauskas. *Hydrolases in Organic Synthesis*. WILEY-VCH, 2 edition, 2006.

L. Brady, A. M. Brzozowski, Z. S. Derewenda, E. Dodson, G. Dodson, S. Tolley, J. P. Turkenburg, L. Christiansen, B. Huge-Jensen, and L. Norskov. A serine protease triad forms the catalytic centre of a triacylglycerol lipase. *Nature*, 343:767–770, Feb 1990.

D.A. Case, T.A. Darden, T.E. Cheatham, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, K.M. Merz, D.A. Pearlman, M. Crowley, R.C. Walker, W. Zhang, B. Wang, S. Hayik, A. Roitberg, G. Seabra, K.F. Wong, F. Paesani, X. Wu, S. Brozell, V. Tsui, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Homak, G. Cui, P. Beroza, D.H. Mathews, C. Schafmeister, W.S. Ross, and P.A. Kollman. Amber 9. Technical report, University of California, San Francisco, 2006.

Claudio N Cavasotto and Andrew J W Orry. Ligand docking and structure-based virtual screening in drug discovery. *Curr Top Med Chem*, 7:1006–1014, 2007.

H. Claussen, C. Buning, M. Rarey, and T. Lengauer. Flexe: Efficient molecular docking considering protein structure variations. *Journal of Molecular Biology*, 308(2):377–395, 2001.

W. D. Cornell, P. Cieplak, C. I. Baylb, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollmann. A 2nd generation force-field for the simulation of proteins, nucleic-acids, and organic-molecule. *J Am Chem Soc*, 117:5179–5197, May 1995.

Pietro Cozzini, Glen E Kellogg, Francesca Spyrakis, Donald J Abraham, Gabriele Costantino, Andrew Emerson, Francesca Fanelli, Holger Gohlke, Leslie A Kuhn, Garrett M Morris, Modesto Orozco, Thelma A Pertinhez, Menico Rizzi, and Christoph A Sotriffer. Target flexibility: an emerging consideration in drug discovery and design. *J Med Chem*, 51:6237–6255, Oct 2008.

A. M. Davis and S. J. Teague. Hydrogen bonding, hydrophobic interactions, and failure of the rigid receptor hypothesis. *Angewandte Chemie-International Edition*, 38(6):737–749, 1999.

J. A. Erickson, M. Jalaie, D. H. Robertson, R. A. Lewis, and M. Vieth. Lessons in molecular recognition: The effects of ligand and protein flexibility on molecular docking accuracy. *J. Med. Chem.*, 47:45–55, 2004.

Angelo D Favia, Irene Nobeli, Fabian Glaser, and Janet M Thornton. Molecular docking for substrate identification: the short-chain dehydrogenases/reductases. *J Mol Biol*, 375:855–874, Jan 2008.

H. A. Gabb, R. M. Jackson, and M. J. E. Sternberg. Modelling protein docking using shape complementarity, electrostatics and biochemical information. *Journal of Molecular Biology*, 272(1): 106–120, 1997.

P. Grochulski, F. Bouthillier, R. J. Kazlauskas, A. N. Serreqi, J. D. Schrag, E. Ziomek, and M. Cygler. Analogs of reaction intermediates identify a unique substrate binding site in candida rugosa lipase. *Biochemistry*, 33:3494–3500, Mar 1994.

N. Guex and M. C. Peitsch. Swiss-model and the swiss-pdbviewer: an environment for comparative protein modeling. *Electrophoresis*, 18(15):2714–2723, Dec 1997.

Erik Hedenström, Ba-Vu Nguyen, and III Louis A. Silks. Do enzymes recognise remotely located stereocentres? highly enantioselective candida rugosa lipase-catalysed esterification of the 2- to 8-methyldecanoic acids. *Tetrahedron: Asymmetry*, 13:835–844, 2002.

J. Heringa and P. Argos. Strain in protein structures as viewed through nonrotameric side chains: Ii. effects upon ligand binding. *Proteins*, 37(1):44–55, Oct 1999.

J. C. Hermann, E. Ghanem, Y. Li, F. M. Raushel, J. J. Irwin, and B. K. Shoichet. Predicting substrates by docking high-energy intermediates to enzyme structures. *J Am Chem Soc*, 128(49):15882–91, 2006.

Johannes C Hermann, Ricardo Marti-Arbona, Alexander A Fedorov, Elena Fedorov, Steven C Almo, Brian K Shoichet, and Frank M Raushel. Structure-based activity prediction for an enzyme of unknown function. *Nature*, 448(7155):775–779, Aug 2007.

S. Y. Huang and X. Q. Zou. Ensemble docking of multiple protein structures: Considering protein structural variations in molecular docking. *Proteins-Structure Function and Bioinformatics*, 66(2): 399–421, 2007.

John J Irwin, Frank M Raushel, and Brian K Shoichet. Virtual screening against metalloenzymes for inhibitors and substrates. *Biochemistry*, 44(37):12316–12328, Sep 2005.

M. Jacobson and A. Sali. Comparative protein structure modeling and its applications to drug discovery. *Annual Reports in Medicinal Chemistry*, 39:259–276, 2004.

WL. Jorgensen, J. Chandrasekhar, JD. Madura, RW. Impey, and ML. Klein. Comparison of simple potential functions for simulating liquid water. *J Chem Phys*, 79:926–935, 1983.

C. Kalyanaraman, K. Bernacki, and M. P. Jacobson. Virtual screening against highly charged active sites: identifying substrates of alpha-beta barrel enzymes. *Biochemistry*, 44(6):2059–71, 2005.

A. Kamper, J. Apostolakis, M. Rarey, C. M. Marian, and T. Lengauer. Fully automated flexible docking of ligands into flexible synthetic receptors using forward and inverse docking strategies. *Journal of Chemical Information and Modeling*, 46(2):903–911, 2006.

Yoh Kodera, Katsunobu Takahashi, Hiroyuki Nishimura, Ayako Matsushima, Yuji Saito, and Yuji Inada. Ester sythesis from alpha-substituted carboxylic acid catalyzed by polyethylene glycol-modified lipase from candida cylindracea in benzene. *Biotechnol. Lett.*, 8:881–884, 1986. The Pseudomonas fluorescens lipase from Amano Pharmaceutical Ltd. used in this work was later reclassified as Burkholderia cepacia lipase.

B. Kramer, M. Rarey, and T. Lengauer. Evaluation of the flexx incremental construction algorithm for protein-ligand docking. *Proteins-Structure Function and Genetics*, 37(2):228–241, 1999.

Marcin Król, Alexander L Tournier, and Paul A Bates. Flexible relaxation of rigid-body docking solutions. *Proteins*, 68(1):159–169, Jul 2007.

A. R. Leach. Ligand docking to proteins with discrete side-chain flexibility. *Journal of Molecular Biology*, 235(1):345–356, 1994.

Antonio Macchiarulo, Irene Nobeli, and Janet M Thornton. Ligand selectivity and competition between enzymes in silico. *Nat Biotechnol*, 22(8):1039–1045, Aug 2004.

A. O. Magnusson, M. Takwa, A. Hamberg, and K. Hult. An s-selective lipase was created by rational redesign and the enantioselectivity increased with temperature. *Angew Chem Int Ed Engl*, 44(29):4582–5, 2005.

S. L. McGovern and B. K. Shoichet. Information decay in molecular docking screens against holo, apo, and modeled conformations of enzymes. *Journal of Medicinal Chemistry*, 46(14):2895–2907, 2003.

A. D. Mclachlan. Rapid comparison of protein structures. *Acta Crystallographica Section A*, 38(NOV):871–873, 1982.

C. B. Millard, G. Kryger, A. Ordentlich, H. M. Greenblatt, M. Harel, M. L. Raves, Y. Segall, D. Barak, A. Shafferman, I. Silman, and J. L. Sussman. Crystal structures of aged phosphonylated acetylcholinesterase: nerve agent reaction products at the atomic level. *Biochemistry*, 38:7032–7039, Jun 1999.

Maria A Miteva, Pierre Tufféry, and Bruno O Villoutreix. Pce: web tools to compute protein continuum electrostatics. *Nucleic Acids Res*, 33(Web Server issue):W372–W375, Jul 2005.

Serge Moralev and Eugene Rozengart. *Comparative Enzymology of Cholinesterases*. International University Line, La Jolla, 1st edition, 2007.

C. W. Murray, C. A. Baxter, and A. D. Frenkel. The sensitivity of the results of molecular docking to induced fit effects: Application to thrombin, thermolysin and neuraminidase. *Journal of Computer-Aided Molecular Design*, 13(6):547–562, 1999.

Y. Nicolet, O. Lockridge, P. Masson, J. C. Fontecilla-Camps, and F. Nachon. Crystal structure of human butyrylcholinesterase and of its complexes with substrate and products. *J Biol Chem*, 278 (42):41141–7, 2003.

Angel R Ortiz, Paulino Gomez-Puertas, Alejandra Leo-Macias, Pedro Lopez-Romero, Eduardo Lopez-Vi nas, Antonio Morreale, Marta Murcia, and Kun Wang. Computational approaches to model ligand selectivity in drug design. *Curr Top Med Chem*, 6:41–55, 2006.

Jenny Ottosson, Linda Fransson, and Karl Hult. Substrate entropy in enzyme enantioselectivity: an experimental and molecular modeling study of a lipase. *Protein Sci*, 11:1462–1471, Jun 2002a.

Jenny Ottosson, Linda Fransson, Jerry W King, and Karl Hult. Size as a parameter for solvent effects on candida antarctica lipase b enantioselectivity. *Biochim Biophys Acta*, 1594:325–334, Feb 2002b.

P. L. Overbeeke, J. A. Jongejan, and J. J. Heijnen. Solvent effect on lipase enantioselectivity. evidence for the presence of two thermodynamic states. *Biotechnol Bioeng*, 70:278–290, Nov 2000.

T. Panda and B. S. Gowrishankar. Production and applications of esterases. *Appl Microbiol Biotechnol*, 67:160–169, Apr 2005.

M. T. Petersen, P. Martel, E. I. Petersen, F. Drabløs, and S. B. Petersen. Surface and electrostatics of cutinases. *Methods Enzymol*, 284:130–154, 1997.

Ursula Pieper, Narayanan Eswar, Fred P Davis, Hannes Braberg, M. S. Madhusudhan, Andrea Rossi, Marc Marti-Renom, Rachel Karchin, Ben M Webb, David Eramian, Min-Yi Shen, Libusha

Kelly, Francisco Melo, and Andrej Sali. Modbase: a database of annotated comparative protein structure models and associated resources. *Nucleic Acids Res*, 34(Database issue):D291–D295, Jan 2006.

M. Rarey, B. Kramer, T. Lengauer, and G. Klebe. A fast flexible docking method using an incremental construction algorithm. *Journal of Molecular Biology*, 261(3):470–489, 1996.

W. M. Rockey and A. H. Elcock. Rapid computational identification of the targets of protein kinase inhibitors. *Journal of Medicinal Chemistry*, 48(12):4138–4152, 2005.

W. M. Rockey and A. H. Elcock. Structure selection for protein kinase docking and virtual screening: Homology models or crystal structures? *Current Protein & Peptide Science*, 7(5):437–457, 2006.

Remo Rohs, Itai Bloch, Heinz Sklenar, and Zippora Shakked. Molecular flexibility in ab initio drug docking to dna: binding-site and binding-mode transitions in all-atom monte carlo simulations. *Nucleic Acids Res*, 33:7048–7057, 2005.

Patrik Rydberg, Sine Myrup Hansen, Jacob Kongsted, Per-Ola Norrby, Lars Olsen, and Ulf Ryde. Transition-state docking of flunitrazepam and progesterone in cytochrome p450. *J Chem Theory Comput*, 4:673–681, 2008.

Keiji Sakaki, Shigeki Hara, and Naotsugu Itoh. Optical resolution of racemic 2-hydroxy octanoic acid using biphasic enzyme membrane reactor. *Desalination*, 149:247–252, 2002.

Andrej Sali and Tom L. Blundell. Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology*, 234(3):779–815, 1993.

B. Sandak, H. J. Wolfson, and R. Nussinov. Flexible docking allowing induced fit in proteins: Insights from an open to closed conformational isomers. *Proteins-Structure Function and Genetics*, 32(2): 159–174, 1998.

A. Schlacher, T. Stanzer, I. Osprian, M. Mischitz, E. Klingsbichel, K. Faber, and H. Schwab. Detection of a new enzyme for stereoselective hydrolysis of linalyl acetate using simple plate assays for the characterization of cloned esterases from burkholderia gladioli. *J Biotechnol*, 62(1):47–54, Jun 1998.

T. Schulz, J. Pleiss, and R. D. Schmid. Stereoselectivity of pseudomonas cepacia lipase toward secondary alcohols: a quantitative model. *Protein Sci*, 9(6):1053–1062, Jun 2000.

Trevor Selwood, Shawn R. Feaster, Michael J. States, Alton N. Pryor, and Daniel M. Quinn. Parallel mechanisms in acetylcholinesterase-catalyzed hydrolysis of choline esters. *J. Am. Chem. Soc.*, 115: 10477–10482, 1993.

U. Chandra Singh and Peter A. Kollman. An approach to computing electrostatic charges for molecules. *J Comput Chem*, 5(2):129–145, 1984.

Adam J T Smith, Roger Müller, Miguel D Toscano, Peter Kast, Homme W Hellinga, Donald Hilvert, and K. N. Houk. Structural reorganization and preorganization in enzyme active sites: comparisons of experimental and theoretically ideal active site geometries

in the multistep serine esterase reaction cycle. *J Am Chem Soc*, 130:15361–15373, Nov 2008.

Ling Song, Chakrapani Kalyanaraman, Alexander A Fedorov, Elena V Fedorov, Margaret E Glasner, Shoshana Brown, Heidi J Imker, Patricia C Babbitt, Steven C Almo, Matthew P Jacobson, and John A Gerlt. Prediction and assignment of function for a divergent n-succinyl amino acid racemase. *Nat Chem Biol*, 3(8): 486–491, Aug 2007.

S. Tyagi and J. Pleiss. Biochemical profiling in silico–predicting substrate specificities of large enzyme families. *J Biotechnol*, 124 (1):108–16, 2006.

A. Warshel and J. Florián. Computer simulations of enzyme catalysis: finding out what has been optimized by evolution. *Proc Natl Acad Sci U S A*, 95(11):5950–5955, May 1998.

F. K. Winkler, A. D'Arcy, and W. Hunziker. Structure of human pancreatic lipase. *Nature*, 343:771–774, Feb 1990.

Maria I Zavodszky and Leslie A Kuhn. Side-chain flexibility in protein-ligand binding: the minimal rotation hypothesis. *Protein Sci*, 14:1104–1114, Apr 2005.

5.10 Figures

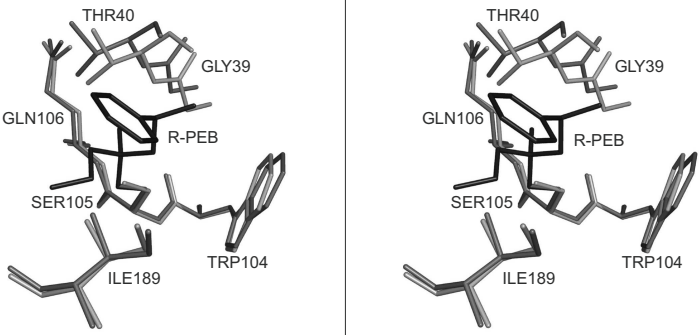

Abbildung 5.1: Binding pocket of *C. antarctica* lipase B with docked substrate (*R*)-PEB: Stereo image of the alcohol binding pocket of 1TCB (light grey), the alcohol pocket of the substrate-imprinted model of 1TCB (dark grey) and the highest scored productive pose of (*R*)-PEB (black), docked into the substrate-imprinted model.

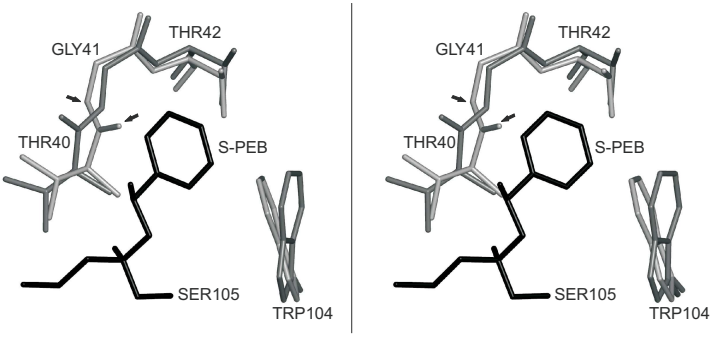

Abbildung 5.2: Binding pocket of C. antarctica lipase B with docked substrate (*S*)-PEB: Stereo image of the binding pocket of the CALB structure 1LBS (light grey) and the substrate-imprinted model 1LBS/(*S*)-PEB (dark grey) with the covalently bound substrate (*S*)-PEB (black). The backbone of T40 and G41 are twisted (arrows), displacing the backbone oxygen by 2.1 Å.

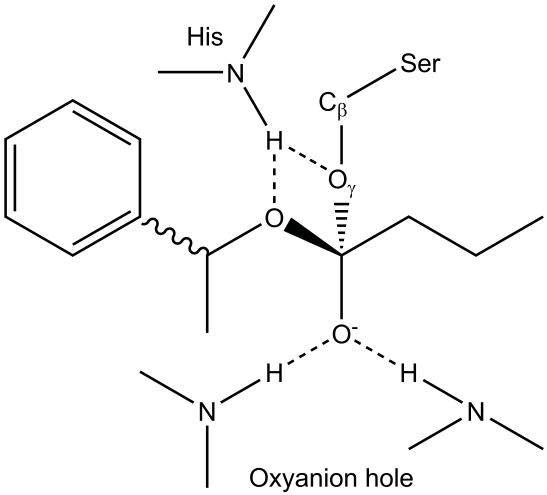

Abbildung 5.3: Substrate in a tetrahedral reaction intermediate form analogous to the transition-state stabilised by the enzyme: A tetrahedral intermediate form of substrate and enzyme. The activated serine O_γ attacks the carbonly oxygen of the substrate. The transition-state is stabilised by four hydrogen bonds (- - -) between the N-H-groups of the oxyanion hole and the substrate oxyanion, the oxygen of the substrate alcohol moiety and a side chain N-H-groups of the catalytic histidine and between the serine O_γ and a side chain N-H-group of the catalytic histidine. The substrate is docked as a tetrahedral intermediate and includes the O_γ and C_β atoms, which are identical to those of the serine residue.

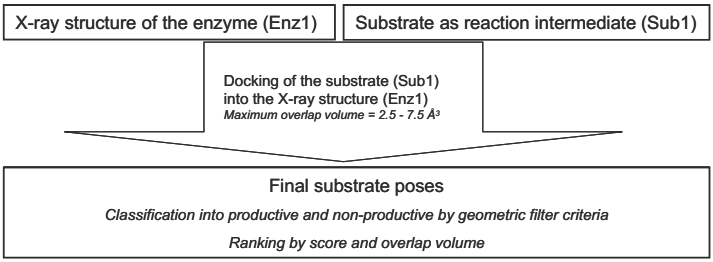

Abbildung 5.4: Flowchart of the conventional docking: Starting from one enzyme structure (Enz1) and one substrate in a reaction intermediate form (Sub1), the substrate is covalently docked into the structure. The resulting substrate poses are classified according to geometric filter criteria into productive and non-productive, and ranked by docking score.

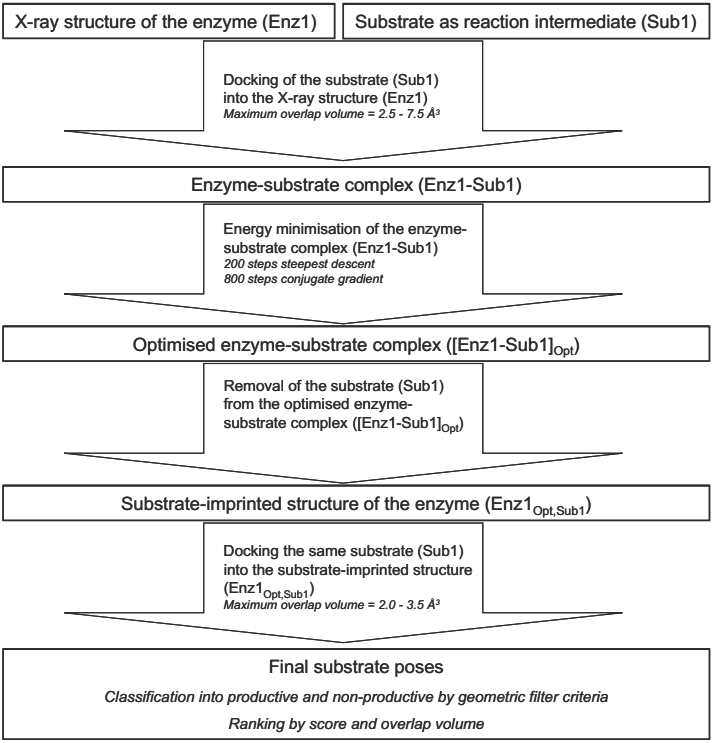

Abbildung 5.5: Flowchart of the substrate-imprinted docking: Starting from one enzyme structure (Enz1) and one substrate in a reaction intermediate form (Sub1), the substrate is covalently docked into the structure in a first round of docking. The best pose from the first docking is used to construct an enzyme-substrate complex (Enz1-Sub1)), which is then energy minimised and provides an optimised enzyme-substrate complex ($[\text{Enz1-Sub1}]_{Opt}$). The substrate is removed from this optimised complex, yielding a substrate-imprinted enzyme structure ($\text{Enz1}_{Opt,Sub1}$). This structure-imprinted structure is used in a second round of docking of the same substrate (Sub1). The resulting substrate poses are classified according to geometric filter criteria into productive and non-productive, and ranked by docking score.

5.11 Tables

Tabelle 5.1: Docking of 1-phenylethyl butyrate: Docking of (*R*)-PEB and (*S*)-PEB into five X-ray structures of CALB and five structure models of a W104A mutant of CALB using FlexX. The substrates were docked into the non-optimised X-ray structures and the substrate-imprinted structures. "+" and "-" indicate that the docking results predict (*R*)-PEB or (*S*)-PEB to be a substrate or a non-substrate. Correct predictions are indicated by bold and large font type. Experimental data (Magnusson et al., 2005) is included for comparison.

	Docking into:			
	X-ray structures		Substrate-imprinted structures	
Structure	(*R*)-PEB	(*S*)-PEB	(*R*)-PEB	(*S*)-PEB
Candida antarctica lipase B				
Experimental data	+	-	+	-
1LBS[a]	**+**	**-**	**+**	+
1LBT[a]	**+**	+	**+**	**-**
1TCA	**+**	**-**	**+**	**-**
1TCB	-	**-**	**+**	**-**
1TCC	**+**	**-**	**+**	**-**
No. false predictions	1	1	0	1
Candida antarctica lipase B, W104A mutant				
Experimental data	+	+	+	+
1LBSW104A[a]	**+**	-	**+**	**+**
1LBTW104A[a]	-	-	**+**	**+**
1TCAW104A	**+**	-	**+**	**+**
1TCBW104A	**+**	-	**+**	**+**
1TCCW104A	**+**	-	**+**	**+**
No. false predictions	1	5	0	0

[a] Structure was resolved with an inhibitor bound.

Tabelle 5.2: RMSD of C. antarctica lipase B alcohol pocket after optimisation: All-atom RMSD of the alcohol pocket of CALB in comparison to the all-atoms RMSD of the whole protein after geometry optimisation. The percentage values in brackets represent the alcohol pocket RMSD as a percentage of the all-atoms RMSD of the whole protein.

	All-atom RMSD [Å]	
	total	alcohol pocket
1LBS/(*R*)-PEB	0.45	0.41 (92%)
1LBS/(*S*)-PEB	0.45	0.68 (150%)[a]
1LBT/(*R*)-PEB	0.39	0.41 (105%)
1LBT/(*S*)-PEB	0.41	0.69 (170%)
1TCA/(*R*)-PEB	0.36	0.40 (110%)
1TCA/(*S*)-PEB	0.36	0.37 (103%)
1TCB/(*R*)-PEB	0.36	0.46 (121%)
1TCB/(*S*)-PEB	0.36	0.43 (118%)
1TCC/(*R*)-PEB	0.40	0.44 (109%)
1TCC/(*S*)-PEB	0.39	0.41 (105%)
1LBSW104A/(*R*)-PEB	0.35	0.52 (149%)
1LBSW104A/(*S*)-PEB	0.33	0.21 (65%)
1LBTW104A/(*R*)-PEB	0.33	0.29 (87%)
1LBTW104A/(*S*)-PEB	0.34	0.28 (83%)
1TCAW104A/(*R*)-PEB	0.28	0.28 (100%)
1TCAW104A/(*S*)-PEB	0.31	0.25 (80%)
1TCBW104A/(*R*)-PEB	0.3	0.33 (109%)
1TCBW104A/(*S*)-PEB	0.32	0.23 (72%)
1TCCW104A/(*R*)-PEB	0.31	0.33 (105%)
1TCCW104A/(*S*)-PEB	0.29	0.24 (84%)

[a] False prediction.

Tabelle 5.3: Docking of methyldecanoic acid butyl esters: Docking of 2- to 8-MDB into seven CRL structures using FlexX. The substrates were docked into the not optimised X-ray structures and the substrate-imprinted structures. "+" and "-" indicate that the docking results predict 2- to 8-MDB to be a substrate or a non-substrate. Correct predictions are indicated by bold and large font type. Experimental data (Hedenström et al., 2002) is included for comparison.

	Experimental data	Docking into X-ray structures							No. of false predictions
Substrate		1CLE[a]	1CRL	1LPM[a]	1LPN[a,b]	1LPO[a]	1LPP[a,b]	1LPS[a]	
(*R*)-2-MDB	+	**+**	-	-	-	**+**	-	-	5
(*S*)-2-MDB	+	**+**	-	-	-	-	-	-	6
(*R*)-3-MDB	+	**+**	**+**	-	-	**+**	-	-	4
(*S*)-3-MDB	+	**+**	**+**	-	-	**+**	-	-	4
(*R*)-4-MDB	+	**+**	**+**	-	-	**+**	-	-	4
(*S*)-4-MDB	+	**+**	**+**	-	-	**+**	-	-	4
(*R*)-5-MDB	+	**+**	**+**	-	-	**+**	-	-	4
(*S*)-5-MDB	+	**+**	**+**	-	-	**+**	-	-	4
(*R*)-6-MDB	+	**+**	**+**	**+**	-	**+**	-	-	3
(*S*)-6-MDB	+	**+**	**+**	**+**	-	**+**	-	-	3
(*R*)-7-MDB	+	**+**	**+**	-	-	**+**	-	-	4
(*S*)-7-MDB	+	**+**	**+**	-	-	**+**	-	-	4
(*R*)-8-MDB	+	**+**	**+**	-	-	**+**	-	-	4
(*S*)-8-MDB	+	**+**	**+**	-	-	**+**	-	-	4
		Docking into substrate-imprinted structures							
(*R*)-2-MDB	+	**+**	**+**	**+**	-	**+**	-	-	3
(*S*)-2-MDB	+	-	**+**	**+**	-	**+**	-	-	4
(*R*)-3-MDB	+	**+**	**+**	**+**	-	**+**	-	-	3
(*S*)-3-MDB	+	-	**+**	**+**	-	**+**	-	-	4
(*R*)-4-MDB	+	**+**	**+**	**+**	-	**+**	-	-	3
(*S*)-4-MDB	+	**+**	**+**	**+**	-	**+**	-	**+**	2
(*R*)-5-MDB	+	**+**	**+**	**+**	-	**+**	-	**+**	2
(*S*)-5-MDB	+	**+**	**+**	**+**	-	**+**	-	**+**	2
(*R*)-6-MDB	+	**+**	**+**	**+**	-	**+**	-	**+**	2
(*S*)-6-MDB	+	**+**	**+**	**+**	-	**+**	-	-	3
(*R*)-7-MDB	+	**+**	**+**	**+**	-	**+**	-	-	3
(*S*)-7-MDB	+	**+**	**+**	**+**	-	**+**	-	-	3
(*R*)-8-MDB	+	**+**	**+**	**+**	-	**+**	-	-	3
(*S*)-8-MDB	+	**+**	**+**	**+**	-	**+**	-	-	3

[a] Structure was resolved with an inhibitor bound. [b] Displaced histidine.

Tabelle 5.4: Docking of 2-hydroxyoctanoic acid butyl esters: Docking of (*R*)-2-HOB and (*S*)-2-HOB into seven BCL and seven CRL structures using FlexX. The substrates were docked into the not optimised X-ray structures and the substrate-imprinted structures. "+" and "-" indicate that the docking results predict (*R*)-2-HOB or (*S*)-2-HOB to be a substrate or a non-substrate. Correct predictions are indicated by bold and large font type. Experimental data (Sakaki et al., 2002) is included for comparison.

	Docking into:			
	X-ray structures		Substrate-imprinted structures	
Structure	(*R*)-2-HOB	(*S*)-2-HOB	(*R*)-2-HOB	(*S*)-2-HOB
Candida rugosa lipase				
Experimental data	+	+	+	+
1CLE[a]	**+**	**+**	**+**	**+**
1CRL	**+**	**+**	**+**	**+**
1LPM[a]	**+**	**+**	**+**	**+**
1LPN[a,b]	-	-	-	-
1LPO[a]	**+**	**+**	**+**	**+**
1LPP[a,b]	-	-	-	-
1LPS[a]	-	-	-	**+**
No. false predictions	3	3	3	2
Burkholderia cepacia lipase				
Experimental data	+	+	+	+
2LIP	-	-	-	**+**
3LIP	-	-	**+**	-
4LIP[a]	**+**	**+**	**+**	**+**
5LIP[a]	**+**	**+**	**+**	**+**
1OIL	**+**	**+**	**+**	**+**
1YS1[a]	**+**	**+**	**+**	**+**
1YS2[a]	**+**	**+**	**+**	**+**
No. false predictions	2	2	1	1

[a] Structure was resolved with an inhibitor bound. [b] Displaced histidine.

Tabelle 5.5: Docking of acetylcholine and butyrylcholine: Docking of ACh and BuCh into six X-ray structures of the acetylcholine esterase from *Torpedo californica* and four X-ray structures of the human butyrylcholine esterase using FlexX. The substrates were docked into the not optimised X-ray structures and the substrate-imprinted structures. "+" and "-" indicate that the docking results predict ACh or BuCh to be a substrate or a non-substrate. Correct predictions are indicated by bold and large font type. Experimental data (Moralev and Rozengart, 2007, Selwood et al., 1993) is included for comparison.

	Docking into:			
	X-ray structures		Substrate-imprinted structures	
Structure	ACh	BuCh	ACh	BuCh
Torpedo californica acetylcholine esterase				
Experimental data	+	-	+	-
1CFJ[a]	**+**	+	**+**	**-**
1DX6[a]	**+**	+	**+**	+
1E3Q[a]	-	**-**	**+**	**-**
1EVE[a]	**+**	+	**+**	**-**
1VXR[a,b]	-	**-**	-	**-**
1QIM	**+**	+	**+**	**-**
No. false predictions	2	4	1	1
human butyrylcholine esterase				
Experimental data	+	+	+	+
1P0M	-	-	-	-
1XLU[a]	**+**	**+**	**+**	**+**
1XLV[a]	**+**	**+**	**+**	**+**
1XLW[a]	-	-	**+**	**+**
No. false predictions	2	2	1	1

[a] Structure was resolved with an inhibitor bound. [b] Displaced histidine.

Tabelle 5.6: RMSD of human butyrylcholine esterase after optimisation: All-atom RMSD of the choline pocket of huBuChE in comparison to the all-atoms RMSD of the whole protein after geometry optimisation. The percentage values in brackets represent the choline pocket RMSD as a percentage of the all-atoms RMSD of the whole protein.

	All-atom RMSD [Å]	
	total	choline pocket
1P0M/ACh	0.49	0.29 (59%)[a]
1P0M/BuCh	0.50	0.33 (66%)[a]
1XLU/ACh	0.50	0.55 (109%)
1XLU/BuCh	0.48	0.54 (113%)
1XLV/ACh	0.52	0.57 (109%)
1XLV/BuCh	0.50	0.57 (113%)
1XLW/ACh	0.48	0.54 (113%)
1XLW/BuCh	0.52	0.56 (107%)

[a] False prediction.

5.12 Supplementary material

Tabelle 5.7: Docking of 2-hydroxyoctanoic acid butyl ester: Docking scores of docking (R)-2-HOB and (S)-2-HOB into seven BCL and seven CRL structures using FlexX. The substrates were docked into the not optimised structures and the substrate-imprinted structures.

	Docking into: non-optimised structures [kJ/mol]		substrate-imprinted structures [kJ/mol]	
Structure	(R)-2-HOB	(S)-2-HOB	(R)-2-HOB	(S)-2-HOB
1CLE	-1.4	-1.0	-7.3	-6.8
1CRL	-1.9	-1.8	-9.0	-7.9
1LPM	-3.0	-1.0	-9.0	-7.2
1LPN[b]	n.s.[a]	n.s.	n.s.	n.s.
1LPO	-1.7	-1.8	-5.2	-6.0
1LPP[b]	n.s.	n.s.	n.s.	n.s.
1LPS	n.s.	n.s.	n.s.	-7.1
2LIP	n.s.	n.s.	n.s.	-1.5
3LIP	n.s.	n.s.	-0.8	n.s.
4LIP	-0.2	-0.4	-3.7	-3.3
5LIP	-2.1	-1.6	-2.9	-2.4
1OIL	-0.5	-1.3	-2.1	-2.5
1YS1	-0.4	-2.4	-3.3	-3.5
1YS2	-0.1	-2.1	-3.5	-3.8

[a] no solution, [b] displaced histidine

Tabelle 5.8: Docking of 2-methylpentanoic acid pentyl ester: Docking scores of docking (R)-2-MPP and (S)-2-MPP into seven BCL and seven CRL structures using FlexX. The substrates were docked into the not optimised structures and the substrate-imprinted structures.

	Docking into:			
	non-optimised structures [kJ/mol]		substrate-imprinted structures [kJ/mol]	
Structure	(R)-2-MPP	(S)-2-MPP	(R)-2-MPP	(S)-2-MPP
1CLE	-4.6	n.s.[a]	n.s.	-9.8
1CRL	-3.0	-1.0	n.s.	n.s.
1LPM	-3.3	-1.7	n.s.	-10.8
1LPN[b]	n.s.	n.s.	n.s.	n.s.
1LPO	-4.9	-1.8	n.s.	n.s.
1LPP[b]	n.s.	n.s.	n.s.	n.s.
1LPS	n.s.	n.s.	n.s.	n.s.
2LIP	n.s.	n.s.	n.s.	n.s.
3LIP	n.s.	n.s.	n.s.	n.s.
4LIP	-0.4	-4.0	-5.4	n.s.
5LIP	-4.0	-1.6	-4.4	n.s.
1OIL	-1.0	-0.9	n.s.	n.s.
1YS1	-1.3	-4.0	-4.6	n.s.
1YS2	-0.8	-3.5	-4.6	n.s.

[a] no solution, [b] displaced histidine

Tabelle 5.9: Docking of 3- and 4-methylpentanoic acid pentyl ester: Docking scores of docking (R)-3-MPP, (S)-3-MPP, and 4-MPP into seven BCL and seven CRL structures using FlexX. The substrates were docked into the not optimised structures and the substrate-imprinted structures.

	Docking into: non-optimised structures [kJ/mol]			substrate-imprinted structures [kJ/mol]		
Structure	(R)-3-MPP	(S)-3-MPP	4-MPP	(R)-3-MPP	(S)-3-MPP	4-MPP
1CLE	-4.8	-4.7	-6.0	n.s.[a]	n.s.	-12.7
1CRL	-4.5	-6.9	-8.2	n.s.	n.s.	-11.8
1LPM	-3.5	-2.4	-4.4	n.s.	n.s.	-11.8
1LPN[b]	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
1LPO	-6.0	-6.4	-7.4	n.s.	n.s.	-12.5
1LPP[b]	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
1LPS	n.s.	n.s.	n.s.	n.s.	n.s.	-9.8
2LIP	n.s.	n.s.	n.s.	n.s.	n.s.	-3.9
3LIP	n.s.	n.s.	-0.0	n.s.	n.s.	-6.8
4LIP	-4.1	-5.4	-4.7	-7.2	-7.3	-5.5
5LIP	-4.6	-3.3	-5.6	n.s.	n.s.	-5.1
1OIL	-1.9	-1.4	-3.5	n.s.	n.s.	-6.2
1YS1	-4.5	-6.1	-5.0	-6.4	-6.7	-5.5
1YS2	-3.9	-4.8	-4.7	-6.4	-6.8	-5.3

[a] no solution, [b] displaced histidine

Tabelle 5.10: Docking of 2-methyldecanoic acid butyl ester: Docking scores of docking (R)-2-MDB and (S)-2-MDBP into seven CRL structures using FlexX. The substrates were docked into the not optimised structures and the substrate-imprinted structures.

	Docking into:			
	non-optimised structures [kJ/mol]		substrate-imprinted structures [kJ/mol]	
Structure	(R)-2-MDB	(S)-2-MDB	(R)-2-MDB	(S)-2-MDB
1CLE	-2.0	-0.1	-7.8	n.s.[a]
1CRL	n.s.	n.s.	-6.5	-1.0
1LPM	n.s.	n.s.	-6.1	-3.4
1LPN[b]	n.s.	n.s.	n.s.	n.s.
1LPO	-2.9	n.s.	-6.2	-5.6
1LPP[b]	n.s.	n.s.	n.s.	n.s.
1LPS	n.s.	n.s.	n.s.	n.s.

[a] no solution, [b] displaced histidine

Tabelle 5.11: Docking of 3-methyldecanoic acid butyl ester: Docking scores of docking (R)-3-MDB and (S)-3-MDBP into seven CRL structures using FlexX. The sub- strates were docked into the not optimised structures and the substrate-imprinted structures.

	Docking into:			
	non-optimised structures [kJ/mol]		substrate-imprinted structures [kJ/mol]	
Structure	(R)-3-MDB	(S)-3-MDB	(R)-3-MDB	(S)-3-MDB
1CLE	-1.5	-1.1	-6.4	n.s.[a]
1CRL	-2.6	-2.9	-7.1	-4.8
1LPM	n.s.	n.s.	-7.7	-7.7
1LPN[b]	n.s.	n.s.	n.s.	n.s.
1LPO	-2.2	-1.2	-7.9	-7.6
1LPP[b]	n.s.	n.s.	n.s.	n.s.
1LPS	n.s.	n.s.	n.s.	n.s.

[a] no solution, [b] displaced histidine

Tabelle 5.12: Docking of 4-methyldecanoic acid butyl ester: Docking scores of docking (R)-4-MDB and (S)-4-MDB into seven CRL structures using FlexX. The substrates were docked into the not optimised structures and the substrate-imprinted structures.

	Docking into: non-optimised structures [kJ/mol]		substrate-imprinted structures [kJ/mol]	
Structure	(R)-4-MDB	(S)-4-MDB	(R)-4-MDB	(S)-4-MDB
1CLE	-2.5	-2.4	-8.2	-8.8
1CRL	-2.1	-2.1	-10.7	-8.0
1LPM	n.s.[a]	n.s.	-10.7	-8.1
1LPN[b]	n.s.	n.s.	n.s.	n.s.
1LPO	-2.7	-4.3	-8.6	-8.7
1LPP[b]	n.s.	n.s.	n.s.	n.s.
1LPS	n.s.	n.s.	n.s.	-3.4

[a] no solution, [b] displaced histidine

Tabelle 5.13: Docking of 5-methyldecanoic acid butyl ester: Docking scores of docking (R)-5-MDB and (S)-5-MDB into seven CRL structures using FlexX. The substrates were docked into the not optimised structures and the substrate-imprinted structures.

	Docking into: non-optimised structures [kJ/mol]		substrate-imprinted structures [kJ/mol]	
Structure	(R)-5-MDB	(S)-5-MDB	(R)-5-MDB	(S)-5-MDB
1CLE	-1.3	-0.7	-8.8	-7.1
1CRL	-1.0	-1.7	-10.8	-11.1
1LPM	n.s.[a]	n.s.	-10.2	-9.9
1LPN[b]	n.s.	n.s.	n.s.	n.s.
1LPO	-2.6	-2.4	-8.7	-7.8
1LPP[b]	n.s.	n.s.	n.s.	n.s.
1LPS	n.s.	n.s.	-8.2	-6.4

[a] no solution, [b] displaced histidine

Tabelle 5.14: Docking of 6-methyldecanoic acid butyl ester: Docking scores of docking (R)-6-MDB and (S)-6-MDB into seven CRL structures using FlexX. The substrates were docked into the not optimised structures and the substrate-imprinted structures.

	Docking into:			
	non-optimised structures [kJ/mol]		substrate-imprinted structures [kJ/mol]	
Structure	(R)-6-MDB	(S)-6-MDB	(R)-6-MDB	(S)-6-MDB
1CLE	-2.0	-0.4	-7.6	-7.8
1CRL	-1.7	-0.9	-10.6	-9.7
1LPM	-0.5	-0.2	-9.7	-10.4
1LPN[b]	n.s.[a]	n.s.	n.s.	n.s.
1LPO	-4.2	-1.9	-8.6	-7.9
1LPP[b]	n.s.	n.s.	n.s.	n.s.
1LPS	n.s.	n.s.	-4.7	n.s.

[a] no solution, [b] displaced histidine

Tabelle 5.15: Docking of 7-methyldecanoic acid butyl ester: Docking scores of docking (R)-7-MDB and (S)-7-MDB into seven CRL structures using FlexX. The substrates were docked into the not optimised structures and the substrate-imprinted structures.

	Docking into:			
	non-optimised structures [kJ/mol]		substrate-imprinted structures [kJ/mol]	
Structure	(R)-7-MDB	(S)-7-MDB	(R)-7-MDB	(S)-7-MDB
1CLE	-0.9	-1.3	-6.1	-7.4
1CRL	-1.2	-1.1	-10.5	-10.3
1LPM	n.s.[a]	n.s.	-9.5	-9.5
1LPN[b]	n.s.	n.s.	n.s.	n.s.
1LPO	-3.1	-2.9	-7.0	-6.7
1LPP[b]	n.s.	n.s.	n.s.	n.s.
1LPS	n.s.	n.s.	n.s.	n.s.

[a] no solution, [b] displaced histidine

Tabelle 5.16: Docking of 8-methyldecanoic acid butyl ester: Docking scores of docking (R)-8-MDB and (S)-8-MDB into seven CRL structures using FlexX. The substrates were docked into the not optimised structures and the substrate-imprinted structures.

	Docking into: non-optimised structures [kJ/mol]		substrate-imprinted structures [kJ/mol]	
Structure	(R)-8-MDB	(S)-8-MDB	(R)-8-MDB	(S)-8-MDB
1CLE	-1.1	-1.7	-6.3	-8.3
1CRL	-1.1	-0.8	-8.5	-10.2
1LPM	n.s.[a]	n.s.	-8.2	-8.8
1LPN[b]	n.s.	n.s.	n.s.	n.s.
1LPO	-2.8	-2.7	-8.2	-8.2
1LPP[b]	n.s.	n.s.	n.s.	n.s.
1LPS	n.s.	n.s.	n.s.	n.s.

[a] no solution, [b] displaced histidine

Tabelle 5.17: Docking of 1-phenylethyl butyrate: Docking scores of docking (R)-PEB and (S)-PEB into five X-ray structures of CALB and five structure models of a W104A mutant of CALB using FlexX. The substrates were docked into the not optimised structures and the substrate-imprinted structures.

	Docking into: non-optimised structures [kJ/mol]		substrate-imprinted structures [kJ/mol]	
Structure	(R)-PEB	(S)-PEB	(R)-PEB	(S)-PEB
1LBS	-5.1	n.s.	-11.8	-14.9
1LBSW104A	-4.1	n.s.	-8.4	-9.2
1LBT	-5.9	-5.4	-9.3	n.s.[a]
1LBTW104A	n.s.	n.s.	-8.7	-8.1
1TCA	-1.5	n.s.	-12.3	n.s.
1TCAW104A	-3.9	n.s.	-8.9	-7.4
1TCB	n.s.	n.s.	-8.3	n.s.
1TCBW104A	-4.4	n.s.	-9	-7.6
1TCC	-5	n.s.	-11.4	n.s.
1TCCW104A	-4.5	n.s.	-8.4	-8.1

[a] no solution

Tabelle 5.18: Docking of acetylcholine and butyrylcholine: Docking scores of docking ACh BuCh into six X-ray structures of the acetylcholine esterase from *Torpedo californica*, four X-ray structures of the human butyrylcholine esterase and four homology models of the human butyrylcholine esterase using FlexX. The substrates were docked into the not optimised structures and the substrate-imprinted structures.

	Docking into: non-optimised structures [kJ/mol]		substrate-imprinted structures [kJ/mol]	
Structure	ACh	BuCh	ACh	BuCh
1CFJ	-16.0	-13.4	-19.4	n.s.[a]
1DX6	-17.1	-13.8	-20.0	-13.9
1E3Q	n.s.	n.s.	-18.4	n.s.
1EVE	-17.2	-13.3	-20.4	n.s.
1VXR[b]	n.s.	n.s.	n.s.	n.s.
1QIM	-5.5	-2.9	-18.2	n.s.
1P0M	n.s.	n.s.	n.s.	n.s.
1XLU	-13.2	-15.1	-17.7	-15.5
1XLV	-11.6	-13.6	-16.4	-15.1
1XLW	n.s.	n.s.	-14.3	-14.4
1P0Mh	-2.4	-3.6	-14.9	-15.4
1XLUh	-10.5	-10.0	n.s.	-15.2
1XLVh	n.s.	n.s.	-16.0	-14.9
1XLWh	n.s.	n.s.	-15.0	-17.8

[a] no solution, [b] displaced histidine

Tabelle 5.19: Docking of methylpentanoic acid pentyl esters: Docking of (*R*)-2-MPP, (*S*)-2-MPP, (*R*)-3-MPP, (*S*)-3-MPP, and 4-MPP into seven BCL and seven CRL structures using FlexX. The substrates were docked into the not optimised X-ray structures and the substrate-imprinted structures. "+" and "-" indicate that the docking results predict MPP to be a substrate or a non-substrate. Correct predictions are indicated by bold and large font type. Experimental dataKodera et al. (1986) is included for comparison.

	Docking into:									
	X-ray structures					Substrate-imprinted structures				
Structure	(*R*)-2-MPP	(*S*)-2-MPP	(*R*)-3-MPP	(*S*)-3-MPP	4-MPP	(*R*)-2-MPP	(*S*)-2-MPP	(*R*)-3-MPP	(*S*)-3-MPP	4-MPP
Candida rugosa lipase										
Experimental data	+	+	-	-	+	+	+	-	-	+
1CLE[a]	**+**	-	+	+	**+**	-	**+**	**-**	**-**	**+**
1CRL	**+**	**+**	+	+	**+**	-	-	**-**	**-**	**+**
1LPM[a]	**+**	**+**	+	+	**+**	-	**+**	**-**	**-**	**+**
1LPN[a,b]	-	-	**-**	**-**	-	-	-	**-**	**-**	-
1LPO[a]	**+**	**+**	+	+	**+**	-	-	**-**	**-**	**+**
1LPP[a,b]	-	-	**-**	**-**	-	-	-	**-**	**-**	-
1LPS[a]	-	-	**-**	**-**	-	-	-	**-**	**-**	**+**
No. false predictions	3	4	4	4	3	7	5	0	0	2
Burkholderia cepacia lipase										
Experimental data	-	-	-	-	+	-	-	-	-	+
2LIP	**-**	**-**	**-**	**-**	-	**-**[c]	**-**[c]	**-**	**-**	**+**
3LIP	**-**	**-**	**-**	**-**	**+**	**-**	**-**	**-**	**-**	**+**
4LIP[a]	+	+	+	+	**+**	+	**-**	+	+	**+**
5LIP[a]	+	+	+	+	**+**	+	**-**	**-**	**-**	**+**
1OIL	+	+	+	+	**+**	**-**	**-**	**-**	**-**	**+**
1YS1[a]	+	+	+	+	**+**	+	**-**	+	+	**+**
1YS2[a]	+	+	+	+	**+**	+	**-**	+	+	**+**
No. false predictions	5	5	5	5	1	4	0	3	3	0

[a] Structure was resolved with an inhibitor bound. [b] Displaced histidine. [c] No substrate pose was found during the first round of docking to construct the substrate-protein complex.

Tabelle 5.20: Charge and protonation of the proteins: Residues were protonated for a pH of 7 according to their calculated pKa values. Most titrateable residues had the same protonation state the would have had with their intrinsic pKa value and are therefore not listed. The catalytic histidine was always protonated as explained in the methods section.

Protein	Total charge	Protonated residues
CALB	0	His224
CRL	-16	His368, His449
BCL	-2	His86, His286
AChE	-5	His422, His437, His468
BuChE	+2	His423, His438

Kapitel 6

Engineering of *Candida antarctica* lipase B for hydrolysis of bulky carboxylic acid esters

P Benjamin Juhl[1], Kai Doderer[2], Frank Hollmann[3], Oliver Thum[4*], Jürgen Pleiss[1*]

[1] Institute of Technical Biochemistry, University of Stuttgart, Allmandring 31, 70569 Stuttgart, Germany

[2] Evonik Degussa GmbH, Rodenbacher Chausee 4, 63457 Hanau-Wolfgang, Germany

[3] Biocatalysis and Organic Chemistry, Delft University of Technology, Julianalaan 136, 2628 BL Delft, Netherlands

[4] Evonik Goldschmidt GmbH, Goldschmidtstrasse 100, 45172

Essen, Germany

6.1 Abstract

Candida antarctica lipase B (CALB) is a widely used biocatalyst with high activity and specificity for a wide range of primary and secondary alcohols. However, the range of converted carboxylic acids is more narrow and mainly limited to unbranched fatty acids. To further broaden the biotechnological applications of CALB it is of interest to expand the range of converted carboxylic acid and extend it to carboxylic acids that are branched or substituted in close proximity of the carboxyl group.

An i*n silico* library of 2400 *Candida antarctica* lipase B variants was built and screened *in silico* by substrate-imprinted docking, a four step docking procedure. First, reaction intermediates of putative substrates are covalently docked into enzyme active sites. Second, the geometry of the resulting enzyme-substrate complex is optimized. Third, the substrate is removed from the complex and then docked again into the optimized structure. Fourth, the resulting substrate poses are rated by geometric filter criteria as productive or non-productive poses.

Eleven enzyme variants resulting from the *in silico* screening were expressed in *Escherichia coli* BL21 and measured in the hydrolysis of two branched fatty acid esters, isononanic acid ethyl ester and 2-ethyl hexanoic acid ethyl esters. Five variants showed an initial increase in activity. The variant with the highest wet mass activity (T138S) was purified and further characterized. It showed a 5-fold increase in hydrolysis of isononanoic acid ethyl ester, but not toward sterically more demanding 2-ethyl hexanoic acid ester.

6.2 Introduction

Candida antarctica lipase B (CALB) is a widely used biocatalyst (Anderson et al., 1998, Gotor-Fernandez et al., 2006, Plou et al., 2002), due to its broad substrate specificity toward esters of primary and secondary alcohols, high stereoselectivity, high activity, and high thermostability. To further optimize the enzyme's properties and to tailor it for different applications a broad range of variants has been developed: To increase expression yield in *E. coli* five amino acid exchanges of hydrophobic residues on the protein surface to aspartate were introduced (Jung and Park, 2008). To increase activity toward *p*-nitrophenyl palmitate the L278P and L219Q/L278P variants of CALB were designed (Kim et al., 2007). To improve thermostability (Patkar et al., 1998) and resistance to irreversible thermal inactivation of the enzyme (Zhang et al., 2003) amino acid exchanges at the positions 103, 221, and 281 were introduced. To lleviate the intracellular proteolytic degradation of CALB when expressed in *E. coli* amino acid exchanges at position L149 and V223 were introduced (Narayanan and Chou, 2009). To increase activity for aldol addition the catalytic serine was changed to an alanine (Branneby et al., 2003). To alter stereoselectivity and substrate specificity of CALB amino acid exchanges at position W104, in the binding pocket for secondary alcohols, were introduced (Magnusson et al., 2005, Patkar et al., 1998).

However, in contrast to the broad spectrum of alcohol moieties ac-

Abbreviations: CALB - *Candida antarctica* lipase B, INA - isonanoic acid, INA-EE - isonanoic acid ethyl ester, EHA - 2-ethyl hexanoic acid, EHA-EE - 2-ethyl hexanoic acid ethyl ester

cepted as substrates, only a limited spectrum of acids is accepted by CALB. For instance, acyl moieties with sterically demanding α- and β-substitutions yield significantly reduced specific activities (Hollmann et al., 2009). It was also shown by molecular modeling and crystallographic studies that the binding pocket close to the active center is rather restricted (Uppenberg et al., 1994, 1995). Therefore CALB can only convert straight chain fatty acids of different length with a preference toward C5 and C12 (Kirk et al., 1992) with high specific activities. However, some esters built from branched and sterically demanding acids are of industrial interest. Branched carboxylic acids such as isononanoic acid (INA, 3,5,5-trimethylhexanoic acid) and 2-ethyl hexanoic acid (EHA) are important raw materials for cosmetic ingredients, such as emollient esters. Esters of aromatic carboxylic acids function as antioxidants (Majekodunmi et al., 2007), while other esters are used as emulsifiers or flavor additives (Su et al., 2010). Because of its high inherent activity and despite the low relative activity toward such substrates, CALB is the best candidate for protein engineering with the goal of developing a biocatalyst for such substrates. To the best of our knowledge no optimization of CALB toward this direction has been described until now.

Therefore, the goal was to engineer a CALB variant with higher activity toward bulky and branched carboxylic acids to achieve high conversion of the racemic esters of the branched substrates INA and EHA. In order to design such a CALB variant, substrate-imprinted docking was used (Juhl et al., 2009). Docking of substrate molecules into enzyme structures and models is frequently used to retrospectively rationalize the molecular basis for experimentally observed enzymatic properties and explain changes caused by specific mutations (El-Hawari et al., 2009, Höst et al., 2006, Padhi et al., 2009). Recently, molecular docking was also used to predict the enzymatic function

of proteins with known structure but unknown function (Hermann et al., 2007, Kalyanaraman et al., 2008, Song et al., 2007, Xiang et al., 2009), and to predict new substrates for enzymes with known structure and function (Favia et al., 2008, Hermann et al., 2006, Irwin et al., 2005, Kalyanaraman et al., 2005, Macchiarulo et al., 2004). The identification of amino acid positions that are relevant for the conversion of a specific substrate has been successfully shown for a carbonyl reductase (Zhu et al., 2008), a fructosyl amine oxidase (Miura et al., 2008), and a glutathione transferase (Kapoli et al., 2008). In contrast to the simple prediction of hotspots by investigating the enzyme structure, a specific amino acid exchanges with increased catalytic activity for specific substrates were predicted for the human anhydrase II (Höst and Jonsson, 2008) and a hydantoinase (Lee et al., 2009) by molecular docking. In this study, molecular docking was first used to identify hotspots for mutagenesis in CALB wild type, then two strategies were followed: (i) Single amino acid exchange variants were modeled, collected in a library I, and selected by their docking scores and (ii) a range of amino acids exchanges for these hotspots were chosen in order to create more space for the substrates in the binding pocket and to establish a biochemical environment favorable for the substrates, an combinatorial *in silico* library II of 2400 CALB variants was built, and substrate-imprinted docking (Juhl et al., 2009) was applied to identify amino acid exchanges that frequently resulted in better docking scores. Enzyme variants that were identified by this method were expressed in *E. coli* BL21, and their activity toward two model esters of branched fatty acids, INA and EHA ethyl esters (INA-EE, EHA-EE), was measured.

6.3 Methods

6.3.1 Modeling

Substrate-imprinted docking

Substrate-imprinted docking follows a four step procedure in order to model substrate binding and conversion by combining molecular docking and molecular mechanics, as described preciously (Juhl et al., 2009). The substrate is docked as a tetrahedral reaction intermediate into an X-ray structure of the enzyme by FlexX (Kramer et al., 1999, Rarey et al., 1996). The resulting enzyme-substrate complexes are energy minimized by Amber (Case et al., 2006). The optimized enzyme geometry is used for a second round of docking of the same substrate. Geometric filter criteria are applied on the resulting substrate poses to discriminate between productive and non-productive poses. In the case of ester hydrolysis, the geometric filter checks for (i) the correct formation of the two hydrogen bonds that form the oxyanion hole, (ii) the formation of a hydrogen bond between the Oγ of the catalytic serine and the catalytic histidine, and (iii) the formation of a hydrogen bond between the oxygen of the ester alcohol moiety and the catalytic histidine. These four hydrogen bonds characterize a productive substrate pose. Parameters during the two docking steps and the energy minimization of the complex were used as described (Juhl et al., 2009). As both substrates are chiral, the two stereoisomers were docked separately.

Identification of hotspots

Hotspots in the enzyme binding pocket were identified by docking tetrahedral reaction intermediates of isononanoic acid ethyl ester (INA-

EE) and 2-ethyl hexanoic acid ethyl ester (EHA-EE) into a CALB X-ray structure with a resolution of 1.6 Å (PDB-ID: 1TCA (Uppenberg et al., 1994)). By analyzing the clashes between substrate and enzyme, amino acids were considered as hotspots, that led to an unfavorable docking score. Amino acids of the catalytic triad and the oxyanion hole were not selected.

Construction of the *in silico* enzyme variant libraries

Two libraries were constructed. Library I consisted of 40 manually built enzyme variants. Amino acid exchanges were selected abd combined to increase space and to reduce the polarity of the acyl pocket. Structure models of these variants were constructed by exchanging the respective amino acids by Swiss-PDB Viewer (Guex and Peitsch, 1997). In contrast, library II is an automatically built combinatorial library. After identifying the hotspots, the respective amino acids were exchanged by less bulky amino acids, to create more space for the substrates in the binding pocket, while maintaining hydrophobicity (Table 6.1). Therefore the amino acids chosen to replace the hotspot residues were (i) of the same size or smaller than the amino acid in the wild type, (ii) not charged, and (iii) not aromatic (Table 6.1). The side chains of these hotspots were removed, and new side chains were built up according to the residue templates in the tLEAP program of the AMBER program suite (Case et al., 2006). The *in silico* library contained all combinations of the hotspots and possible amino acids exchanges and wild type amino acids, resulting in a total of 2400 variant models.

In silico screening of enzyme variant libraries

Library I was screened by docking INA-EE and EHA-EE into the enzyme models and selecting the variants that displayed the largest improvement of the docking score in comparison to wild type CALB. Library II was screened by docking INA-EE and EHA-EE into all structure models of the combinatorial library II. The respective variants were ranked by docking score of a productive substrate pose found with them for a specific substrate. For each docked substrate, up to the 600 best ranked enzyme variants, which corresponds to the best scoring 25% of library II, were analyzed for their amino acid composition in the hotspots. The occurrence of every amino acid at a hotspot position was counted and is listed as a percentage (frequency) of the total number of analyzed variants. As the aim of this study was to achieve better conversion of esters with bulky acyl moieties, but not be stereoselective, the frequencies of an amino acid from both enantiomers of the substrates at one position were added up. The screening results were interpreted in the following way: (i) No change or only a small change in activity was predicted by the screening when the frequency of an amino acid at one position differed by less than 15% from the frequency of the wild type amino acid, (ii) an increase in activity was predicted when the frequency was more than 15% larger than the wild type frequency, and (iii) a decrease in activity was predicted when the freuqnecy was more than 15% smaller than the wild type. The most frequent amino acid exchange at a position was suggested for expression in *E. coli*, if the prediction for this amino acid was either an increase in activity or if no or only a small change was predicted. If more than one amino acid exchange occurred more frequent than the wild type amino acid all of them were considered for expression in *E. coli*.

6.3.2 Experimental

Enzyme expression and purification

CALB wild type and eleven mutant genes were cloned into a pET22b+ vector and expressed in *E. coli* BL21(de3) for an initial screening by wet mass activity. After the initial screening wild type CALB and one promising variant were cloned into pET22b+ in fusion with the C-terminal His-tag for subsequent purification and determination of specific activity. After fermentation and cell disruption by glass bead milling, the His-tagged CALB and its T138S variant were purified on a Ni-NTA column from Quiagen using the standard protocol for native proteins. Protein concentration in the fractions was determined by the Bradford method (Bradford, 1976) and the purity of the fraction was estimated by SDS-PAGE.

Activity measurements

Enzymatic activity of crude cell extracts and purified enzyme solutions was measured using the pH-stat method (Jacobsen et al., 1957). 40 ml of a 5% substrate solution, 1 g/l gum arabic, 3 g/l NaCl, 0,068 g/l KH_2PO_4, 9%(v/v) glycerol were adjusted to a pH of 7.5 by titration of 0.1 M NaOH at 40°C. Autohydrolysis of the substrates was measured at pH 7.5 for 5 minutes. The reaction was then started by addition of 100 μl purified enzyme solution or 2 ml crude cell extract and measured for up to 15 minutes.

6.4 Results

6.4.1 CALB wild type activity toward various esters with α- and β-substituted acyl moieties

In order to assess the loss in specific activity for esters with sterically demanding acyl moieties the hydrolysis of several esters by commercial CALB (Novozyme CALB L) was measured (Table 6.2). INA-EE, EHA-EE, and other sterically hindered substrates show a steep decrease in activity. INA-EE and EHA-EE are hydrolyzed at very low activities (<0.1%) as compared to saturated and unbranched fatty acid esters like lauric acid methyl ester or caprylic acid methyl ester. The activity toward EHA-EE was the lowest of all measured substrates.

6.4.2 Identification of hotspots

INA-EE and EHA-EE were covalently docked into X-ray structures of CALB wild type (PDB-ID: 1TCA), in order to identify hotspots for mutagenesis. This led to a total of five hotspots: D134, T138, Q157, I189, and V190. Side chain atoms from these six amino acids clashed with substrate atoms of INA-EE, EHA-EE, or both (Figure 1). As a control, the docking of methyl caprylate, a good substrate for CALB, did not result in any clashes between the enzyme and the best scoring productive pose of the substrate.

The residue T40 also resulted in clashes with the two docked substrates, but was not considered a hotspot because it is part of the oxyanion hole and it was previously reported that CALB variants, where T40 was exchanged, lost 99% of their catalytic activity (Magnusson et al., 2001).

6.4.3 Prediction derived from screening library I

Based on the identified hotspots and the assumption that smaller and uncharged amino acids in the acyl binding site might improve the hydrolysis of aliphatic esters with a bulky acyl moiety, library I was constructed. The variants T138A, Q157L, Q157I and , I189L as well as the double variants T138A/Q157L and Q157I/I189L led to slightly better docking results than the wild type and were therefore selected for expression in *E. coli*.

6.4.4 *In silico* screening of the combinatorial enzyme variant library II

The combinatorial *in silico* library (Table 6.1) was screened with INA-EE and EHA-EE by using the substrate-imprinted docking methodology. The screening of INA-EE resulted in productive poses for 1680 CALB variants with the R-enantiomer and for 859 variants with the S-enantiomer. The screening of the sterically more hindered EHA-EE resulted in productive poses for 59 CALB variants with the R-enantiomer and for 327 variants with the S-enantiomer. As the analysis was limited to a maximum of the best scoring 25% of the library, this related to 600 variants for each of the INA-EE enantiomers and all variants with productive poses for both EHA-EE enantiomers. The occurrence of every amino acid at a hotspot position was counted and listed as a percentage (frequency) of the total number of analyzed variants (Table 6.3).
For INA-EE the screening resulted in variants with a higher or similar frequency than the wild type at four of the five hotspot position.

Q157T, Q157A, Q157V, Q157S, and I189V all showed a higher frequency than thee wild type amino acid and D134N and T138S showed a similar frequency as the wild type, but were the variants with the highest frequency at this two positions. Therefore, these seven variants were predicted to be beneficial for the hydrolysis of INA-EE. For EHA-EE D134T, D134N, D134S, T138N, T138Q, Q157T, Q157N, Q157A, Q157V, Q157S, I189T, I189A, and V190A showed a higher frequency than the wild type amino acid and were therefore predicted to be beneficial for the hydrolysis of EHA-EE.

For the positions 134, 138 and 157 The variants that were predicted to be beneficial for one substrate were also predicted to be beneficial for the other substrate, or at least to be of similar activity as the wild type for the other substrate. Therefore, these nine variants were selected for expression in *E. coli*: D134S, D134N, T138Q, T138S, Q157T, Q157N, Q157A, Q157V, and Q157S.

The position Q157 is the one where most amino acid exchanges were predicted to increase catalytic activity for both substrates, INA-EE and EHA-EE. Q157T, Q157N, Q157A, Q157V, and Q157S were predicted to improve catalysis for EHA-EE and Q157T, Q157A, Q157V, and Q157S for INA-EE. Therefore these five variants were considered as candidates for expression in *E. coli*. The positions I189 and V190 are close to the catalytic serine, and the screening results for both substrates are contrary to each other. The best prediction for INA-EE at position I189 is I189V, which is predicted to have a decreased activity for EHA-EE. Similarly, the best prediction for EHA-EE at position I189 is I189T, which is predicted to have a reduced activity for INA-EE than the wild type amino acid. Therefore, no amino acid exchanges at these positions were considered for expression in *E. coli* based on the *in silico* screening of library II.

6.4.5 Activity measurements of CALB wild type and variants

Wet mass activities of CALB wild type and variants

In an initial activity screen, enzymatic hydrolysis INA-EE and EHA-EE was measured by using bio wet mass from fermentation (Table 6.4). The four single and two double amino acid exchange variants that were selected from library I did not show an increased activity. Four of the variants (T138A: 1.1 U/g, Q157L: 1.1 U/g, Q157I: 1.0 U/g, Q157L/I189L: 0.9 U/g) even showed a slight decrease in activity for INA-EE compared to the wild type (INA-EE: 1.7 U/g, EHA-EE: 1.1 U/g) and four (T138A: 0.9 U/g, Q157L: 0.9 U/g, T138A/Q157L: 0.9 U/g, Q157L/I189L: 0.8 U/g) showed a decrease in activity for EHA-EE.

The five variants that were successfully expressed based on the *in silico* screening of library II all showed a substantially higher activity for INA-EE (D134S: 5.8 U/g, T138S: 30.6 U/g, Q157N: 2.9 U/g, Q157S: 7.1 U/g, Q157T: 8.6 U/g). For EHA-EE, three of them showed a higher activity (D134S: 3.1 U/g, T138S: 2.1 U/g, Q157S: 1.5 U/g), one showed no change in activity (Q157N), and one showed a decrease in activity (Q157T: 0.8 U/g). From the nine CALB variants that were selected according to the *in silico* screening of library II the variants D134N, T138Q, Q157A, and Q157V could not be expressed in *E. coli*.

Specific activities of CALB wild type and the T140S variant

The CALB wild type and the variant that showed an 18-fold increase in wet mass activity toward INA-EE (T138S) were expressed

with a C-terminal His-tag and prepared with 90% purity (estimated by SDS-PAGE, Figure 6.2). The first elute fraction contained 1 mg/ml protein for wild type CALB and 0.72 mg/ml protein for the T138S variant. The unbranched methyl caprylate was used as reference substrate. The wild type displayed a specific activity of 741 U/mg toward methylcaprylate (Table 6.5), while the T138S variant only had 58% of the wild type activity (427 U/mg toward methylcaprylate). The specific activity of CALB wild type toward INA-EE was 3.1 U/mg, while the specific activity of the T138S variant was increased five times (15.3 U/mg). The specific activity of CALB wild type and the T138S variant toward EHA-EE was similar (0.5 U/mg and 0.6 U/mg).
Compared to EHA the sterical demand of INA is significantly lower as the branching point is only at β-position, and only a methyl group is attached there, thus giving a very good explanation for the higher efficiency regarding INA-EHA hydrolysis. Still, the T130S variant shows a substrate specificity that is shifted toward both substrates. Relative activity (normalized for activity toward methylcaprylate) of this variant is with 3.6% toward INA-EE and 0.14% toward EHA-EE 8.6 times, respecctively 2 times higher than the relative wild type activity (0.42% and 0.07%).

6.5 Discussion

In this study, substrate-imprinted docking was used to identify five CALB variants with increased activity by systematically screening an *in silico* library of enzyme variants. The method starts with 2400 *in silico* enzyme variants, which are systematically screened for amino acid exchanges the occur at a high frequency in the variants that

perform well in docking. From these 2400 variants we predicted an increased enzymatic activity toward the screened substrates INA-EE and EHA-EE for nine single amino acid exchanges. Five of the nine variant could be expressed successfully in *E. coli* and were initially characterized by measuring bio wet mass activity toward the two substrates. In this initial experimental characterization one variant (T138S) clearly displayed the highest wet mass activity and was therefore selected for further characterization. The T138S variant and the wild type enzyme were purified and their specific activity was measured. The activity increase observed for the wet mass activity was much more pronounced than the increase observed for the specific activity. Reasons for this difference can be varying expression levels, which could not be quantified by SDS-PAGE, and varying soluble or active fractions of expressed CALB.

While a rational amino acid exchange at a position identified by molecular docking was successful in the past (Höst and Jonsson, 2008, Lee et al., 2009), the amino acid exchanges from library I, which was constructed in a similar way, did not successfully increase the activity for bulky acyl substrates in the case of CALB. With a bulky and hydrophobic substrate we modeled enzyme variant with a single amino acid exchange that featured smaller and uncharged amino acids in the hotspot positions and expressed selected enzyme variants that displayed small improvements in the docking score in *E. coli*. However, these enzyme variants did not display an increased activity for INA-EE and EHA-EE. In contrast, library II follows a new approach and systematically investigates a large number of *in silico* enzyme variants. The automatic and combinatorial approach demonstrated here has the advantage of initially allowing for a wider variety of enzyme variants that are systematically screened and does not rely on docking into a single or limited number of enzyme variants, which

might not be able to identify a beneficial variant, as seen in library I.

Five hotspots for mutagenesis in the acyl binding pocket of CALB were suggested in this study. For four hotspots amino acid exchanges were predicted to improve activity toward INA-EE and EHA-EE, while for the potential hotspot V190 neither the screening of library I, nor the *in silico* screening of library II led to predictions of increased activity. Expressed enzyme variants with amino acid exchanges at three of the four hotspots with beneficial predictions (D134, T138, Q157) showed significant improvements in bio wet mass activity toward INA-EE and EHA-EE. Amino acid exchanges at the fourth hotspot (I189) did not significantly change the enzyme activity. Thus, molecular docking is able to successfully and reliably (75%) identify amino acid positions that are important for the catalytic activity toward the docked substrate. This is in good agreement with other studies that successfully identified relevant amino acid positions for substrate conversion for a carbonyl reductase (Zhu et al., 2008), a fructosyl amine oxidase (Miura et al., 2008), and a glutathione transferase (Kapoli et al., 2008) in the past.

The number of productive poses found during the screening of library II for EHA-EE was notably smaller than for INA-EE (386 vs. 2593). This reflects very well the fact, that the α-ethyl group of EHA-EE is much more sterically demanding than the β-methyl group of INA-EE, which is most likely the reason for the differences in CALB activity toward INA-EE and EHA-EE (Table 6.2). Therefore, it can be assumed, that the overall geometry of the docking results is good. This is also supported by the high reliability of the identified mutagenesis hotspots. However, due to the inability of the predictions from library I to correctly identify an improved variant, the scoring and ranking for these docking results with a reliable geometry seems

to be much less reliable.

To assess the reliability the second approach, all measured bio wet mass activities for single amino acid exchange variants were compared to the predictions for the respective amino acid exchange derived from screening library II. Three of the expressed CALB variants (T138S, Q157S, Q157T) were predicted to have an increased activity for INA-EE, and an increased activity was found for all of them, including the T138S variant that displayed the highest activity of all variants toward INA-EE. Five of the expressed CALB variants were predicted to have an increased activity for EHA-EE, and an increased activity was confirmed for three of them, including the D134S variant, that displayed the highest activity of all variants toward EHA-EE. Six of the expressed enzyme variants were not predicted to be beneficial for the activity toward INA-EE, and no increase was confirmed for four of them, while two of these variants displayed an increased activity toward INA-EE. Four of the expressed enzyme variants were not predicted to be beneficial for the activity toward EHA-EE, and this was confirmed for all four of them. Interestingly, all four single amino acid variants selected from library I were predicted to have no change or even a decrease in activity according to the screening of library II. In total the comparison of the *in silico* screening with the experimentally measured wet mass activities resulted in 78% true and 22% false predictions. The success rate for the beneficial predictions is 75%, and those are typically the ones investigated further by experiments. Such a success rate makes it very likely to find an enzyme variant with increased activity by only expressing a few enzyme variants selected by *in silico* screening.

While directed evolution is a very powerful tool for protein engineering it very much relies on a strong selection or screening method for the property of interest (Jäckel et al., 2008). The development

of such screening or selection methods is very time consuming and not always possible. Therefore, rational design approaches are very useful if no high-throughput selection or screening method exists, as was the case for INA-EE and EHA-EE. A rational design approach can lead to one or few enzyme variants to be experimentally tested for a specific activity, or a focused library useful for finding a wider range of activities (Seifert et al., 2009). However, the rational design approach is restricted to the molecular principles that are understood well and are adequately represented in the theoretical models used. Directed evolution is not restricted in such a manner and can therefore also find amino acid exchanges that or located far away from the active site or influence the entropy of the reaction.
All variants derived from library II were single amino acid exchanges. However, previous works have established, that a maximum activity is often not achieved with just one amino acid exchange, but require two, three, or even more amino acid exchanges to achieve a maximum activity toward a specific substrate (Reetz et al., 2009, Seifert et al., 2009). Due to the screening process that already examines the T138S amino acid exchange in combination with many other amino acid exchanges in library II, the T138S variant is well suited to act as a first step for further studies to improve the activity. These could include a combination of beneficial amino acid exchanges at different positions found in library II, or an analysis of the subset of library II that is positive for the T138S amino acid exchange.

6.6 Conclusions

The T138S CALB variant showed shift in activity toward sterically demanding acyl substrates demonstrated by the 8.6-fold increase

in relative activity and 5-fold increase in specific activity toward INA-EE. Other expressed CALB variants identified in this study also showed altered activity toward the examined substrates. Therefore, substrate-imprinted was successful in identifying hotspots for mutagenesis with regard to activity toward INA-EE and EHA-EE, and *in silico* screening of an enzyme variant library was successful in identifying a small, highly enriched set of enzyme variants with a predicted and experimentally verified increase in activity toward the substrates INA-EE and EHA-EE. It was unexpected, that enzyme variants featuring smaller aliphatic amino acids in the identified hotspots than the wild type often led to worse scores in substrate-imprinted docking of bulky and aliphatic substrates than the wild type. Expression and activity measurements of these enzyme variants confirmed the predicted decrease in activity.
Therefore, the use of an *in silico* library to perform virtual screening is a suitable tool for protein engineering of enzymes. The enzyme variant found in this study unlocks new substrates for the enzymatic synthesis with CALB that have not been previously described, and provide a starting point for further optimization of CALB for bulky substrates.

6.7 Acknowledgements

This work was supported by the FFachagentur Nachwachsende Rohstoffe e. V.FFKZ 22003005.

References

E. M. Anderson, M. Karin, and O. Kirk. One biocatalyst - many applications: The use of candida antarctica b-lipase in organic synthesis. *Biocatalysis Biotransform*, 16(3):181–204, 1998.

M. M. Bradford. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem*, 72:248–254, May 1976.

Cecilia Branneby, Peter Carlqvist, Anders Magnusson, Karl Hult, Tore Brinck, and Per Berglund. Carbon-carbon bonds by hydrolytic enzymes. *J Am Chem Soc*, 125(4):874–875, Jan 2003.

D.A. Case, T.A. Darden, T.E. Cheatham, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, K.M. Merz, D.A. Pearlman, M. Crowley, R.C. Walker, W. Zhang, B. Wang, S. Hayik, A. Roitberg, G. Seabra, K.F. Wong, F. Paesani, X. Wu, S. Brozell, V. Tsui, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Homak, G. Cui, P. Beroza, D.H. Mathews, C. Schafmeister, W.S. Ross, and P.A. Kollman. Amber 9. Technical report, University of California, San Francisco, 2006.

Yasser El-Hawari, Angelo D Favia, Ewa S Pilka, Michael Kisiela, Udo Oppermann, Hans-Jörg Martin, and Edmund Maser. Analysis of the substrate-binding site of human carbonyl reductases cbr1 and cbr3 by site-directed mutagenesis. *Chem Biol Interact*, 178(1-3): 234–241, Mar 2009.

Angelo D Favia, Irene Nobeli, Fabian Glaser, and Janet M Thornton. Molecular docking for substrate identification: the short-chain dehydrogenases/reductases. *J Mol Biol*, 375:855–874, Jan 2008.

V. Gotor-Fernandez, E. Busto, and V. Gotor. Candida antarctica lipase b: An ideal biocatalyst for the preparation of nitrogenated

organic compounds. *Adv. Synth. Catal.*, 348(7-8):797–812, May 2006.

N. Guex and M. C. Peitsch. Swiss-model and the swiss-pdbviewer: an environment for comparative protein modeling. *Electrophoresis*, 18(15):2714–2723, Dec 1997.

J. C. Hermann, E. Ghanem, Y. Li, F. M. Raushel, J. J. Irwin, and B. K. Shoichet. Predicting substrates by docking high-energy intermediates to enzyme structures. *J Am Chem Soc*, 128(49):15882–91, 2006.

Johannes C Hermann, Ricardo Marti-Arbona, Alexander A Fedorov, Elena Fedorov, Steven C Almo, Brian K Shoichet, and Frank M Raushel. Structure-based activity prediction for an enzyme of unknown function. *Nature*, 448(7155):775–779, Aug 2007.

F. Hollmann, P. Grzebyk, V. Heinrichs, K. Doderer, and O. Thum. On the inactivity of candida antartica lipase b towards strong acids. *J. Mol. Catal. B.*, 57(1-4):257–261, May 2009. ISSN 1381-1177.

Gunnar Höst, Lars-Göran Mårtensson, and Bengt-Harald Jonsson. Redesign of human carbonic anhydrase ii for increased esterase activity and specificity towards esters with long acyl chains. *Biochim Biophys Acta*, 1764(10):1601–1606, Oct 2006.

Gunnar E Höst and Bengt-Harald Jonsson. Converting human carbonic anhydrase ii into a benzoate ester hydrolase through rational redesign. *Biochim Biophys Acta*, 1784(5):811–815, May 2008.

John J Irwin, Frank M Raushel, and Brian K Shoichet. Virtual screening against metalloenzymes for inhibitors and substrates. *Biochemistry*, 44(37):12316–12328, Sep 2005.

Christian Jäckel, Peter Kast, and Donald Hilvert. Protein design by directed evolution. *Annu Rev Biophys*, 37:153–173, 2008.

C. F. Jacobsen, J. Léonis, K. Linderstrøm-Lang, and M. Ottesen. The ph-stat and its use in biochemistry. In David Glick, editor, *Methods of Biochemical Analysis*, pages 171–210. Interscience Publishers, Inc., 1957.

Benjamin Juhl, Peter Trodler, Sadhna Tyagi, and Jurgen Pleiss. Modelling substrate specificity and enantioselectivity for lipases and esterases by substrate-imprinted docking. *BMC Struct. Biol.*, 9 (1):39, 2009.

Suhyun Jung and Seongsoon Park. Improving the expression yield of candida antarctica lipase b in escherichia coli by mutagenesis. *Biotechnol Lett*, 30(4):717–722, Apr 2008.

C. Kalyanaraman, K. Bernacki, and M. P. Jacobson. Virtual screening against highly charged active sites: identifying substrates of alpha-beta barrel enzymes. *Biochemistry*, 44(6):2059–71, 2005.

Chakrapani Kalyanaraman, Heidi J Imker, Alexander A Fedorov, Elena V Fedorov, Margaret E Glasner, Patricia C Babbitt, Steven C Almo, John A Gerlt, and Matthew P Jacobson. Discovery of a dipeptide epimerase enzymatic function guided by homology modeling and virtual screening. *Structure*, 16(11):1668–1677, Nov 2008.

Panagiota Kapoli, Irene A Axarli, Dimitris Platis, Maria Fragoulaki, Mark Paine, Janet Hemingway, John Vontas, and Nikolaos E Labrou. Engineering sensitive glutathione transferase for the detection of xenobiotics. *Biosens Bioelectron*, 24(3):498–503, Nov 2008.

So-Young Kim, Jung-Hoon Sohn, Yu-Ryang Pyun, In Seok Yang, Kyung-Hyun Kim, and Eui-Sung Choi. In vitro evolution of lipase b from candida antarctica using surface display in hansenula polymorpha. *J Microbiol Biotechnol*, 17(8):1308–1315, Aug 2007.

Ole Kirk, Fredrik Björkling, Sven Godtfredsen, and Thomas Larsen. Fatty acid specificity in lipase-catalyzed synthesis of glucoside esters. *Biocatal. Biotransform.*, 6:127–134, 1992.

B. Kramer, M. Rarey, and T. Lengauer. Evaluation of the flexx incremental construction algorithm for protein-ligand docking. *Proteins-Structure Function and Genetics*, 37(2):228–241, 1999.

S. C. Lee, Y. J. Chang, D. M. Shin, J. Han, M. H. Seo, H. Fazelinia, C. D. Maranas, and H. S. Kim. Designing the substrate specificity of d-hydantoinase using a rational approach. *Enzyme And Microbial Technology*, 44(3):170–175, March 2009.

Antonio Macchiarulo, Irene Nobeli, and Janet M Thornton. Ligand selectivity and competition between enzymes in silico. *Nat Biotechnol*, 22(8):1039–1045, Aug 2004.

A. Magnusson, K. Hult, and M. Holmquist. Creation of an enantioselective hydrolase by engineered substrate-assisted catalysis. *J Am Chem Soc*, 123(18):4354–4355, May 2001.

Anders O Magnusson, Johanna C Rotticci-Mulder, Alberto Santagostino, and Karl Hult. Creating space for large secondary alcohols by rational redesign of candida antarctica lipase b. *Chembiochem*, 6(6):1051–1056, Jun 2005.

Bola D Majekodunmi, Cesar A Lau-Cam, and Robert A Nash. Stability of benzoyl peroxide in aromatic ester-containing topical formulations. *Pharm Dev Technol*, 12(6):609–620, 2007.

Seiji Miura, Stefano Ferri, Wakako Tsugawa, Seungsu Kim, and Koji Sode. Development of fructosyl amine oxidase specific to fructosyl valine by site-directed mutagenesis. *Protein Eng Des Sel*, 21(4): 233–239, Apr 2008.

Niju Narayanan and C. Perry Chou. Alleviation of proteolytic sensitivity to enhance recombinant lipase production in escherichia coli. *Appl Environ Microbiol*, 75(16):5424–5427, Aug 2009.

Santosh Kumar Padhi, Despina J Bougioukou, and Jon D Stewart. Site-saturation mutagenesis of tryptophan 116 of saccharomyces pastorianus old yellow enzyme uncovers stereocomplementary variants. *J Am Chem Soc*, 131:3271–3280, 2009.

S. Patkar, J. Vind, E. Kelstrup, M. W. Christensen, A. Svendsen, K. Borch, and O. Kirk. Effect of mutations in candida antarctica b lipase. *Chem Phys Lipids*, 93(1-2):95–101, Jun 1998.

Francisco J Plou, M. Angeles Cruces, Manuel Ferrer, Gloria Fuentes, Eitel Pastor, Manuel Bernabé, Morten Christensen, Francisco Comelles, José L Parra, and Antonio Ballesteros. Enzymatic acylation of di- and trisaccharides with fatty acids: choosing the appropriate enzyme, support and solvent. *J Biotechnol*, 96(1):55–66, Jun 2002.

M. Rarey, B. Kramer, T. Lengauer, and G. Klebe. A fast flexible docking method using an incremental construction algorithm. *Journal of Molecular Biology*, 261(3):470–489, 1996.

Manfred T Reetz, Daniel Kahakeaw, and Joaquin Sanchis. Shedding light on the efficacy of laboratory evolution based on iterative saturation mutagenesis. *Mol Biosyst*, 5(2):115–122, Feb 2009.

Alexander Seifert, Sandra Vomund, Katrin Grohmann, Sebastian Kriening, Vlada B Urlacher, Sabine Laschat, and Jürgen Pleiss. Rational design of a minimal and highly enriched cyp102a1 mutant library with improved regio-, stereo- and chemoselectivity. *Chembiochem*, 10(5):853–861, Mar 2009.

Ling Song, Chakrapani Kalyanaraman, Alexander A Fedorov, Elena V Fedorov, Margaret E Glasner, Shoshana Brown, Heidi J Imker, Patricia C Babbitt, Steven C Almo, Matthew P Jacobson, and John A Gerlt. Prediction and assignment of function for a divergent n-succinyl amino acid racemase. *Nat Chem Biol*, 3(8): 486–491, Aug 2007.

Guo-Dong Su, Deng-Feng Huang, Shuang-Yan Han, Sui-Ping Zheng, and Ying Lin. Display of candida antarctica lipase b on pichia pastoris and its application to flavor ester synthesis. *Appl Microbiol Biotechnol*, 86(5):1493–1501, May 2010.

J. Uppenberg, M. T. Hansen, S. Patkar, and T. A. Jones. The sequence, crystal structure determination and refinement of two crystal forms of lipase b from candida antarctica. *Structure*, 2(4): 293–308, 1994.

J. Uppenberg, N. Ohrner, M. Norin, K. Hult, G. J. Kleywegt, S. Patkar, V. Waagen, T. Anthonsen, and T. A. Jones. Crystallographic and molecular-modeling studies of lipase b from candida antarctica reveal a stereospecificity pocket for secondary alcohols. *Biochemistry*, 34(51):16838–51, 1995.

Dao Feng Xiang, Peter Kolb, Alexander A Fedorov, Monika M Meier, Lena V Fedorov, T. Tinh Nguyen, Reinhard Sterner, Steven C Almo, Brian K Shoichet, and Frank M Raushel. Functional annotation and three-dimensional structure of dr0930 from deinococcus

radiodurans, a close relative of phosphotriesterase in the amidohydrolase superfamily. *Biochemistry*, 48(10):2237–2247, Mar 2009.

Ningyan Zhang, Wen-Chen Suen, William Windsor, Li Xiao, Vincent Madison, and Aleksey Zaks. Improving tolerance of candida antarctica lipase b towards irreversible thermal inactivation through directed evolution. *Protein Eng*, 16(8):599–605, Aug 2003.

Dunming Zhu, Yan Yang, Stephanie Majkowicz, Thoris Hsin-Yuan Pan, Katherine Kantardjieff, and Ling Hua. Inverting the enantioselectivity of a carbonyl reductase via substrate-enzyme docking-guided point mutation. *Org Lett*, 10(4):525–528, Feb 2008.

6.8 Figures

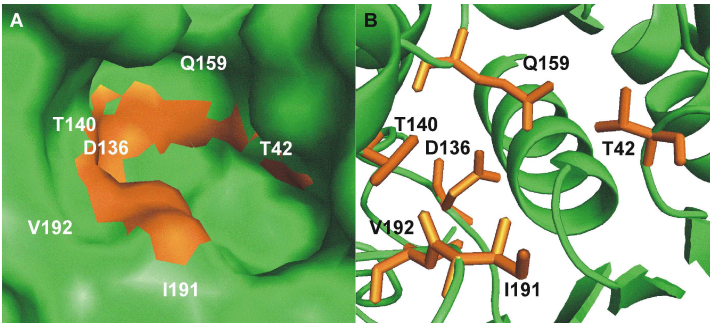

Abbildung 6.1: Binding pocket of Candida antarctica lipase B (green) with the hotspots colored in orange. A Surface representation of the pocket, B CALB shown as cartoon (green) and the hotspot amino acids as sticks (orange)

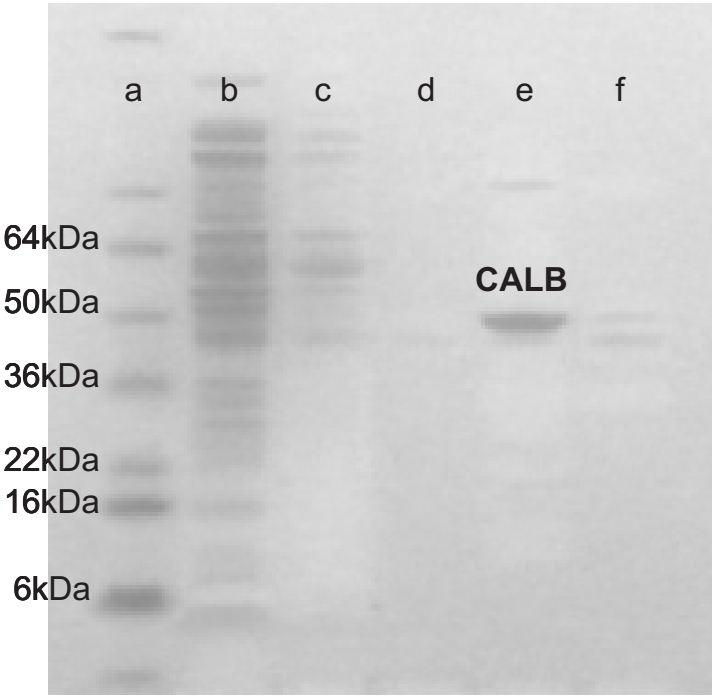

Abbildung 6.2: SDS-PAGE of the His-tagged Candida antarctica lipase B purification. a Marker, b Flow through, c Wash I, d Wash II, e Elute I, f Elute II. Elute I is 90% - 95% pure CALB and carries almost all enzymatic activity.

6.9 Tables

Tabelle 6.1: Identified hotspots, wild type amino acids, and amino acid exchanged for the combinatorial *in silico* library II

Position	Wild type amino acid	Exchanged by
134	Asp	Ala Ser Thr Asn
138	Thr	Ala Val Ser Asn Gln
157	Gln	Ala Val Leu Ile Ser Thr Asn
189	Ile	Ala Val Leu Thr
190	Val	Ala

Tabelle 6.2: Hydrolysis of bulky esters by *C. antarctica* lipase B wild type.

Substrate	Conversion rate [U/ml]
Methylcaprylate	15 000±700
Dimethylsuiccinate	11 150±500
Methyllaurate	3 650±100
Diethylmalate	2 400±350
Ethyl-L-lactate	280±51
Isononanoic acid ethyl ester (INA-EE)	10±0.9
Diethyltartrate	10±0.5
L-Lysine ethyl ester	1.8±1.3
Triethylcitrate	1.3±0.5
2-Ethyl hexanoic acid ethyl ester (EHA-EE)	0.8±0.04

Tabelle 6.3: *In silico* screening of the combinatorial enzyme variant library II by substrate-imprinted docking. The percentage values represent how frequently a respective amino acid occurs at the respective position in the top 600 docking results for the respective substrate and stereoisomer. The "Effect on activity" refers to experimentally determined wet mass activity (specific activity where available) for the respective single mutant enzyme in comparison to wild type CALB.

	(*R*)-3,5,5-Trimethyl-hexanoic acid ethyl ester	(*S*)-3,5,5-Trimethyl-hexanoic acid ethyl ester	(*R*)-2-Ethyl hexanoic acid ethyl ester	(*S*)-2-Ethyl hexanoic acid ethyl ester	Effect on activity
Position 134					
Asp	23%	17%	15%	13%	wild type
Ala	20%	21%	10%	20%	ND[a]
Asn	20%	23%	20%	24%	ND
Ser	17%	22%	14%	25%	increase
Thr	20%	17%	41%	18%	ND
Position 138					
Thr	24%	16%	19%	11%	wild type
Ala	17%	16%	14%	16%	decrease
Asn	11%	15%	15%	21%	ND
Gln	11%	22%	25%	18%	ND
Ser	20%	15%	10%	21%	increase
Val	17%	17%	17%	13%	ND
Position 157					
Gln	8%	4%	0%	0%	wild type
Ala	20%	28%	17%	24%	ND
Asn	1%	2%	44%	15%	increase
Ile	2%	1%	0%	0%	decrease
Leu	1%	5%	0%	0%	decrease
Ser	22%	30%	8%	25%	increase
Thr	23%	17%	15%	22%	increase
Val	23%	14%	15%	15%	ND
Position 189					
Ile	32%	20%	0%	0%	wild type
Ala	16%	42%	36%	49%	ND
Leu	14%	0%	0%	0%	no change
Thr	2%	2%	64%	51%	ND
Val	36%	36%	0%	0%	ND
Position 190					
Val	63%	46%	47%	39%	wild type
Ala	37%	54%	53%	61%	ND

[a] ND not determined

Tabelle 6.4: Wet mass activities of CALB variants

Variant	Hydrolysis of isononaoic acid ethyl ester (INA-EE) [U/g]	Hydrolysis of 2-Ethyl hexanoic acid ethyl ester (EHA-EE) [U/g]
CALB wild type	1.7	1.1
T138A	1.1	0.9
Q157L	1.1	0.9
Q157I	1.0	1.0
I189L	1.7	1.0
T138A, Q157L	1.5	0.9
Q157I, I189L	0.9	0.8
D134S	5.8	3.1
T138S	30.6	2.1
Q157N	2.9	1.1
Q157S	7.1	1.5
Q157T	8.6	0.8

Tabelle 6.5: Specific activities of His-tagged and purified *C. antarctica* lipase B wild type and the T138S variant.

	CALB wild type	T138S variant
Methylcaprylate	741 U/mg	427 U/mg
Isonanoic acid ethyl ester	3.1 U/mg	15.3 U/mg
2-ethyl hexanoic acid ethyl ester	0.5 U/mg	0.6 U/mg

Kapitel 7

Biocatalytic Synthesis of Polyesters from Sugar-Based building blocks using immobilized Candida antarctica lipase B

David I. Habeych[a,b], P Benjamin Juhl[c], Jürgen Pleiss[c], Gerrit Eggink[a,b], Diana Vanegas[d], Carmen G. Boeriu[b]

[a] Food and Bioprocess Engineering Group, Wageningen University and Research Centre, P.O. Box 8129, 6700 AA Wageningen, The

Netherlands
[b] Agrotechnology and Food Sciences Group, Division Biobased Products, Wageningen University & Research Centre, P.O. Box 17, 6700 AA, Wageningen, The Netherlands
[c] Institute of Technical Biochemistry, University of Stuttgart, Allmandring 31, 70569 Stuttgart, Germany
[d] Universidad Pontificia Bolivariana, Chemical Engineering Department, Circular 1 70-01, Medellín, Colombia

7.1 Abstract

Biocatalyzed synthesis of linear ester oligomers (LEOs) and cyclic ester oligomers (CEOs) from non-activated succinic acid (A) and dianhydro hexitols (B, DAH) in toluene based medium using *Candida antarctica* lipase B (CAL B), was studied. CAL B exhibited higher affinity for isomannide. These experimental results were corroborated by substrate-imprinted docking, indicating that the hydroxyl group oriented inwards the "V"-shaped plane of the DAHs (*endo*-hydroxyl bond) is preferred over the outwards oriented hydroxyl group (*exo*-hydroxyl) by CAL B. The maximum conversions were 88.2% and 93.7% for succinic acid and isomannide, respectively. The initial reaction rate for succinic acid (A) was 1.53 mole.(h.mg)$^{-1}$ and 0.68 for isomannide mole.(h.mg)$^{-1}$. MALDI-TOF detected products were a mixture of cyclic (32.2%) and linear ester oligomers (67.8%).

7.2 Introduction

The sugar based economy, together with technical developments is leading to the production of novel raw materials that could be ap-

plied for innovative oil-independent polymer synthesis. Di-anhydro hexitols (DAH): 1,4:3,6-dianhydro-D-glucitol or isosorbide; 1,4:3,6-dianhydro-D-mannitol or isomannide; and 1,4:3,6-dianhydro-L-iditol or isoidide are good examples of such raw materials, which are produced from the glucose of natural feedstock (i.e. starch). Currently isosorbide is the only bulk-produced and the least high-priced among the DAH. Nevertheless, the growing application of the DAH isomers has trigger efforts of industrial producers to implement the synthesis of all three isomers at reasonable prices (Fenouillot et al., 2010). The most significant differences among these DAHs are the two hydroxyl and proton groups that confer variations in spatial configuration, physical and chemical properties (see Figure 7.1). The bond orientation (chirality) in the DAHs is a fundamental parameter for modulation of polyesters properties based on them (Fenouillot et al., 2010, Kricheldorf, 1997, Noordover et al., 2006, 2007).

Polyesters derivative from DAH can be synthesized via polycondensation or ring opening polymerization. The latter has several advantages over the quality of the final product, but the limited availability of raw materials (cyclic ester oligomers - CEOs -) has reduced its application for synthesis of polyesters. Therefore, study of new synthetic routes to CEOs is of great interest. The condensation of aliphatic dicarboxylic acid dichloride with isosorbide using pyridin as catalyst and chloride acceptor under pseudo-high-diluted conditions have been studied (Kricheldorf et al., 2003). Interestingly, condensation of adipoyl dichloride led to formation of 100% CEOs with reaction yields above 96%. The authors used MALDI-TOF as a mean to estimate the relative occurrence of the different species (Kricheldorf et al., 2003, Williams et al., 1997). The occurrence of cyclic ester oligomers is not only a feature of halo-derivatives. The formation of CEOs have been also found during chemical synthe-

sis of linear poly(isosorbide succinate) (Noordover et al., 2006). The first known enzymatic-catalyzed synthesis of DAH was reported in 1993. Immobilized lipase (Lipozyme IM-20) and oleic acid in a tubular reactor in a solvent-free system was used. After 13 h of the operation of the system 95% of the initial hydroxyl derivatives were consumed8. linear ester oligomers (LEOs) and cyclic ester oligomers (CEOs) can be synthesized in presence of lipases and it has been proposed by Jacobson and Stockmayer that the chain - ring equilibrium (LEOs ¡=¿ CEOs) obeys a proposed statistical mechanics model in a thermodynamically controlled reaction (Jacobson and Stockmayer, 1950). In a kinetically controlled reaction and in the absence of side reactions, however, all the products will be cycles when conversions approaches 100% (Kricheldorf and Schwarz, 2003). Results in agreement with the Kricheldorf's theory have been found recently by other authors (Chatti et al., 2009). They found that condensation of 4, 4-difluorodiphenylsulfone with DAH in DMSO/toluene system at 140 °C and K_2CO_3 as catalyst, led to the formation of CEOs after 24 h and with a conversion above 95%. To the best of our knowledge nothing has been reported over *in vitro* enzymatic synthesis of ester oligomers based on DAH and non-activated di-acids. Therefore, the purpose of the present work was to explore the synthesis of CEOs and LEOs from sugar derivatives (DAHs) together with a non-activated di-acid (succinic acid) in the presence of a suitable biocatalyst as basis for synthesis of biodegradable polyesters. We illustrate and discuss the origin of reactivity differences found among the DAH isomers. The reaction conditions were improved by experimental design of the most relevant factors and further water removal from the reaction medium. Furthermore, substrate-imprinted docking was used to model the formation of the esters by the enzyme and to understand the molecular basis of observed CAL B specificity (Juhl et al., 2009).

7.3 Results and Discussion

7.3.1 Factors that influence the reaction of CAL B with DAHs and non-activated succinic acid

The reaction between non-activated succinic acid and the three DAHs was catalyzed by *C. antactica* lipase B, CAL B Novozym® 435. Initial experiments with isoidide (Scheme 1C) as substrate did not show positive outcome. Conversion with *Candida antarctica* lipase B were less than 5% and further experiments with other enzymes such as *Candida rugosa*, porcine pancreas, wheat germ, and *Thermomyces lanuginose* lipase showed almost no conversion (Data not shown). Isosorbide and Isomannide (Figure 7.1A and B) showed better results in the presence of CAL B (conversion higher than 15%). Furthermore, an experimental design was applied to find an improvement in the reaction when CAL B was the biocatalyst. Three variables were considered as the most relevant to find the best conditions of the esterification. We chose temperature, initial substrate concentration, and enzyme concentration as the main factors and substrate (DAH) consumption after 24 h as response variable to evaluate the performance of CAL B. The reaction was carried out in 20 mL tubes under the similar conditions described in the Experimental section. The formation of the ester bond, when using DAH and non-activated succinic acid as substrate, was confirmed by FT-IR. Initial succinic acid carbonyl (C=O) stretching at 1680 cm^{-1} shifts after reaction to 1710 cm^{-1} and 1737 cm^{-1} corresponding to carbonyl from the ester bonds formed at LEOs and CEOs respectively (Fig. 7.2). A new strong C-O stretching band at 1175 cm^{-1} confirms also the formation

of esters after 24 h of enzymatic reaction. Effects of the initial substrate and enzyme concentration on consumption of the chosen DAH at 50 and 65 °C after 24 h reaction are shown in Fig. 7.3 and Fig. 7.4 respectively. Isomannide showed higher conversions than isosorbide at 50 °C (Fig. 7.3) and 65 °C (Fig. 7.4). At the levels evaluated in this experimental design initial enzyme concentration did not have a significant effect, while temperature and initial substrate concentration did. The highest isomannide conversion was 47.4% when the initial isomannide concentration was 10 mM and the enzyme concentration was 4 g/L. This is an excellent conversion value if it is taken into account that water produced during the reaction was not removed and isomannide is a sterically demanding substrate. The highest isosorbide conversion was 16.3% when the initial isosorbide concentration was 10 mM and the enzyme concentration was 4 g/L. In both cases there were no significant differences in conversions at 50 and 65 °C. However, the reaction conducted at 80 °C exhibited significant reduction of substrate conversion for both, isomannide and isosorbide. The isomannide conversion at 80 °C ranged from 3.7% to 24.2%, while the isosorbide conversion at the same temperature ranged from 0% to 4% (Data not shown). This finding demonstrates the limitations of CAL B thermal stability in the reaction system used, which has also been demonstrated by others (Feng et al., 1999a,b). The ratio of isomannide consumption divided by succinic acid consumption varied from 1.5 to 2.0 (data not shown), while the same ratio for the case of isosorbide was always less than 1 (data not shown). This fact indicates that CAL B must have a preference for *endo*-hydroxyl groups (attached to C5 in isosorbide and C2 and C5 in isomanide, Figure 7.1) rather than for *exo*-hydroxyl group (attached to C2 in isosorbide and C2 and C5 in isoidide, Figure 7.1). This will explain also the low reactivity of isoidide found in earlier expe-

riments (two *exo*-hydroxyl groups, Figure 7.1). The steric configuration of the DAHs has been also suggested as the main factor of the reactivity for polyester synthesis (Fenouillot et al., 2010, Kricheldorf, 1997, Noordover et al., 2006, 2007). In contrast, an opposite effect to what we found with CAL B has been observed when chemical catalyst was used. For example, Noordover *et al* prepared succinic acid and isosorbide polyesters in the presence of titanium (IV) *n*-butoxide in the melt. They found a limitation in the final molecular weight, reaching a value up to 2500 $g.mol^{-1}$. Such limitation in the chemical catalyzed polymerization of isosorbide is thought due to the low reactivity of the *endo*-hydroxyl group (Figure 7.1A, position 5). Such low reactivity is owing to the formation of hydrogen bonds in the *endo* position, while the *exo* hydroxyl group is exposed and ready to react. In addition, they found by different means that the ratio in the terminal hydroxyl groups (*endo/exo*) was 6:4. This confirmed the hypothesis that *exo*-hydroxyl group in isosorbide is significantly more reactive than its *endo* counterpart under chemical catalyzed conditions. This feature of the *exo*-hydroxyl group has been observed by other authors not only with chemical catalyst but also with biological degradation of polyesters (Fenouillot et al., 2010).

A more thorough analysis of the DAH structure could explain our results. The rings are *cis* orientated in V shape and inclined to one another in an angle of approximately 120°. The orientation of the *endo*-hydroxyl group or (R)-configurated hydroxyl group attached to C5 in isosorbide, present as well in C2 and C5 in isomannide, and absent in isoidide (Figure 7.1A, B, and C) must play a key role on the enzymatic stereospecificity of CAL B. Moreover, due that the *endo*-hydroxyl group ((R)-configurated) generates an intramolecular hydrogen bond with the adjacent oxygen atom in the cyclic ether ring, the carbon attached to it must be more positively charged than the carbon atta-

ched to an *exo*-hydroxyl group ((S)-configurated) (Noordover et al., 2006, Wright and Brandner, 1964). Hence, the positive charge in the C5 of isosorbide (C2 and C5 in isomannide) must favor the nucleophilic attack that leads to the formation of the ester bond. This explanation is in agreement with the enzymatic esterification mechanism described by Zaks and Klibanov (1985) (Zaks and Klibanov, 1985). Moreover, Boulif *et al* (2010) (Boulifi et al., 2010) reported recently the monosubstitution of isosorbide with ricinoleic acid with 95% conversion of the *endo*-hydroxyl group ((R)-configurated) and less than 1% conversion of the *exo*-hydroxyl group ((S)-configurated) in the presence of CAL B at 63 °C. Reetz *et al* (Reetz and Schimossek, 1996) use CAL B for N-acylation of (R,S)-phenylethylamine with ethyl acetate. CAL B only reacted with the (R)-phenylethylamine reaching 77% conversion and synthesizing the (R)-configurated amide enantiomerically pure (ee=99%). Likewise, Overmeyer *et al* used CAL B for transesterification of (R,S)-1-phenylethanol with vinylacetate in supercritical CO_2 environment. They reached a conversion above 45% of the (R)-enatiomer in few hours with enantiomeric excess above 99%. Interestingly, Hilker and collaborators (2006) (Hilker et al., 2006) used CAL B-catalyzed dynamic kinetic resolution with *in situ* chemical racemization to produce chiral polyesters. They started from the racemic mixture of 2,2-dimethyl-1,4-benzenedimethanol as the di-ol and dimethyl adipate as the di-acyl donor. The reaction led to the formation of polyesters containing only the (R)-enatiomer with a conversion of 92% after 70 h. Hence, these evidences together with our results suggest that *Candida antarctica* lipase B (CAL B) exhibits higher reactivity for (R)-configurated groups than for (S)-configurated ones.

These results are also in agreement with ours, where *exo*-hydroxyl group exhibits less reactivity than *endo*-hydroxyl group. Therefore,

the reactivity of DAH is featured by the type of catalyst used and also due to the mechanism involved in the reaction. As a result, the esterification in the presence of CAL B biocatalyst shows a potential in the synthesis of novel biobased polyesters beyond the possibilities of the chemical catalyst.

The analysis of variance (Table 7.1) for non-activated succinic acid consumed in the esterification with isosorbide in the presence of CAL B showed that the temperature is the only important variable. A similar analysis of variance (Table 7.2) for non-activated succinic acid in the esterification with isomannide showed that not only the reaction temperature, but also the initial substrate concentration are important variables. These results, in both cases, were also corroborated by Probability test (P-test) and Fisher-Snedecor test (F-test). No interactions among the studied variables were noticed in the range evaluated. In order to elucidate the stereopreference found for CAL B on the DAHs a substrate-imprinted docking analysis was carried out.

7.3.2 Docking of the substrates into the active site

To rationalize the molecular basis for the observed CAL B preference of isomannide over isosorbide and over isoidide, substrate-imprinted docking was used to dock butyric acid esters of the DAHs into five CAL B structures (Figure 7.5). Productive docking solutions were always found if the ester bond is formed with an *endo*-hydroxyl group, as found in isomannide and one of the isosorbide hydroxyl groups (C5, Figure 7.1 and 7.5), while no productive docking solutions could be found for esters with *exo*-hydroxyl groups, as found in isoidide and one of the isosorbide hydroxyl groups (C2 Figure 7.1 and 7.5). This is in agreement with the experimentally observed pre-

ference of CAL B for the *endo*-hydroxyl group. An analysis of the substrate placements in the active site generated by docking shows, that esterification at the *endo*-hydroxyl group leads to a reaction intermediate that fits into the CAL B alcohol pocket without clashes of the two rings with the enzyme (Fig. 7.6C) and forms the hydrogen bonds required for stabilization of the transition state (Schulz et al., 2000) (Fig. 7.6A).
The analysis of the non-productive substrate placements achieved for C (Fig. 7.6B) shows that C can be fitted into the binding pocket of CAL B, but not in a pose that forms the required hydrogen bonds for catalysis (Fig. 7.6B, 7.6D). C is placed in the CAL B binding pocket in a way that puts the alcohol oxygen of the ester beyond hydrogen bond distance to the NH-group of H224 (3.5 Å) and additionally points the free electron pairs of the alcohol oxygen away from H224, making a hydrogen bond even unlikely if the distance was smaller. Therefore, the CAL B preference of isomannide over isosorbide over isoidide is a direct result of the preference for *endo*-hydroxyl groups, which is due to the transition state of esters with *exo*-hydroxyl groups not forming all the required hydrogen bonds for catalysis. If these bonds are not formed properly, catalysis is either very slow or does not occur at all.

7.3.3 Product characterization

Since CAL B exhibits more reactivity toward isomannide than toward its homologous isomers, we continued our work on characterization of the reaction products only for reactions with isomannide. The thermodynamic controlled synthesis of isomannide succinate oligomers from non-activated succinic acid (A) and isomannide (B) was

conducted at a 50 mL scale as described in Experimental section. No conversion occurred in the absence of the enzyme (data not shown). Samples were taken every hour up to 24 hours of reaction and a MALDI-TOF analysis was performed. The synthesized products are shown in Figure 7.7 and summarized in Table 7.3. Oligomers up to octadecamers were detected in the reaction mixture. Low molecular weight CEOs and LEOs were produced during all the time course analysis. Oligomers with high molecular weight were detected only at longer reaction times (Table 7.3, CCEO_7, CEO_8, CEO_9, AB_7, AB_8, AB_9, AA_7, AA_8, BB_7, and BB_8). The permanent presence of low molecular weight oligomers is an indication that small molecules become the pool of building blocks that lead to formation of larger ones. In addition, the formation and consumption of larger oligomers during the time course analysis suggests that the existent equilibrium involves the simultaneous hydrolysis of the latter. Products formed at 24 h are shown in Fig. 7.8. The total cyclic ester oligomers are 32.2% of the total detected products. Cyclic tetramer (CEO_2, 1.03 mM) and decamer (CEO_5, 0.5 mM) are the 20% and 10% of the total synthesized products, respectively. Hydroxy-carbonyl terminal oligomers are 21% of the total oligomers produced, where the tetramer (AB_2, 0.56 mM) counts for 11% of the total detected products. Similarly, di-hydroxy terminal oligomers are 30.2% of the total oligomers produced, and the pentamer member (BB_2, 0.59 mM) makes up 11% of all products formed. Di-carbonyl terminal LEOs (AA_n) count for 16.6% of the products with higher concentration also for the pentamer (AA_2, 0.35 mM). Interestingly, the concentration of the first member cyclic ester oligomer (CEO_1, 0.03 mM) was one of the lowest in the homologous series. This is an indication that the capacity of CAL B to carry out the cyclization process is hindered by steric effect coming from the spatial configuration of the di-anhydro hexitol

ring and the conformational requirement imposed by CAL B binding pocket. This finding is in agreement with results found in different enzymatic-catalyzed reactions (Berkane et al., 1997, Lalot and Marechal, 2001). The authors found in ring-chain equilibrium studies that formation of cyclic dimers from non-aromatic compounds occurs at lower molar cyclization equilibrium constant than formation of cyclic tetramers. Moreover, they found that when an aryl di-ol was esterified with an alkyl di-acid, no formation of dimers was found. While DAHs are not aromatic compounds, the rigidity of the rings and the configuration of the hydroxyl groups results in a similar behaviour, as the formation of cyclic dimers is limited.

7.3.4 Reaction yield optimization

Additional experiments were conducted to increase the conversion yield in the biocatalyzed reaction of non-activated succinic acid and isomannide. We mentioned before that water produced in the reaction plays a key role in the biosynthesis of esters, shifting the equilibrium from ester formation towards ester hydrolysis with consequent reaction yield diminishing. Hence, we decided to perform a reaction improvement in terms of water control in the presence of molecular sieves. Enzymatic reactions were carried out in 200 mL round bottom glass reactor with a working volume of 50 mL. The reactor was equipped with a flat paddle impeller at 200 rpm and the same conditions mentioned at the Experimental section. Molecular sieves (5 g) dried for at least 24h at 120 °C were added at the beginning of the reaction. Samples were withdrawn from the reactor every hour during the first 8 h and a final sample was taken at 24 h of reaction for further analysis. Two control reactions were set up. The first con-

trol included both substrates without CAL B and without molecular sieves, while the second control contained substrate with molecular sieves and without CAL B. In both control reactions we did not observe quantifiable substrate conversion or product synthesis. 43.7% of initial succinic acid was converted together with 47.4% isomannide when the water produced was not removed (no molecular sieves added, see Fig. 7.4). Conversion of 88.2% of succinic acid and 93.7% of isomannide were observed after 24 h enzymatic reaction in the presence of molecular sieves (Fig. 7.9). In the latter case the initial reaction rate in mol of substrate consumed per hour and mg of CAL B were 1.53 mol.(h.mg)$^{-1}$ and 0.68 mol.(h.mg)$^{-1}$ for succinic acid and isomannide, respectively. The ratio between the initial reaction rates is higher than 2, indicating that the limiting step in the isomannide esterification is not the formation of the acyl-enzyme complex, but the synthesis of the final ester bond (deacylation). It is worthwhile to mention that the conversions in the presence of molecular sieves were twice the value reached than with no molecular sieves added. That is an indication of the tremendous effect of the water removal within the system.

7.4 Experimental

7.4.1 Materials

Lipases and solvents were obtained from Sigma - Aldrich. Novozym® 435 was provided by Novozyme (The Netherlands). Chirazyme® enzymatic kit was purchased from Roche Molecular Biochemicals. The rest of enzymes were purchased from Fluka. All solvents were equilibrated in 4 Å molecular sieves (Sigma-Aldrich) for at least 24 hours

before use. *Trans*-2-[3-(4-tert-butylphenyl)-2-methyl-2-propenylidene]-malononitrile (DCTB) and potassium trifluoroacetate were obtained from Sigma - Aldrich. All solvents were purchased from Merck. Isosorbide, isomannide and isoidide were gently provided by Roquette Frères S. A. (France).

7.4.2 Enzymatic Esterification

Non-activated succinic acid (50 mol) of and DAH (50 mol) were dissolved in 5 mL of toluene and *tert*-butanol (70:30, %wt), in 10 mL glass tubes. The reaction was performed at 65 °C and 24 h, unless otherwise stated. The reaction was initiated by addition of the lipase (20 mg). Reactions were stopped by rapid cooling on ice bath. Subsequently, the enzyme was separated by filtration and the supernatant was collected and stored for further analysis.

7.4.3 Matrix Assisted Laser desorption/Ionisation Time-of-Flight Mass spectrometry (MALDI-TOF MS)

Samples were analysed by MALDI-TOF spectrometry in an Ultraflex workstation (Bruker Daltonics, Bremen, Germany) in the positive mode and outfitted with a laser (337 nm). The matrix, *trans*-2-[3-(4-tert-butylphenyl)-2-methyl-2-propenylidene]malononitrile (DCTB), was prepared at 40 mg mL^{-1} and the dopant, potassium trifluoroacetate at 5 mg mL^{-1}, both in tetrahydrofuran. Sample, matrix and dopant solutions were mixed (10:5:5) and 0.3 μL of mixture was placed in MALDI-TOF plates (5 mm in diameter) and dried in air. Maltod-

extrins and polyethylene glycol (600, 1000, and 2000 Da) werer used for calibration. Ions were accelerated with a 25 kV voltage (delayed extraction time of 200 ns). At least three spectrums with a total of 200 shoots per spot were collected using the lowest possible laser intensity that led to a good quality spectrum (Laine et al., 2001, Mezoul et al., 1995). Within each MALDI-TOF spectrum, the intensity of all signals obtained by the potassium-adduct per chemical specie were added up and subsequently the relative contribution of each species was calculated. It was assumed that there are no significant differences between the response factors of the molecules (Laine et al., 2001, Williams et al., 1997).

7.4.4 Substrate analysis

Unreacted substrate was followed with a Waters 1525 binary HPLC pump implemented with a refraction index detector and an ionic exchange column Alltech OA-100 (sulfonated polystyrene-divinylbenzene) eluted with 3 mM sulfuric acid aqueous solution (0.4 mL min^{-1}). Samples and standards solutions (1000 μL) were dried under reduced pressure. The remained solid was suspended in Milli-Q water (100 μL) and warm up in a bath (70-80 °C). Insoluble solid was separated by centrifugation and the supernatant (10 μL) was injected in the column and data were recorded. Similar procedure was done for standards of succinic acid and DAHs.

7.4.5 Oligomer analysis

The reaction samples (10 μL) were injected in a HPLC Waters 2690, outfitted with a PDA detector and a 250 x 3 mm ODS-2 Inter-

sil (Varian Inc.) column at 40 °C. Elution of oligomers was done in a gradient flow, phase A: acetonitrile (Merck) and phase B: water. Both mobile phases were supplemented with trifluoroacetic acid (0.1% v/v). Gradient elution started at 5:95 (v/v) A/B and ended at 95:5 (v/v) A/B with a flow rate of 0.5 mL min^{-1} and the sample injection volume of 10 μL.

7.4.6 Fourier Transform Infrared analysis (FT-IR)

Between 2 to 10 μL liquid samples were dried on a zinc selenite microplate of a Varian 1000, Scimitar TM series FT-IR system. A total of sixty-four infrared scans were recorded and averaged with a resolution of 2 cm^{-1}.

7.4.7 Experimental design and statistical analysis

A three level full factorial design with additional three center points was used. The three factors chosen were temperature (A, 50-80 °C), initial concentration of non-activated succinic acid and DAH (B, 10-100 mM), and enzyme concentration (C, 4-10 mg ml^{-1}). Response variable was the substrate conversion after 24 h. The substrate ratio was 1:1 (mol). All the reactions were conducted as independent duplicates and each duplicate was done twice. Statistic results were analyzed by Statgraphics Plus 5.0 at 95% confidence interval ($\alpha = 0.05$). Analysis of variance (ANOVA) was used to evaluate significance of factors, interactions among factors, and presence of autocorrelation in the residuals of the regression analysis. Hence, Probability test (P-test), Fisher-Snedecor test (F-test) and Durbin-Watson test

(DW-test) were applied.

7.4.8 Substrate-imprinted docking

The docking procedure consists of five steps. First, a tetrahedral reaction intermediate is covalently docked into an enzyme structure. Second, the best scoring substrate placement is used to construct an enzyme-substrate complex. Third, the geometry of the enzyme-substrate complex is optimized by energy minimization (200 steps steepest descent, 800 steps conjugate gradient). Fourth, the substrate is removed from the optimized complex resulting in an optimized enzyme structure. Fifth, the optimized enzyme structure is used for a second covalent docking of the same substrate. Final docking results are analyzed for geometric filter criteria necessary for the catalytic mechanism, their docking score, and the maximum overlap volume used during the docking.
The CAL B X-ray structures 1LBS, 1LBT (Uppenberg et al., 1995), 1TCA, 1TCB, and 1TCC (Uppenberg et al., 1994) were used in this study. The substrates were modeled as tetrahedral reaction intermediates of isosorbide, isomannide, and isoidide esters with butyric acid. Because of symmetry, one ester was modeled for isomannide and isoidide, while it was necessary to model two different esters for isosorbide, as the two stereocenters that feature the hydroxyl groups are not identical. Butyric acid and succinic acid are both readily converted by CAL B, and therefore butyric acid was used during the docking as a less complex model substrate.

7.5 Conclusions

The reaction esterification of di-anhydro hexitols and non-activated succinic acid in a toluene based medium in the presence of *Candida antarctica* lipase B (CAL B) was studied. Temperature and initial substrate concentration were the most important factors over the reaction yields. Interestingly, CAL B had preference for isomannide over isosorbide and over isoidide; such an enantiopreference was not only found experimentally, but also supported by substrate-imprinted docking analysis. This fact set isomannide as a preferable substrate among the DAHs for further enzymatic studies due to difficulties in reactivity found with chemical catalyst. Finally, a characterization of products in the thermodynamically controlled system was conducted with the positive outcome that the studied system promotes the formation of CEOs as well as the synthesis of di-hydroxy terminal LEOs. Therefore, esterification of non-activated succinic acid and isomannide in the presence of CAL B shows potential in the biosynthesis of novel building blocks (CEOs and LEOs) from ready-to-use biobased raw materials and is advantageous in respect to chemical ones due to the specificity already discussed.

7.6 Acknowledgement

This work was financially supported by the Dutch Ministry Economic Affairs and the B-Basic partner organization through B-Basic, a public-private Dutch Organization for Science and Research (NWO) - Advanced Chemical Technology for Sustainability (ACTS) program, the IP-project Sustainable Microbial and Biocatalytic Production of Advanced Functional Materials (Bioproduction/NMP-2-CT-2007-026515) funded by the European Commission, and AFSG (Wagenin-

gen University). We also want to thank Professor Robert Duchateau and Mrs. Carin Dietz from Technical University of Eindhoven for their valuable help in MALDI-TOF analysis.

References

Christophe Berkane, Gilles Mezoul, Thierry Lalot, Maryvonne Brigodiot, and Ernest Marechal. Lipase-catalyzed polyester synthesis in organic medium. study of ring-chain equilibrium. *Macromolecules*, 30(25):7729–7734, 1997.

Noureddin El Boulifi, José Aracil, and Mercedes Martínez. Lipase-catalyzed synthesis of isosorbide monoricinoleate: Process optimization by response surface methodology. *Bioresource Technology*, 101(22):8520–8525, 2010.

Saber Chatti, Moez A. Hani, Kirstin Bornhorst, and Hans R. Kricheldorf. Poly(ether sulfone) of isosorbide, isomannide and isoidide. *High Performance Polymers*, 21(1):105–118, 2009.

Yakai Feng, Jens Knüfermann, Doris Klee, and Hartwig Höcker. Enzyme-catalyzed ring-opening polymerization of 3(¡i¿s¡/i¿)-isopropylmorpholine-2,5-dione. *Macromolecular Rapid Communications*, 20(2):88–90, 1999a.

Yakai Feng, Jens Knüfermann, Doris Klee, and Hartwig Höcker. Lipase-catalyzed ring-opening polymerization of 3(¡i¿s¡/i¿)-isopropylmorpholine-2,5-dione. *Macromolecular Chemistry and Physics*, 200(6):1506–1514, 1999b.

F. Fenouillot, A. Rousseau, G. Colomines, R. Saint-Loup, and J.-P. Pascault. Polymers from renewable 1,4:3,6-dianhydrohexitols (isosorbide, isomannide and isoidide): A review. *Progress in Polymer Science*, 35(5):578–622, 2010.

Iris Hilker, Gouher Rabani, Gerard K. M. Verzijl, Anja R. A. Palmans, and Andreas Heise. Chiral polyesters by dynamic kinetic resolution polymerization13. *Angewandte Chemie International Edition*, 45(13):2130–2132, 2006.

Homer Jacobson and Walter H. Stockmayer. Intramolecular reaction in polycondensations. i. the theory of linear systems. *J. Chem. Phys.*, 18(12):1600–1606, 1950.

Benjamin Juhl, Peter Trodler, Sadhna Tyagi, and Jurgen Pleiss. Modelling substrate specificity and enantioselectivity for lipases and esterases by substrate-imprinted docking. *BMC Struct. Biol.*, 9 (1):39, 2009.

Hans R. Kricheldorf. "sugar diols" as building blocks of polycondensates. *Polymer Reviews*, 37(4):599–631, 1997.

Hans R. Kricheldorf, Saber Chatti, Gert Schwarz, and Ralf-P. Krüger. Macrocycles 27: Cyclic aliphatic polyesters of isosorbide. *Journal of Polymer Science Part A: Polymer Chemistry*, 41 (21):3414–3424, 2003.

HR Kricheldorf and G Schwarz. Cyclic polymers by kinetically controlled step-growth polymerization. *Macromolecular Rapid Communications*, 24(5-6):359–381, 2003.

Olli Laine, Heidi Österholm, Maaria Seläntaus, Hannele Järvinen, and Pirjo Vainiotalo. Determination of cyclic polyester oligo-

mers by gel permeation chromatography and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry*, 15(20):1931–1935, 2001.

Thierry Lalot and Ernest Marechal. Enzyme-catalyzed polyester synthesis. *International Journal of Polymeric Materials*, 50(3):267–286, 2001.

Gilles Mezoul, Thierry Lalot, Maryvonne Brigodiot, and Ernest Maréchal. Enzyme-catalyzed syntheses of poly(1,6-hexanediyl maleate) and poly(1,6-hexanediyl fumarate) in organic medium. *Macromolecular Rapid Communications*, 16(8):613–620, 1995.

Bart A. J. Noordover, Viola G. van Staalduinen, Robbert Duchateau, Cor E. Koning, van Benthem, Manon Mak, Andreas Heise, August E. Frissen, and Jacco van Haveren. Co- and terpolyesters based on isosorbide and succinic acid for coating applications: Synthesis and characterization. *Biomacromolecules*, 7(12):3406–3416, 2006.

Bart A. J. Noordover, Robbert Duchateau, Rolf A. T. M. van Benthem, Weihua Ming, and Cor E. Koning. Enhancing the functionality of biobased polyester coating resins through modification with citric acid. *Biomacromolecules*, 8(12):3860–3870, 2007.

M. T. Reetz and K. Schimossek. Lipase-catalyzed dynamic kinetic resolution of chiral amines: Use of palladium as the racemization catalyst. *Chimia*, 50(12):668–669, 1996.

T. Schulz, J. Pleiss, and R. D. Schmid. Stereoselectivity of pseudomonas cepacia lipase toward secondary alcohols: a quantitative model. *Protein Sci*, 9(6):1053–1062, Jun 2000.

J. Uppenberg, M. T. Hansen, S. Patkar, and T. A. Jones. The sequence, crystal structure determination and refinement of two crystal forms of lipase b from candida antarctica. *Structure*, 2(4): 293–308, 1994.

J. Uppenberg, N. Ohrner, M. Norin, K. Hult, G. J. Kleywegt, S. Patkar, V. Waagen, T. Anthonsen, and T. A. Jones. Crystallographic and molecular-modeling studies of lipase b from candida antarctica reveal a stereospecificity pocket for secondary alcohols. *Biochemistry*, 34(51):16838–51, 1995.

John B. Williams, Arkady I. Gusev, and David M. Hercules. Characterization of polyesters by matrix-assisted laser desorption ionization mass spectrometry. *Macromolecules*, 30(13):3781–3787, 1997.

L. W. Wright and J. D. Brandner. Catalytic isomerization of polyhydric alcohols.1 ii. the isomerization of isosorbide to isomannide and isoidide. *The Journal of Organic Chemistry*, 29(10):2979–2982, 1964.

A Zaks and A M Klibanov. Enzyme-catalyzed processes in organic solvents. *Proceedings of the National Academy of Sciences of the United States of America*, 82(10):3192–3196, 1985.

7.7 Figures

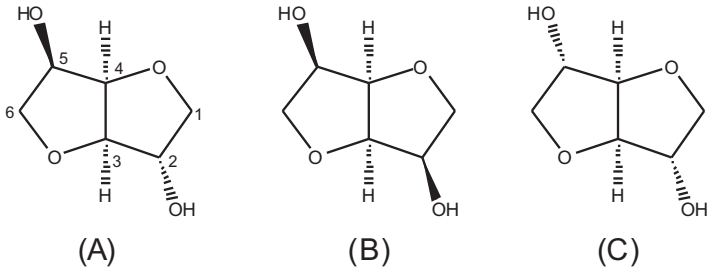

Abbildung 7.1: (A). (3R,3aR,6S,6aR)-Hexahydrofuro[3,2-b] furan-3,6-diol; 1,4:3,6-dianhydro-D-sorbitol, or isosorbide; (B). (3R,3aR,6R,6aR)-Hexahydrofuro[3,2-b] furan-3,6-diol; 1,4:3,6-dianhydro-D-mannitol, or isomannide; (C). (3S,3aR,6S,6aR)-Hexahydrofuro[3,2-b] furan-3,6-diol; 1,4:3,6-dianhydro-L-iditol, or isoidide.

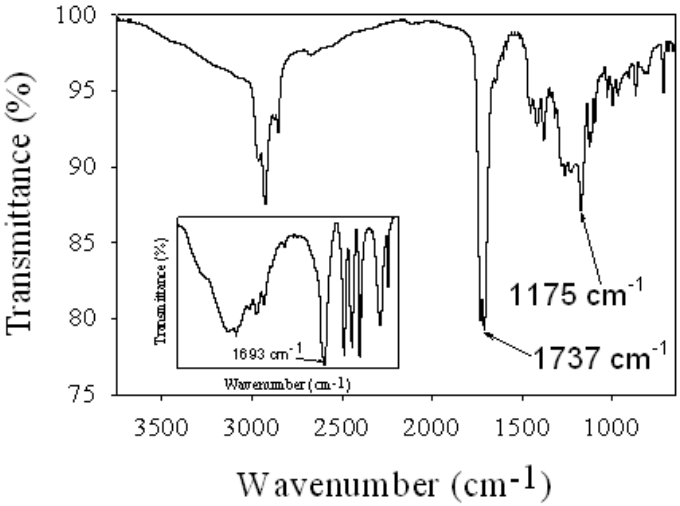

Abbildung 7.2: FT-IR spectrum of non-activated succinic acid (A) and isomannide (B) reaction mixture in the presence of Candida antarctica lipase CAL B after 24 hours (Main figure). Succinic acid spectrum (Inserted figure).

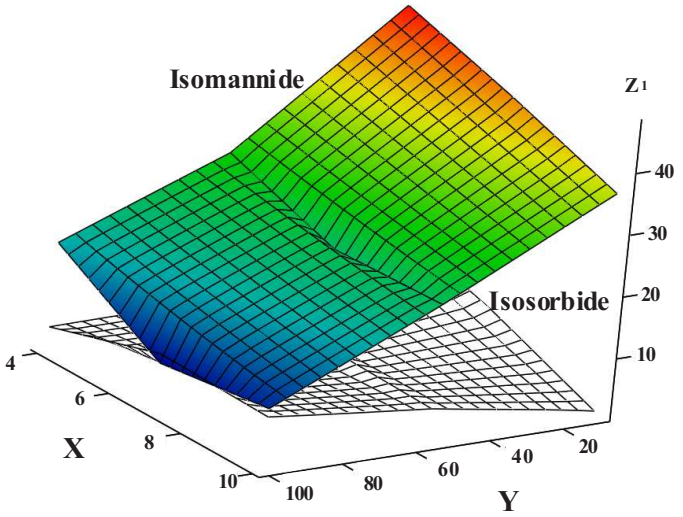

Abbildung 7.3: Response surface for isomannide and isosorbide at 50°C. X: Enzyme concentration (g/L); Y: Initial substrate concentration (g/L); Z: DAH conversion (%).

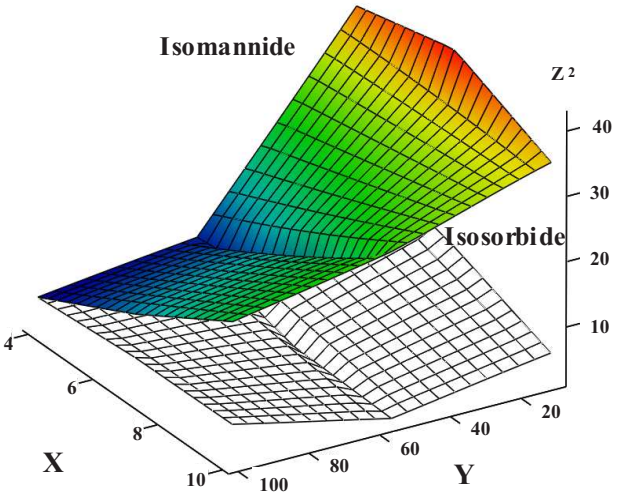

Abbildung 7.4: Response surface for isomannide and isosorbide at 65°C. X: Enzyme concentration (g/L); Y: Initial substrate concentration (g/L); Z: DAH conversion (%).

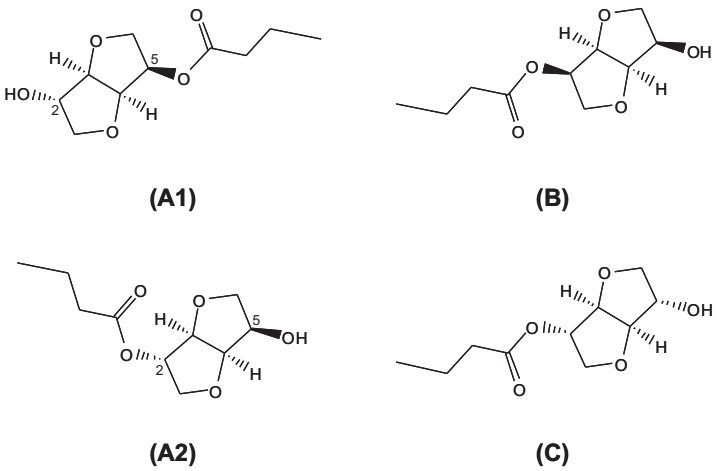

Abbildung 7.5: (AE1) (3R,3aR,6S,6aR)-6-hydroxyhexahydrofuro[3,2-b]furan-3-yl butyrate, (AE2) (3S,3aR,6R,6aR)-6-hydroxyhexahydrofuro[3,2-b]furan-3-yl butyrate, (BE) (3R,3aR-,6R,6aR)-6-hydroxyhexahydrofuro[3,2-b]furan-3-yl butyrate, (CE) (3S,3aR,6S,6aR)-6-hydroxyhexahydrofuro[3,2-b]furan-3-yl butyrate.

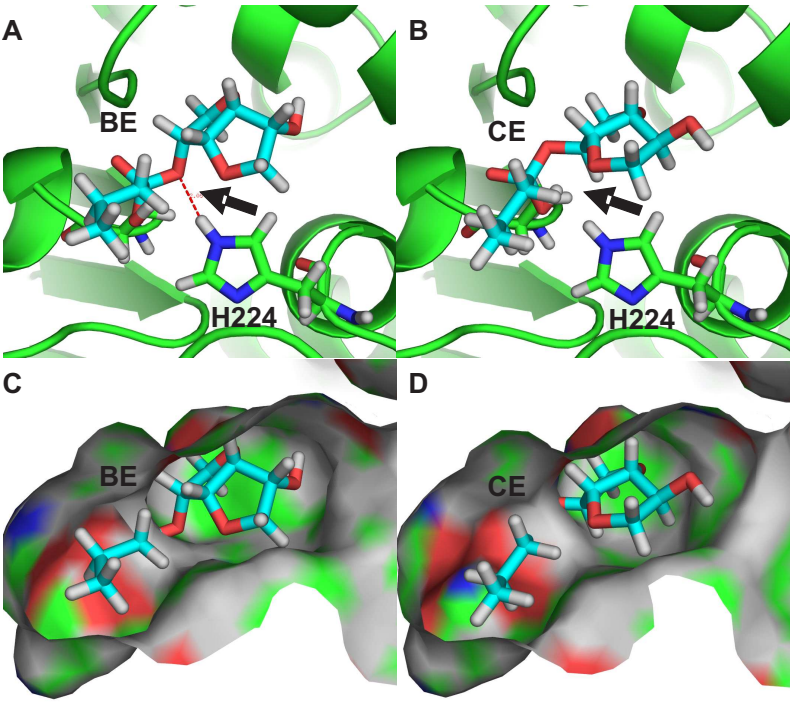

Abbildung 7.6: (A, C) (3R,3aR,6R,6aR)-6-hydroxyhexahydrofuro[3,2-b]furan-3-yl butyrate (BE) docked into CAL B (PDB-ID: 1LBS). H224 forms a hydrogen bond with the alcohol oxygen of the ester (red dashed line, black arrow). (B,D) (3S,3aR,6S,6aR)-6-hydroxyhexahydrofuro[3,2-b]furan-3-yl butyrate (CE) docked into CAL B (PDB-ID: 1LBS). The necessary hydrogen bond between H224 and the alcohol oxygen of the ester cannot be formed (black arrow).

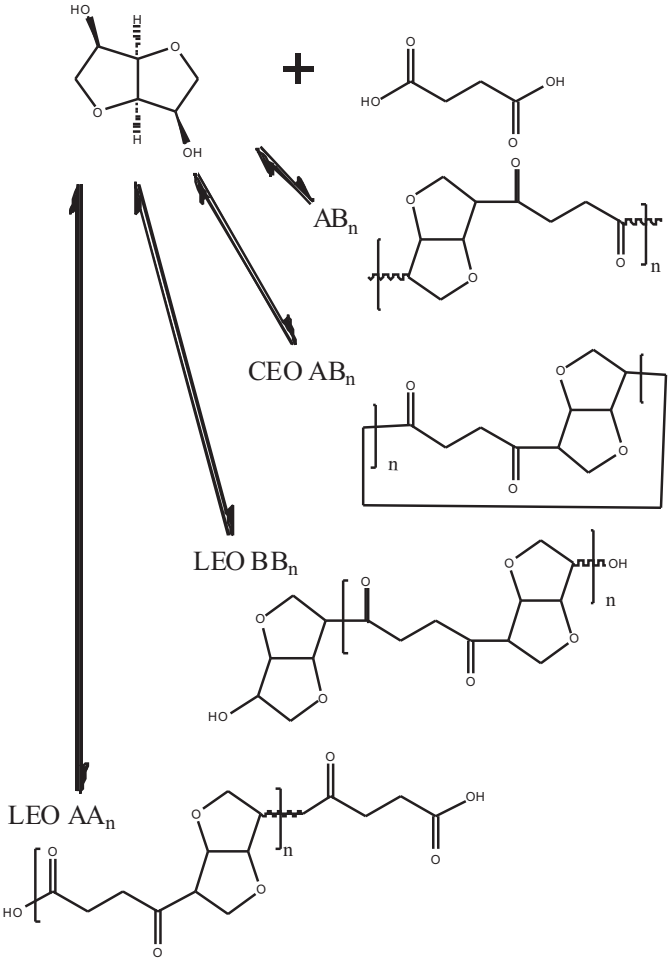

Abbildung 7.7: Isomannide succinate oligomers from non-activated succinic acid (A) and isomannide (B) in the presence of Candida antarctica CAL B lipase. Example LEO BB2 stands for the di-hyrdoxyl terminal oligomer BABAB.

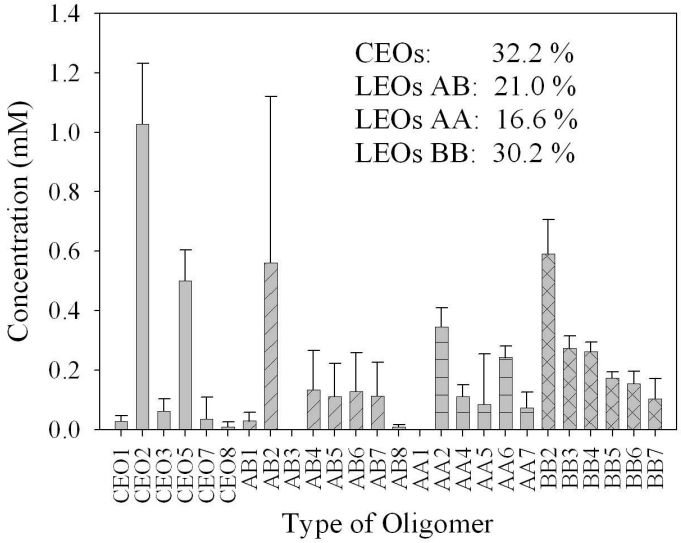

Abbildung 7.8: Ester oligomers formed after 24 h enzymatic conversion of non-activated succinic acid and isomannide in the presence of CAL B. (CEO) Cyclic ester oligomers; (AB) linear ester oligomers hydroxy-carbonyl terminal; (AA) linear ester oligomers di-carbonyl terminal; (BB) linear ester oligomers di-hydroxy terminal. Error bars indicates the standard deviation between duplicates.

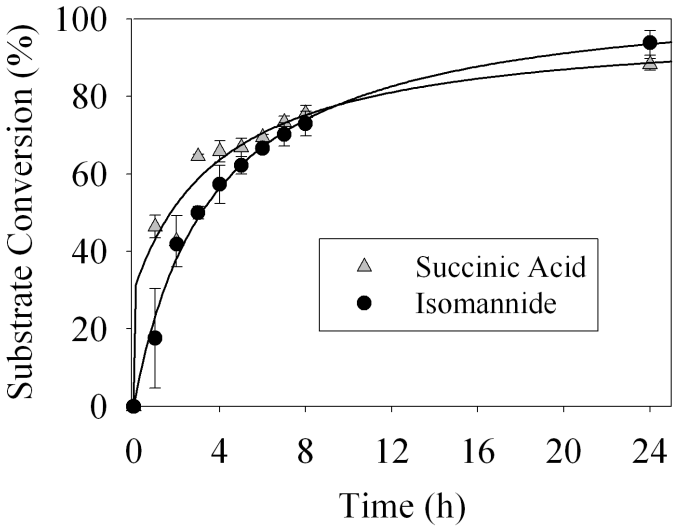

Abbildung 7.9: Non-activated succinic acid and isomannide conversion in the presence of CAL B and molecular sieves. Isomannide (•), Succinic acid (△). Error bars represent the standard deviation between duplicates.

7.8 Tables

Tabelle 7.1: ANOVA analysis for reaction of isosorbide and non-activated succinic acid in the presence of *C. antarctica* lipase CAL B. DF: degree of freedom; F-ratio: Fisher-Snedecor test; P-value: Probability test.

Isosorbide	**Sum of squares**	**DF**	**Mean square**	**F-ratio**	**P-value**
A: Temperature	252.8	1	252.8	13.69	0.0005
B: Substrate Conc.	27.2	1	27.2	1.45	0.232
C: Enzyme Conc.	48	1	48	2.58	0.114
AB	0.9	1	0.88	0.05	0.828
AC	1.9	1	1.85	0.1	0.754
BC	187.4	1	187.4	10.07	0.003
Blocks	3.9	1	3.9	0.21	0.649
Total Error	967.5	52	18.6		

Tabelle 7.2: ANOVA analysis for reaction of isomannide and non-activated succinic acid in the presence of *C. antarctica* lipase B. DF: degree of freedom; F-ratio: Fisher-Snedecor test; P-value: Probability test.

Isomanide	**Sum of squares**	**DF**	**Mean square**	**F-ratio**	**P-value**
A: Temperature	1284.2	1	1284.2	10.09	0.003
B: Substrate Conc.	4554.5	1	4554.5	35.78	0
C: Enzyme Conc.	0.6	1	0.6	0	0.945
AB	426.3	1	426.3	3.35	0.073
AC	47.2	1	47.2	0.37	0.545
BC	188.69	1	188.69	1.48	0.229
Blocks	70.36	1	70.36	0.55	0.461
Total Error	13190.2	52	127.3		

Tabelle 7.3: MALDI-TOF time course analysis of oligo(isomannide succinate) biosynthesized. Example. AA_n states for $(AB)_nA$, the linear (n + 1)-emer.

		Cyclic ester oligomers								
Entry	Reactiontime	CEO_1	CEO_2	CEO_3	CEO_4	CEO_5	CEO_6	CEO_7	CEO_8	CEO_9
	(h)	228.23*	456.5	684.7	912.9	1141.2	1369.4	1597.6	1825.8	2054.1
1	1	+	+	+		+	+		+	
2	2	+	+	+		+	+	+	+	
3	3	+	+	+	+	+		+	+	
4	4	+	+	+	+	+	+			+
5	5	+	+	+		+	+	+	+	+
6	6	+	+	+	+	+	+	+	+	
7	7	+	+	+	+	+	+	+		
8	8	+	+	+		+	+	+	+	+
9	24	+	+	+	+	+		+	+	
		Linear ester oligomers (AB)								
Entry	Reactiontime	AB_1	AB_2	AB_3	AB_4	AB_5	AB_6	AB_7	AB_8	AB_9
	(h)	246.2*	474.5	702.7	930.9	1159.2	1387.4	1615.6	1843.8	2072.1
10	1	+	+	+	+		+	+		
11	2	+	+	+	+	+	+	+		
12	3	+	+	+	+	+	+			
13	4	+	+	+	+	+	+			
14	5	+	+	+	+	+	+			
15	6	+	+	+	+	+	+			
16	7	+	+	+	+	+	+			
17	8	+	+	+	+	+	+			+
18	24	+	+	+	+	+	+	+	+	
		Linear ester oligomers (AA)								
Entry	Reactiontime	AA_1	AA_2	AA_3	AA_4	AA_5	AA_6	AA_7	AA_8	AA_9
	(h)	346.3*	574.6	802.8	1031	1259.2	1487.5	1715.7	1943.9	2172.2
19	1	+			+	+	+		+	
20	2	+	+		+	+	+	+		
21	3	+	+	+	+	+	+	+	+	
22	4	+	+			+	+	+	+	
23	5	+	+	+	+	+	+	+	+	
24	6	+	+	+	+	+	+	+		
25	7	+	+		+	+	+	+	+	
26	8	+	+		+	+	+	+	+	
27	24	+	+		+	+	+	+		
		Linear ester oligomers (BB)								
Entry	Reactiontime	BB_1	BB_2	BB_3	BB_4	BB_5	BB_6	BB_7	BB_8	BB_9
	(h)	374.4*	602.6	830.8	1059.1	1287.3	1515.5	1743.8	1972	2200.2
28	1		+	+		+	+	+	+	
29	2	+	+	+	+	+	+	+	+	
30	3	+	+	+	+	+	+	+	+	
31	4	+	+	+	+	+	+	+		
32	5		+	+	+	+	+	+		
33	6		+	+	+	+	+	+	+	
34	7	+	+	+	+	+	+	+	+	
35	8		+	+	+	+	+	+		
36	24		+	+	+	+	+	+		

* Molecular weight in Da

Kapitel 8

Structural classification by the Lipase Engineering Database: a case study of Candida antarctica lipase A

Michael Widmann[1], **P Benjamin Juhl**[1], **Jürgen Pleiss**[1*]

[1] Institute of Technical Biochemistry, University of Stuttgart, Allmandring 31, 70569 Stuttgart, Germany

8.1 Abstract

Background: The Lipase Engineering Database (LED) integrates information on sequence, structure and function of lipases, esterases and related proteins with the α/β hydrolase fold. A new superfamily for Candida antarctica lipase A (CALA) was introduced including the recently published crystal structure of CALA. Since CALA has a highly divergent sequence in comparison to other α/β hydrolases, the Lipase Engineering Database was used to classify CALA in the frame of the already established classification system. This involved the comparison of CALA to similar structures as well as sequence-based comparisons against the content of the LED.

Results: The new release 3.0 (December 2009) of the Lipase Engineering Database contains 24783 sequence entries for 18585 proteins as well as 656 experimentally determined protein structures, including the structure of CALA. In comparison to the previous release [1] with 4322 protein and 167 structure entries this update represents a significant increase in data volume. By comparing CALA to representative structures from all superfamilies, a structure from the deacetylase superfamily was found to be most similar to the structure of CALA. While the α/β hydrolase fold is conserved in both proteins, the major difference is found in the cap region. Sequence alignments between both proteins show a sequence similarity of only 15%. A multisequence alignment of both protein families was used to create hidden Markov models for the cap region of CALA and showed that the cap region of CALA is unique among all other proteins of the α/β hydrolase fold. By specifically comparing the substrate binding pocket of CALA to other binding pockets of α/β hydrolases, the binding pocket of Candida rugosa lipase was identified as being

highly similar. This similarity also applied to the lid of Candida rugosa lipase in comparison to the potential lid of CALA.

Conclusions: TThe LED serves as a valuable tool for the systematic analysis of single proteins or protein families. The updated release 3.0 was used for the evaluation of α/β hydrolases. The HTML version of the database with new features is available at http://www.led.uni-stuttgart.de/ and provides sequences, structures and a set of analysis tools including phylogenetic trees and HMM profiles.

8.2 Background

Lipases (triacylglycerol hydrolases E.C. 3.1.1.3) are a versatile group of enzymes which catalyze the hydrolysis or synthesis of a broad range of water insoluble esters. They belong to the class of α/β-hydrolases which also contains esterases, acetylcholinesterases, cutinases, carboxylesterases and epoxide hydrolases. Despite their high diversity in sequence and function, the α/β-hydrolases share a common architecture, the α/β- hydrolase fold (Ollis et al., 1992) and conserved active site signatures, the GxSxG and GxDxG motifs (Barth et al., 2004, Pleiss et al., 2000). Two conserved features found in all α/β-hydrolases are the active site, consisting of the catalytic triad of S-D(E)-H, and the oxyanion hole. Depending on the amino acids involved in forming the oxyanion hole, the enzymes can be classified into three classes, the GGGX-, GX-, and the Y-class (Pleiss et al., 2000). The Lipase Engineering Database (LED) (Fischer and Pleiss, 2003) is a resource of fully and consistently annotated superfamilies and homologous families of α/β hydrolases including multisequence alignments of all families. The curation and annotation process for the LED is supported by DWARF (Fischer et al., 2006), an inhouse

data warehouse system for protein families. The LED is accessible by a web interface at http://www.led.uni-stuttgart.de/. It can be browsed on the level of families, organisms, or structures, and BLAST searches can be performed against all sequence entries.
Prominent members of the α/β hydrolases are the two lipases from Candida antarctica. Lipase B is a versatile and well characterized biocatalyst in many organic syntheses and biotransformations (Gotor-Fernandez et al., 2006, Orrenius et al., 1995a,b) and shows a low sequence similarity to other α/β hydrolases. The second lipase from Candida antarctica, lipase A (CALA), shows a number of unique biocatalytic properties among hydrolases, e.g. high thermostability and stability at acidic pH ranges and the acceptance of tertiary and sterically hindered alcohols (Domínguez de María et al., 2005). CALA also has a low sequence similarity to other members of the α/β hydrolase fold including lipase B. Therefore it was not included in previous versions of the LED. Only after its structure was recently determined (Ericsson et al., 2008), a detailed analysis of its structure identified CALA unambiguously as a member of the α/β hydrolase family. However, in this structure the active site is not accessible to a substrate, therefore the molecular details of substrate binding or the existence of a possible lid are still elusive.

8.3 Results

8.3.1 Database content and layout

Release 3.0 of the Lipase Engineering Database (LED) contains 18582 proteins with 24782 sequence and 656 structure entries of which about 14000 protein and 540 structure entries are new. Six new ho-

mologous families and one new superfamily (the "Candida antarctica lipase A like" superfamily) have been added to the LED in the update process. Seed sequences for the new "Candida antarctica lipase A like" superfamily (LED identifier: abH38.01) included the sequence from the resolved crystal structure and three sequences of homologous lipases from other organisms which showed high sequence similarity to Candida antarctica lipase A (CALA) (Ericsson et al., 2008) (Table 8.1). The four largest superfamilies in release 3.0 contain 50% of all proteins in the LED: The "Cytosolic Hydrolases" superfamily (LED identifier: abh08) with 3188 proteins, containing epoxide hydrolases and haloalkane dehalogenases, the "Carboxylesterases" superfamily (LED identifier: abh01) with 2998 proteins, containing a wide range of carboxylesterases, such as acetylcholine esterases and bile salt activated lipases, the "Moraxella lipase 2 like" superfamily (LED identifier: abh04) with 1781 proteins containing mainly lipases and carboxylesterases, and the "Microsomal Hydrolases" superfamily (LED identifier: abh09) with 1336 proteins, containing microsomal epoxide hydrolases and peptidases. The "Cytosolic Hydrolases" and "Microsomal Hydrolases" superfamilies (abh08 and abh09) belong to the GX-class of α/β hydrolases, the "Carboxylesterases" and "Moraxella lipase 2 like" superfamilies (abh01 and abh04) belong to the GGGX-class of α/β hydrolases.

8.3.2 Candida antarctica lipase A protein family

The "Candida antarctica lipase A like" superfamily contains one crystal structure and 39 sequences, assigned to 32 proteins. They were grouped into four homologous families based on sequence similarity (Figure 8.1): The "Candida antarctica lipase A like" homologous

family consisting of Lipase A from C. antarctica, the "Malassezia lipase like" homologous family consisting entirely of lipases and esterases from Malassezia globosa or Malassezia furfur, the "Candida albicans lipase like" homologous family consisting of various isoforms of the secretory lipase from Candida albicans, and the "Aspergillus lipase like" homologous family consisting mainly of hypothetical or putative lipases, mostly from Aspergillus. All 32 proteins are from organisms belonging to the subkingdom Dikarya of the kingdom Fungi. 12 proteins are classified as either lipases or esterases in GenBank (Benson et al., 2008) while 20 proteins are classified as putative or hypothetical. The only structure entry in this superfamily is from the recently resolved crystal structure of CALA (Ericsson et al., 2008). Based on the structure of the oxyanion hole, CALA can be classified as a Y-class lipase, and Tyr 93 was identified as the oxyanion hole forming amino acid. A structural comparison with other structures from the LED identified a structure from the deacetylase superfamily (LED identifier: abH26) as most closely related. A detailed structural alignment of CALA (PDB: 2VEO) with the structure of the Bacillus subtilis deacetylase (PDB: 1L7A) from the deacetylase superfamily showed a superimposition of the common α/β hydrolase fold including the catalytic triad (2VEO: Ser184, Asp334, His366; 1L7A: Ser181, Asp269, His298), despite having a low overall sequence identity of only 15%. Structural differences between the two structures are found in the cap and the C-terminal region. The cap region, located between â-strands 6 and 7, often confers substrate specificity or additional functions to the enzyme (Wei et al., 1999). In the case of CALA, the cap region is involved in forming the tunnel like binding site for the acyl moiety (Ericsson et al., 2008). For B. subtilis deacetylase, the cap region partially shields the active site from the solvent (Vincent et al., 2003). The cap region of CALA con-

sists of six á-helices, while the cap region of B. subtilis deacetylase consists of only four á-helices (Figure 8.2). Three á- helices in both proteins are found at identical positions. CALA shows an insert of three additional á-helices after the first two conserved á-helices and is missing the last á-helix of the cap present in B. subtilis deacetylase (Figure 8.2). In order to identify residues in the cap region which are conserved between the "Candida antarctica lipase A like" and "Deacetylases" superfamilies, a multisequence alignment of each family was performed. The two family alignments were aligned using a structural alignment of the two protein structures (Figure 8.3). The alignment demonstrated that despite the high structural similarity, there are no conserved residues in the cap region of both protein families. Two hidden Markov models of the three inserted á-helices of CALA and the á-helices shared by both proteins were created and used to search against all other protein families of the LED. No sequences with a significant similarity were found in the entire database, demonstrating that the sequence of the cap region of the "Candida antarctica lipase A like" superfamily is unique.

In comparison to the B. subtilis deacetylase, CALA has two additional â.strands (9 and 10) in the C-terminal region (Figure 8.2). They are positioned directly above the active site and prevent a direct access of the substrate to the active site. We assume that the â.strands 9 and 10 perform a lid like function for CALA since movement of the two â-strands would allow substrate access to the active site of CALA from a similar direction as for the B. subtilisdeacetylases (Vincent et al., 2003). A comparison of the substrate binding sites of both proteins showed that the alcohol binding site is similar in both proteins and provides ample space for alcohol moieties of substrates (Figure 8.4). Therefore, both proteins are expected to accept a variety of bulky alcohols. The binding sites for the acyl moieties

are highly different. CALA has a long, tunnel like binding site, while the B. subtilis deacetylase has a small cavity which is part of a cleft on the protein surface. Therefore, the acyl moieties of the substrates are expected to differ significantly between both enzymes. CALA is expected to accept mediumt to long chain fatty acids, while the B. subtilis deacetylase is limited to short-chain acyl moieties. Thus, despite the overall similarity between CALA and the B. subtilis deacetylase, the acyl binding site is fundamentally different.
However, the binding site of CALA shows surprising similarity to another lipase, Candida rugosa lipase (CRL). For CRL, two different structural confirmations have been resolved, an open conformation (1CRL) (Grochulski et al., 1993), and a closed conformation (1TRH) (Grochulski et al., 1994) where the lid of CRL is blocking the substrate access to the active site. CRL has a cap region between â-strands 6 and 7, consisting of four á-helices (Figure 8.2). The substrate binding site of CRL consists of a long tunnel for the acyl moiety of the substrate and provides ample space for the alcohol moiety of the substrate (Figure 8.4). Despite having a lower overall structure similarity to CALA than the B. subtilis deacetylase, the binding sites of CALA and CRL are highly similar (Figure 8.4). Both provide space for large, bulky alcohol moieties of the substrate and have a tunnel like binding site for the acyl moiety. Both proteins posses a lid which covers the active site and prevents direct access to the substrate binding site in its closed state. The lid of CRL lipase is formed by a á-helix between â-strands 1 and 2 and is located in the N-terminal region while the putative lid in CALA is formed by the two C-terminal â-strands 9 and 10 (Figure 8.2).

8.4 Discussion

The LED contains annotated and systematically classified protein families of α/β hydrolases. It has been shown to be a useful tool for the systematic analysis of protein families. Previous work employed the LED and BLAST in order to identify novel enzymes belonging to the α/β hydrolase fold (Kim et al., 2009, Lämmle et al., 2007). A model for the prediction of protein solubility was developed and refined by performing a comprehensive analysis of the protein families of the LED (Koschorreck et al., 2005). A further study involved the systematic analysis of protein families of the LED in regard to the distribution and conservation of functionally relevant rare codons (Widmann et al., 2008).

Since the first release of the LED (Pleiss et al., 2000), more than 14000 new α/β hydrolases became available and were integrated in the release 3.0. As a case study for the utility of the highly enriched and annotated database, the newly introduced superfamily of CALA was analysed and compared to other protein structures in the LED. The goal was to characterise the sequence and structure of CALA in comparison to other α/β hydrolases despite its low sequence similarity and to understand the molecular basis of substrate recognition. While CALA shows structural similarity to the deacetylase family, the substrate specificity of both enzymes differs, which is consistent with the differences observed in the substrate binding sites of both proteins. In contrast, the lipase from C. rugosa (CRL), which shows a lower overall structural similarity to CALA, is remarkably similar in regard to the substrate binding site. The structural similarities and differences are in accordance with experimentally observed substrate specificities of the three enzymes. All three proteins have a spacious alcohol binding site. The B. subtilis deacetylase accepts a wide varie-

ty of bulky substrates like cephalosporin C and xylose (Vincent et al., 2003). CALA and CRL also accept bulky substrates, ranging from primary alcohols to sterically hindered secondary alcohols and even tertiary alcohols (Akoh et al., 2004, Kirk and Christensen, 2002). The tunnel like binding site of CALA allows the enzyme to accept esters of long chain fatty acids (Kirk and Christensen, 2002, Pfeffer et al., 2006). The similar tunnel like acyl binding site of CRL also accepts fatty acids up to a chain length of 18 (Pfeffer et al., 2006). In contrast, the small acyl binding site of the B. subtilis deacetylase is unable to accept large acyl groups and is restricted towards acetyl moieties (Vincent et al., 2003). Experimentally, CALA and CRL have been shown to display interfacial activation (Grochulski et al., 1993, Martinelle et al., 1995). While a lid in CRL has beenlocalized and the open and closed form of CRL has been crystallized (Grochulski et al., 1993, 1994), the lid function of the â-strands 9-10 in CALA remains to be experimentally verified. However, the similarities to CRL suggest a substrate access involving the movement of â-strands 9-10.

8.5 Conclusions

The analysis of the newly introduced protein family of Candida antarctica lipase A demonstrates the strength of our database approach by providing a large set of protein families which share a common protein fold despite an overall low sequence similarity. By combining both, structural and sequential information of a large number of proteins a thorough analysis and classification of proteins of interest is made possible.

8.6 Availability and Requirements

The Lipase Engineering Database (LED) is online accessible at http://www.led.uni-stuttgart.de/. All information on families of sequence and structure data, as well as alignments, phylogenetic trees, and family-specific profiles can be accessed by manual download.

8.7 Methods

8.7.1 Structural comparisons and alignments

Comparison of structures were carried out using DALI (Holm and Sander, 1995). The structure of CALA was compared against 28 representative structures from all superfamilies. To identify the most closely related superfamilies, only structures which could be aligned to more than 50% of the residues of CALA were considered. Structural alignments of proteins were performed by STAMP (Russell and Barton, 1992).

For superfamilies which share a close structural relationship but a low overall sequence identity, a two step strategy was used in order to obtain a more significant multisequence alignment. First, a multisequence alignment for each of the two superfamilies was carried out separately. Then a structural alignment, between reference structures from each protein family was performed using STAMP (Russell and Barton, 1992). The multisequence alignments where then aligned against the structure from their respective protein family.

8.7.2 Sequence analysis

Multisequence alignments for all protein families were generated using ClustalW (Thompson et al., 1994) with a gap opening and extension penalties of 10 and 0.2, respectively. Hidden Markov models were created using HMMER (Eddy, 1998).

8.7.3 Database model

The implemented data model is based on Firebird (Fir) and is based on the previously published (Fischer et al., 2006) data model (Supplementary material 8.5). Protein families are organised on the level of homologous families and superfamilies based on their sequence similarity. The database is updated by an automated Perl (PER) script. It performs a BLAST (Altschul et al., 1990) search against the current version of the non-redundant sequence database an NCBI (Benson et al., 2008) for each sequence entry with an E-value cut-off of 10^{-50}. Crystal structure information referring to new sequence entries is updated as well. New sequence and structure entries are assigned to homologous families and superfamilies based on sequence similarity. New families which consisted of only one putative protein entry where not included. Annotation information of residues is either taken directly from the according GenBank entry or is transferred to new sequences using the DWARF graphical user interface. Annotation information is then transferred to the newly integrated sequences.

8.8 Authors' contribution

MW performed all analyses regarding CALA and drafted the manuscript. MW and BJ performed the update of the database and the manual curation process. JP supervised the project and finalized the manuscript.

8.9 Acknowledgements

We acknowledge the valuable contribution of Robert Radloff for help in the annotation process and of Florian Wagner for the programming of the dynamic user interface. The work was carried out in the framework of the IP-project "Sustainable Microbial and Biocatalytic Production of Advanced Functional Materials" (BIOPRODUCTION / NMP-2-CT-2007- 026515) funded by the European Commission.

References

URL http://sourceforge.net/projects/firebird.

URL http://www.perl.org/.

Casimir C Akoh, Guan-Chiun Lee, and Jei-Fu Shaw. Protein engineering and applications of candida rugosa lipase isoforms. *Lipids*, 39(6):513–526, Jun 2004.

S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. Basic local alignment search tool. *J Mol Biol*, 215(3):403–410, Oct 1990.

Sandra Barth, Markus Fischer, Rolf D Schmid, and Jürgen Pleiss. The database of epoxide hydrolases and haloalkane dehalogenases: one structure, many functions. *Bioinformatics*, 20(16):2845–2847, Nov 2004.

Dennis A. Benson, Ilene Karsch-Mizrachi, David J. Lipman, James Ostell, and David L. Wheeler. Genbank. *Nucl. Acids Res.*, 36 (suppl1):D25–30, 2008.

Pablo Domínguez de María, Chiara Carboni-Oerlemans, Bernard Tuin, Gerrald Bargeman, Ab van der Meer, and Robert van Gemert. Biotechnological applications of candida antarctica lipase a: State-of-the-art. *Journal of Molecular Catalysis B: Enzymatic*, 37(1-6): 36–46, December 2005. ISSN 1381-1177.

S. R. Eddy. Profile hidden markov models. *Bioinformatics*, 14(9): 755–763, 1998.

Daniel J Ericsson, Alex Kasrayan, Patrik Johansson, Terese Bergfors, Anders G Sandström, Jan-E. Bäckvall, and Sherry L Mowbray. X-ray structure of candida antarctica lipase a shows a novel lid structure and a likely mode of interfacial activation. *J Mol Biol*, 376(1):109–119, Feb 2008.

Markus Fischer and Jürgen Pleiss. The lipase engineering database: a navigation and analysis tool for protein families. *Nucleic Acids Res*, 31(1):319–321, Jan 2003.

Markus Fischer, Quan K Thai, Melanie Grieb, and Jürgen Pleiss. Dwarf–a data warehouse system for analyzing protein families. *BMC Bioinformatics*, 7:495, 2006.

V. Gotor-Fernandez, E. Busto, and V. Gotor. Candida antarctica lipase b: An ideal biocatalyst for the preparation of nitrogenated

organic compounds. *Adv. Synth. Catal.*, 348(7-8):797–812, May 2006.

P. Grochulski, Y. Li, J. D. Schrag, F. Bouthillier, P. Smith, D. Harrison, B. Rubin, and M. Cygler. Insights into interfacial activation from an open structure of candida rugosa lipase. *J Biol Chem*, 268 (17):12843–12847, Jun 1993.

P. Grochulski, Y. Li, J. D. Schrag, and M. Cygler. Two conformational states of candida rugosa lipase. *Protein Sci*, 3(1):82–91, Jan 1994.

L. Holm and C. Sander. Dali: a network tool for protein structure comparison. *Trends Biochem Sci*, 20(11):478–480, Nov 1995.

Eun-Young Kim, Ki-Hoon Oh, Mi-Hwa Lee, Chul-Hyung Kang, Tae-Kwang Oh, and Jung-Hoon Yoon. Novel cold-adapted alkaline lipase from an intertidal flat metagenome and proposal for a new family of bacterial lipases. *Appl Environ Microbiol*, 75(1):257–260, Jan 2009.

Ole Kirk and Morten WÃ$\frac{1}{4}$rtz Christensen. Lipases from candida antarctica: Unique biocatalysts from a unique origin. *Organic Process Research & Development*, 6(4):446–451, July 2002. ISSN 1083-6160.

Markus Koschorreck, Markus Fischer, Sandra Barth, and Jürgen Pleiss. How to find soluble proteins: a comprehensive analysis of alpha/beta hydrolases for recombinant expression in e. coli. *BMC Genomics*, 6(1):49, 2005.

Katrin Lämmle, Hubert Zipper, Michael Breuer, Bernhard Hauer, Christiane Buta, Herwig Brunner, and Steffen Rupp. Identifica-

tion of novel enzymes with different hydrolytic activities by metagenome expression cloning. *J Biotechnol*, 127(4):575–592, Jan 2007.

Mats Martinelle, Mats Holmquist, and Karl Hult. On the interfacial activation of candida antarctica lipase a and b as compared with humicola lanuginosa lipase. *Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism*, 1258(3):272–276, October 1995. ISSN 0005-2760.

D. L. Ollis, E. Cheah, M. Cygler, B. Dijkstra, F. Frolow, S. M. Franken, M. Harel, S. J. Remington, I. Silman, and J. Schrag. The alpha/beta hydrolase fold. *Protein Eng*, 5(3):197–211, Apr 1992.

Christian Orrenius, Niklas Öhrner, Didier Rotticci, Anders Mattson, Karl Hult, and Torbjörn Norin. Candida antarctica lipase b catalysed kinetic resolutions: Substrate structure requirements for the preparation of enantiomerically enriched secondary alcanols. *Tetrahedron: Asymmetry*, 6(5):1217–1220, May 1995a. ISSN 0957-4166.

Christian Orrenius, Torbjörn Norin, Karl Hult, and Giacomo Carrea. The candida antarctica lipase b catalysed kinetic resolution of seudenol in non-aqueous media of controlled water activity. *Tetrahedron: Asymmetry*, 6(12):3023–3030, December 1995b. ISSN 0957-4166.

Jan Pfeffer, Sven Richter, Jens Nieveler, Carl-Erik Hansen, Rachid Bel Rhlid, Rolf D Schmid, and Monika Rusnak. High yield expression of lipase a from candida antarctica in the methylotrophic yeast pichia pastoris and its purification and characterisation. *Appl Microbiol Biotechnol*, 72(5):931–938, Oct 2006.

J. Pleiss, M. Fischer, M. Peiker, C. Thiele, and R. D. Schmid. Lipase engineering database - understanding and exploiting sequence-structure-function relationships. *Journal Of Molecular Catalysis B-Enzymatic*, 10(5):491–508, October 2000.

R. B. Russell and G. J. Barton. Multiple protein sequence alignment from tertiary structure comparison: assignment of global and residue confidence levels. *Proteins*, 14(2):309–323, Oct 1992.

J. D. Thompson, D. G. Higgins, and T. J. Gibson. Clustal w: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*, 22(22):4673–4680, Nov 1994.

Florence Vincent, Simon J Charnock, Koen H G Verschueren, Johan P Turkenburg, David J Scott, Wendy A Offen, Shirley Roberts, Gavin Pell, Harry J Gilbert, Gideon J Davies, and James A Brannigan. Multifunctional xylooligosaccharide/cephalosporin c deacetylase revealed by the hexameric structure of the bacillus subtilis enzyme at 1.9a resolution. *J Mol Biol*, 330(3):593–606, Jul 2003.

Y. Wei, J. A. Contreras, P. Sheffield, T. Osterlund, U. Derewenda, R. E. Kneusel, U. Matern, C. Holm, and Z. S. Derewenda. Crystal structure of brefeldin a esterase, a bacterial homolog of the mammalian hormone-sensitive lipase. *Nat Struct Biol*, 6(4):340–345, Apr 1999.

Michael Widmann, Marie Clairo, Jürgen Dippon, and Jürgen Pleiss. Analysis of the distribution of functionally relevant rare codons. *BMC Genomics*, 9:207, 2008.

8.10 Figures

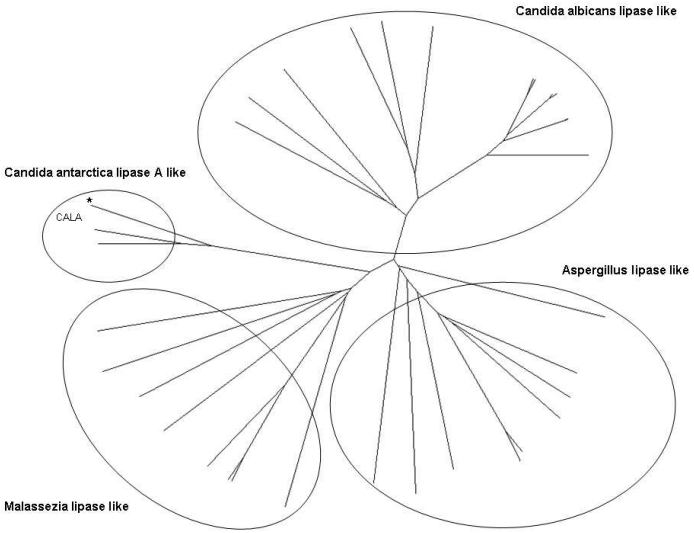

Abbildung 8.1: Phylogenetic tree of the "Candida antarctica lipase A like" superfamily: The superfamily consists of 4 homologous families based on sequence similarity. The sequence of C. antarctica lipase A is indicated.

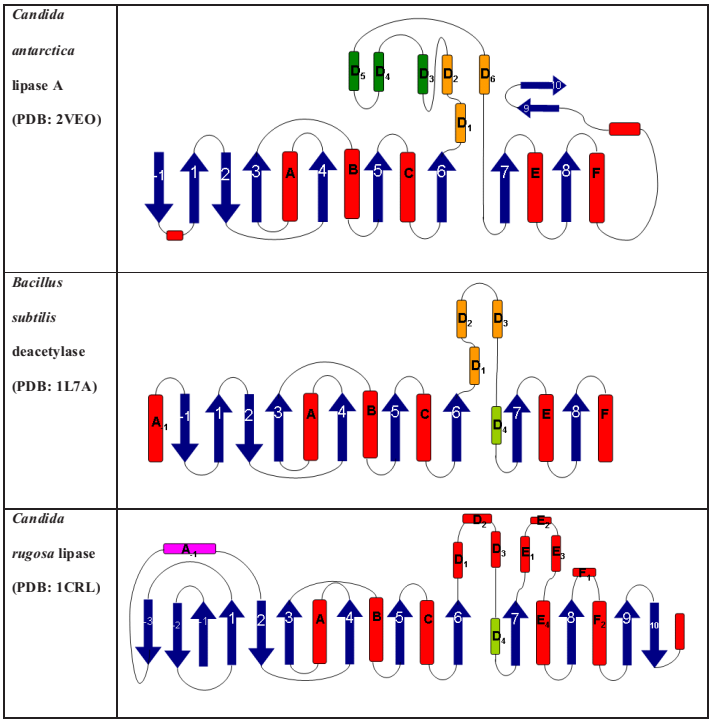

Abbildung 8.2: Topology diagrams of Candida antarctica lipase A, Bacillus subtilis deacetylase and Candida rugosa lipase: The shared cap region between C. antarctica lipase A and B. subtilis deacetylase is colored orange. The additional 3 á-helices of the cap region in CALA are labelled D3, D4 and D5 and colored in dark green. The C-terminal, presumably lid forming â-strands of CALA are labelled 9 and 10. The lid forming á-helix of C. rugosa lipase is labelled A-1.

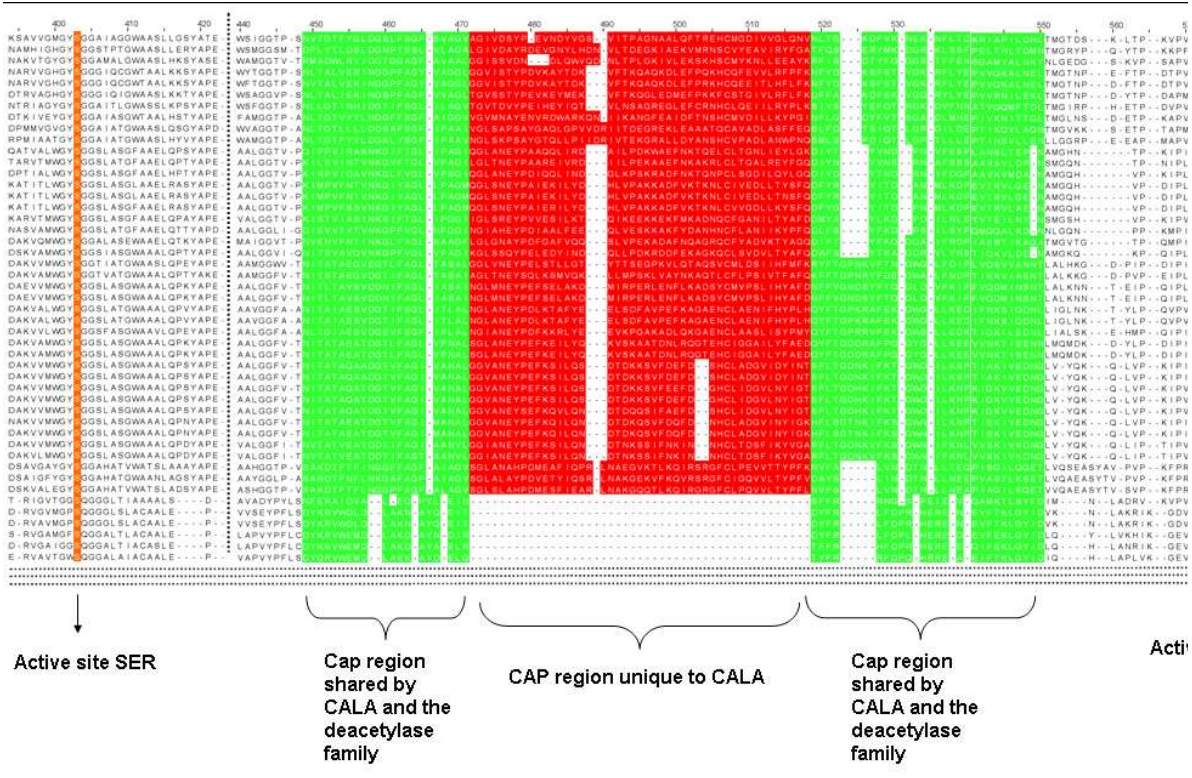

Abbildung 8.3: Multisequence alignment between the "Candida antarctica lipase A ike" and "Deacetylases" superfamilies: Shown is an excerpt of the alignment containing the active site Ser and Asp and the cap region. the cap region of both superfamilies is color coded: shared regions green, the additional 3 α-helices of the "Candida antarctica lipase A like" superfamily red. Columns containing active site residues are coded orange.

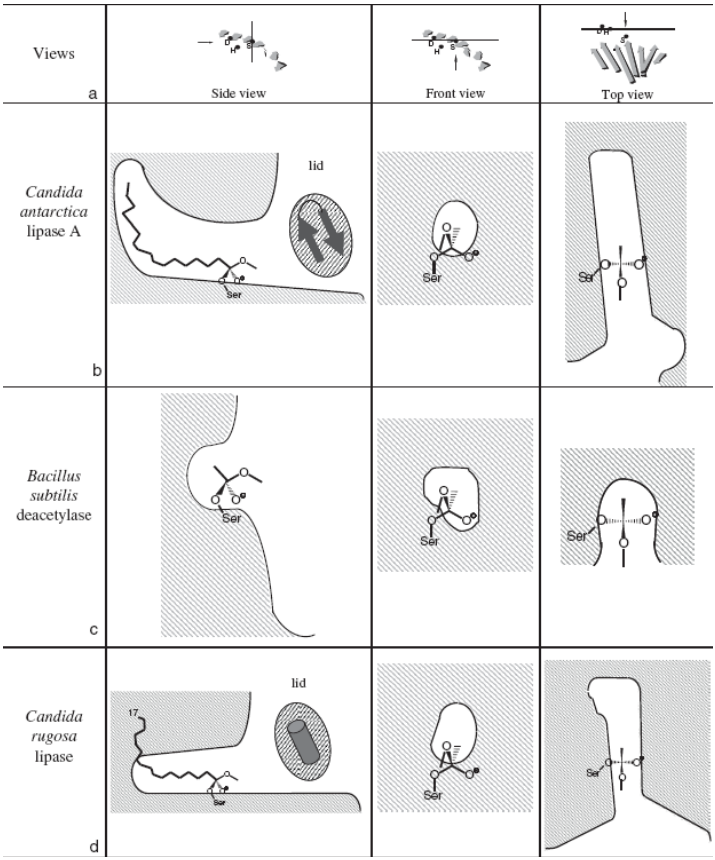

Abbildung 8.4: Shape of the binding site of Candida antarctica lipase A, Bacillus subtilis deacetylase and Candida rugosa lipase: (a) Orientation of the cross-sections which are planes perpendicular to the paper plane and indicated by a straight line. The direction of the view is indicated by an arrow. Shape of the binding sites is displayed in side, front and top view for (b) Candida antarctica lipase A, (c) Bacillus subtilis deacetylase, and (d) Candida rugosa lipase. A model of the acyl moiety of the substrate is displayed, the alcohol moiety is not shown for clarity. The position of the lid in a closed state for Candida antarctica lipase A and Candida rugosa lipase is indicated.

8.11 Tables

Tabelle 8.1: Seed sequences for the "Candida antarctica lipase A like" protein family.

Accession number (gi)	**Organism**	**Homologous Family**
160286179	Candida antarctica	Candida antarctica lipase A like
20429169	Kurtzmanomyces sp. I-11	Candida antarctica lipase A like
73765555	Malassezia furfur	Malassezia lipase like
71018653	Ustilago maydis 521	Candida antarctica lipase A like

8.12 Supplementary material

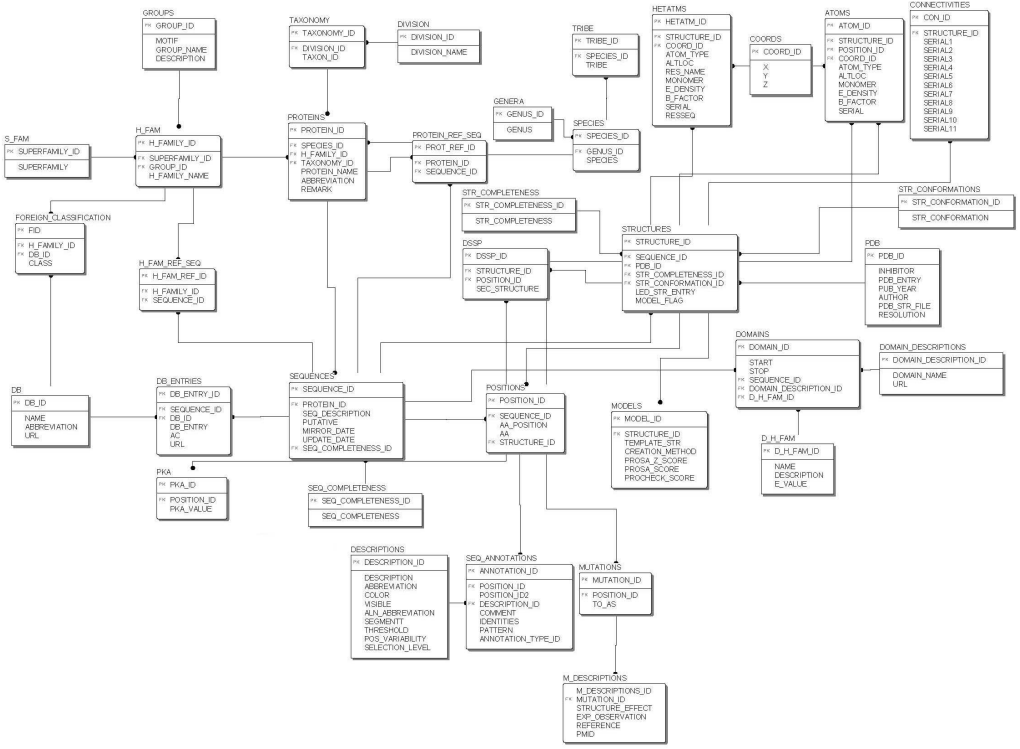

Abbildung 8.5: Conceptual data scheme for the LED using Logical Data Structure (LDS) notation: Each database table is represented by a separate table. Primary key attributes are displayed in the header of the respective table.

Kapitel 9

Literatur

URL http://sourceforge.net/projects/firebird.

URL http://www.perl.org/.

S. N. Ahmed, R. J. Kazlauskas, A. H. Morinville, P. Grochulski, J. D. Schrag, and M. Cygler. Enantioselectivity of candida-rugosa lipase toward carboxylic-acids - a predictive rule from substrate mapping and x-ray crystallography. *Biocatalysis*, 9(1-4):209–225, 1994.

Ajay and M. A. Murcko. Computational methods to predict binding free energy in ligand-receptor complexes. *J Med Chem*, 38(26): 4953–4967, Dec 1995.

Casimir C Akoh, Guan-Chiun Lee, and Jei-Fu Shaw. Protein engineering and applications of candida rugosa lipase isoforms. *Lipids*, 39(6):513–526, Jun 2004.

I. L. Alberts, N. P. Todorov, and P. M. Dean. Receptor flexibility in de novo ligand design and docking. *J. Med. Chem.*, 48:6585–6596, 2005.

S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. Basic local alignment search tool. *J Mol Biol*, 215(3):403–410, Oct 1990.

E. M. Anderson, M. Karin, and O. Kirk. One biocatalyst - many applications: The use of candida antarctica b-lipase in organic synthesis. *Biocatalysis Biotransform*, 16(3):181–204, 1998.

X. Barril and S. D. Morley. Unveiling the full potential of flexible receptor docking using multiple crystallographic structures. *Journal of Medicinal Chemistry*, 48(13):4432–4443, 2005.

Sandra Barth, Markus Fischer, Rolf D Schmid, and Jürgen Pleiss. The database of epoxide hydrolases and haloalkane dehalogenases: one structure, many functions. *Bioinformatics*, 20(16):2845–2847, Nov 2004.

C.I. Bayly, P. Cieplak, W.D. Cornell, and P.A. Kollman. A well-behaved electrostatic potential based method using charge restraints for determining atom-centered charges: The resp model. *J.Phys.Chem*, 97:10269–10280, 1993.

Dennis A. Benson, Ilene Karsch-Mizrachi, David J. Lipman, James Ostell, and David L. Wheeler. Genbank. *Nucl. Acids Res.*, 36 (suppl1):D25–30, 2008.

Christophe Berkane, Gilles Mezoul, Thierry Lalot, Maryvonne Brigodiot, and Ernest Marechal. Lipase-catalyzed polyester synthesis in organic medium. study of ring-chain equilibrium. *Macromolecules*, 30(25):7729–7734, 1997.

H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. The protein data bank. *Nucleic Acids Res*, 28(1):235–42, 2000.

Brent H. Besler, Kenneth M. Merz Jr., and Peter A. Kollman. Atomic charges derived from semiempirical methods. *J Comput Chem*, 11 (4):431–439, 1990.

H. J. Böhm. The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *J Comput Aided Mol Des*, 8: 243–256, Jun 1994.

Uwe T. Bornscheuer and Romas J. Kazlauskas. *Hydrolases in Organic Synthesis*. WILEY-VCH, 2 edition, 2006.

Noureddin El Boulifi, José Aracil, and Mercedes Martínez. Lipase-catalyzed synthesis of isosorbide monoricinoleate: Process optimization by response surface methodology. *Bioresource Technology*, 101(22):8520–8525, 2010.

M. M. Bradford. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem*, 72:248–254, May 1976.

L. Brady, A. M. Brzozowski, Z. S. Derewenda, E. Dodson, G. Dodson, S. Tolley, J. P. Turkenburg, L. Christiansen, B. Huge-Jensen, and L. Norskov. A serine protease triad forms the catalytic centre of a triacylglycerol lipase. *Nature*, 343:767–770, Feb 1990.

Cecilia Branneby, Peter Carlqvist, Anders Magnusson, Karl Hult, Tore Brinck, and Per Berglund. Carbon-carbon bonds by hydrolytic enzymes. *J Am Chem Soc*, 125(4):874–875, Jan 2003.

S. Brenner. The molecular evolution of genes and proteins: a tale of two serines. *Nature*, 334(6182):528–530, Aug 1988.

Natasja Brooijmans and Irwin D Kuntz. Molecular recognition and docking algorithms. *Annu Rev Biophys Biomol Struct*, 32:335–373, 2003.

Andreas Buthe, Alice Kapitain, Winfried Hartmeier, and Marion B. Ansorge-Schumacher. Generation of lipase-containing static emulsions in silicone spheres for synthesis in organic media. *Journal of Molecular Catalysis B: Enzymatic*, 35(4-6):93–99, 2005.

D.A. Case, T.A. Darden, T.E. Cheatham, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, K.M. Merz, D.A. Pearlman, M. Crowley, R.C. Walker, W. Zhang, B. Wang, S. Hayik, A. Roitberg, G. Seabra, K.F. Wong, F. Paesani, X. Wu, S. Brozell, V. Tsui, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Homak, G. Cui, P. Beroza, D.H. Mathews, C. Schafmeister, W.S. Ross, and P.A. Kollman. Amber 9. Technical report, University of California, San Francisco, 2006.

Claudio N Cavasotto and Andrew J W Orry. Ligand docking and structure-based virtual screening in drug discovery. *Curr Top Med Chem*, 7:1006–1014, 2007.

Donald D. Chamberlin and Raymond F. Boyce. Sequel: A structured english query language. In *Proceedings of the 1974 ACM SIGFIDET (now SIGMOD) workshop on Data description, access and control*, pages 249–264, Ann Arbor, Michigan, 1974. ACM.

Saber Chatti, Moez A. Hani, Kirstin Bornhorst, and Hans R. Kricheldorf. Poly(ether sulfone) of isosorbide, isomannide and isoidide. *High Performance Polymers*, 21(1):105–118, 2009.

Matthew Clark, Richard D. Cramer III, and Nicole Van Opdenbosch. Validation of the general purpose tripos 5.2 force field. *Journal of Computational Chemistry*, 10(8):982–1012, 1989.

H. Claussen, C. Buning, M. Rarey, and T. Lengauer. Flexe: Efficient molecular docking considering protein structure variations. *Journal of Molecular Biology*, 308(2):377–395, 2001.

E. F. Codd. A relational model of data for large shared data banks. *Commun. ACM*, 13(6):377–387, 1970.

W. D. Cornell, P. Cieplak, C. I. Baylb, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollmann. A 2nd generation force-field for the simulation of proteins, nucleic-acids, and organic-molecule. *J Am Chem Soc*, 117:5179–5197, May 1995.

Pietro Cozzini, Glen E Kellogg, Francesca Spyrakis, Donald J Abraham, Gabriele Costantino, Andrew Emerson, Francesca Fanelli, Holger Gohlke, Leslie A Kuhn, Garrett M Morris, Modesto Orozco, Thelma A Pertinhez, Menico Rizzi, and Christoph A Sotriffer. Target flexibility: an emerging consideration in drug discovery and design. *J Med Chem*, 51:6237–6255, Oct 2008.

A. M. Davis and S. J. Teague. Hydrogen bonding, hydrophobic interactions, and failure of the rigid receptor hypothesis. *Angewandte Chemie-International Edition*, 38(6):737–749, 1999.

Raquel Dias and Walter Filgueira de Azevedo. Molecular docking algorithms. *Curr Drug Targets*, 9(12):1040–1047, 2008.

Pablo Domínguez de María, Chiara Carboni-Oerlemans, Bernard Tuin, Gerrald Bargeman, Ab van der Meer, and Robert van Gemert. Biotechnological applications of candida antarctica lipase a: State-of-the-art. *Journal of Molecular Catalysis B: Enzymatic*, 37(1-6): 36–46, December 2005. ISSN 1381-1177.

Yong Duan, Chun Wu, Shibasish Chowdhury, Mathew C. Lee, Guoming Xiong, Wei Zhang, Rong Yang, Piotr Cieplak, Ray Luo, Taisung Lee, James Caldwell, Junmei Wang, and Peter Kollman. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *Journal of Computational Chemistry*, 24(16):1999–2012, 2003.

S. R. Eddy. Profile hidden markov models. *Bioinformatics*, 14(9): 755–763, 1998.

Yasser El-Hawari, Angelo D Favia, Ewa S Pilka, Michael Kisiela, Udo Oppermann, Hans-Jörg Martin, and Edmund Maser. Analysis of the substrate-binding site of human carbonyl reductases cbr1 and cbr3 by site-directed mutagenesis. *Chem Biol Interact*, 178(1-3): 234–241, Mar 2009.

M. El Khattabi, P. Van Gelder, W. Bitter, and J. Tommassen. Role of the calcium ion and the disulfide bond in the burkholderia glumae lipase. *Journal Of Molecular Catalysis B-Enzymatic*, 22(5-6):329–338, jul 2003.

J. A. Erickson, M. Jalaie, D. H. Robertson, R. A. Lewis, and M. Vieth. Lessons in molecular recognition: The effects of ligand and protein flexibility on molecular docking accuracy. *J. Med. Chem.*, 47:45–55, 2004.

Daniel J Ericsson, Alex Kasrayan, Patrik Johansson, Terese Bergfors, Anders G Sandström, Jan-E. Bäckvall, and Sherry L Mowbray. X-ray structure of candida antarctica lipase a shows a novel lid structure and a likely mode of interfacial activation. *J Mol Biol*, 376(1):109–119, Feb 2008.

Angelo D Favia, Irene Nobeli, Fabian Glaser, and Janet M Thornton. Molecular docking for substrate identification: the short-chain dehydrogenases/reductases. *J Mol Biol*, 375:855–874, Jan 2008.

Yakai Feng, Jens Knüfermann, Doris Klee, and Hartwig Höcker. Enzyme-catalyzed ring-opening polymerization of 3(¡i¿s¡/i¿)-isopropylmorpholine-2,5-dione. *Macromolecular Rapid Communications*, 20(2):88–90, 1999a.

Yakai Feng, Jens Knüfermann, Doris Klee, and Hartwig Höcker. Lipase-catalyzed ring-opening polymerization of 3(¡i¿s¡/i¿)-isopropylmorpholine-2,5-dione. *Macromolecular Chemistry and Physics*, 200(6):1506–1514, 1999b.

F. Fenouillot, A. Rousseau, G. Colomines, R. Saint-Loup, and J.-P. Pascault. Polymers from renewable 1,4:3,6-dianhydrohexitols (isosorbide, isomannide and isoidide): A review. *Progress in Polymer Science*, 35(5):578–622, 2010.

Robert D Finn, Jaina Mistry, John Tate, Penny Coggill, Andreas Heger, Joanne E Pollington, O. Luke Gavin, Prasad Gunasekaran, Goran Ceric, Kristoffer Forslund, Liisa Holm, Erik L L Sonnhammer, Sean R Eddy, and Alex Bateman. The pfam protein families database. *Nucleic Acids Res*, 38(Database issue):D211–D222, Jan 2010.

M Fischer. *The Lipase Engineering Database: Systematic analysis of family-specific properties and the sequence-structure-function relationship of alpha/beta-Hydrolases*. PhD thesis, Fakultät Chemie der Universität Stuttgart, 2004.

Markus Fischer and Jürgen Pleiss. The lipase engineering database:

a navigation and analysis tool for protein families. *Nucleic Acids Res*, 31(1):319–321, Jan 2003.

Markus Fischer, Quan K Thai, Melanie Grieb, and Jürgen Pleiss. Dwarf–a data warehouse system for analyzing protein families. *BMC Bioinformatics*, 7:495, 2006.

M. C. R. Franssen, H. Jongejan, H. Kooijman, A. L. Spek, L. F. L. Nuno, N. L. F. L. C. Mondril, P. M. A. C. B. dosSantos, and A. de-Groot. Resolution of a tetrahydrofuran ester by candida rugosa lipase (crl) and an examination of crl's stereochemical preference in organic media. *Tetrahedron-Asymmetry*, 7(2):497–510, February 1996.

H. A. Gabb, R. M. Jackson, and M. J. E. Sternberg. Modelling protein docking using shape complementarity, electrostatics and biochemical information. *Journal of Molecular Biology*, 272(1): 106–120, 1997.

D. Ghosh, Z. Wawrzak, V. Z. Pletnev, N. Li, R. Kaiser, W. Pangborn, H. Jörnvall, M. Erman, and W. L. Duax. Structure of uncomplexed and linoleate-bound candida cylindracea cholesterol esterase. *Structure*, 3(3):279–288, Mar 1995.

V. Gotor-Fernandez, E. Busto, and V. Gotor. Candida antarctica lipase b: An ideal biocatalyst for the preparation of nitrogenated organic compounds. *Adv. Synth. Catal.*, 348(7-8):797–812, May 2006.

P. Grochulski, Y. Li, J. D. Schrag, F. Bouthillier, P. Smith, D. Harrison, B. Rubin, and M. Cygler. Insights into interfacial activation from an open structure of candida rugosa lipase. *J Biol Chem*, 268 (17):12843–12847, Jun 1993.

P. Grochulski, F. Bouthillier, R. J. Kazlauskas, A. N. Serreqi, J. D. Schrag, E. Ziomek, and M. Cygler. Analogs of reaction intermediates identify a unique substrate binding site in candida rugosa lipase. *Biochemistry*, 33:3494–3500, Mar 1994a.

P. Grochulski, Y. Li, J. D. Schrag, and M. Cygler. Two conformational states of candida rugosa lipase. *Protein Sci*, 3(1):82–91, Jan 1994b.

N. Guex and M. C. Peitsch. Swiss-model and the swiss-pdbviewer: an environment for comparative protein modeling. *Electrophoresis*, 18(15):2714–2723, Dec 1997.

G. D. Haki and S. K. Rakshit. Developments in industrially important thermostable enzymes: a review. *Bioresour Technol*, 89(1): 17–34, 2003.

G.G. Haraldsson. *The Chemistry of the Functional Groups*, chapter Supplement B2: The Application of Lipases in Organic Synthesis, pages 1395–1473. John Wiley and Sons, 1992.

Fariha Hasan, Aamer Ali Shah, and Abdul Hameed. Methods for detection and characterization of lipases: A comprehensive review. *Biotechnol Adv*, 27(6):782–798, 2009.

Erik Hedenström, Ba-Vu Nguyen, and III Louis A. Silks. Do enzymes recognise remotely located stereocentres? highly enantioselective candida rugosa lipase-catalysed esterification of the 2- to 8-methyldecanoic acids. *Tetrahedron: Asymmetry*, 13:835–844, 2002.

Erik Henke, Jürgen Pleiss, and Uwe T Bornscheuer. Activity of lipases and esterases towards tertiary alcohols: insights into structure-

function relationships. *Angew Chem Int Ed Engl*, 41(17):3211–3213, Sep 2002.

Erik Henke, Uwe T Bornscheuer, Rolf D Schmid, and Jürgen Pleiss. A molecular mechanism of enantiorecognition of tertiary alcohols by carboxylesterases. *Chembiochem*, 4(6):485–493, Jun 2003.

J. Heringa and P. Argos. Strain in protein structures as viewed through nonrotameric side chains: Ii. effects upon ligand binding. *Proteins*, 37(1):44–55, Oct 1999.

J. C. Hermann, E. Ghanem, Y. Li, F. M. Raushel, J. J. Irwin, and B. K. Shoichet. Predicting substrates by docking high-energy intermediates to enzyme structures. *J Am Chem Soc*, 128(49):15882–91, 2006.

Johannes C Hermann, Ricardo Marti-Arbona, Alexander A Fedorov, Elena Fedorov, Steven C Almo, Brian K Shoichet, and Frank M Raushel. Structure-based activity prediction for an enzyme of unknown function. *Nature*, 448(7155):775–779, Aug 2007.

Iris Hilker, Gouher Rabani, Gerard K. M. Verzijl, Anja R. A. Palmans, and Andreas Heise. Chiral polyesters by dynamic kinetic resolution polymerization13. *Angewandte Chemie International Edition*, 45(13):2130–2132, 2006.

I Hoegh, S Patkar, T Halkier, and MT Hansen. 2 lipases from candida antarctica - cloning and expression in aspergillus oryzae. *Canadian Journal of botany - revue Canadienne de botanique*, 73(Suppl. 1 E-H):S869–S875, 1995. ISSN 0008-4026.

F. Hollmann, P. Grzebyk, V. Heinrichs, K. Doderer, and O. Thum. On the inactivity of candida antartica lipase b towards strong

acids. *J. Mol. Catal. B.*, 57(1-4):257–261, May 2009. ISSN 1381-1177.

L. Holm and C. Sander. Dali: a network tool for protein structure comparison. *Trends Biochem Sci*, 20(11):478–480, Nov 1995.

Thierry Hotelier, Ludovic Renault, Xavier Cousin, Vincent Negre, Pascale Marchot, and Arnaud Chatonnet. Esther, the database of the alpha/beta-hydrolase fold superfamily of proteins. *Nucleic Acids Res*, 32(Database issue):D145–D147, Jan 2004.

Gunnar Höst, Lars-Göran Mårtensson, and Bengt-Harald Jonsson. Redesign of human carbonic anhydrase ii for increased esterase activity and specificity towards esters with long acyl chains. *Biochim Biophys Acta*, 1764(10):1601–1606, Oct 2006.

Gunnar E Höst and Bengt-Harald Jonsson. Converting human carbonic anhydrase ii into a benzoate ester hydrolase through rational redesign. *Biochim Biophys Acta*, 1784(5):811–815, May 2008.

S. Y. Huang and X. Q. Zou. Ensemble docking of multiple protein structures: Considering protein structural variations in molecular docking. *Proteins-Structure Function and Bioinformatics*, 66(2): 399–421, 2007.

Sarah Hunter, Rolf Apweiler, Teresa K Attwood, Amos Bairoch, Alex Bateman, David Binns, Peer Bork, Ujjwal Das, Louise Daugherty, Lauranne Duquenne, Robert D Finn, Julian Gough, Daniel Haft, Nicolas Hulo, Daniel Kahn, Elizabeth Kelly, Aurélie Laugraud, Ivica Letunic, David Lonsdale, Rodrigo Lopez, Martin Madera, John Maslen, Craig McAnulla, Jennifer McDowall, Jaina Mistry, Alex Mitchell, Nicola Mulder, Darren Natale, Christine Orengo,

Antony F Quinn, Jeremy D Selengut, Christian J A Sigrist, Manjula Thimma, Paul D Thomas, Franck Valentin, Derek Wilson, Cathy H Wu, and Corin Yeats. Interpro: the integrative protein signature database. *Nucleic Acids Res*, 37(Database issue):D211–D215, Jan 2009.

John J Irwin, Frank M Raushel, and Brian K Shoichet. Virtual screening against metalloenzymes for inhibitors and substrates. *Biochemistry*, 44(37):12316–12328, Sep 2005.

Christian Jäckel, Peter Kast, and Donald Hilvert. Protein design by directed evolution. *Annu Rev Biophys*, 37:153–173, 2008.

C. F. Jacobsen, J. Léonis, K. Linderstrøm-Lang, and M. Ottesen. The ph-stat and its use in biochemistry. In David Glick, editor, *Methods of Biochemical Analysis*, pages 171–210. Interscience Publishers, Inc., 1957.

Homer Jacobson and Walter H. Stockmayer. Intramolecular reaction in polycondensations. i. the theory of linear systems. *J. Chem. Phys.*, 18(12):1600–1606, 1950.

M. Jacobson and A. Sali. Comparative protein structure modeling and its applications to drug discovery. *Annual Reports in Medicinal Chemistry*, 39:259–276, 2004.

Karl-Erich Jaeger, Stéphane Ransac, Bauke W. Dijkstra, Charles Colson, Margreet van Heuvel, and Onno Misset. Bacterial lipases. *FEMS Microbiology Reviews*, 15(1):29–63, 1994.

G. Jones, P. Willett, R. C. Glen, A. R. Leach, and R. Taylor. Development and validation of a genetic algorithm for flexible docking. *J Mol Biol*, 267(3):727–748, 1997.

William L. Jorgensen and Julian. Tirado-Rives. The opls [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *Journal of the American Chemical Society*, 110(6):1657–1666, mar 1988. ISSN 0002-7863.

WL. Jorgensen, J. Chandrasekhar, JD. Madura, RW. Impey, and ML. Klein. Comparison of simple potential functions for simulating liquid water. *J Chem Phys*, 79:926–935, 1983.

Benjamin Juhl, Peter Trodler, Sadhna Tyagi, and Jurgen Pleiss. Modelling substrate specificity and enantioselectivity for lipases and esterases by substrate-imprinted docking. *BMC Struct. Biol.*, 9 (1):39, 2009.

Suhyun Jung and Seongsoon Park. Improving the expression yield of candida antarctica lipase b in escherichia coli by mutagenesis. *Biotechnol Lett*, 30(4):717–722, Apr 2008.

C. Kalyanaraman, K. Bernacki, and M. P. Jacobson. Virtual screening against highly charged active sites: identifying substrates of alpha-beta barrel enzymes. *Biochemistry*, 44(6):2059–71, 2005.

Chakrapani Kalyanaraman, Heidi J Imker, Alexander A Fedorov, Elena V Fedorov, Margaret E Glasner, Patricia C Babbitt, Steven C Almo, John A Gerlt, and Matthew P Jacobson. Discovery of a dipeptide epimerase enzymatic function guided by homology modeling and virtual screening. *Structure*, 16(11):1668–1677, Nov 2008.

A. Kamper, J. Apostolakis, M. Rarey, C. M. Marian, and T. Lengauer. Fully automated flexible docking of ligands into flexible synthe-

tic receptors using forward and inverse docking strategies. *Journal of Chemical Information and Modeling*, 46(2):903–911, 2006.

Panagiota Kapoli, Irene A Axarli, Dimitris Platis, Maria Fragoulaki, Mark Paine, Janet Hemingway, John Vontas, and Nikolaos E Labrou. Engineering sensitive glutathione transferase for the detection of xenobiotics. *Biosens Bioelectron*, 24(3):498–503, Nov 2008.

R. J. Kazlauskas and H. K. Weber. Improving hydrolases for organic synthesis. *Current Opinion In Chemical Biology*, 2(1):121–126, feb 1998.

RJ Kazlauskas, ANE Weissfloch, AT Rappaport, and LA Cuccia. A rule to predict which enantiomer of a secondary alcohol reacts faster in reactions catalyzed by cholesterol esterase. lipase from pseudomonas cepacia, and lipase from candida rugosa. *J Org Chem*, 56(8):2656–2665, Apr 12 1991. ISSN 0022-3263.

Eun-Young Kim, Ki-Hoon Oh, Mi-Hwa Lee, Chul-Hyung Kang, Tae-Kwang Oh, and Jung-Hoon Yoon. Novel cold-adapted alkaline lipase from an intertidal flat metagenome and proposal for a new family of bacterial lipases. *Appl Environ Microbiol*, 75(1):257–260, Jan 2009.

So-Young Kim, Jung-Hoon Sohn, Yu-Ryang Pyun, In Seok Yang, Kyung-Hyun Kim, and Eui-Sung Choi. In vitro evolution of lipase b from candida antarctica using surface display in hansenula polymorpha. *J Microbiol Biotechnol*, 17(8):1308–1315, Aug 2007.

Ole Kirk and Morten WÃ$\frac{1}{4}$rtz Christensen. Lipases from candida antarctica: Unique biocatalysts from a unique origin. *Organic*

Process Research & Development, 6(4):446–451, July 2002. ISSN 1083-6160.

Ole Kirk, Fredrik Björkling, Sven Godtfredsen, and Thomas Larsen. Fatty acid specificity in lipase-catalyzed synthesis of glucoside esters. *Biocatal. Biotransform.*, 6:127–134, 1992.

Yoh Kodera, Katsunobu Takahashi, Hiroyuki Nishimura, Ayako Matsushima, Yuji Saito, and Yuji Inada. Ester sythesis from alpha-substituted carboxylic acid catalyzed by polyethylene glycol-modified lipase from candida cylindracea in benzene. *Biotechnol. Lett.*, 8:881–884, 1986. The Pseudomonas fluorescens lipase from Amano Pharmaceutical Ltd. used in this work was later reclassified as Burkholderia cepacia lipase.

Markus Koschorreck, Markus Fischer, Sandra Barth, and Jürgen Pleiss. How to find soluble proteins: a comprehensive analysis of alpha/beta hydrolases for recombinant expression in e. coli. *BMC Genomics*, 6(1):49, 2005.

B. Kramer, M. Rarey, and T. Lengauer. Evaluation of the flexx incremental construction algorithm for protein-ligand docking. *Proteins-Structure Function and Genetics*, 37(2):228–241, 1999.

Hans R. Kricheldorf. "sugar diols" as building blocks of polycondensates. *Polymer Reviews*, 37(4):599–631, 1997.

Hans R. Kricheldorf, Saber Chatti, Gert Schwarz, and Ralf-P. Krüger. Macrocycles 27: Cyclic aliphatic polyesters of isosorbide. *Journal of Polymer Science Part A: Polymer Chemistry*, 41 (21):3414–3424, 2003.

HR Kricheldorf and G Schwarz. Cyclic polymers by kinetically controlled step-growth polymerization. *Macromolecular Rapid Communications*, 24(5-6):359–381, 2003.

Marcin Król, Alexander L Tournier, and Paul A Bates. Flexible relaxation of rigid-body docking solutions. *Proteins*, 68(1):159–169, Jul 2007.

Olli Laine, Heidi Österholm, Maaria Seläntaus, Hannele Järvinen, and Pirjo Vainiotalo. Determination of cyclic polyester oligomers by gel permeation chromatography and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry*, 15(20):1931–1935, 2001.

Thierry Lalot and Ernest Marechal. Enzyme-catalyzed polyester synthesis. *International Journal of Polymeric Materials*, 50(3):267–286, 2001.

P. Therese Lang, Scott R Brozell, Sudipto Mukherjee, Eric F Pettersen, Elaine C Meng, Veena Thomas, Robert C Rizzo, David A Case, Thomas L James, and Irwin D Kuntz. Dock 6: combining techniques to model rna-small molecule complexes. *RNA*, 15(6): 1219–1230, 2009.

A. R. Leach. Ligand docking to proteins with discrete side-chain flexibility. *Journal of Molecular Biology*, 235(1):345–356, 1994.

S. C. Lee, Y. J. Chang, D. M. Shin, J. Han, M. H. Seo, H. Fazelinia, C. D. Maranas, and H. S. Kim. Designing the substrate specificity of d-hydantoinase using a rational approach. *Enzyme And Microbial Technology*, 44(3):170–175, March 2009.

M. Liu and S. Wang. Mcdock: a monte carlo simulation approach to the molecular docking problem. *J Comput Aided Mol Des*, 13(5): 435–451, 1999.

Katrin Lämmle, Hubert Zipper, Michael Breuer, Bernhard Hauer, Christiane Buta, Herwig Brunner, and Steffen Rupp. Identification of novel enzymes with different hydrolytic activities by metagenome expression cloning. *J Biotechnol*, 127(4):575–592, Jan 2007.

O. Lockridge and P. Masson. Pesticides and susceptible populations: people with butyrylcholinesterase genetic variants may be at risk. *Neurotoxicology*, 21(1-2):113–126, 2000.

M. Lotti, R. Grandori, F. Fusetti, S. Longhi, S. Brocca, A. Tramontano, and L. Alberghina. Cloning and analysis of candida cylindracea lipase sequences. *Gene*, 124(1):45–55, Feb 1993.

Antonio Macchiarulo, Irene Nobeli, and Janet M Thornton. Ligand selectivity and competition between enzymes in silico. *Nat Biotechnol*, 22(8):1039–1045, Aug 2004.

A.D. MacKerel Jr., C.L. Brooks III, L. Nilsson, B. Roux, Y. Won, and M. Karplus. *CHARMM: The Energy Function and Its Parameterization with an Overview of the Program*, volume 1 of *The Encyclopedia of Computational Chemistry*, pages 271–277. John Wiley & Sons: Chichester, 1998.

A. Magnusson, K. Hult, and M. Holmquist. Creation of an enantioselective hydrolase by engineered substrate-assisted catalysis. *J Am Chem Soc*, 123(18):4354–4355, May 2001.

A. O. Magnusson, M. Takwa, A. Hamberg, and K. Hult. An s-selective lipase was created by rational redesign and the enantio-

selectivity increased with temperature. *Angew Chem Int Ed Engl*, 44(29):4582–5, 2005a.

Anders O Magnusson, Johanna C Rotticci-Mulder, Alberto Santagostino, and Karl Hult. Creating space for large secondary alcohols by rational redesign of candida antarctica lipase b. *Chembiochem*, 6(6):1051–1056, Jun 2005b.

Bola D Majekodunmi, Cesar A Lau-Cam, and Robert A Nash. Stability of benzoyl peroxide in aromatic ester-containing topical formulations. *Pharm Dev Technol*, 12(6):609–620, 2007.

José M Mancheño, María A Pernas, María J Martínez, Begoña Ochoa, M. Luisa Rúa, and Juan A Hermoso. Structural insights into the lipase/esterase behavior in the candida rugosa lipases family: crystal structure of the lipase 2 isoenzyme at 1.97a resolution. *J Mol Biol*, 332(5):1059–1069, Oct 2003.

Mats Martinelle, Mats Holmquist, and Karl Hult. On the interfacial activation of candida antarctica lipase a and b as compared with humicola lanuginosa lipase. *Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism*, 1258(3):272–276, October 1995. ISSN 0005-2760.

S. L. McGovern and B. K. Shoichet. Information decay in molecular docking screens against holo, apo, and modeled conformations of enzymes. *Journal of Medicinal Chemistry*, 46(14):2895–2907, 2003.

A. D. Mclachlan. Rapid comparison of protein structures. *Acta Crystallographica Section A*, 38(NOV):871–873, 1982.

Gilles Mezoul, Thierry Lalot, Maryvonne Brigodiot, and Ernest Maréchal. Enzyme-catalyzed syntheses of poly(1,6-hexanediyl ma-

leate) and poly(1,6-hexanediyl fumarate) in organic medium. *Macromolecular Rapid Communications*, 16(8):613–620, 1995.

C. B. Millard, G. Kryger, A. Ordentlich, H. M. Greenblatt, M. Harel, M. L. Raves, Y. Segall, D. Barak, A. Shafferman, I. Silman, and J. L. Sussman. Crystal structures of aged phosphonylated acetylcholinesterase: nerve agent reaction products at the atomic level. *Biochemistry*, 38:7032–7039, Jun 1999.

M. D. Miller, S. K. Kearsley, D. J. Underwood, and R. P. Sheridan. Flog: a system to select 'quasi-flexible' ligands complementary to a receptor of known three-dimensional structure. *J Comput Aided Mol Des*, 8(2):153–174, 1994.

Maria A Miteva, Pierre Tufféry, and Bruno O Villoutreix. Pce: web tools to compute protein continuum electrostatics. *Nucleic Acids Res*, 33(Web Server issue):W372–W375, Jul 2005.

Seiji Miura, Stefano Ferri, Wakako Tsugawa, Seungsu Kim, and Koji Sode. Development of fructosyl amine oxidase specific to fructosyl valine by site-directed mutagenesis. *Protein Eng Des Sel*, 21(4): 233–239, Apr 2008.

Serge Moralev and Eugene Rozengart. *Comparative Enzymology of Cholinesterases*. International University Line, La Jolla, 1st edition, 2007.

G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell, and A. J. Olson. Autodock4 and autodocktools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.*, 30(16):2785–2791, 2009.

R. Morrone, G. Nicolosi, A. Patti, and M. Piattelli. Resolution of racemic flurbiprofen by lipase-mediated esterification in organic-solvent. *Tetrahedron-Asymmetry*, 6(7):1773–1778, July 1995.

C. W. Murray, C. A. Baxter, and A. D. Frenkel. The sensitivity of the results of molecular docking to induced fit effects: Application to thrombin, thermolysin and neuraminidase. *Journal of Computer-Aided Molecular Design*, 13(6):547–562, 1999.

V Nagy, ER Toke, LC Keong, G Szatzker, D Ibrahim, IC Omar, G Szakacs, and L Poppe. Kinetic resolutions with novel, highly enantioselective fungal lipases produced by solid state fermentation. *J. Mol. Catal. B.*, 39(1-4, Sp. Iss. SI):141–148, May 2 2006. ISSN 1381-1177.

Niju Narayanan and C. Perry Chou. Alleviation of proteolytic sensitivity to enhance recombinant lipase production in escherichia coli. *Appl Environ Microbiol*, 75(16):5424–5427, Aug 2009.

Y. Nicolet, O. Lockridge, P. Masson, J. C. Fontecilla-Camps, and F. Nachon. Crystal structure of human butyrylcholinesterase and of its complexes with substrate and products. *J Biol Chem*, 278 (42):41141–7, 2003.

Bart A. J. Noordover, Viola G. van Staalduinen, Robbert Duchateau, Cor E. Koning, van Benthem, Manon Mak, Andreas Heise, August E. Frissen, and Jacco van Haveren. Co- and terpolyesters based on isosorbide and succinic acid for coating applications: Synthesis and characterization. *Biomacromolecules*, 7(12):3406–3416, 2006.

Bart A. J. Noordover, Robbert Duchateau, Rolf A. T. M. van Benthem, Weihua Ming, and Cor E. Koning. Enhancing the func-

tionality of biobased polyester coating resins through modification with citric acid. *Biomacromolecules*, 8(12):3860–3870, 2007.

D. OHagan and N. A. Zaidi. The resolution of tertiary alpha-acetylene-acetate esters by the lipase from candida-cylindracea. *Tetrahedron-Asymmetry*, 5(6):1111–1118, June 1994.

D. L. Ollis, E. Cheah, M. Cygler, B. Dijkstra, F. Frolow, S. M. Franken, M. Harel, S. J. Remington, I. Silman, and J. Schrag. The alpha/beta hydrolase fold. *Protein Eng*, 5(3):197–211, Apr 1992.

Christian Orrenius, Niklas Öhrner, Didier Rotticci, Anders Mattson, Karl Hult, and Torbjörn Norin. Candida antarctica lipase b catalysed kinetic resolutions: Substrate structure requirements for the preparation of enantiomerically enriched secondary alcanols. *Tetrahedron: Asymmetry*, 6(5):1217–1220, May 1995a. ISSN 0957-4166.

Christian Orrenius, Torbjörn Norin, Karl Hult, and Giacomo Carrea. The candida antarctica lipase b catalysed kinetic resolution of seudenol in non-aqueous media of controlled water activity. *Tetrahedron: Asymmetry*, 6(12):3023–3030, December 1995b. ISSN 0957-4166.

Angel R Ortiz, Paulino Gomez-Puertas, Alejandra Leo-Macias, Pedro Lopez-Romero, Eduardo Lopez-Vi nas, Antonio Morreale, Marta Murcia, and Kun Wang. Computational approaches to model ligand selectivity in drug design. *Curr Top Med Chem*, 6:41–55, 2006.

Jenny Ottosson, Linda Fransson, and Karl Hult. Substrate entropy in enzyme enantioselectivity: an experimental and molecular modeling study of a lipase. *Protein Sci*, 11:1462–1471, Jun 2002a.

Jenny Ottosson, Linda Fransson, Jerry W King, and Karl Hult. Size as a parameter for solvent effects on candida antarctica lipase b enantioselectivity. *Biochim Biophys Acta*, 1594:325–334, Feb 2002b.

P. L. Overbeeke, J. A. Jongejan, and J. J. Heijnen. Solvent effect on lipase enantioselectivity. evidence for the presence of two thermodynamic states. *Biotechnol Bioeng*, 70:278–290, Nov 2000.

Santosh Kumar Padhi, Despina J Bougioukou, and Jon D Stewart. Site-saturation mutagenesis of tryptophan 116 of saccharomyces pastorianus old yellow enzyme uncovers stereocomplementary variants. *J Am Chem Soc*, 131:3271–3280, 2009.

T. Panda and B. S. Gowrishankar. Production and applications of esterases. *Appl Microbiol Biotechnol*, 67:160–169, Apr 2005.

Yuan-Ping Pang, Emanuele Perola, Kun Xu, and Franklyn G. Prendergast. Eudoc: a computer program for identification of drug interaction sites in macromolecules and drug leads from chemical databases. *J Comput Chem*, 22(15):1750–1771, 2001.

S. Patkar, J. Vind, E. Kelstrup, M. W. Christensen, A. Svendsen, K. Borch, and O. Kirk. Effect of mutations in candida antarctica b lipase. *Chem Phys Lipids*, 93(1-2):95–101, Jun 1998.

SA Patkar, F Bjorking, M Zundel, M Schulein, A Svendsen, HP Heldthansen, and E Gormsen. Purification of 2 lipases from candida antarctica and their inhibition by various inhibitors. *Indian Journal of chemistry Section B-Organic Chemistry including Medicinal Chemistry*, 32(1):76–80, Jan 1993. ISSN 0376-4699.

M. T. Petersen, P. Martel, E. I. Petersen, F. Drabløs, and S. B. Petersen. Surface and electrostatics of cutinases. *Methods Enzymol*, 284:130–154, 1997.

Jan Pfeffer, Sven Richter, Jens Nieveler, Carl-Erik Hansen, Rachid Bel Rhlid, Rolf D Schmid, and Monika Rusnak. High yield expression of lipase a from candida antarctica in the methylotrophic yeast pichia pastoris and its purification and characterisation. *Appl Microbiol Biotechnol*, 72(5):931–938, Oct 2006.

Ursula Pieper, Narayanan Eswar, Fred P Davis, Hannes Braberg, M. S. Madhusudhan, Andrea Rossi, Marc Marti-Renom, Rachel Karchin, Ben M Webb, David Eramian, Min-Yi Shen, Libusha Kelly, Francisco Melo, and Andrej Sali. Modbase: a database of annotated comparative protein structure models and associated resources. *Nucleic Acids Res*, 34(Database issue):D291–D295, Jan 2006.

J. Pleiss, M. Fischer, M. Peiker, C. Thiele, and R. D. Schmid. Lipase engineering database - understanding and exploiting sequence-structure-function relationships. *Journal Of Molecular Catalysis B-Enzymatic*, 10(5):491–508, October 2000.

Francisco J Plou, M. Angeles Cruces, Manuel Ferrer, Gloria Fuentes, Eitel Pastor, Manuel Bernabé, Morten Christensen, Francisco Comelles, José L Parra, and Antonio Ballesteros. Enzymatic acylation of di- and trisaccharides with fatty acids: choosing the appropriate enzyme, support and solvent. *J Biotechnol*, 96(1):55–66, Jun 2002.

M. Rarey, B. Kramer, T. Lengauer, and G. Klebe. A fast flexible docking method using an incremental construction algorithm. *Journal of Molecular Biology*, 261(3):470–489, 1996.

M. Rarey, B. Kramer, and T. Lengauer. Multiple automatic base selection: protein-ligand docking based on incremental constructi-

on without manual intervention. *J Comput Aided Mol Des*, 11(4): 369–384, Jul 1997.

M. T. Reetz and K. Schimossek. Lipase-catalyzed dynamic kinetic resolution of chiral amines: Use of palladium as the racemization catalyst. *Chimia*, 50(12):668–669, 1996.

Manfred T Reetz, Daniel Kahakeaw, and Joaquin Sanchis. Shedding light on the efficacy of laboratory evolution based on iterative saturation mutagenesis. *Mol Biosyst*, 5(2):115–122, Feb 2009.

W. M. Rockey and A. H. Elcock. Rapid computational identification of the targets of protein kinase inhibitors. *Journal of Medicinal Chemistry*, 48(12):4138–4152, 2005.

W. M. Rockey and A. H. Elcock. Structure selection for protein kinase docking and virtual screening: Homology models or crystal structures? *Current Protein & Peptide Science*, 7(5):437–457, 2006.

E. Rogalska, C. Cudrey, F. Ferrato, and R. Verger. Stereoselective hydrolysis of triglycerides by animal and microbial lipases. *Chirality*, 5(1):24–30, 1993.

Remo Rohs, Itai Bloch, Heinz Sklenar, and Zippora Shakked. Molecular flexibility in ab initio drug docking to dna: binding-site and binding-mode transitions in all-atom monte carlo simulations. *Nucleic Acids Res*, 33:7048–7057, 2005.

D. Rotticci, C. Orrenius, K. Hult, and T. Norin. Enantiomerically enriched bifunctional sec-alcohols prepared by candida antarctica lipase b catalysis. evidence of non-steric interactions. *Tetrahedron-Asymmetry*, 8(3):359–362, February 1997.

R. B. Russell and G. J. Barton. Multiple protein sequence alignment from tertiary structure comparison: assignment of global and residue confidence levels. *Proteins*, 14(2):309–323, Oct 1992.

Patrik Rydberg, Sine Myrup Hansen, Jacob Kongsted, Per-Ola Norrby, Lars Olsen, and Ulf Ryde. Transition-state docking of flunitrazepam and progesterone in cytochrome p450. *J Chem Theory Comput*, 4:673–681, 2008.

Keiji Sakaki, Shigeki Hara, and Naotsugu Itoh. Optical resolution of racemic 2-hydroxy octanoic acid using biphasic enzyme membrane reactor. *Desalination*, 149:247–252, 2002.

Andrej Sali and Tom L. Blundell. Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology*, 234(3):779–815, 1993.

B. Sandak, H. J. Wolfson, and R. Nussinov. Flexible docking allowing induced fit in proteins: Insights from an open to closed conformational isomers. *Proteins-Structure Function and Genetics*, 32(2): 159–174, 1998.

A. Schlacher, T. Stanzer, I. Osprian, M. Mischitz, E. Klingsbichel, K. Faber, and H. Schwab. Detection of a new enzyme for stereoselective hydrolysis of linalyl acetate using simple plate assays for the characterization of cloned esterases from burkholderia gladioli. *J Biotechnol*, 62(1):47–54, Jun 1998.

J. D. Schrag, Y. Li, M. Cygler, D. Lang, T. Burgdorf, H. J. Hecht, R. Schmid, D. Schomburg, T. J. Rydel, J. D. Oliver, L. C. Strickland, C. M. Dunaway, S. B. Larson, J. Day, and A. McPherson. The open conformation of a pseudomonas lipase. *Structure*, 5(2): 187–202, Feb 1997.

T. Schulz, J. Pleiss, and R. D. Schmid. Stereoselectivity of pseudomonas cepacia lipase toward secondary alcohols: a quantitative model. *Protein Sci*, 9(6):1053–1062, Jun 2000.

Alexander Seifert, Sandra Vomund, Katrin Grohmann, Sebastian Kriening, Vlada B Urlacher, Sabine Laschat, and Jürgen Pleiss. Rational design of a minimal and highly enriched cyp102a1 mutant library with improved regio-, stereo- and chemoselectivity. *Chembiochem*, 10(5):853–861, Mar 2009.

Trevor Selwood, Shawn R. Feaster, Michael J. States, Alton N. Pryor, and Daniel M. Quinn. Parallel mechanisms in acetylcholinesterase-catalyzed hydrolysis of choline esters. *J. Am. Chem. Soc.*, 115: 10477–10482, 1993.

U. Chandra Singh and Peter A. Kollman. An approach to computing electrostatic charges for molecules. *J Comput Chem*, 5(2):129–145, 1984.

Adam J T Smith, Roger Müller, Miguel D Toscano, Peter Kast, Homme W Hellinga, Donald Hilvert, and K. N. Houk. Structural reorganization and preorganization in enzyme active sites: comparisons of experimental and theoretically ideal active site geometries in the multistep serine esterase reaction cycle. *J Am Chem Soc*, 130:15361–15373, Nov 2008.

Ling Song, Chakrapani Kalyanaraman, Alexander A Fedorov, Elena V Fedorov, Margaret E Glasner, Shoshana Brown, Heidi J Imker, Patricia C Babbitt, Steven C Almo, Matthew P Jacobson, and John A Gerlt. Prediction and assignment of function for a divergent n-succinyl amino acid racemase. *Nat Chem Biol*, 3(8): 486–491, Aug 2007.

Guo-Dong Su, Deng-Feng Huang, Shuang-Yan Han, Sui-Ping Zheng, and Ying Lin. Display of candida antarctica lipase b on pichia pastoris and its application to flavor ester synthesis. *Appl Microbiol Biotechnol*, 86(5):1493–1501, May 2010.

J. L. Sussman, M. Harel, F. Frolow, C. Oefner, A. Goldman, L. Toker, and I. Silman. Atomic structure of acetylcholinesterase from torpedo californica: a prototypic acetylcholine-binding protein. *Science*, 253(5022):872–879, Aug 1991.

Mariann Szigeti, Eniko R. Toke, Maria C. Turoczi, Viviana Nagy, Gyoergy Szakacs, and Laszlo Poppe. Lipase-catalyzed kinetic resolution of 4-aryl- and 4-heteroarylbut-3-en-2-ols. *ARKIVOC*, (Part 3):54–65, 2008. ISSN 1424-6376.

A. Tanaka, H. Sugimoto, Y. Muta, T. Mizuno, K. Senoo, H. Obata, and K. Inouye. Differential scanning calorimetry of the effects of ca2+ on the thermal unfolding of pseudomonas cepacia lipase. *Bioscience Biotechnology And Biochemistry*, 67(1):207–210, jan 2003.

R. D. Taylor, P. J. Jewsbury, and J. W. Essex. A review of protein-small molecule docking methods. *J Comput Aided Mol Des*, 16(3): 151–166, Mar 2002.

Joachim Thiem and Harald Lüders. Synthesis of polyterephthalates derived from dianhydrohexitols. *Polymer Bulletin*, 11(4):365–369, 1984-04-01.

J. D. Thompson, D. G. Higgins, and T. J. Gibson. Clustal w: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and

weight matrix choice. *Nucleic Acids Res*, 22(22):4673–4680, Nov 1994.

S. Tyagi and J. Pleiss. Biochemical profiling in silico–predicting substrate specificities of large enzyme families. *J Biotechnol*, 124 (1):108–16, 2006.

J. Uppenberg, M. T. Hansen, S. Patkar, and T. A. Jones. The sequence, crystal structure determination and refinement of two crystal forms of lipase b from candida antarctica. *Structure*, 2(4): 293–308, 1994.

J. Uppenberg, N. Ohrner, M. Norin, K. Hult, G. J. Kleywegt, S. Patkar, V. Waagen, T. Anthonsen, and T. A. Jones. Crystallographic and molecular-modeling studies of lipase b from candida antarctica reveal a stereospecificity pocket for secondary alcohols. *Biochemistry*, 34(51):16838–51, 1995.

Wilfred F. van Gunsteren and Herman J. C. Berendsen. Moleküldynamik-computersimulationen; methodik, anwendungen und perspektiven in der chemie. *Angewandte Chemie*, 102(9):1020–1055, 1990.

R. Verger. 'interfacial activation' of lipases: Facts and artifacts. *Trends In Biotechnology*, 15(1):32–38, 1997.

Florence Vincent, Simon J Charnock, Koen H G Verschueren, Johan P Turkenburg, David J Scott, Wendy A Offen, Shirley Roberts, Gavin Pell, Harry J Gilbert, Gideon J Davies, and James A Brannigan. Multifunctional xylooligosaccharide/cephalosporin c deacetylase revealed by the hexameric structure of the bacillus subtilis enzyme at 1.9a resolution. *J Mol Biol*, 330(3):593–606, Jul 2003.

A. Warshel and J. Florián. Computer simulations of enzyme catalysis: finding out what has been optimized by evolution. *Proc Natl Acad Sci U S A*, 95(11):5950–5955, May 1998.

Y. Wei, J. A. Contreras, P. Sheffield, T. Osterlund, U. Derewenda, R. E. Kneusel, U. Matern, C. Holm, and Z. S. Derewenda. Crystal structure of brefeldin a esterase, a bacterial homolog of the mammalian hormone-sensitive lipase. *Nat Struct Biol*, 6(4):340–345, Apr 1999.

Michael Widmann, Marie Clairo, Jürgen Dippon, and Jürgen Pleiss. Analysis of the distribution of functionally relevant rare codons. *BMC Genomics*, 9:207, 2008.

Michael Widmann, P. Benjamin Juhl, and Jürgen Pleiss. Structural classification by the lipase engineering database: a case study of candida antarctica lipase a. *BMC Genomics*, 11:123, 2010.

John B. Williams, Arkady I. Gusev, and David M. Hercules. Characterization of polyesters by matrix-assisted laser desorption ionization mass spectrometry. *Macromolecules*, 30(13):3781–3787, 1997.

F. K. Winkler, A. D'Arcy, and W. Hunziker. Structure of human pancreatic lipase. *Nature*, 343:771–774, Feb 1990.

L. W. Wright and J. D. Brandner. Catalytic isomerization of polyhydric alcohols.1 ii. the isomerization of isosorbide to isomannide and isoidide. *The Journal of Organic Chemistry*, 29(10):2979–2982, 1964.

Xiaole Xia, Yong-Hua Wang, Bo Yang, and Xiaoning Wang. Wheat germ lipase catalyzed kinetic resolution of secondary alcohols in non-aqueous media. *Biotechnol Lett*, 31(1):83–87, Jan 2009.

Dao Feng Xiang, Peter Kolb, Alexander A Fedorov, Monika M Meier, Lena V Fedorov, T. Tinh Nguyen, Reinhard Sterner, Steven C Almo, Brian K Shoichet, and Frank M Raushel. Functional annotation and three-dimensional structure of dr0930 from deinococcus radiodurans, a close relative of phosphotriesterase in the amidohydrolase superfamily. *Biochemistry*, 48(10):2237–2247, Mar 2009.

A Zaks and A M Klibanov. Enzyme-catalyzed processes in organic solvents. *Proceedings of the National Academy of Sciences of the United States of America*, 82(10):3192–3196, 1985.

Maria I Zavodszky and Leslie A Kuhn. Side-chain flexibility in protein-ligand binding: the minimal rotation hypothesis. *Protein Sci*, 14:1104–1114, Apr 2005.

Ningyan Zhang, Wen-Chen Suen, William Windsor, Li Xiao, Vincent Madison, and Aleksey Zaks. Improving tolerance of candida antarctica lipase b towards irreversible thermal inactivation through directed evolution. *Protein Eng*, 16(8):599–605, Aug 2003.

Dunming Zhu, Yan Yang, Stephanie Majkowicz, Thoris Hsin-Yuan Pan, Katherine Kantardjieff, and Ling Hua. Inverting the enantioselectivity of a carbonyl reductase via substrate-enzyme docking-guided point mutation. *Org Lett*, 10(4):525–528, Feb 2008.

Erklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Verwendung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht.

Stuttgart, September 2010

Benjamin Juhl