

Von der Sequenz zur Funktion:
systematische Modellierung verschiedener
Proteinfamilien auf Sequenz- und Strukturebene

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

Michael Knoll

aus Bietigheim-Bissingen

Hauptberichter: Prof. Dr. Jürgen Pleiss

Mitberichter: Prof. Dr. Rolf D. Schmid

Tag der mündlichen Prüfung: 13.06.2008

Institut für Technische Biochemie der
Universität Stuttgart

2008

Inhaltsverzeichnis

Zusammenfassung	5
Summary	9
Publikationen und Publikationsmanuskripte	13
1 Einleitung	15
1.1 Systematische Untersuchungen von Proteinfamilien	15
1.1.1 Proteine und deren Bedeutung	15
1.1.2 Integration biologischer Daten	16
1.2 Enzyme in der Industrie	20
1.2.1 Thiamindiphosphat-abhängige Enzyme	21
1.2.2 Alkoholdehydrogenasen	24
1.2.3 Cytochrom P450 Monooxygenasen	27
2 Motivation	29
3 Ergebnisse und Diskussion	31
3.1 Thiamindiphosphat-abhängige Enzyme	31
3.1.1 Faktoren die Aktivität, Selektivität und Substratspezifität der Thiamindiphosphat-abhängigen Enzyme Benzaldehydlyase und Benzoylformiatdecarboxylase beeinflussen.....	31
3.1.2 Rationales Proteindesign von ThDP-abhängigen Enzymen: Beeinflussung der Stereoselektivität	34
3.1.3 Pyruvatdecarboxylase aus <i>Acetobacter pasteurianus</i> : biochemische und strukturelle Charakterisierung	38
3.2 Alkoholdehydrogenasen	40
3.2.1 Die <i>Medium-Chain Dehydrogenase/Reductase Engineering</i> <i>Database</i> : Eine systematische Analyse einer diversen Proteinfamilie zum Verständnis der Sequenz-Struktur- Funktionsbeziehung.	40

3.3 Cytochrom P450 Monooxygenasen	43
3.3.1 Die Cytochrome P450 Engineering Database: Ein Orientierungs- und Vorhersagewerkzeug für die Cytochrom P450 Proteinfamilie.....	43
4 Darstellung der Ergebnisse als Publikations-manuskripte und Publikationen in englischer Sprache	47
4.1 Factors Mediating Activity, Selectivity, and Substrate Specificity for the Thiamin Diphosphate-Dependent Enzymes Benzaldehyde Lyase and Benzoylformate Decarboxylase.....	49
4.2 Rational Protein Design of ThDP-Dependent Enzymes: Engineering Stereoselectivity.....	67
4.3 Pyruvate Decarboxylase from <i>Acetobacter pasteurianus</i> : BIOCHEMICAL AND STRUCTURAL CHARACTERISATION	85
4.4 The Medium-Chain Dehydrogenase/Reductase Engineering Database: A systematic analysis of a diverse protein family to understand sequence structure-function relationship.....	123
4.5 The Cytochrome P450 Engineering Database: a navigation and prediction tool for the cytochrome P450 protein family.....	133
5 Abschließende Betrachtung.....	137
6 Gesamtliteraturverzeichnis	139
Danksagung.....	153
Erklärung.....	155

Zusammenfassung

Das Verständnis der Sequenz-Struktur-Funktionsbeziehung ist Voraussetzung, um experimentell beobachtete Unterschiede bei Enzymen hinsichtlich Selektivität und Substratspezifität erklären und verstehen zu können. Eine systematische Analyse der Enzyme innerhalb einer umfassenden Proteinfamilie verwandter Proteine ist zum Verständnis dieser Beziehung von Vorteil. Dadurch können in Sequenz und Struktur konservierte Bereiche aufgedeckt und familienspezifische Parameter abgeleitet werden. Im Rahmen dieser Arbeit wurde am Beispiel von drei verschiedenen Proteinfamilien eine systematische Untersuchung der Proteinfamilie oder einzelner Vertreter durchgeführt, um familienspezifische Regeln aufzustellen und diese für das Proteindesign zu nutzen. Unterschiedliche Substratspezifitäten und Selektivitäten konnten mit einfachen Modellen beschrieben werden. Für zwei der Proteinfamilien wurden Proteinfamiliendatenbanken etabliert, die eine erweiterte und umfassendere systematische Analyse dieser Proteinfamilien ermöglichen.

Für die Familie der Thiamindiphosphat-abhängigen Enzyme wurden systematische Strukturvergleiche verwandter Proteine durchgeführt, um Unterschiede bezüglich Selektivität und Substratspezifität zu erklären. Für die Familien der mittelkettigen Alkoholdehydrogenasen und der Cytochrom P450 Monooxygenasen wurden umfassende Proteinfamiliendatenbanken etabliert, die eine systematische Analyse dieser Familien ermöglichen.

Anhand der Strukturen der Benzaldehydlyase aus *Pseudomonas fluorescens* (BAL) und der Benzoylformiatdecarboxylase aus *Pseudomonas putida* (BFD) wurde für die Familie der Thiamindiphosphat-abhängigen Enzyme ein systematischer Vergleich der Substratbindetaschen durchgeführt. Als Ergebnis konnten einfache schematische Modelle der Form der Substratbindetaschen aufgestellt werden, die es erlaubten, die im Experiment beobachteten Unterschiede in Stereoselektivität und Substratspezifität hauptsächlich auf sterische Gründe zurückzuführen. In der Struktur der BFD konnte eine kleine Tasche identifiziert werden (*S-Pocket*), in welcher die Methylgruppe im Falle von Acetaldehyd als Akzeptoraldehyd Platz findet. Dies erlaubt eine Anordnung der Substrate, die in BFD hinsichtlich der Bildung von (*S*)-2-Hydroxypropiophenon zur Bildung des (*S*)-Produkts führen. Die Tasche ist in BFD hauptsächlich durch die Aminosäuren Pro24, Gly25, Ser26 und

Leu461 begrenzt. Da in der Struktur der BAL keine solche Tasche existiert, konnten die beobachteten Unterschiede auf die Existenz dieser Tasche zurückgeführt werden. Des Weiteren konnte die bei der BAL beobachtete Fähigkeit der enantioselektiven Spaltung von 2-Hydroxyketonen auf sterische Gründe zurückgeführt werden. Um das Konzept des *S-pocket* zu verifizieren, wurden für die BFD Mutanten vorgeschlagen (Leu461Gly und Leu461Ala), die eine Aufweitung des *S-pocket* zur Folge haben sollten, um die Bildung von (*S*)-2-Hydroxyphenylbutanon ((*S*)-2-HPB) zu ermöglichen. Die Partner des Instituts für Molekulare Enzymtechnologie der Heinrich-Heine-Universität Düsseldorf am Forschungszentrum Jülich konnten experimentell zeigen, dass ein Austausch zu Alanin in der Tat zu einer Enzymvariante führte, die in der Lage war (*S*)-2-HPB mit einem Enantiomerenüberschuss von 90% zu bilden, während im Wildtyp (*R*)-2-HPB mit einem Enantiomerenüberschuss von 21% gebildet wird.

Um dieses erfolgreiche Konzept auf weitere Familienmitglieder der Thiamindiphosphat-abhängigen Enzyme zu übertragen, wurden die Strukturen der Pyruvatdecarboxylase (PDC) aus *Zymomonas mobilis* (*ZmPDC*), der PDC aus *Acetobacter pasteurianus* (*ApPDC*) und der verzweigtkettigen Ketosäuredecarboxylase aus *Lactococcus lactis* (*KdcA*) hinsichtlich der Existenz eines *S-pocket* untersucht. Tatsächlich konnte in diesen Strukturen jeweils ein *S-pocket* gefunden werden, dessen Zugang jedoch stets durch sperrige Aminosäuren blockiert war, was die strikte (*R*)-Spezifität dieser Enzyme erklärt. Um das Konzept des *S-pocket* auf weitere Enzyme anzuwenden, muss dieses insofern erweitert werden, dass auch der Zugang zu dieser Tasche mit in die Untersuchungen einbezogen wird.

Um das Verständnis der Beziehung zwischen der Form der Bindetasche und den experimentellen Beobachtungen auszuweiten, wurde am Beispiel der *ApPDC* ein ausführlicher Vergleich der Bindetasche zu Strukturen der Pyruvatdecarboxylase aus *Saccharomyces cerevisiae* (*ScPDC*), der *ZmPDC*, der BAL, der BFD und der *KdcA* durchgeführt. Es konnte wiederum anhand einfacher Modelle gezeigt werden, dass in der *ApPDC* und in der *ZmPDC* ein Tryptophanrest (Trp388, beziehungsweise Trp392) den Raum der Donorbindestelle maßgeblich einschränkt. Dies erklärt die für diese Enzyme beobachtete Präferenz gegenüber aliphatischen Donoraldehyden. Für die *ApPDC* konnte zusätzlich die geringere Stereoselektivität im Vergleich zu den andern untersuchten Enzymen dadurch erklärt werden, dass der Zugang zum *S-pocket* der *ApPDC* aufgrund von lokalen strukturellen Unterschieden nicht völlig blockiert ist, und so ein geringes Maß an (*S*)-Produkten gebildet werden kann.

Durch die systematischen Untersuchungen der Bindetaschen konnte ein besseres Verständnis der Sequenz-Struktur-Funktionsbeziehung dieser Enzymfamilie erzielt werden. Diese Kenntnisse können helfen neue Enzymvarianten zu generieren, die in der Lage sind unterschiedliche (S)-Produkte darzustellen.

Um eine systematische Analyse einer kompletten Enzymfamilie in einem einheitlichen Format zu ermöglichen, wurde das Datenbanksystem DWARF (Fischer et al. 2006) im Rahmen dieser Arbeit weiterentwickelt und die *Medium-Chain Dehydrogenase/Reductase Engineering Database* (MDRED, <http://www.mdred.uni-stuttgart.de>) etabliert. Die MDRED umfasst 6420 Sequenzeinträge für 2684 Proteine, für die funktionell wichtige Aminosäuren in der Datenbank annotiert sind. Für 42 Proteine wurden 257 Strukturdatensätze gespeichert. Die Proteine wurden nach Sequenzähnlichkeit in 29 Superfamilien und 199 homologe Familien klassifiziert. Die Superfamilien wurden abhängig vom Vorhandensein eines katalytischen Zinkatoms in zwei Klassen eingeteilt (*zinc-containing MDRs* und *non-zinc-containing MDRs*). Die Klasse der *zinc-containing MDRs* konnte, basierend auf Unterschieden in der Länge eines strukturell variablen Elements (QSDL = *quaternary structure determining loop*), weiter unterteilt werden (kurze QSDL, mittlere QSDL und lange QSDL). Ein Zusammenhang der Proteine mit bekannter Struktur der jeweiligen Klassen und der Quartärstruktur der Proteine konnte bis auf eine Ausnahme nachgewiesen werden. Somit ermöglicht die Klassifizierung eine Vorhersage der Quartärstruktur für mehr als 2000 weitere Proteine innerhalb der Familien. Zusätzlich konnten durch die Klassifizierung in die einzelnen Klassen abhängig von der Länge des QSDL Sequenzmotive aufgedeckt werden, die innerhalb von QSDL lokalisiert sind, sich aber in den beiden Klassen unterscheiden. In diesen Sequenzmotiven befinden sich Aminosäuren, die wesentlich an der Form der Bindetasche beteiligt sind. Für die bekannte Alkoholdehydrogenase der Pferdeleber, die zur Klasse der langen QSDL gehört, sind dies zum Beispiel die Aminosäuren Phe140 und Leu141.

Für die Klasse der langen QSDL konnte außerdem eine weitere konservierte Aminosäure identifiziert werden, die für die Form der Bindetasche einer zusätzlichen Region verantwortlich ist. Die relevanten Aminosäuren wurden in den Multisequenzalignments der Familien annotiert und können so auf eine große Anzahl an Enzymen innerhalb der Familien übertragen werden. Dies ermöglicht ein einfaches Auffinden von *hotspots* für Mutationsvorschläge zur Aufweitung der Bindetasche zur Veränderung der Substratspezifität.

Als weitere Proteinfamiliendatenbank wurde die *Cytochrome P450 Engineering Database* (CYPED, <http://www.cyped.uni-stuttgart.de>) etabliert. Die CYPED enthält Sequenzdaten zu 3911 Cytochrom P450 Monooxygenasen. Die Proteine sind nach der Klassifizierung von Nelson (Nelson 2006) in Superfamilien und homologe Familien unterteilt. Funktionell wichtige Aminosäuren wurden in den Multisequenzalignments annotiert. Die CYPED ermöglicht durch die Integration von Sequenz- und Strukturdaten in einem einheitlichen Format eine systematische Analyse dieser großen Enzymfamilie und trägt daher zum Verständnis der familienspezifischen Eigenschaften bei. Wie auch für die MDRED sind die Multisequenzalignments der Familien, die phylogenetischen Bäume, familienspezifische Profile und eine BLAST-Schnittstelle zur Klassifizierung neuer Proteine über das Webinterface zugänglich.

Abschließend betrachtet konnte im Rahmen dieser Arbeit gezeigt werden, dass der Ansatz einer systematischen Familienanalyse oder der Analyse eines bestimmten Enzyms im Vergleich zu nah verwandten Proteinen zum Verständnis der Sequenz-Struktur-Funktionsbeziehung wesentlich beitragen kann. Abgeleitete Regeln und Erkenntnisse können direkt zur Vorhersage von Enzymvarianten mit verbesserten Eigenschaften herangezogen werden.

Summary

The knowledge of sequence, structure, and function relationship of enzymes is prerequisite to understand and explain experimentally observed differences in substrate specificity and selectivity. This is best achieved by investigating an enzyme in the framework of its complete enzyme family in order to derive family-specific parameters. Derived rules from a systematic analysis of protein families can then subsequently be applied for biocatalyst design. In this study, systematic family analysis was performed for three different protein families, and family-specific parameters were derived and applied to predict enzymes with improved properties. Simple models were established, explaining differences in substrate specificity and selectivity. For two of the investigated families protein family databases were established to enable a systematic family-specific analysis of these huge families.

For the family of thiamin diphosphate (ThDP)-dependent enzymes systematic structure comparisons were performed, whereas protein family databases were established for the family of medium-chain dehydrogenases/reductases (MDRs) and for the family of cytochrome P450 monooxygenases.

Systematic and comprehensive comparison and characterization of the binding sites of benzaldehyde lyase (BAL) from *Pseudomonas fluorescens* and benzoylformate decarboxylase (BFD) from *Pseudomonas putida* was performed for the family of thiamin diphosphate (ThDP)-dependent enzymes, in order to explain experimentally observed differences in stereoselectivity and substrate specificity. Molecular models of the shape of the binding site of each protein were derived, demonstrating that predominantly sterical reasons are responsible for the observed differences in 2-hydroxypropiophenone formation. Within the binding site of BFD, a small pocket (*S-pocket*) was identified that offers just enough space for a methyl group. This pocket allows the substrate in the case of acetaldehyde as acceptor aldehyde to be oriented relative to the donor aldehyde in a way, that the (*S*)-product can be formed. The *S-pocket* in BFD is mainly shaped by Pro24, Gly25, Ser26 and Leu461. In BAL no such *S-pocket* was observed, which explains the strict (*R*)-specificity of this enzyme. Due to the characterization of the binding sites, the ability of enantioselective cleavage of 2-hydroxy ketones with BAL could also be attributed to the architecture of the binding site. To prove the concept of the *S-pocket*, mutations for BFD (Leu461Ala and Leu461Gly) were

suggested in order to increase the space of the *S-pocket* to allow (*S*)-2-hydroxy-1-phenylbutan-1-one to be built. It was experimentally shown by the partners from the Institute of Molecular Enzyme Technology at the Heinrich-Heine University of Düsseldorf, that substitution of leucine at position 461 to alanine indeed resulted in a variant, which was able to ligate benzaldehyde and propionaldehyde to the (*S*)-product with an enantiomeric excess of 90 %, which is not observed for BFD wild-type. In order to transfer this successful concept to other ThDP-dependent enzymes, available structures from benzaldehyde lyase from *Pseudomonas fluorescens* (BAL), pyruvate decarboxylase from *Zymomonas mobilis* (*ZmPDC*), pyruvate decarboxylase from *Acetobacter pasteurianus* (*ApPDC*), and branched-chain 2-keto acid decarboxylase from *Lactococcus lactis* (*KdcA*) were compared to examine the existence of a *S-pocket*. In fact, for all enzymes except of BAL, the existence of a *S-pocket* could be verified, although the entrance to the *S-pocket* is blocked by bulky residues in all investigated structures. This explained the observed (*R*)-specificity of these enzymes. Thus, the concept of the *S-pocket* was extended, as the entrance to the *S-pocket* has to be included for future approaches.

For deeper understanding of these proteins apart from the concept of the *S-pocket*, a detailed characterization of *ApPDC* has been carried out by systematic comparison to pyruvate decarboxylase from *Saccharomyces cerevisiae* (*ScPDC*), *ZmPDC*, BAL, BFD, and *KdcA*. All amino acids shaping the binding site were derived, where a bulky tryptophan residue within the donor binding site was found for *ApPDC* (Trp388) and *ZmPDC* (Trp392) limiting the space for bulky donor aldehydes. This explained sufficiently the generally observed preference of these PDCs for aliphatic aldehydes as donor. A detailed comparison of the *S-pocket* of *ApPDC* with others could further explain the higher amount of (*S*)-products observed for the wild-type of *ApPDC*, as the *S-pocket* revealed not completely blocked as compared to other enzyme structures due to side chain flexibility.

The insight gained from this systematic investigations helps to understand the sequence, structure, and function relationship of this family, and can be applied to design new enzyme variants.

To facilitate a systematic analysis of data in a consistent format of a complete protein family, the database system DWARF (Fischer et al. 2006) was applied and further extended in this study, to establish the Medium-Chain Dehydrogenase/Reductase Engineering Database (MDRED, <http://www.mdred.uni-stuttgart.de>). The MDRED comprises 6420 sequence entries of 2684 proteins with consistently annotated functionally relevant residues. For 42 proteins

257 structure data sets from the Protein Data Bank were included. The proteins were assigned to 29 superfamilies and 199 homologous families based on sequence similarity. The superfamilies were assigned either to the class of zinc-containing MDRs or non-zinc-containing MDRs dependent on the presence of a catalytic zinc atom. Superfamilies of the class of zinc-containing MDRs were further sub-classified depending on the length of a structural variable region (QSDL = quaternary structure determining loop). A correlation of multimeric architecture and the members of one of the classes (short QSDL, medium QSDL, and long QSDL) was found for proteins with known structures with only one exception. Thus, multimeric state can be predicted for more than 2000 MDR proteins. It has been shown, that residues shaping the binding site are located within QSDL are embedded in different sequence motifs for each class. Thus, sub-dividing of zinc-containing MDRs revealed family-specific properties for the QSDL classes, which would have been difficult to detect without the performed systematic classification. These relevant residues are for example Phe140 and Leu141 for the well known horse liver alcohol dehydrogenase. The binding site shaping residues were annotated within the multiple sequence alignments, and thus can be transferred to other family members. This enables a fast prediction of mutation hotspots for a huge number of MDR proteins in order to broaden the binding site for altered substrate specificity.

Further, the Cytochrome P450 Engineering Database (CYPED, <http://www.cyped.uni-stuttgart.de>) has been established. The CYPED comprises sequence data for 3911 cytochrome P450 monooxygenases. Proteins are assigned to superfamilies and homologous families according to Nelson (Nelson 2006), and functionally relevant residues are annotated and transferred to other family members. Since the CYPED integrates data on sequence and structure in a consistent format, the CYPED enables a systematic analysis of this huge family to understand sequence, structure, and function relationship. As for the MDRED, family multiple sequence alignments, phylogenetic trees, family-specific profiles, and a BLAST interface to classify new proteins are online accessible.

In conclusion it could be shown, that a systematic approach of protein family analysis and enzyme comparison is valuable to deepen the understanding of sequence, structure, and function relationship of large and diverse protein families. Derived rules can be applied to predict mutation hotspots to pave the way for designing enzymes with improved properties.

Publikationen und Publikationsmanuskripte

Die vorliegende Arbeit umfasst folgende Publikationen und Publikationsmanuskripte:

1. Knoll, M., Müller, M., Pleiss, J., Pohl, M., 2006. Factors mediating activity, selectivity, and substrate specificity for the thiamin diphosphate-dependent enzymes benzaldehyde lyase and benzoylformate decarboxylase. *ChemBioChem* **7**: 1928-1934
2. Gocke, D., Walter, L., Gauchenova, E., Kolter, G.K., Knoll, M., Berthold, C.L., Schneider, G., Pleiss, J., Müller, M., Pohl, M., 2008. Rational protein design of ThDP-dependent enzymes: engineering stereoselectivity. *ChemBioChem* **9**: 406-412
3. Gocke, D., Berthold C. L., Graf, T., Brosi, B., Frindi-Wosch, I., Knoll, M., Stillger, T., Walter, L., Müller, M., Pleiss, J., Schneider, G., Pohl, M., Pyruvate Decarboxylase from *Acetobacter pasteurianus*: BIOCHEMICAL AND STRUCTURAL CHARACTERISATION. (eingereicht bei *BBA: Proteins & Proteomics*)
4. Knoll, M., Pleiss, J., 2008. The Medium-Chain Dehydrogenase/Reductase Engineering Database: A systematic analysis of a diverse protein family to understand sequence-structure-function relationship. *Protein Sci* **17**: 1689-1697
5. Fischer, M. Knoll, M., Sirim, D., Wagner, F., Funke, S., Pleiss, J., 2007. The Cytochrome P450 Engineering Database: a navigation and prediction tool for the cytochrome P450 protein family. *Bioinformatics* **23**: 2015-2017

Zusätzliche Publikationen, die nicht in dieser Arbeit enthalten sind:

1. Messerschmidt, S.K.E, Kolbe, A., Müller, D., Knoll, M., Pleiss, J, Kontermann, R.E., 2008. Novel single-chain Fv' formats for the generation of immunoliposomes by site-directed coupling. *Bioconjugate Chem.* **19**: 362-369
2. Soliman, N.A., Knoll, M., Abdel-Fattah, Y.R., Schmid, R.D., Lange, S., 2007. Molecular cloning and characterization of thermostable esterase and lipase from *Geobacillus thermoleovorans* YN isolated from desert soil in Egypt. *Process Biochem* **42**: 1090-1100

1 Einleitung

1.1 Systematische Untersuchungen von Proteinfamilien

1.1.1 Proteine und deren Bedeutung

Proteine spielen durch ihre charakteristische Struktur oder ihre katalytischen Fähigkeiten (Enzyme) in nahezu allen biochemischen und physiologischen Prozessen eines lebenden Organismus eine bedeutende Rolle. Besonders die Enzyme haben in der letzten Zeit als Biokatalysatoren in der industriellen Anwendung aufgrund ihrer Substratspezifitäten und Enantioselektivitäten zunehmend an Bedeutung gewonnen. Ein wichtiges Anwendungsgebiet der industriell verwendeten Enzyme findet sich in der chemischen Industrie, wo sie als Biokatalysatoren die Synthese verschiedener (Fein-) Chemikalien und deren Vorstufen katalysieren.

Die Natur stellt zwar bereits eine eindrucksvolle Bandbreite an verschiedenen Biokatalysatoren bereit, diese sind aber meist von der Natur so optimiert, dass sie dem jeweiligen Organismus eine bestmögliche Anpassung an die Lebensbedingungen bieten. Für den Einsatz in großtechnischen Prozessen werden jedoch häufig Biokatalysatoren mit spezifischen Eigenschaften benötigt. Hierzu zählen zum Beispiel hohe spezifische Aktivitäten und Selektivitäten auch gegenüber unnatürlichen Substraten sowie breite Substratspektren. Erfolgreiche Beispiele der Anpassung von Enzymeigenschaften an industrielle Anforderungen durch *protein engineering* sind zahlreich beschrieben worden (Arnold 2001; Bornscheuer and Pohl 2001). So sind zum Beispiel wesentliche Verbesserungen und Veränderungen der Enantioselektivität (Liebeton et al. 2000; May et al. 2000; Reetz and Jaeger 2000), der Stereoselektivität (Scheib et al. 1998), des Substratspektrums (Yano et al. 1998) sowie die Einführung neuartiger Funktionen in Enzyme gelungen (Arnold 2001).

Um jedoch Enzyme gezielt an verschiedene Anforderungen anzupassen, ist das Verständnis der Sequenz-Struktur-Funktionsbeziehung essentiell, damit funktionsrelevante Aminosäuren aufgedeckt werden, und das Wissen auf verwandte Proteine übertragen werden kann. Die Übertragung und Zuweisung von Funktion einzelner Aminosäuren oder aber ganzer Proteine bleibt allerdings bis heute eine der wichtigsten Aufgaben der modernen Molekularbiologie

(Kim et al. 2003; Wolfson et al. 2005; Lee et al. 2007). Zum Verständnis der Sequenz-Struktur-Funktionsbeziehung eignet sich eine systematische Analyse einer Proteinfamilie. Die hierbei abgeleiteten Regeln können dann gezielt zur Vorhersage von Enzymvarianten eingesetzt werden.

1.1.2 Integration biologischer Daten

Die rasant anwachsende Datenflut an biologischen Sequenz- und Strukturdaten bedingt durch zahlreiche Genomsequenzierungsprojekte und das Forschungsfeld der strukturellen Genomik (Gaasterland 1998) stellt die Wissenschaft in den letzten Jahren vor eine große Herausforderung (Gerstein 2000). Die großen Datenmengen sind jedoch eine gute Grundlage für die systematische Analyse ganzer Proteinfamilien zur Identifizierung von Funktionen bestimmter Bereiche eines Proteins oder ganzer Proteine. Um verlässlich die Funktionen vorhersagen zu können, und diese einzelnen Bereichen zuweisen zu können (Annotationen), ist das Verständnis der Sequenz-Struktur-Funktionsbeziehung wichtig. Dies kann nur durch den Vergleich verwandter (homologer) Proteine innerhalb einer definierten Proteinfamilie erreicht werden. Die Analysen innerhalb von Proteinfamilien stützen sich dabei auf die Annahmen, dass Proteine, die eine Sequenzähnlichkeit zueinander aufweisen und miteinander verwandt sind, auch eine ähnliche Struktur besitzen (Chothia and Lesk 1986) und funktionelle Verwandtschaft zeigen (Lee et al. 2007). Die Klassifizierung in biologisch relevante Familien ist dabei ein wichtiger Schritt. So kann zum Beispiel anhand von Struktureigenschaften, Sequenzähnlichkeit oder Funktion klassifiziert werden.

Die anfallenden Sequenz- und Strukturdaten sind in öffentlichen Sequenz- und Strukturdatenbanken abgelegt. Die wohl umfassendste Datenbank stellt die *Genbank* dar (Benson et al. 2007), die insgesamt über 52 Millionen abgelegte Nukleotid- und Proteinsequenzen enthält (Stand Januar 2008). Neben der *Genbank*, die eine reine Sequenzdatenbank darstellt, sind in der *Uniprot*-Datenbank (Wu et al. 2006) Querverweise auf andere Datenbankeinträge, Literaturverweise und Funktionszuweisungen für die jeweiligen Einträge zusammengeführt. In der UniProt/Swiss-Prot zum Beispiel findet man über 290 000 annotierte Proteinsequenzen (Stand Januar 2008). Die Zahl der in der *Protein Data Bank* (PDB, <http://www.pdb.org>, (Berman et al. 2000)) hinterlegten Strukturen ist dabei mit ungefähr 44 000 Proteinstrukturen (Stand Januar 2008) sehr gering. Gerade diese Lücke zwischen der Zahl der bekannten Proteinstrukturen und der großen Anzahl bekannter

Proteinsequenzen stellt für die Funktionszuweisung und das Verständnis der Sequenz-Struktur-Funktionsbeziehung ein Problem dar.

Neben den reinen Sequenz- und Strukturdatenbanken existieren Datenbanken, die Familien anhand von Sequenzähnlichkeiten oder familienspezifischer Diskriptoren, wie zum Beispiel Sequenzmotive, Sequenzprofile oder funktionellen Gruppen (Proteindomänen) beschreiben.

Eine umfassende Datenbank an Proteinfamilien und Proteindomänen stellt die *InterPro* (Mulder et al. 2007) dar. Die *InterPro* teilt Proteineinträge anhand konservierter Domänen oder funktioneller Bereiche ein. Die *InterPro* umfasst unter anderem den Datenbestand von Pfam (*Protein Families Database*), PROSITE, PRINTS und ProDom (Servant et al. 2002). In der Bibliothek von PROSITE (Hulo et al. 2006) befinden sich Sequenzmuster hoch konservierter Aminosäuren, welche anhand regulärer Ausdrücke funktionell wichtige Bereiche beschreiben. Diese Sequenzmuster geben allerdings nur Auskunft, ob ein bestimmtes Protein dieses Muster enthält oder nicht. Eine Bewertung der Relevanz eines gefundenen Musters fehlt dabei. Ein erweiterter Ansatz zur Beschreibung einer Proteinfamilie ist zum Beispiel die Datenbank PRINTS (Attwood et al. 2003). PRINTS ist eine Sammlung von sogenannten *fingerprints*, die konservierte Motive aus Multisequenzalignments beschreiben und auch Aminosäurevariationen innerhalb einer Spalte des Multisequenzalignments berücksichtigen. Durch die größere Flexibilität der *fingerprint*-Motive ist es möglich, auch entfernter verwandte Proteinsequenzen zu identifizieren. Einen Schritt weiter geht die Beschreibung verwandter Proteine mithilfe von Profilen, die Wahrscheinlichkeiten für eine Position für das Auftreten einer bestimmten Aminosäure, einer Insertion oder einer Deletion wiedergeben. Diese Profile können zur Klassifizierung von Proteinen verwendet werden (Krogh et al. 1994). Ein Beispiel der Anwendung ist die Datenbank *Pfam* (Finn et al. 2008), die konservierte Proteindomänen beinhaltet.

Neben den beschriebenen Ansätzen zur Klassifizierung von Proteinen über familienspezifische Diskriptoren existieren auch Datenbanken, die globale Sequenzvergleiche zur Einteilung der Proteine in Familien heranziehen. Hier wäre insbesondere die PROTOMAP (Yona et al. 2000) und die SYSTERS Datenbank (Krause et al. 2005) zu nennen. Beide versuchen Proteinsequenzen mithilfe von Distanzgraphen in definierte Familien zu klassifizieren. SYSTERS verwendet zusätzlich einen *single-linkage* Clusteralgorithmus, der ebenfalls Ähnlichkeiten zwischen Proteinsequenzen abschätzt. Ein

strukturbasierter Ansatz zur Klassifizierung bekannter Proteinstrukturen ist zum Beispiel in der SCOP (*Structural Classification of Proteins*) (Murzin et al. 1995) verwirklicht. Die SCOP bietet eine manuell validierte Einteilung aller bekannter Proteinstrukturen nach evolutionärer und struktureller Verwandtschaft. Ein weiterer Ansatz zur Klassifizierung von bekannten Proteinstrukturen ist die CATH-Datenbank (Orengo et al. 1997). Darin werden alle bekannten Strukturen in vier Klassen, abhängig ihrer Sekundärstrukturelemente eingeteilt.

Nicht allein die große Datenmenge und die Klassifizierung der Sequenz- und Strukturdaten ist ein Problem, sondern auch die ständige Zunahme der Diversität der Daten. So wächst die Menge an biochemischen Daten (Aktivitäten, Selektivitäten, Mutationseffekte, Löslichkeit,...) für bestimmte Enzyme ständig. Um komplexe biologische Zusammenhänge zu verstehen, ist es zwar generell für die systematische Analyse einer Proteinfamilie äußerst hilfreich verschiedene Arten biologischer Information zu haben, doch gerade die Integration solcher Daten stellt eine große Herausforderung dar. Einen Ansatz zur Integration solcher Daten ist die BRENDA-Datenbank (BRAunschweig ENzyme DAtabase) (Schomburg et al. 2004). Die BRENDA-Datenbank umfasst unter anderem Daten zu Substratspezifitäten, Selektivitäten und Aktivitäten für eine Vielzahl bisher klassifizierter Enzyme. Darüber hinaus sind Informationen über Mutationseffekte und Literaturverweise verfügbar.

Eine systematische Analyse einer definierten Proteinfamilie zum Verständnis familienspezifischer Eigenschaften anhand der existierenden Datenbanken wird allerdings dadurch erschwert, dass das Ziel dieser Datenbanken darin liegt, einen möglichst großen Sequenzraum abzudecken. Darüber hinaus liegen die vorhandenen Informationen über Aktivitäten, Substratspezifitäten oder Selektivitäten in einem uneinheitlichen Format, und daher nicht systematisch vergleichbar vor (Gerstein 2000; Hansen 2006). Die Konzentration auf die Integration der notwendigen Informationen, die für das Verständnis der Sequenz-Struktur-Funktionsbeziehung einer definierten Proteinfamilie notwendig sind, kann dieses Problem lösen.

Ein Ansatz ist die lokale Integration der benötigten Daten einer definierten Proteinfamilie in einem *Data Warehouse System*. Durch das Verwenden eines solchen Systems wird eine lokale Speicherung der benötigten Daten in einem einheitlichen Format ermöglicht. Zwar muss ein einheitliches Datenmodell erst entwickelt und die Daten aus den Quelldatenbanken erst extrahiert und in ein einheitliches Format übersetzt werden, das System ist dann aber

unabhängig von den Quelldatenbanken. Am Institut für Technische Biochemie der Universität Stuttgart wurde zu diesem Zwecke das Datenbanksystem DWARF (Data Warehouse for Analysing Protein Families) (Fischer et al. 2006) entwickelt, das es ermöglicht Sequenz- und Strukturdaten in einem lokalen System zu integrieren. Dadurch wird eine systematische Analyse der Proteinfamilien und damit das Verständnis der Zusammenhänge ermöglicht. Das Datenbanksystem wurde bereits erfolgreich zur Erstellung der *Lipase Engineering Database* (LED, <http://www.led.uni-stuttgart.de>) (Pleiss et al. 2000; Fischer and Pleiss 2003) eingesetzt, die eine umfassende Anzahl an Proteinen des Faltungsmusters der α/β -Hydrolasen beinhaltet.

Durch Untersuchungen eines Proteins im Kontext einer wohl definierten Proteinfamilie oder zumindest im Vergleich zu homologen Proteinen lassen sich familienspezifische Diskriptoren und konservierte Bereiche auf Struktur- und Sequenzebene identifizieren und gegebenenfalls mit einer bestimmten Funktion in Verbindung bringen, was zum Verständnis familienspezifischer Eigenschaften beiträgt. Von Vorteil ist bei einem solchen Ansatz das Vorhandensein einer aufgeklärten Proteinstruktur innerhalb der Familie (Wolfson et al. 2005; Leisola and Turunen 2007) um funktionelle Annotationen einzelner Aminosäuren zu ermöglichen (zum Beispiel Zuweisung der katalytisch aktiven Aminosäuren) (Gerstein 2000; Thornton et al. 2000).

Oft können komplexe experimentell beobachtete Phänomene schon durch sehr einfache Regeln beschrieben werden. Ein Beispiel hierfür sind die durch systematische Familienanalyse identifizierten Sequenzmotive GX und GGGX der Lipasen. Diese Sequenzmotive beschreiben die Architektur des *Oxyanion Hole*, welches das Substrat im Übergangszustand stabilisiert. Untersuchungen zeigten, dass das Vorhandensein des GGGX-Motivs bei Lipasen entscheidend für deren Fähigkeit ist, tertiäre Alkohole umzusetzen (Henke et al. 2002). Durch Kombination von computergestützter Vorhersage und experimenteller Validierung können Informationen über die Gültigkeit solcher Vorhersagen gewonnen werden, die dazu beitragen, abgeleitete Regeln zur Unterstützung des Proteindesigns weiter zu etablieren (Wolfson et al. 2005). Diese Regeln, die zum Verständnis der Sequenz-Struktur-Funktionsbeziehung beitragen, können dann prädiktiv zur gezielten Veränderung der Substratspezifitäten und Selektivitäten der in industriellen Prozessen eingesetzten Enzyme angewandt werden.

1.2 Enzyme in der Industrie

Alle Enzyme werden abhängig von den von ihnen katalysierten Reaktionen nach dem System der *International Union of Pure and Applied Chemistry* (IUPAC) beziehungsweise der *International Union of Biochemistry and Molecular Biology* (IUBMB) in sechs Klassen gruppiert (Tabelle 1.1). Die Einteilung in die verschiedenen Klassen erfolgt mithilfe des von der *Enzyme Commission* (EC) etablierten EC-Nummern-Systems, welches die vom Enzym katalysierten Reaktionen katalogisiert (Liese et al. 2006).

Tabelle 1.1: Klassifizierung der Enzyme

EC-Nummer	Enzymklasse	Reaktion
1.x.x.x	Oxidoreduktasen	Redoxreaktionen
2.x.x.x	Transferasen	Transfer funktioneller Gruppen
3.x.x.x	Hydrolasen	Hydrolytische Spaltung
4.x.x.x	Lyasen	Hinzufügen und Entfernung funktioneller Gruppen (zur Bildung von Doppelbindungen)
5.x.x.x	Isomerasen	Isomerisierungsreaktionen
6.x.x.x	Ligasen	Bindungsknüpfung (unter Triosephosphatspaltung)

Neben den Hydrolasen, die für den industriellen Einsatz die bedeutendste Enzymklasse darstellen, sind vor allem die Oxidoreduktasen und Lyasen von großer Bedeutung (Straathof et al. 2002). Oxidoreduktasen reagieren mit einem Substrat mittels Elektronentransfer. In den meisten Fällen wird ein Cofaktor oder ein Coenzym benötigt, das als Akzeptor fungiert. Eine wichtige Gruppe von Oxidoreduktasen für die industrielle Anwendung ist die Gruppe der Dehydrogenasen, welche aufgrund ihrer Enantio- und Regioselektivität, und ihres breiten Substratspektrums zur Darstellung zahlreicher chiraler Ketone oder Alkohole eingesetzt werden. Neben den Dehydrogenasen spielen die Cytochrom P450 Monooxygenasen als weitere Gruppe der Oxidoreduktasen eine immer größere Rolle. Die Cytochrom P450 Monooxygenasen, die ebenfalls im Rahmen dieser Arbeit untersucht wurden, sind für die industrielle Anwendung vor allem zur regio- und stereospezifischen Hydroxylierung von aromatischen Substraten wichtig (Steroidsynthese) (Schmid 2002; Liese et al. 2006). Darüber

hinaus spielen sie im Abbau von körperfremden Stoffen (zum Beispiel Pharmaka) eine zentrale Rolle. Die ebenfalls im Rahmen dieser Arbeit behandelten Thiamindiphosphat-abhängigen Enzyme gehören zu der Klasse der Lyasen. Sie sind in vielen industriellen Prozessen aufgrund ihrer Fähigkeit C-C-, C-O- oder C-N-Bindungen spalten und knüpfen zu können von großer Bedeutung.

1.2.1 Thiamindiphosphat-abhängige Enzyme

Thiamindiphosphat (ThDP)-abhängige Enzyme kommen in den unterschiedlichsten Organismen vor und katalysieren eine Vielzahl an Reaktionen. Von besonderer Bedeutung ist ihre Fähigkeit C-C-Knüpfungen (enantioselektiv) bilden zu können. Durch die Verknüpfung zweier Aldehyde können mittels ThDP-abhängiger Enzyme so 2-Hydroxyketone gebildet werden. 2-Hydroxyketone stellen wichtige Synthesebausteine für pharmakologisch aktive Verbindungen dar (Jordan 2003; Dünkermann et al. 2004; Faber and Kroutil 2005). Ein bekanntes Beispiel ist das biotechnologische Verfahren zur Synthese des 2-Hydroxyketons (*R*)-Phenylacetylcarbinol ((*R*)-PAC) mittels ThDP-abhängigen Enzymen aus Hefezellen. (*R*)-PAC dient als chirale Vorstufe zur Herstellung des Antihistaminikum L-Ephedrin (Hildebrandt and Klavehn 1930).

Der Cofaktor ThDP leitet sich vom Thiamin (Vitamin B1) ab. Viele Organismen (darunter der Mensch) müssen täglich solche Cofaktoren oder ihre Vorstufen in Form von Vitaminen mit der Nahrung aufnehmen. Die Mangelerkrankung von Vitamin B1 zum Beispiel äußert sich unter anderem im Krankheitsbild der seit mehreren hundert Jahren bekannten Beri-Beri-Krankheit. Das ThDP ist in ThDP-abhängigen Enzymen tief in einer hydrophoben Bindetasche gebunden. Dabei ist das ThDP nicht kovalent an das Protein gebunden, sondern wird durch Komplexbildung mit zweiwertigen Metallionen (meistens Mg^{2+} und Ca^{2+}) stabilisiert. In der Bindetasche liegt das ThDP in einer typischen V-Konformation vor, welche durch Wechselwirkungen mit hoch konservierten hydrophoben Seitenketten bedingt wird. Für die katalytische Reaktion ragt so lediglich der Thiazoliumring in die Bindetasche hinein, wodurch das reaktive Kohlenstoffatom (C2-Atom) gut zugänglich in der Bindetasche positioniert ist (Abb. 1.1) (Jordan 2003).

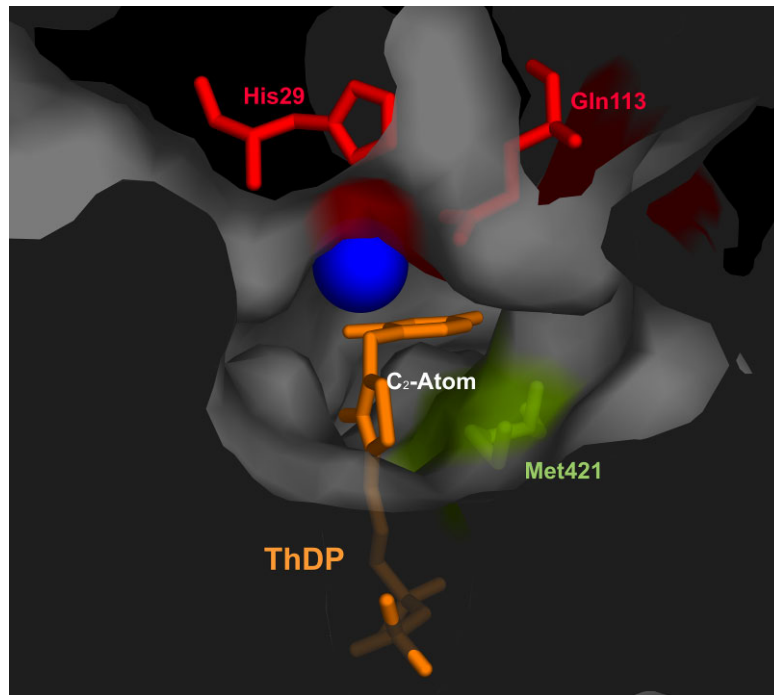


Abbildung 1.1: Substratbindetasche der Thiamindiphosphat-abhängigen Enzyme am Beispiel der Benzaldehydlyase (BAL). Die Abbildung zeigt einen Ausschnitt der Bindetasche der BAL (PDB-Eintrag 2AG0 (Mosbacher et al. 2005)). Die katalytisch wichtigen Aminosäuren His29 und Gln113 sind rot dargestellt, die grün dargestellte Aminosäure M421 stabilisiert die V-Konformation des Cofaktors ThDP (orange).

Trotz der zahlreichen unterschiedlichen Reaktionen, die durch ThDP-abhängige Enzyme katalysiert werden, liegt ihnen ein gemeinsames Reaktionsprinzip zugrunde (Abb.1.2) (Schellenberger 1998). Für die Katalyse der Knüpfung von C-C-Bindungen entsteht durch Bindung des Carbonylkohlenstoffs des Substrats an das C2-Atom des Thiazoliumrings des ThDP ein hoch aktives Intermediat – der so genannte „aktivierte Aldehyd“. Die Bindung des Substrats an das ThDP erfolgt unter Abspaltung der Abgangsgruppe. Der „aktivierte Aldehyd“ dient als Donor und kann im Folgenden unter Freisetzung des Cofaktors mit einem Akzeptor(aldehyd) reagieren (Jordan 2003).

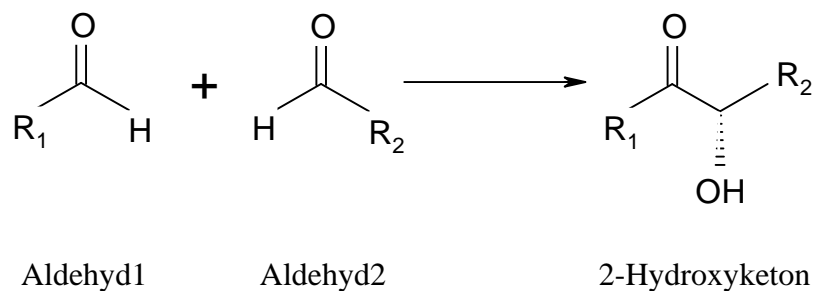


Abbildung 1.2: Schematische Darstellung des Reaktionsprinzips zur Verknüpfung zweier Aldehyde

Für die katalysierten Reaktionen müssen also zwei Substrate (Donor und Akzeptor) gleichzeitig in der Bindetasche Platz finden. Durch die räumlichen Gegebenheiten der Bindetasche kann man eine Donorbindestelle sowie eine Akzeptorbindestelle unterscheiden. Die Substratspezifität der einzelnen Enzyme kann daher durch die Aminosäuren, die diese Bindestellen räumlich begrenzen, beeinflusst werden.

Zur gezielten chemo- und enantioselektiven Herstellung von 2-Hydroxyketonen sind bereits einige Enzyme, darunter die Benzoylformiatdecarboxylase (BFD), die verzweigt-kettige Ketosäuredecarboxylase (KdcA) und Benzaldehydlyase (BAL) beschrieben (Demir et al. 2002; Pohl et al. 2002; Demir et al. 2003; Dünkermann et al. 2004; Pohl et al. 2004; Siegert et al. 2005; Dominguez de Maria et al. 2006; Stillger et al. 2006; Dominguez de Maria et al. 2007; Gocke et al. 2007). Aufgrund ihrer Fähigkeit, eine Vielzahl an Aldehyden zu verknüpfen, stellen die BFD und die BAL wertvolle Biokatalysatoren in der organischen Synthese dar (Demir et al. 1999; Dünnwald et al. 2000; Dünnwald and Müller 2000; Iding et al. 2000; Demir et al. 2001; Demir et al. 2002; Dünkermann et al. 2002; Lingen et al. 2002; Lingen et al. 2003; Sanchez-Gonzalez and Rosazza 2003; Hischer et al. 2005; Siegert et al. 2005). Darüber hinaus sind Pyruvatdecarboxylasen (PDCs) aus verschiedenen Organismen beschrieben. PDCs stellen die Schlüsselenzyme im fermentativen Stoffwechsel dar, da sie die Spaltung von Pyruvat zu Acetaldehyd und Kohlendioxid katalysieren. Bei der Herstellung von Biodiesel aus leicht zugänglichen pflanzlichen Materialien wie Zuckermoleküle oder Zellulose spielen PDCs eine große Rolle. Zurzeit werden vor allem Stämme mit angepassten Ethanolstoffwechselwegen für die industrielle Anwendung untersucht (Kalscheuer et al. 2006).

Aufgrund der hohen Chemo- und Enantioselektivität der genannten Enzyme, und der daraus resultierenden Einsetzbarkeit zur Produktion von Feinchemikalien, wurden in den letzten Jahren ThDP-abhängige Enzyme hinsichtlich der Darstellung von 2-Hydroxyketonen genauer untersucht. Durch ihre unterschiedlichen Substratspezifitäten ist es bereits möglich eine ganze Reihe verschiedener 2-Hydroxyketone zu erhalten. Das Ziel ist jedoch eine große Plattform verschiedener Enzyme und Enzymvarianten zur Verfügung zu stellen, um ein Werkzeug zur Darstellung einer Vielzahl unterschiedlicher 2-Hydroxyketone zu schaffen. Dazu ist es zwingend notwendig die Sequenz-Struktur-Funktionsbeziehungen zu verstehen und molekulare Grundlagen aufzudecken, die zu den unterschiedlichen Spezifitäten und Selektivitäten der einzelnen Enzyme führen. Mithilfe dieses Verständnisses kann dann die

Selektivität und Spezifität einzelner Enzyme gezielt beeinflusst werden. Zusätzliche Aldehyde können dann durch ThDP-abhängige Enzyme umgesetzt werden, um somit zu neuen Synthesebausteinen für die angewandte organische Chemie zu gelangen.

1.2.2 Alkoholdehydrogenasen

Die Alkoholdehydrogenasen (ADHs) bilden eine Untergruppe der Oxidoreduktasen. Alle Oxidoreduktasen sind abhängig von Cofaktoren und katalysieren die Übertragung von Reduktionsäquivalenten zwischen den Redoxpartnern (Substrat und Cofaktor). Aufgrund ihres biotechnologischen Potentials haben Oxidoreduktasen eine besondere Bedeutung. Alkoholdehydrogenasen gehören innerhalb der Oxidoreduktasen zur Klasse der NAD(P)-abhängigen Dehydrogenasen.

Alkoholdehydrogenasen sind in der Natur weit verbreitet. Sie kommen sowohl in Mikroorganismen, als auch in Pflanzen und Säugetieren vor und setzen ein breites Spektrum unterschiedlichster Substrate um (Reid and Fewson 1994). Aufgrund ihrer Vielzahl an Funktionen findet man sie in allen Bereichen des Lebens. Alkoholdehydrogenasen sind in zahlreichen Alkohol-, Zucker- und Fettstoffwechselwegen involviert und sorgen zum Beispiel bei Gärungsprozessen für die lebenswichtige Regenerierung der Cofaktoren NAD und NADP in der Zelle. Eine weitere wichtige Bedeutung kommt den Alkoholdehydrogenasen beim Schutz der Zelle gegenüber organismusfremden Alkoholen und Aldehyden durch Beteiligung an oxidativen Abbauprozessen zu (Danielsson et al. 1994). Von großer industrieller Bedeutung sind Alkoholdehydrogenasen bei der Herstellung chiraler Alkoholverbindungen durch stereospezifische Reduktion von Ketonen und bei der Racemat-Spaltung. Chirale Alkohole sind wichtige Bausteine für Produkte der pharmazeutischen, kosmetischen und landwirtschaftlichen Industrie (Nakamura and Matsuda 1998; Bornscheuer and Pohl 2001). Neben der Pferdeleberalkoholdehydrogenase (HLADH) werden bislang vor allem Alkoholdehydrogenasen aus Hefen für industrielle Prozesse eingesetzt (Nakamura et al. 2003).

Alkoholdehydrogenasen werden basierend auf physiologischen und katalytischen Eigenschaften in vier Klassen eingeteilt. Dabei spielt die Länge der Aminosäuresequenz der Proteine eine entscheidende Rolle. Es wird abhängig von der Aminosäurekettenlänge zwischen kurzkettigen (*short-chain*), mittelkettigen (*medium-chain*) und langkettigen (*long-*

chain) Alkoholdehydrogenasen unterschieden. Die Klasse der langkettigen Alkoholdehydrogenasen mit Kettenlängen von bis zu 750 Aminosäuren stellen neben den eisenabhängigen Alkoholdehydrogenasen (*Fe-activated*) die kleinste und am wenigsten untersuchte Klasse dar. Weitaus besser untersucht sind die Klassen der kurzkettigen und mittelkettigen Alkoholdehydrogenasen. Kurzkettige Alkoholdehydrogenasen haben eine Aminosäurekettenlänge von ungefähr 250 Aminosäuren. Zu dieser Klasse gehören zum Beispiel die bekannte Alkoholdehydrogenase aus *Drosophila* oder die industriell wichtige (*R*)-spezifische Alkoholdehydrogenase aus *Lactobacillus brevis* (Chase 1999). Die am besten untersuchte Klasse stellen jedoch die mittelkettigen Alkoholdehydrogenasen dar.

1.2.2.1 Mittelkettige Alkoholdehydrogenasen

Die Klasse der mittelkettigen Alkoholdehydrogenasen (*medium-chain dehydrogenases/reductases*, MDRs) umfasst eine große Anzahl an Alkoholdehydrogenasen, die im Vergleich zu anderen Klassen Enzyme mit unterschiedlichsten Funktionen beinhaltet. Mittelkettige Alkoholdehydrogenasen sind aus circa 350 Aminosäuren aufgebaut und setzen sich aus zwei kovalent verbundenen Domänen zusammen (Abb. 1.3) (Jörnvall et al. 1999; Nordling et al. 2002). Diese sind eine Coenzym-Bindedomäne und eine katalytische Domäne. Die Coenzym-Bindedomäne ist für die Bindung des Cofaktors NAD(P) verantwortlich. Die katalytische Domäne beherbergt das aktive Zentrum der Enzyme und hat bei den meisten mittelkettigen Alkoholdehydrogenasen ein katalytisches Zinkatom im aktiven Zentrum gebunden. Darüber hinaus findet man in vielen mittelkettigen Alkoholdehydrogenasen ein zweites Zinkatom vor, welches strukturell stabilisierende Eigenschaften hat. Die Proteine liegen im nativen Zustand meist als Tetramer oder Dimer vor. Selten werden nativ aktive Trimere beobachtet (Norin et al. 1997).

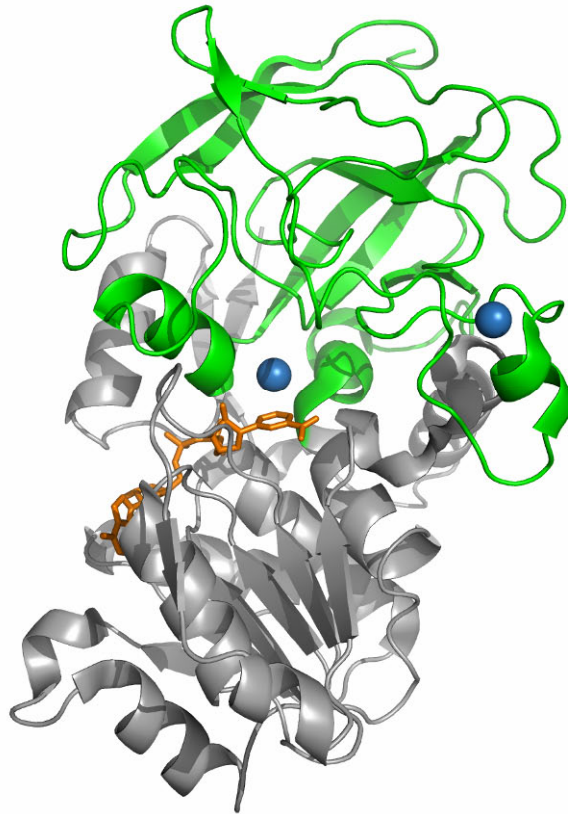


Abbildung 1.3: Sekundärstrukturdarstellung der Pferdeleberalkoholdehydrogenase. PDB-Eintrag 1HLD (Ramaswamy et al. 1994). Die katalytische Domäne ist grün, die Cofaktor-Bindedomäne grau eingefärbt. Die beiden Zinkatome sind blau, der Cofaktor NAD/NADH ist orange dargestellt. Das näher am Cofaktor liegende Zinkatom ist das katalytisch relevante, das entfernt davon gebundene das strukturelle Zinkatom.

Einer der wichtigsten Vertreter der mittelkettigen Alkoholdehydrogenasen ist die Pferdeleberalkoholdehydrogenase (HLADH), die vor allem zur Darstellung von chiralen Alkoholen industriell eingesetzt wird (Bradshaw et al. 1992). Um Substratspezifitäten für den industriellen Einsatz von mittelkettigen Alkoholdehydrogenasen gezielt verändern zu können, muss die Sequenz-Struktur-Funktionsbeziehung verstanden werden. Um dies zu erreichen, ist eine systematische Untersuchung familienspezifischer Eigenschaften nötig.

1.2.3 Cytochrom P450 Monooxygenasen

Die Cytochrom P450 Monooxygenasen (CYPs) bezeichnen eine Enzymfamilie, die zur Klasse der Monooxygenasen gehört (EC 1.14.x.y). Die Enzyme enthalten ein Eisenporphyrin als prosthetische Gruppe, dessen Bindung über das Eisen an das Protein der Grund für das charakteristische Absorptionsmaximum bei 450 nm ist, welches der Enzymgruppe seinen Namen gab. CYPs sind in der Natur weit verbreitet und kommen in nahezu allen Organismen von Bakterien über Pilze, Pflanzen, Insekten und Säugetieren vor. Heute sind mehr als 4000 Proteine der CYP-Familie bekannt. Sie katalysieren eine Vielzahl unterschiedlicher Reaktionen wie zum Beispiel die Hydroxylierung aromatischer oder aliphatischer Verbindungen. Dabei zeigen sie ein breites Substratspektrum und stellen die wichtigste Enzymgruppe beim Metabolismus von Xenobiotika dar, wobei sie für 75% des Phase-I-Metabolismus verantwortlich sind (Ingelman-Sundberg 2004).

Ein systematisches Nomenklaturschema wurde von Nelson etabliert (Nelson 2006), welches die CYPs in eine Hierarchie von Proteinfamilien basierend auf ihrer Sequenzähnlichkeit klassifiziert. Proteine innerhalb einer Superfamilie zeigen dabei eine Sequenzidentität >40% zueinander. Innerhalb der Superfamilien werden die Proteine weiter in homologe Familien unterteilt, welche jeweils Proteine mit einer Sequenzidentität >55% beinhalten. Die Bezeichnung der einzelnen Proteine der CYP-Familie (zum Beispiel CYP2C9) gibt dabei die Klassifizierung des Proteins in der beschriebenen Hierarchie wieder. Dabei folgt der allgemeinen Bezeichnung *CYP* eine Zahl für die Superfamilie (2), gefolgt von der Bezeichnung der homologen Familie (C). Die darauf folgende Zahl (9) bezeichnet das individuelle Enzym dieser Familie. Mit Ausnahme einiger hoch konservierter Sequenzbereiche wie das ExxR-Motiv in der Helix K (Hasemann et al. 1995) oder dem GxRxCxG-Motiv der Hämbindestelle, besitzen die Proteine unterschiedlicher Superfamilien nur sehr geringe Sequenzidentitäten zueinander (meist <20%).

Die bisher bekannten Proteinstrukturen zeigen jedoch einen konservierten Aufbau des Proteinrückgrates und einzelner Aminosäureseitenketten. Funktionell wichtige Elemente wie zum Beispiel die Hämbindestelle liegen daher in Sequenz und Struktur hoch konserviert in allen CYP-Familien vor. Kleine Unterschiede, verursacht durch Mutationen in diesen funktionell wichtigen Bereichen auf Sequenzebene können sich auf die Struktur auswirken und sind hauptverantwortlich für Enzyme mit veränderter Substratspezifität, Regioselektivität und Aktivität. Bei CYPs spielen dabei vor allem die auftretenden Polymorphismen, die zu

unterschiedlichen Isoenzymen, und dadurch zu unterschiedlichen phänotypischen Eigenschaften führen, eine bedeutende Rolle. Die unterschiedlichen Allele können zum Beispiel die Verträglichkeit und Wirksamkeit vieler Arzneistoffe beeinflussen. Für die gezielte Wirkstoffentwicklung zum Beispiel ist daher die Kenntnis der Sequenz-Struktur-Funktionsbeziehung dieser Enzymklasse von großer Bedeutung. Einige Ansätze zur Sammlung von CYP-Sequenzen in Datenbanken und zur systematischen Nomenklatur existieren bereits (Lisitsa et al. 2001; Nelson 2002). So ist zum Beispiel das beschriebene systematische Nomenklaturschema nach Nelson (Nelson 2006) weitgehend anerkannt. Eine einheitlich strukturierte Sammlung an Informationen, die es ermöglicht die Sequenz-Struktur-Funktionsbeziehung dieser Familie zu verstehen, fehlte allerdings bislang.

2 Motivation

Im Rahmen der vorliegenden Arbeit wurden am Beispiel von drei verschiedenen Proteinfamilien (mittelkettigen Alkoholdehydrogenasen, Thiamindiphosphat-abhängigen Enzymen und Cytochrom P450 Monooxygenasen) Methoden der Bioinformatik angewandt um systematische Familienanalysen zu ermöglichen und durchzuführen. Die Analysen sollten zu einem besseren Verständnis der Sequenz-Struktur-Funktionsbeziehungen der einzelnen Familien beitragen. Die gewonnenen Einsichten wurden an Beispielen zur gezielten Vorhersage von Mutationen zur Verbesserung der Enzymeigenschaften angewandt. Darüber hinaus sollten einfache Regeln abgeleitet werden, welche experimentelle Beobachtungen erklären und als Werkzeug für das *protein engineering* dienen können.

Am Beispiel der Thiamindiphosphat-abhängigen Enzyme sollten durch Strukturvergleiche verschiedener Vertreter der Proteinfamilie einfache schematische Modelle zur Erklärung der experimentell beobachteten unterschiedlichen Substratspezifitäten und Selektivitäten erarbeitet werden. Die schematischen Modelle sollten zur gezielten Vorhersage von Mutationen zur Veränderung der Selektivität und des Substratspektrums beitragen. (Seite 49, Seite 67, Seite 85)

Der Aufbau einer umfassenden Proteinfamiliendatenbank über mittelkettige Alkoholdehydrogenasen/Reduktasen (MDR) und deren systematische Analyse sollten einfache Diskriptoren innerhalb dieser vielfältigen Proteinfamilie charakterisieren, die eine systematische Klassifizierung der Familienmitglieder erlauben. Die Untersuchung familienspezifischer Eigenschaften sollte eine einfache Identifizierung von Mutationsstellen zur Vorhersage von Enzymvarianten mit verändertem Substratspektrum ermöglichen. (Seite 123)

Zum besseren Verständnis der Sequenz-Struktur-Funktionsbeziehung innerhalb der Familie der Cytochrom P450 Monooxygenasen sollte eine familienspezifische Datenbank etabliert werden, in der große Datenmengen an Sequenz- und Strukturinformation organisiert abgelegt werden können. Die Datenbank soll als nützliches Werkzeug dienen, welches es ermöglicht, systematische Analysen familienspezifischer Eigenschaften durch umfassende Sequenz- und Strukturvergleiche durchzuführen. (Seite 133)

3 Ergebnisse und Diskussion

3.1 Thiamindiphosphat-abhängige Enzyme

3.1.1 Faktoren die Aktivität, Selektivität und Substratspezifität der Thiamindiphosphat-abhängigen Enzyme Benzaldehydlyase und Benzoylformiatdecarboxylase beeinflussen.

(siehe: Factors Mediating Activity, Selectivity, and Substrate Specificity for the Thiamin Diphosphate-Dependent Enzymes Benzaldehyde Lyase and Benzoylformate Decarboxylase, Seite 49)

Vergleicht man die Benzaldehydlyase aus *Pseudomonas fluorescens* (BAL) und Benzoylformiatdecarboxylase aus *Pseudomonas putida* (BFD) hinsichtlich der Bildung von (*S*)-2-Hydroxypropiophenon ((*S*)-2-HPP) und Benzoin, so bilden beide Enzyme enantiomerenreines (*R*)-Benzoin, wohingegen die Carboligasereaktion mit Acetaldehyd als Akzeptor in BAL zu (*R*)-2-HPP und in BFD zu (*S*)-2-HPP führt. Die BAL ist außerdem in der Lage 2-Hydroxyketone enantioselektiv zu spalten, was für die BFD nicht beobachtet wird. Bei der durch BAL katalysierten Spaltung wird ausschließlich (*R*)-2-HPP und (*R*)-Benzoin umgesetzt. Zur Erklärung der im Experiment beobachteten Unterschiede wurden detaillierte struktur- und sequenzbasierte Charakterisierungen der Bindetaschen durchgeführt, und ein molekulares Modell zur Veranschaulichung der experimentellen Unterschiede erstellt.

Strukturbasierte Untersuchungen und Vergleiche der Bindetaschen ermöglichten es zunächst die unterschiedlichen Orientierungsmöglichkeiten der Substrate (Benzaldehyd und Acetaldehyd) in der Bindetasche auf wenige mögliche einzuschränken. Für die Bildung chiraler 2-Hydroxyketone ist es notwendig, dass ein Donoraldehyd zunächst durch Bindung an den Cofaktor Thiamindiphosphat (ThDP) aktiviert wird, um dann nachfolgend mit einem Akzeptoraldehyd verbunden zu werden. Die Substratbindetasche wird durch die Anordnung des Thiazoliumrings des ThDP-Cofaktors in zwei Teile unterteilt. Im Falle von Benzaldehyd als Donor, kann sich der Akzeptoraldehyd von zwei Seiten an den Donor anlagern, wenn eine

koplanare Anordnung des Phenylrings zum Thiazoliumring des ThDP angenommen wird. Aufgrund der Form der Bindetasche und der vorliegenden V-Konformation des ThDP ist allerdings für die Anlagerung des Akzeptoraldehyds an den Donor eine Seite auszuschließen. Unter Beachtung der weiteren Vorgaben, dass die zu verknüpfenden C-Atome der Substrate (C=O des Akzeptors und C-OH des Donors) zueinander in einem reaktionsfähigen Abstand angeordnet sein müssen, und die Sauerstoffatome in Richtung der als Protonenrelais dienenden Aminosäuren (His70 in BFD (Polovnikova et al. 2003), His29 und Gln113 in BAL (Kneen et al. 2005)) zeigen müssen, wird die Anzahl der möglichen Anordnungen der Substrate in der Bindetasche wie folgt limitiert: 1. Die Reste sowohl des Donor- als auch des Akzeptoraldehyds zeigen (parallel) in die Substratbindetasche hinein. 2. Die Reste zeigen in entgegengesetzte Richtungen (Abb. 3.1).

Es konnte gezeigt werden, dass in der BFD eine kleine Tasche existiert (*S-pocket*), die eine entgegengesetzte Anordnung der Substrate erlaubt. Diese Tasche bietet in der BFD Platz für eine Methylgruppe, sodass Acetaldehyd optimal Platz findet, nicht aber größere Akzeptoraldehyde. Dieses *S-pocket* wird in der BFD hauptsächlich durch die Aminosäuren Pro24 (Rückgratatom), Gly25, Ser26 und Leu461 begrenzt. In der BAL konnte eine solche Tasche, die ausreichend groß ist um zumindest eine Methylgruppe aufzunehmen nicht nachgewiesen werden.

Somit ließen sich die experimentellen Beobachtungen hinsichtlich der Bildung von (*S*)-2-HPP anhand der Form der Bindetasche hinreichend erklären. (*R*)-Produkte entstehen daher bei einer parallelen Ausrichtung der Reste von Donor- und Akzeptoraldehyd, (*S*)-Produkte bei antiparalleler Ausrichtung. Im Falle von Benzaldehyd als Akzeptoraldehyd entsteht folglich in BFD und BAL immer das (*R*)-Produkt, da eine antiparallele Anordnung aus sterischen Gründen nicht möglich ist. Die Relevanz des *S-pocket* hinsichtlich der stereoselektiven Bildung von (*S*)-2-HPP wurde durch experimentelle Arbeiten der Partner am Institut für Molekulare Enzymtechnologie der Heinrich-Heine-Universität Düsseldorf im Forschungszentrum Jülich mittels einer Gly25Ala-Variante der BFD bestätigt (Siegert 2000). Mit dieser Enzymvariante wurde ein reduzierter Enantiomerenüberschuss des gebildeten (*S*)-2-HPP (74% im Vergleich zu >90% im Wildtyp) beobachtet.

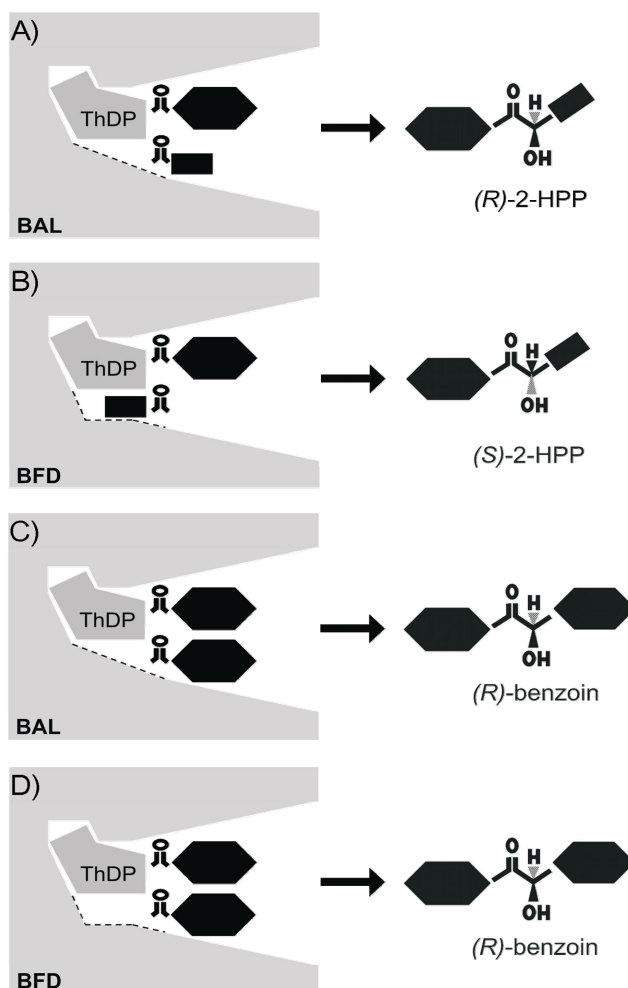


Abbildung 3.1: Schematische Darstellung der Form der Bindetaschen in BAL und BFD. Die möglichen Anordnungen des Akzeptoraldehyds Acetaldehyd (A, B) bzw. Benzaldehyd (C, D) relativ zum Donoraldehyd (Benzaldehyd) sind für die jeweiligen Reaktionen in Abhängigkeit der jeweiligen Bindetaschenform dargestellt.

Die im Experiment beobachtete enantioselektive Spaltung von 2-Hydroxyketonen mit der BAL konnte ebenfalls durch die Form der Substratbindetasche erklärt werden. So konnte aus sterischen Gründen weder (*R*)-Benzoin noch (*R*)-2-HPP in einer reaktiven Anordnung in die Substratbindetasche der BFD modelliert werden. Eine Platzierung in einer reaktiven Anordnung war hingegen in BAL mit beiden Substraten möglich. Die (*S*)-Enantiomere von 2-HPP und Benzoin fanden weder in der Substratbindetasche von BAL, noch in der von BFD Platz.

Durch den detaillierten Vergleich der Substratbindetaschen von BAL und BFD und den Überlegungen zur möglichen Anordnung der Substrate konnte gezeigt werden, dass vor allem sterische Gründe für die unterschiedlichen Enantiomere der Produkte bei der Bildung von

2-HPP und Benzoin verantwortlich sind. Basierend auf den gewonnenen Einsichten in die Sequenz-Struktur-Funktionsbeziehung der Enzyme können gezielt Enzymvarianten mit verbesserter Substratspezifität und Selektivität vorgeschlagen werden.

3.1.2 Rationales Proteindesign von ThDP-abhängigen Enzymen: Beeinflussung der Stereoselektivität.

(siehe: Rational Protein Design of ThDP-Dependent Enzymes: Engineering Stereoselectivity, Seite 67)

Um das beschriebene Konzept des *S-pocket* (siehe Seite 49) zur Darstellung von (*S*)-2-Hydroxypropiophenon ((*S*)-2-HPP) zu verifizieren, wurden basierend auf Strukturuntersuchungen Mutationsvarianten zur Aufweitung des *S-pocket* der Benzoylformiatdecarboxylase aus *Pseudomonas putida* (BFD) vorgeschlagen, um aliphatische Aldehyde unterschiedlicher Größe als Akzeptoraldehyd zur Darstellung weiterer (*S*)-Hydroxyketone zu testen. Es existiert zwar mit den vorhandenen gut charakterisierten Enzymen Benzaldehydlyase (BAL), Benzoylformiatdecarboxylase (BFD), verzweigtkettige Ketosäuredecarboxylase (KdcA) und verschiedenen Pyruvatdecarboxylasen (PDC) beziehungsweise deren bekannte Varianten, eine große Auswahl an Biokatalysatoren zur Darstellung von 2-Hydroxyketonen, diese bilden jedoch hoch enantioselektiv (*R*)-2-Hydroxyketone ausgehend von verschiedensten aliphatischen und aromatischen Aldehyden. Die entsprechenden (*S*)-Enantiomere der Produkte sind bis auf wenige Ausnahmen schwer zu erhalten. Ziel ist es, eine Auswahl verschiedener ThDP-abhängiger Enzyme zu generieren, um eine Plattform zur Darstellung unterschiedlich substituierter und enantiokomplementärer 2-Hydroxyketone zu erhalten.

Mittels molekularer Modellierung wurden Varianten zur Darstellung von (*S*)-2-Hydroxyphenylbutanon ((*S*)-2-HPB) in BFD untersucht. Da sich Leu461 als räumlich meist limitierender Faktor des *S-pocket* herausstellte, wurden Leu461Ala und Leu461Gly als Varianten vorgeschlagen. Die beiden Mutationen sollten jeweils eine gezielte Aufweitung des *S-pocket* zur Folge haben, um die Darstellung von (*S*)-2-HPB mit Propanal als Akzeptoraldehyd zu ermöglichen. Benzaldehyd wurde als Donor und Propanal im Vergleich

zu Acetaldehyd als Akzeptor *in silico* in die Substratbindetaschen der einzelnen BFD-Varianten modelliert. Es konnte durch Arbeiten der Projektpartner am Institut für Molekulare Enzymtechnologie der Heinrich-Heine-Universität Düsseldorf experimentell gezeigt werden, dass wie vorhergesagt das *S-pocket* der Leu461Ala-Variante ausreichend Platz für die Ethylgruppe des Propanal bietet (Abb. 3.2). (*S*)-2-HPB wurde mit beiden Varianten im Experiment mit hoher Stereoselektivität (93%-97% Enantiomerenüberschuss) gebildet, während mit dem Wildtyp der BFD (*R*)-2-HPB mit einem Enantiomerenüberschuss von 21% gebildet wird.

Um zu belegen, dass die Beobachtungen im Experiment ausschließlich auf die Aufweitung des *S-pocket* durch Mutation der sperrigen Aminosäure Leucin beruhen, wurde die Struktur der Leu461Ala-Variante durch einen der Projektpartner mit einer Auflösung von 2,2 Ångström bestimmt. Die Struktur zeigte, dass außer der gewünschten Aufweitung des *S-pocket* durch den Austausch der Aminosäure Leucin an Position 461 keine strukturellen Abweichungen im Vergleich zur Struktur des Wildtyps der BFD existierten, und somit die experimentell beobachteten Unterschiede zum Wildtyp ausschließlich durch den Austausch der Aminosäure Leucin 461 und die damit verbundene Aufweitung des *S-pocket* zu erklären sind.

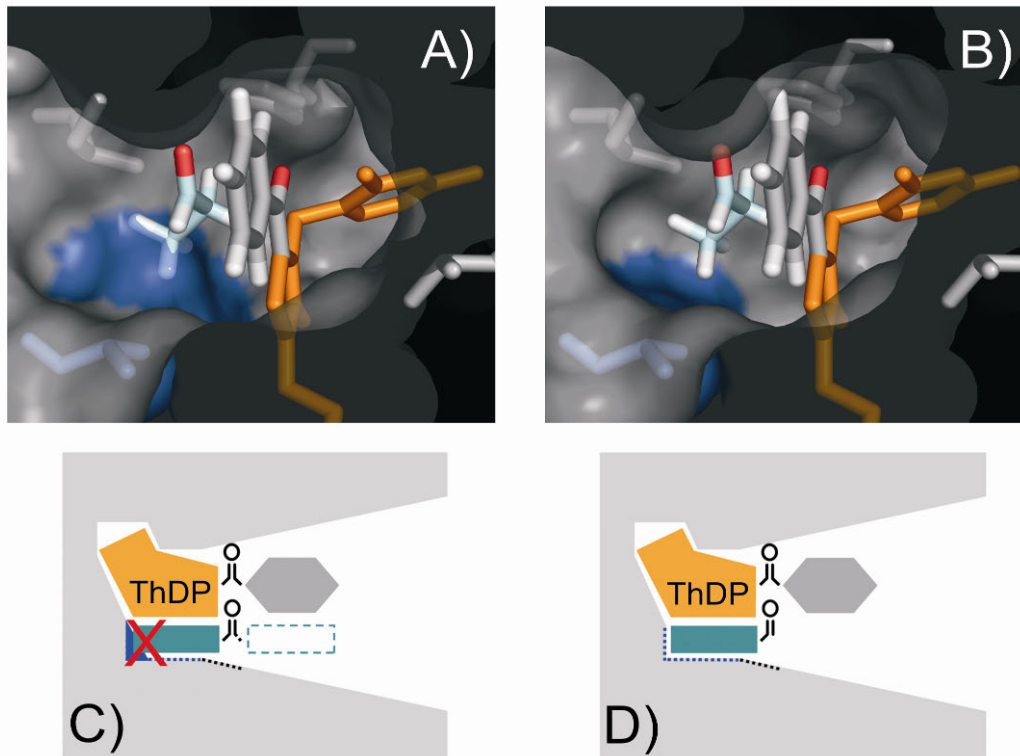


Abbildung 3.2: Struktur des aktiven Zentrums der BFD und deren Variante. Dargestellt ist der Wildtyp (A) und die Leu461Ala-Variante (B) mit Propanal (hellblau) und Benzaldehyd (grau) im aktiven Zentrum. Benzaldehyd ist am C2-Atom des Thiazoliumrings des ThDP gebunden, und Propanal ist im *S-Pocket* lokalisiert. Die Aminosäure Leucin (A) bzw. Alanin (B) an der Position 461 ist blau markiert. Die schematischen Darstellungen (C, D) zeigen die optimale Stabilisierung des Propanal im *S-Pocket* der Leu461Ala-Variante (B, D), während beim Wildtyp das Propanal aus sterischen Gründen nur „parallel“ zum Benzaldehyd ausgerichtet sein kann (A, C gestricheltes Rechteck), was letztlich zum (*R*)-Produkt führt.

Da das *S-pocket* in der BFD hauptsächlich durch die Aminosäuren Pro24, Gly25, Ser26 und Leu461, ferner durch Ala460 begrenzt ist, wurden zur Aufweitung des *S-pockets* weitere Mutanten (Pro24Ala, Ala460Gly, Leu461Gly, Leu461Ala und Leu461Val) vorgeschlagen, die von den Projektpartnern in weiteren Experimenten untersucht wurden.

Es konnten erfolgreich (*S*)-spezifische Varianten der BFD durch einen strukturbasierten Ansatz konstruiert werden, welche zeigen, dass es möglich ist, die Substratbindetaschen in relevanten Bereichen so zu formen, dass sperrigere Substrate umgesetzt, und Enantioselektivitäten beeinflusst werden können. Im vorliegenden Beispiel der BFD entsprachen die experimentellen Befunde sehr gut den Vorhersagen *in silico*. Um das Konzept zu erweitern und auf weitere ThDP-abhängigen Proteine übertragbar zu machen, wurden die

Strukturen weiterer Proteine (BAL aus *Pseudomonas fluorescens* (Mosbacher et al. 2005), PDC aus *Zymomonas mobilis* (ZmPDC) (Dobritsch et al. 1998), PDC aus *Saccharomyces cerevisiae* (ScPDC) (Arjunan et al. 1996), PDC aus *Acetobacter pasteurianus* (ApPDC) (Gocke et al.) und KdcA aus *Lactococcus lactis* (Berthold et al. 2007; Gocke et al. 2007)) mit der Struktur des Wildtyps der BFD (Hasson et al. 1998) hinsichtlich der Existenz eines *S-Pocket* verglichen. In allen genannten Proteinen mit Ausnahme der BAL konnte tatsächlich die Existenz eines *S-Pocket* nachgewiesen werden. Die strikte (*R*)-Selektivität dieser Enzyme erklärt sich wiederum anhand struktureller Eigenschaften. Mit Ausnahme der BFD ist bei allen anderen untersuchten Strukturen der Zugang zum *S-Pocket* durch eine sperrige Aminosäure blockiert. Im Falle der ZmPDC handelt es sich hierbei zum Beispiel um Isoleucin an der Position 472. Experimentelle Arbeiten haben gezeigt, dass ein Austausch zu Alanin die Bildung von (*S*)-HPP mit einem Enantiomerenüberschuss von 70% bewirkt, während der Wildtyp ausschließlich (*R*)-PAC (*ee* > 98%) bildet (Siegert et al. 2005). Das Konzept des *S-Pocket* kann daher erweitert auf andere ThDP-abhängige Enzyme übertragen werden, wenn gleichzeitig der Zugang zum *S-Pocket* mit in die Betrachtungen einbezogen wird.

Das Konzept des *S-Pocket* konnte in dieser Arbeit verifiziert und erweitert werden. Das erweiterte Konzept trug weiter zum tieferen Verständnis der Enzymfamilie der ThDP-abhängigen Enzyme bei, und stellt ein mächtiges Werkzeug zur gezielten enzymatischen Darstellung einer Vielzahl von (*S*)-2-Hydroxyketonen dar. Damit kann eine wertvolle Plattform an Enzymvarianten zur chemoenzymatischen Synthese unterschiedlichste 2-Hydroxyketone bereitgestellt werden.

3.1.3 Pyruvatdecarboxylase aus *Acetobacter pasteurianus*: biochemische und strukturelle Charakterisierung.

(siehe: *Pyruvate Decarboxylase from Acetobacter pasteurianus: BIOCHEMICAL AND STRUCTURAL CHARACTERISATION, Seite 85*)

Im Rahmen dieser Arbeit wurde die Pyruvatdecarboxylase (PDC) aus *Acetobacter pasteurianus* (*ApPDC*) von den Projektpartnern am Institut für Molekulare Enzymtechnologie der Heinrich-Heine-Universität Düsseldorf experimentell charakterisiert und die Substratspezifität bestimmt. Außerdem wurde die dreidimensionale Struktur des Enzyms mit einer Auflösung von 2,75 Ångström von den Partnern experimentell aufgeklärt. Die *ApPDC* bevorzugt als Donoraldehyd aliphatische Aldehyde und kurzkettige 2-Ketosäuren. In Übereinstimmung mit anderen PDCs wurde daher ausgehend von Benzaldehyd eine äußerst geringe Bildung von Benzoin beobachtet. Während andere ThDP-abhängigen Enzyme meist strikt (*R*)-spezifisch sind, zeigt die *ApPDC* im Vergleich jedoch eine geringere Stereoselektivität bei der Carboligation.

Um die im Experiment beobachteten Unterschiede in Substratspezifität und Stereoselektivität zu verstehen, wurde die Struktur der *ApPDC* mit den Strukturen der Pyruvatdecarboxylase aus *Zymomonas mobilis* (*ZmPDC*) (Dobritzsch et al. 1998), der Pyruvatdecarboxylase aus *Saccharomyces cerevisiae* (*ScPDC*) (Arjunan et al. 1996), der verzweigtkettigen Ketosäuredecarboxylase (*KdcA*) (Berthold et al. 2007), der Benzoylformiatdecarboxylase aus *Pseudomonas putida* (*BFD*) (Hasson et al. 1998) und der Benzaldehydlyase aus *Pseudomonas fluorescens* (*BAL*) (Mosbacher et al. 2005) überlagert und verglichen.

Die Überlagerung zeigte eine gute Übereinstimmung des Proteinrückgrates der Strukturen, welche bei den Strukturen der *ApPDC* und der *ZmPDC* besonders hoch war, was sich durch die hohe Sequenzähnlichkeit der beiden Enzyme erklären lässt. Beim Vergleich der Strukturen fiel zunächst auf, dass nur die Struktur der *BFD* keine Helix am C-terminalen Ende der Strukturen aufweist. Alle anderen besitzen eine solche Helix in unterschiedlicher Länge. Diese Beobachtung scheint relevant, da im Rahmen früherer Arbeiten diesem C-terminalen Bereich eine wichtige Funktion im Zusammenhang mit der Aktivität des Enzyms zugesprochen wurde (Dobritzsch et al. 1998; Chang et al. 2000).

Für alle Enzyme konnten die Aminosäuren, die maßgeblich an der Form der Bindetasche beteiligt sind bestimmt werden. Unterschiede in der Form der Bindetasche konnten hierbei vor allem im Bereich der Donorbindestelle beobachtet werden. Während die BFD, die KdcA und die BAL großen aromatischen Donoraldehyden Platz bieten, ist der Bereich der Donorbindestelle in der *ApPDC* und *ZmPDC* durch einen Tryptophanrest räumlich stark eingeschränkt (Trp388 in *ApPDC*). Dadurch lässt sich die beobachtete Präferenz der *ApPDC* gegenüber aliphatischen Donoraldehyden und kurzkettige 2-Ketosäuren erklären. Es konnte gezeigt werden, dass dieser Bereich der Bindetasche in der *ApPDC* dennoch etwas mehr Platz bietet als zum Beispiel in der eng verwandten *ZmPDC*, wodurch erklärt werden kann, weshalb *ApPDC* im Vergleich zu andern PDCs die Decarboxylierung von Benzoylformiat mit höherer Aktivität katalysiert. Der Platz der Donorbindestelle der *ApPDC* ist gerade so groß, dass ein Benzaldehydmolekül darin Platz finden kann, was die sehr geringe Bildungsrate von Benzoin erklärt.

Die niedrige Stereoselektivität der *ApPDC* bei der Carboligation lässt sich ebenfalls aufgrund struktureller Unterschiede der Bindetasche erklären. BFD ist das einzige Enzym, welches bei der Carboligation von Benzaldehyd (als Donor) und Acetaldehyd (als Akzeptor) zu (*S*)-Produkten führt. Dies wird durch ein besonderes strukturelles Merkmal, dem *S-pocket* möglich (siehe Seite 49, (Knoll et al. 2006)). Aufgrund dessen kann sich die Methylgruppe des Acetaldehyds so orientieren (in das *S-pocket*), dass das (*S*)-Produkt entsteht. Auch in den Strukturen der *ZmPDC*, *ScPDC* und *KdcA* konnte ein *S-pocket* gefunden werden (siehe Seite 67, (Gocke et al. 2008)). Im Gegensatz zum *S-pocket* in BFD sind diese aber durch sperrige Aminosäuren blockiert und daher für Substrate nicht zugänglich, was deren strikte (*R*)-Spezifität erklärt. Das *S-pocket* der *ApPDC* bietet im Vergleich zu den zuvor Beschriebenen Platz für ein aromatisches Substrat, sodass unter Annahme von optimal ausgerichteten Seitenketten (vor allem am Eingangsbereich des *S-pocket*) eine Orientierung von Benzaldehyd im *S-pocket* denkbar ist. Die beobachtete niedrige Stereoselektivität der *ApPDC* beruht daher auf der Größe und der Zugänglichkeit des *S-pocket*.

Die vergleichenden Untersuchungen der einzelnen Strukturen ermöglichten ein tieferes Verständnis der Struktur-Funktionsbeziehung ThDP-abhängiger Enzyme. In Kombination mit den vorliegenden biochemischen Daten konnten experimentelle Befunde erklärt, und mit strukturellen Gegebenheiten der *ApPDC* korreliert werden.

3.2 Alkoholdehydrogenasen

3.2.1 Die *Medium-Chain Dehydrogenase/Reductase Engineering Database*: Eine systematische Analyse einer diversen Proteinfamilie zum Verständnis der Sequenz-Struktur-Funktionsbeziehung.

(siehe: *The Medium-Chain Dehydrogenase/Reductase Engineering Database: A systematic analysis of a diverse protein family to understand sequence-structure-function relationship*, Seite 123)

Die mittelkettigen Alkoholdehydrogenasen/Reduktasen (MDRs) sind eine große Enzymfamilie, deren Vertreter eine Vielzahl unterschiedlicher Reaktionen katalysieren. Trotz ihrer großen Unterschiede in der Aminosäuresequenz besitzen die Vertreter der Familie eine sehr ähnliche dreidimensionale Struktur. Da eine umfassende Zusammenstellung aller MDRs bislang nicht existierte, wurde zur systematischen Analyse dieser großen und variablen Enzymfamilie die *Medium-Chain Dehydrogenase/Reductase Engineering Database* (MDRED) erstellt. Die MDRED ist unter <http://www.mdred.uni-stuttgart.de> zugänglich und kann entweder auf der Stufe der Organismen, der Sequenzen oder der Strukturen durchsucht werden. Alle Alignments der einzelnen Familien sowie die phylogenetischen Bäume zur Familienanalyse werden neben familienspezifischen Sequenzprofilen online zur Verfügung gestellt. Eine vorhandene BLAST-Schnittstelle der MDRED ermöglicht eine Klassifizierung neuer oder bisher nicht klassifizierter Familienmitglieder.

Die Datenbank beinhaltet Informationen zu Sequenz, Struktur und funktioneller Annotation von insgesamt 6420 Aminosäuresequenzen, die 2684 individuelle Proteine codieren. Für 42 Proteine sind Strukturinformationen in der Datenbank vorhanden. Die Proteineinträge wurden nach ihrer Sequenzähnlichkeit in 29 Superfamilien und insgesamt darin enthaltenen 199 homologen Familien klassifiziert. Für die einzelnen Familien in der MDRED wurde die systematische Nomenklatur *mdrx.y* eingeführt. Dabei steht *x* für die Superfamilie und *y* für die homologe Familie.

Die Superfamilien der MDRED konnten abhängig vom Vorhandensein eines katalytischen Zinkatoms zunächst in zwei Klassen eingeteilt werden, den *zinc-containing MDRs* und den *non-zinc-containing MDRs*. Letztere beinhaltet insgesamt vier Superfamilien, die vor allem Reduktasen umfassen, deren Aminosäuren der Zinkbindestellen durch andere, nicht Zink-assoziiierende ersetzt sind. Die Klasse der *zinc-containing MDRs* wurde basierend auf strukturellen Unterschieden einer variablen Region weiter unterteilt. Systematische Strukturvergleiche von 37 Proteinstrukturen ergaben, dass sich die Strukturen der *zinc-containing MDRs* hauptsächlich in dieser variablen Region, welche direkt im Anschluss an die Bindestelle des strukturellen Zinkatoms lokalisiert ist, unterscheiden. Diese variable Region wurde zuvor in der Literatur anhand von Sequenzvergleichen einzelner MDRs beschrieben und ein Zusammenhang derer mit der Quartärstruktur postuliert (Norin et al. 1997). Eine tatsächliche Korrelation der Quartärstruktur mit den Eigenschaften der variablen Region wurde allerdings systematisch noch nicht untersucht. Aufgrund des postulierten Zusammenhangs der variablen Region und der Architektur der Quartärstruktur der Proteine wurde diese variable Region *quaternary structure determining loop* (QSDL) genannt. Basierend auf den Strukturvergleichen konnten strukturell hoch konservierte Positionen als eindeutiger Anfang (*QSDL-start*), sowie als eindeutiges Ende (*QSDL-end*) der variablen Regionen definiert werden, die somit erstmals systematisch vergleichbar war.

Die Positionen für *QSDL-start* und *QSDL-end* konnten in den Multisequenzalignments der Familien aufgrund ihrer hohen Konserviertheit auch auf Sequenzebene annotiert werden. Bei der Position *QSDL-start* handelte es sich in 98% der Fälle um ein Cystein, in 2% um ein Serin. Für die Position *QSDL-end* konnte für 75% aller Proteine ein Glycin, für 22% ein Serin annotiert werden. In der Struktur der Pferdeleberalkoholdehydrogenase findet man zum Beispiel für *QSDL-start* ein Cystein an Position 111 und ein Serin an Position 144 für *QSDL-end*. Aufgrund der hohen Konserviertheit der begrenzenden Positionen des QSDL konnte für 92% aller Proteineinträge der *zinc-containing MDRs* der QSDL vollständig annotiert werden. Eine Analyse der Häufigkeit der Länge des QSDL zeigte zwei eindeutig voneinander getrennte Klassen. Die erste Klasse (lange QSDL) umfasst 33% aller Proteine. Das QSDL-Element besteht bei dieser Klasse aus 32 oder mehr Aminosäuren. Die zweite Klasse (kurze QSDL) beinhaltet Proteine, deren QSDL aus bis zu 18 Aminosäuren besteht, und umfasst 52% aller Proteine. 15% der Proteine mit annotiertem QSDL wurden einer weiteren Klasse (mittleren QSDL) zugewiesen.

Unter den untersuchten Strukturen konnte bis auf eine Ausnahme ein eindeutiger Zusammenhang zwischen der Quartärstruktur und deren Zugehörigkeit zu einer der QSDL-Klassen beobachtet werden. Alle Strukturen, die aufgrund ihrer Länge des QSDL zur Klasse der langen QSDL gehören, sind aktiv als Dimere, alle die ein QSDL von 18 Aminosäuren und weniger aufweisen sind Tetramere. Die einzige Ausnahme stellt die Benzylalkoholdehydrogenase aus *Acinetobacter calcoaceticus* (PDB-Eintrag: 1F8F, <http://dx.doi.org/10.2210/pdb1f8f/pdb>) dar, welche ein QSDL-Element von 36 Aminosäuren Länge besitzt, jedoch als Tetramer aktiv ist (MacKintosh and Fewson 1988). Die Zuweisung der Superfamilien zu den QSDL-Klassen konnte den beobachteten Zusammenhang weiter bestätigen. So wurden zum Beispiel die als Tetramer aktiven sekundären Alkoholdehydrogenasen (*mdr23*) zur Klasse der kurzen QSDL zugewiesen. Innerhalb der einzelnen QSDL-Klassen wurden allerdings auch einzelne Vertreter ohne bekannte Struktur gefunden, die nicht diesen Zusammenhang zu zeigen scheinen. Eine (R,R)-Butanedioldehydrogenase aus *Saccharomyces cerevisiae* wurde als Dimer beschrieben (Gonzalez et al. 2000), obwohl das Protein ein QSDL von 14 Aminosäuren Länge besitzt. Aus der Klasse der langen QSDL sind die Galactitol-1-Phosphate Dehydrogenasen als Tetramer beschrieben (Heidlas and Tressl 1990), obwohl ein Dimerzustand aufgrund der Klassenzugehörigkeit erwartet wird. Der beobachtete Zusammenhang des Quartärzustandes und der Länge des QSDL ermöglicht es somit bis auf wenige Ausnahmen für eine Vielzahl an Proteinen ohne experimentell bestimmte Struktur die wahrscheinliche Quartärstruktur vorherzusagen.

Für die zwei QSDL-Klassen konnten jeweils Sequenzmotive innerhalb des QSDL identifiziert werden, die Aminosäuren beinhalten, welche die Form der Bindetasche in einem bestimmten Bereich beeinflussen. Die Relevanz der einzelnen Aminosäuren, die innerhalb der identifizierten Sequenzmotive liegen, wurde bereits zuvor durch Mutationsvarianten bestätigt (Xie and Hurley 1999; Ziegelmann-Fjeld et al. 2007). Innerhalb der Klasse der langen QSDL wurde außerdem eine zusätzliche hoch konservierte Aminosäure gefunden, die so einheitlich nur in dieser Klasse vorkommt und eine zusätzliche Region der Substratbindetasche beschreibt. Eine Identifizierung funktionell wichtiger Aminosäuren ist durch die Annotation der Aminosäuren die für die Form der Bindetasche relevant sind somit innerhalb der Klassen einfach, und für eine Vielzahl an Proteinen möglich. Eine genauere Untersuchung des QSDL-Elements ergab, dass innerhalb der Klasse der langen QSDL Salzbrücken-bildende

Aminosäuren innerhalb des QSDL-Elements und in der katalytischen Domäne hoch konserviert sind, sodass eine Stabilisierung des langen QSDL damit erklärt werden kann.

Die Klassifizierung abhängig vom vorhandenen katalytischen Zinkatom stimmt mit vorangegangenen Klassifizierungsansätzen überein (Nordling et al. 2002). Die zusätzliche Unterteilung der *zinc-containing MDRs* basierend auf Unterschiede einer variablen Region (QSDL) ermöglichte das Auffinden lokal ähnlicher Sequenzmotive innerhalb der Klassen, die funktionell und strukturell von Bedeutung sind. Diese wichtigen Aminosäuren können anhand der familienspezifischen Multisequenzalignments leicht auf weitere Familienmitglieder übertragen werden. So wird eine schnelle und systematische Anreicherung von Information für die Familie der MDRs, und damit die Möglichkeit zur Vorhersage von Mutanten zur Enzymoptimierung ermöglicht. Der große Vorteil der MDRED ist das lokale Ablegen der verschiedenen Struktur-, Sequenz- und Funktionsinformationen in einem einheitlichen Format sowie die öffentliche Zugänglichkeit der Datenbank. Damit wird eine umfassende systematische Analyse der Proteinfamilie ermöglicht.

3.3 Cytochrom P450 Monooxygenasen

3.3.1 Die Cytochrome P450 Engineering Database: Ein Orientierungs- und Vorhersagewerkzeug für die Cytochrom P450 Proteinfamilie.

(siehe: The Cytochrome P450 Engineering Database: A Navigation and Prediction Tool for the Cytochrome P450 Protein Family, Seite 133)

Die Cytochrom P450 Monooxygenasen (CYPs) stellen eine große Familie dar, die eine große Bandbreite unterschiedlicher Reaktionen katalysieren. Beim Menschen sind sie vor allem bei der Umsetzung von Xenobiotika wie zum Beispiel Medikamenten und Umweltschadstoffen von besonderer Bedeutung. Ein Verständnis der Sequenz-Struktur-Funktionsbeziehung ist daher für diese Familie besonders wichtig. Eine einheitlich strukturierte Datenbank, die es ermöglicht Sequenz-, Struktur- und Funktionsinformation in einem einheitlichen Format zu

integrieren, fehlte allerdings bislang. Daher wurde die *Cytochrome P450 Engineering Database* (CYPED, <http://www.cyped.uni-stuttgart.de>) etabliert.

Basierend auf Startsequenzen der *Cytochrome P450 Homepage* von Nelson (<http://drnelson.utmem.edu/CytochromeP450.html>) wurden Sequenz-, Struktur- und Annotationsinformationen aus der *Genbank* (Benson et al. 2007) und der *Protein Data Bank* (PDB, (Berman et al. 2000)) automatisch heruntergeladen. Die CYPED beinhaltet 3911 Proteineinträge. Für 25 verschiedene Proteine ist ein Struktureintrag in der Datenbank vorhanden. Die Proteine sind in 1111 homologe Familien aufgeteilt, die sich auf 531 Superfamilien aufteilen. Die Struktureinträge erstrecken sich über insgesamt 20 verschiedenen homologe Familien.

Für die Funktion relevante Aminosäuren sowie die Sekundärstruktur in Proteinen mit Struktureintrag wurden in den Multisequenzalignments der Familien annotiert. Neben den vorhandenen Annotationen der *Genbank* wurden bekannte funktionsrelevante Sequenzmotive aus der Literatur in den Sequenzen der Datenbank annotiert. Dies sind die Prolin-reiche Region am N-Terminus der CYPs, das Motiv am C-terminalen Ende der Helix K, das AGXXT-Motiv der Helix I sowie die funktionell wichtigen Cysteinreste (Kemper 2004; Mestres 2005). Alle Annotationsinformationen wurden, sofern die Aminosäuren konserviert im Multisequenzalignment vorlagen, auf alle Vertreter einer Familie übertragen. Dadurch bietet die CYPED ein mächtiges Werkzeug zur Informationsanreicherung für bisher wenig charakterisierte CYPs.

Die CYPED ist unter <http://www.cyped.uni-stuttgart.de> zugänglich und kann auf verschiedenen Einstiegsebenen (basierend auf Organismen, Sequenzen oder Strukturen) durchsucht werden. Die *Accession Codes* der jeweiligen Einträge sind dabei zu den Original Datenbankeinträgen bei *Genbank* verlinkt. Die CYPED bietet eine Plattform zur Klassifizierung neuer oder unbekannter CYPs durch die Möglichkeit eine BLAST-Suche (Altschul et al. 1997) innerhalb der CYPED online durchführen zu können. Außerdem können für jede Familie Sequenzprofile heruntergeladen werden, um Proteine einer Familie zuzuordnen. Alle Annotationen sind in den zur Verfügung stehenden Multisequenzalignments farblich markiert und können online betrachtet werden. Des Weiteren stehen vorgefertigte phylogenetische Bäume zur Verfügung, die eine einfache Übersicht der Verwandtschaftsbeziehung ermöglichen. Alle Alignments, phylogenetischen Bäume,

Sequenzprofile, Struktur- und Sequenzdaten werden für den Anwender zum Herunterladen angeboten. Die CYPED wird regelmäßig aktualisiert, wodurch neu hinzugekommene CYP-Sequenzen in die Datenbank aufgenommen und Annotationsinformationen in die Datenbank übertragen werden.

Im Gegensatz zu den bisher existierenden Cytochrom P450 Datenquellen ist die CYPED die erste Datenbank, die es ermöglicht, lokale Informationen über Sequenz und Struktur mit Funktion in einem einheitlichen Format zu kombinieren. Dies ermöglicht eine systematische Analyse dieser großen Proteinfamilie und die daraus gewonnenen Informationen helfen die Familie besser zu verstehen und unterstützen das Design von Enzymvarianten mit verbesserten Eigenschaften.

4 Darstellung der Ergebnisse als Publikationsmanuskripte und Publikationen in englischer Sprache

1. Factors Mediating Activity, Selectivity, and Substrate Specificity for the Thiamin Diphosphate-Dependent Enzymes Benzaldehyde Lyase and Benzoylformate Decarboxylase, Seite 49
2. Rational Protein Design of ThDP-Dependent Enzymes: Engineering Stereoselectivity, Seite 67
3. Pyruvate Decarboxylase from *Acetobacter pasteurianus*: BIOCHEMICAL AND STRUCTURAL CHARACTERISATION, Seite 85
4. The Medium-Chain Dehydrogenase/Reductase Engineering Database: A systematic analysis of a diverse protein family to understand sequence-structure-function relationship, Seite 123
5. The Cytochrome P450 Engineering Database: A Navigation and Prediction Tool for the Cytochrome P450 Protein Family, Seite 133

4.1

Publikation erschienen in *ChemBioChem* 7: 1928-1934

Knoll, M., Müller, M., Pleiss, J., Pohl, M., 2006. Factors mediating activity, selectivity, and substrate specificity for the thiamin diphosphate-dependent enzymes benzaldehyde lyase and benzoylformate decarboxylase.

Factors mediating activity, selectivity, and substrate specificity for the thiamin diphosphate-dependent enzymes benzaldehyde lyase and benzoylformate decarboxylase

Michael Knoll^[a], Michael Müller^[b], Jürgen Pleiss^[a], and Martina Pohl^{[c]*}

^[a]M. Knoll and Priv.-Doz. J. Pleiss
Institute of Technical Biochemistry
University of Stuttgart
Allmandring 31
70569 Stuttgart (Germany)
Fax: (+49)711-68563196

^[b]Prof. Dr. M. Müller
Institute of Pharmaceutical Sciences,
Albert-Ludwigs-University Freiburg
Albertstr. 25
79104 Freiburg (Germany)
Fax: (+49)761-2036351

^{[c]*}Priv.-Doz. Dr. Martina Pohl
Institute of Molecular Enzyme Technology
Heinrich-Heine University Duesseldorf
Research Center Jülich
52426 Jülich (Germany)
Fax: (+49)2461-612490
ma.pohl@fz-juelich.de

Abstract

Benzaldehyde lyase from *Pseudomonas fluorescens* and benzoylformate decarboxylase from *Pseudomonas putida* are homologous thiamin diphosphate-dependent enzymes catalyzing carboligase and carbolyase reactions. Both enzymes catalyze the formation of chiral 2-hydroxy ketones from aldehydes. However, the reverse reaction has only been observed with benzaldehyde lyase. Whereas benzaldehyde lyase is strictly (*R*)-specific, the stereoselectivity of benzoylformate decarboxylase from *P. putida* is dependent on the structure and orientation of the substrate aldehydes. In this study the binding sites of both enzymes were investigated using molecular modelling studies to explain experimentally observed differences in activity, stereo- and enantioselectivity, and specificity of both enzymes.

We designed a detailed illustration, describing the shape of the binding site of both enzymes, which sufficiently explains the experimental effects obtained with the wild type enzymes and different variants. These findings demonstrate that steric reasons are predominantly responsible for the differences observed in (*R*)-benzoin cleavage and the formation of chiral 2-hydroxy ketones.

Keywords: structure-property relationship, molecular modelling, carboligation, substrate specificity, stereoselectivity, enantioselectivity.

Introduction

Benzaldehyde lyase from *Pseudomonas fluorescens* (BAL, E.C. 4.1.2.38) and benzoylformate decarboxylase from *Pseudomonas putida* (BFD, E.C. 4.1.1.7) are well characterized thiamin diphosphate (ThDP)-dependent enzymes catalyzing carbolyase and carboligase reactions. Both enzymes prefer aromatic substrates. The main reaction of BFD, which is found in the mandelate pathway of some bacteria, is the decarboxylation of benzoylformate to benzaldehyde.^[1,2] BAL catalyzes the cleavage of aromatic 2-hydroxy ketones like benzoin thereby allowing *P. fluorescens* to grow on media with benzoin as the only C-source.^[3,4] In 1991 a BFD catalyzed carboligase reaction yielding 2-hydroxypropiophenone (2-HPP) has first been described by Wilcocks and Ward using benzoylformate and acetaldehyde as substrates.^[5] Further studies revealed that BFD^[6-12] and BAL^[9,13-19] are both able to catalyze the carboligation of a broad range of aldehydes, which makes them valuable catalysts for bioorganic syntheses.

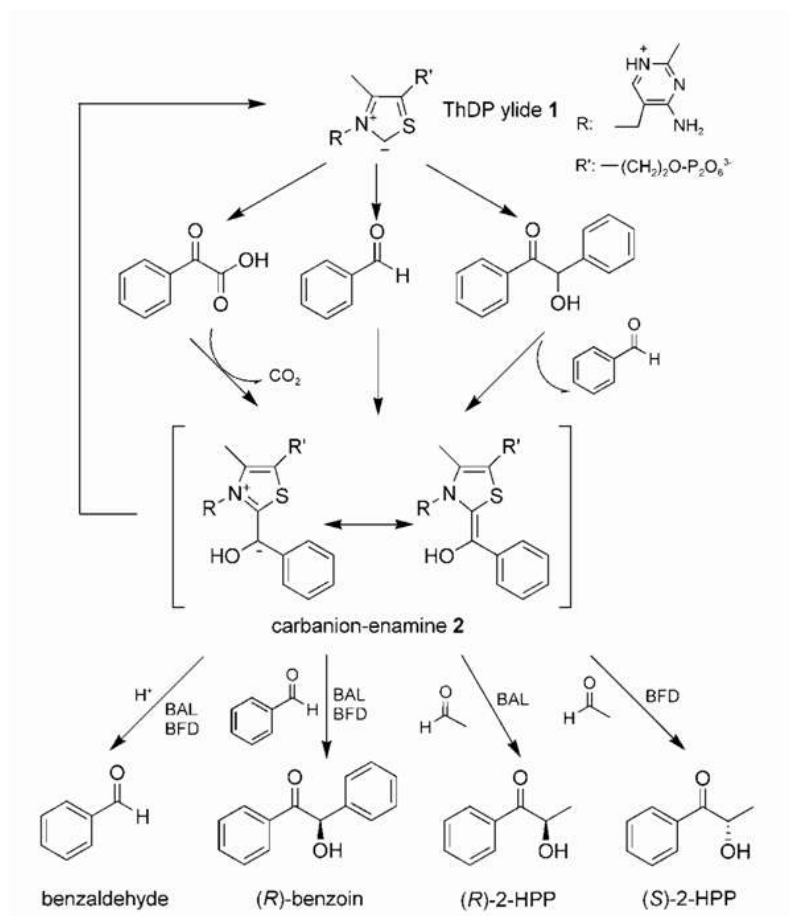
In this study, we focused on the formation of 2-HPP and benzoin catalysed by BAL and BFD, as well as on the BAL-catalysed enantioselective cleavage of (*R*)-benzoin and (*R*)-HPP. Comparison of both enzymes demonstrates that BAL is the more active catalyst, catalyzing the carboligation of benzoin by a factor of 1000 faster than BFD (Tab. 1). With respect to the formation of 2-HPP the catalytic activity of BFD is by factor of 2 higher than with BAL (Tab.1).^[20] While both enzymes catalyse the formation of enantiopure (*R*)-benzoin, carboligations using acetaldehyde as an acceptor result in enantiomeric products: (*R*)-2-HPP with BAL, and (*S*)-2-HPP with BFD. Apart from these common carboligase reactions, the carbolyase reactions differ between both enzymes: with BFD the decarboxylation of benzoylformate but not the cleavage of 2-hydroxy ketones is performed. On the other hand BAL catalyzes the cleavage of 2-hydroxy ketones but not the decarboxylation reaction.

Carboligation yielding chiral 2-hydroxy ketones requires the activation of a donor aldehyde by addition to C2-ThDP and subsequent ligation to an acceptor aldehyde. Although the complete catalytic cycles of BFD and BAL have not been analyzed in detail so far, analogous steps to other ThDP-dependent enzymes can be assumed (scheme 1).^[21-24]

Catalytically important amino acids in the active site have already been determined based on the 3D structure of BFD^[24, 25] and site-directed mutagenesis studies (Tab.1).^[8-12,24] Since the 3D structure of BAL has only recently been published,^[26] mutations in BAL have been planned so far based on sequence alignments with BFD and a homology model published by

Kneen et al. in 2005.^[27] Important residues which were identified based on the homology model could be validated in this study by investigating the 3D structure of BAL.

In order to explain the formation of (*S*)-2-HPP and (*R*)-benzoin catalyzed by BFD we have earlier developed a mechanistic model based on the different variants and further known important residues.^[8] Based on the now available structure of BAL, we have analyzed structural differences between BAL and BFD in order to explain the general (*R*)-specificity of BAL with respect to 2-hydroxyketone formation and cleavage in this study. The main reasons for formation of (*S*)-2-HPP by BFD is explained by the shape of the binding sites. The mechanistic model is discussed in the light of biochemical data obtained with various variants of BAL and BFD.



Scheme 1: Reaction mechanism of BAL and BFD. The first step of the catalytic cycle is the deprotonation of the C2-atom yielding a ThDP-ylide **1** which subsequently attacks a carbonyl compound (e.g. benzoylformate, benzaldehyde or benzoin). The intermediate products are stabilized by decarboxylation (in the case of benzoylformate), deprotonation (in the case of benzaldehyde), or acyloin cleavage (in the case of benzoin) yielding a carbanion-enamine **2**, which can be regarded as a ThDP-bound activated donor aldehyde. Subsequently, **2** can attack an acceptor aldehyde forming a 2-hydroxy ketone. Alternatively, protonation of **2** results in release of benzaldehyde. In either case ThDP is regenerated during the catalytic cycle.

Results and Discussion

Overall structure of BAL and BFD

The tetrameric 3D structures of BFD (528 aa) (PDB accession number 1MCZ)^[24] and BAL (563 aa) (PDB accession number 2AG0)^[26] are structurally highly similar, although sequence identity is only about 24%. The two active sites are created at the interfaces of two monomers forming one dimeric halve of the enzyme. In both enzymes, the cofactor ThDP is located at the bottom of a narrow channel. Thus, the channel is divided in two halves by the thiazolium ring of the ThDP, and the C2-atom of the thiazolium ring as the actual active centre is directed towards the entrance of the substrate channel (Fig. 2, 3). ThDP is bound in the typical “V”-conformation which is stabilized by M421 in BAL and L403 in BFD (Fig. 2, 3, Tab. 2).

A structure-based sequence alignment revealed a high variability of amino acids lining the active sites of the two proteins (Tab. 2). Out of 26 amino acids in similar structural elements of both enzymes only 10 are identical. Further 6 residues which are forming the shape of the overall binding site are found in different structural orientation with no corresponding residue in the other respective structure (Tab. 2). Some of these 32 residues in total are defining mainly the donor aldehyde binding site part, others the part of the binding site for the acceptor aldehyde.

The main differences of the overall structure between BAL and BFD are the wider overall binding site in BAL compared to BFD and a positional shift of the substrate channel entrance on the protein surface. These effects are a result of different lengths of the C-termini and differences in backbone orientation around H281 in BFD compared to the region around the corresponding H286 in BAL. Due to this longer loop and different backbone orientation, H281 in BFD extends further into the binding site than H286 in BAL. Thus, H281, which influences the binding site of both, the donor and acceptor aldehyde in BFD, narrows the space within the binding site in this region (Fig. 3). These steric restraints explain the differences in activity observed with (*R*)-benzoin formation. It is known, that H281 in BFD further influences the protonation of the carbanion-enamine **2** (scheme 1).^[24] Substitution of H281 in BFD for alanine resulted in a variant with very interesting properties: although the decarboxylase activity is diminished, the carboligase activity with respect to (*R*)-benzoin formation is 125-times increased compared to wild-type BFD (wtBFD) (Tab. 1).^[28] Further, the variant accepts a broad range of substituted benzaldehyde derivatives as acyl acceptor, which are not accepted by wtBFD,^[8,11] allowing asymmetric cross-benzoin condensation.^[9] We showed that beside (*S*)-2-HPP (ee 70%) significant amounts (15% relative to 2-HPP) of

(*R*)-phenylacetylcarbinol (PAC) are formed,^[28] which is confirmed by the fact that H281A leads to more space in the region of the acceptor binding site allowing bulkier acceptor aldehydes to bind. Results of site-directed mutagenesis of the corresponding residue H286 in BAL to alanine revealed a decreased benzoin forming activity to about 50% of wild-type BAL (wtBAL), whereas the 2-HPP-forming activity and the lyase activity towards benzoin were hardly affected (Tab. 1). This effect can be explained by the longer distance of H286 to the active site, thus having less influence on the reaction itself, which is consistent with the structural data.^[20]

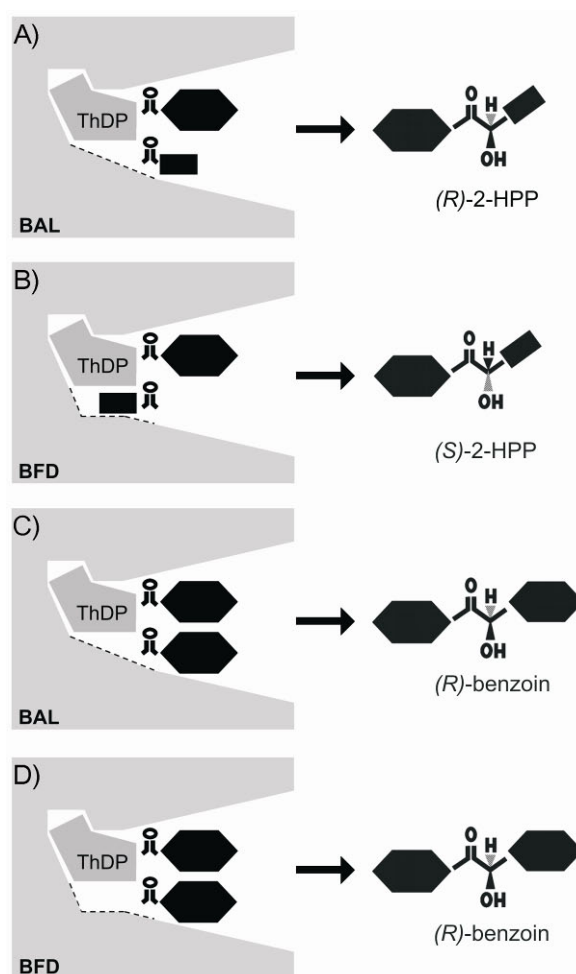


Figure 1: Schematic presentation of the detailed illustrations showing the shape of the binding site for BAL (left) and BFD (right). The possible orientations of the acceptor aldehyde (A,B: acetaldehyde; C,D: benzaldehyde) relative to the donor aldehyde (benzaldehyde) prior to the formation of 2-HPP (A,B) and benzoin (C,D) is shown for the two different binding site architectures. For details see also Fig. 2.

In order to explain the observed differences in stereoselectivity of BAL and BFD during formation of chiral 2-hydroxy ketones, further investigations presented below focus especially on the region of the binding sites which mediates this selectivity. This region includes the parts of the binding site relevant for donor and acceptor binding, but neither the channel entrance of the overall binding site, nor the overall space within the binding sites which are not relevant for stereoselective formation of chiral 2-hydroxy ketones as investigated in this study.

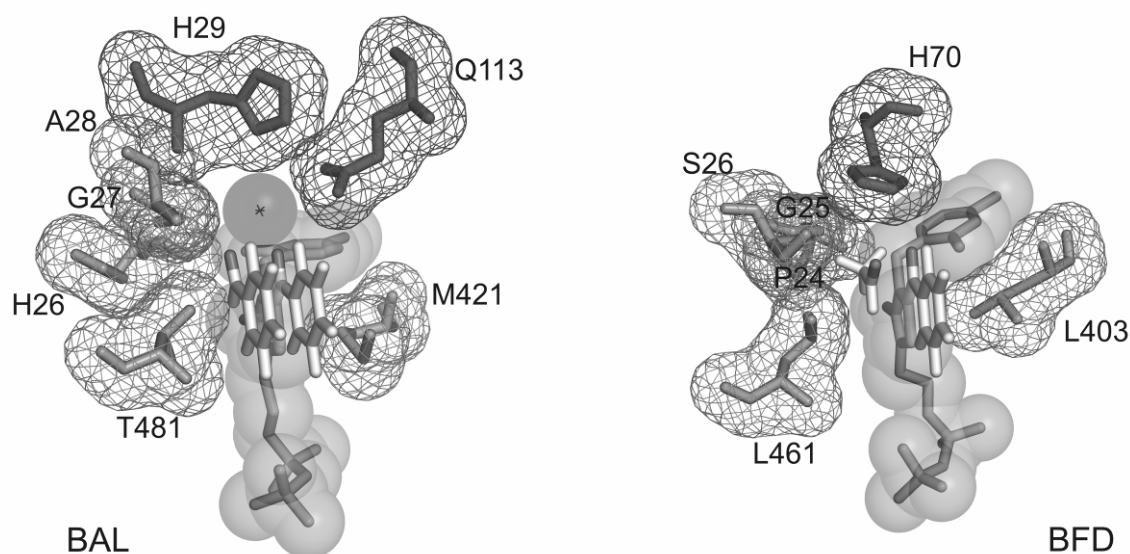


Figure 2: Molecular model related to scheme C and B in Fig. 1 showing the assumptions for the orientation of the acceptor aldehyde in BAL (left) with benzaldehyde, and in BFD (right) with acetaldehyde in detail. For BAL the formation of (*R*)-benzoin is shown, and for BFD the formation of (*S*)-2-HPP is shown. Additionally, the residues forming the small pocket in BFD mediating the differences in substrate selectivity are displayed. ThDP is shown in a transparent spacefilling presentation; the cross marks a water molecule in BAL. Further, the residues stabilizing the conformation of the ThDP are displayed (M421 in BAL and L403 in BFD).

Substrate orientation in the binding site

The preferred donor aldehydes of BAL and BFD are aromatic aldehydes. The donor aldehyde can either bind directly to the ThDP-ylide **1** yielding carbanion-enamine **2** or formation of intermediate **2** is a result of a previous C-C-cleavage reaction (scheme 1). It can be assumed that the phenyl- and the thiazolium ring in **2** are coplanarly arranged in order to allow electron delocalisation.^[29]

In principle the acceptor aldehyde can attack the planar carbanion-enamine **2** from both sides, however one side is shielded, which is caused by the pyrimidine ring due to the “V”-

conformation of enzyme bound ThDP (Fig. 1). It is further assumed that both sp^2 -hybridised C-atoms (C=O of the acceptor and C-OH of the ThDP-bound donor) approach in one line with a distance of about one C-C-bond length before a new C-C-bond of a tetrahedral intermediate is formed. Additionally we orientated the C1-C2-bonds of both aldehydes in consideration of the work of Bürgi and Dunitz,^[31] assuming an almost parallel orientation of the two substrates (Fig. 1). However, only one orientation of the oxygen atoms ensures the contact of both oxygen atoms to a histidine residue acting as a proton relay. Both enzymes contain catalytically important histidine serving as proton acceptor/donor for the several protonation and deprotonation steps in the catalytic cycle (scheme 1, Tab. 1, 2). These are H70 in BFD,^[24] and H29 in BAL.^[27] In BAL also Q113 is discussed as an catalytically important residue,^[27] which is close to H29 (Fig. 2, 3). As demonstrated in Tab. 1, Q113 has a significant impact on the catalytic activity and it can be speculated that this residue does not directly exhibit acid/base properties but influences the orientation of H29 or influences the stability of the local structure or of substrate derived intermediates.

The protonation steps in BAL are probably mediated by a water molecule which bridges the distance between the oxygen atom and H29 (Fig. 2, 3). It has been shown earlier that substitution of H70 in BFD for alanine or glutamine diminished both, the decarboxylase and the carboligase activity drastically (Tab. 1).^[24,27,28] As shown in Tab. 1, the exchange H70A in BFD has also some impact on the stereoselectivity of the 2-HPP formation.^[12, 28] For the corresponding residues in BAL Kneen et al.^[27] showed that exchange of H29 for alanine decreased the lyase activity towards benzoin by a factor of 17, however the variant is still able to catalyze the cleavage of benzoin, and the K_M value is not altered. The exchange of Q113 for alanine or histidine both reduced the catalytic activity of the benzoin lyase reaction by a factor of > 250 compared to wtBAL. The K_M -value for benzoin was also decreased by a factor of 2 (Tab. 1), thus underlining the importance of Q113 for catalytic activity.^[27]

The above mentioned restrictions for the orientation of the acceptor aldehyde with respect to the ThDP-bound donor aldehyde limit the possible orientations of the substrates to:

- orientation of both substrate side chains (acceptor and donor aldehyde) in the same direction, pointing into the substrate channel
- orientation of substrate side chains in opposite directions: acceptor side chain directing into the small pocket at the end of the channel, and donor side chain directing into the substrate channel (Fig. 1).

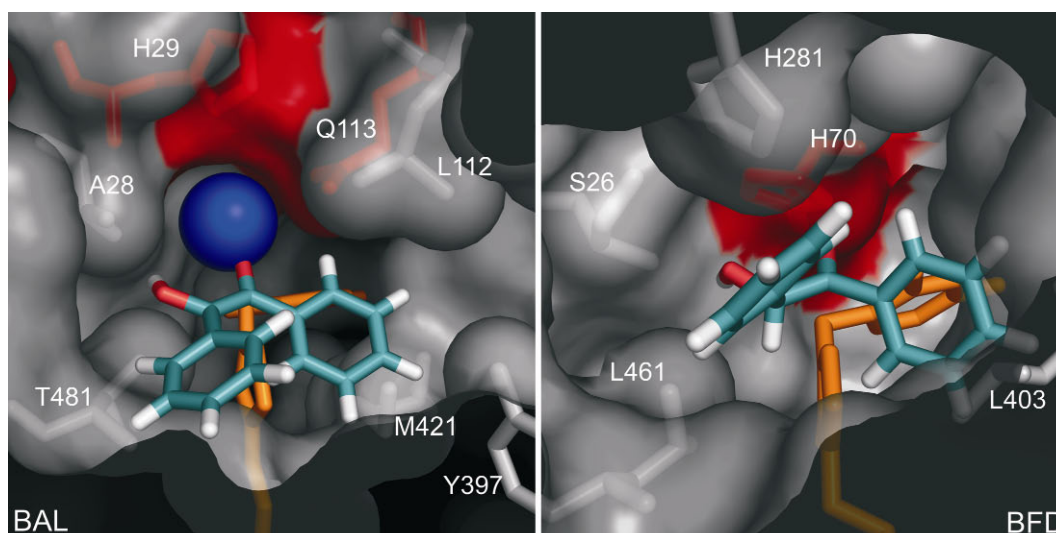


Figure 3: (*R*)-benzoin is modelled into the active sites of BAL (left) and BFD (right), showing that only (*R*)-benzoin fits into the binding site of BAL but not into BFD because of steric hindrance. Mechanistically relevant residues are indicated in red (BALH29, BALQ113, BFDH70). ThDP is presented in orange. The main space limiting residues are shown as sticks. A water molecule (blue) is displayed in the active site of BAL, which is probably involved in the protonation and deprotonation steps during catalysis.

Residues mainly defining the binding site of the donor aldehyde

Residues mainly defining the region of the binding site for the donor aldehyde in BFD are L109, L110, H281, T377, F397, C398, G401, and L403. In BAL this donor binding site region is defined by L112, Q113, A394, Y397, L398, H415, and M421 (Tab. 2). Note that these residues are part of similar structural regions in both enzymes and that equal positions are occupied with different residues, although the substrate specificity for accepted donor aldehydes is similar.

Site-directed mutagenesis studies have been performed with F397, L110 and H281 in BFD (Tab. 1). All mutations resulted in a loss of activity and, with the exception of F397A, impaired stereoselectivity. The variant H281A in BFD shows most pronounced effects also with respect to modification of the substrate range of the carboligation reaction (see above, Tab.1). In BAL only Q113, has so far been mutated in this part of the enzyme (Tab. 1, and see below).

Residues mainly defining the binding site of the acceptor aldehyde

In BFD the region of the binding site for the acceptor aldehyde is mainly formed by backbone atoms of P24, and the side chains of residues G25, S26, A460, L461, and F464. Among these S26 is essential for the decarboxylase activity of BFD (Tab. 1).^[24] A460 in BFD has recently

been identified as an important side chain influencing the acyl donor spectrum, and the stereoselectivity of the carboligation among others.^[12] Carboligation of benzaldehyde and acetaldehyde catalyzed by the BFD-variant A460I yields (*S*)-2-HPP with an enantiomeric excess (ee) of 42-10% (depending on the reaction conditions). Thus, the amount of (*R*)-2-HPP is increased by the mutation A460I compared to wtBFD. This could be explained by the additional hydrophobic residue introduced, forming a stabilizing effect for the acceptor side chain orientated in the same direction as the donor aldehyde side chain, and thus producing more (*R*)-2-HPP than wtBFD. Similar effects are described with variants of F464.^[12] As F464 is directly located next to A460 in the channel leading to the active centre, similar explanation of the effect can be assumed. Combination of both single mutations resulted in a BFD-variant A460I/F464I which catalyzes the formation of (*R*)-2-HPP predominantly.^[12]

The analogous region in BAL is predominantly formed by backbone atoms of H26, and the side chains of residues L25, G27, A28, A480, T481, and F484 (Tab. 2). Exchange of A28, which is located in equivalent position to S26 in BFD, for serine resulted in a variant with a significant decarboxylase activity towards benzoylformate (Tab. 1), which is not observed with wtBAL.^[20,27] A480 and F484 in BAL are in similar positions to A460 and F464 in BFD. As with BFD the exchange of A480 for isoleucine has more pronounced effects on the catalytic activity than the equivalent exchange of F484. Although the K_M for benzoin is not influenced by the mutation, the K_M for (*R*)-2-HPP increased by a factor of 16 with the variant A480I in BAL (Tab. 1), indicating that an increase of this side chain reduces the affinity of (*R*)-2-HPP to the active site.^[20]

Differences in stereoselectivity of 2-hydroxy ketone formation The formation of (*S*)-2-HPP and (*R*)-benzoin by BFD and (*R*)-2-HPP and (*R*)-benzoin by BAL (Tab. 1) can be explained by the shape of the binding sites (Fig. 1). Based on the considerations given above, (*R*)-products are a result of the orientation of both substrate side chains (donor and acceptor aldehyde) in the same direction, whereas (*S*)-products result from a diametrically opposed orientation of the substrate's side chains (Fig. 1). Since an (*S*)-product is formed only in the case of BFD, if acetaldehyde is the acceptor aldehyde this suggests a small binding pocket for the methyl group of acetaldehyde, but not for bigger side chains. Detailed inspection of the binding sites of both structures revealed that such a small pocket in fact exists in BFD but not in BAL (Fig. 1). In BFD this pocket is mainly defined by the side chain of L461. Further residues are G25, S26 and the backbone of P24 (Tab. 2, Fig. 2). In BAL this part of the

binding site is formed by G27, A28, T481, and the backbone of H26 (Tab. 2, Fig. 2). The main limiting effect is caused by the side chain of T481, which causes more steric hindrance than the corresponding residue L461 in BFD, thus resulting in a smaller pocket in this region offering not enough space for a methyl group (or bigger) (Fig. 1). The relevance of this small pocket for the stereoselective production of (*S*)-2-HPP was shown by site-directed mutagenesis of G25 in BFD. Exchanging this position to bulkier alanine results in a variant with significantly reduced activity and selectivity of the carboligase reaction of (*S*)-2-HPP (Tab. 1).^[28] The fact that the enantiomeric excess of (*S*)-2-HPP is reduced from 90% (wtBFD) to 74% with the variant G25A in BFD, hints to steric influences in the additional small pocket of the acceptor binding site. However, the size of the alanine side chain seems to be not sufficient to avoid (*S*)-2-HPP-formation completely.

Specificity of (*R*)-hydroxy ketone cleavage

The enantioselective cleavage of (*R*)-2-hydroxy ketones is only catalyzed by BAL but not by any of the BFD variants tested so far (Tab. 1). Our studies of the architecture of the binding sites revealed mainly steric reasons for the observed differences. (*R*)-benzoin and (*R*)-2-HPP fit only into the active site of BAL and not into BFD, which is demonstrated for (*R*)-benzoin in Fig.4. The main space limiting residue in BFD is L403 which is also important for stabilization of the “V”-conformation of ThDP. There is no possibility to fit the (*S*)-enantiomers either in the active site of BAL or BFD by mimicking a situation prior to the transition state formation (data not shown). In this orientation the 2-hydroxy ketone is approaching the C2-atom of the thiazolium ring. The (*R*)-hydroxy ketones were placed such that the oxygen atoms are pointing in direction of the residues important for protonation steps: H70 in BFD and H29 in BAL.

By contrast, comparison of the tetrahedral intermediates of the ThDP-bound 2-hydroxy ketones did not lead to any steric preference of the (*R*)- or the (*S*)-enantiomer, due to the fact, that both enantiomers could be fitted into the binding sites of both enzymes. The latter also explains the fact that BFD is able to catalyze the formation of 2-hydroxy ketones but not the reverse reaction, which would require an approach of a 2-hydroxy ketone in binding distance to the C2-ThDP.

Conclusions

The shape of the binding sites for two thiamin diphosphate-dependent enzymes, BAL and BFD, which are important biocatalysts to access chiral 2-hydroxy ketones, were compared.

The shape of the binding sites can explain sufficiently the experimental data observed with both wild type enzymes and different variants concerning the formation of chiral 2-hydroxy ketones. The main differences observed with both enzymes, the formation of (*S*)-HPP catalyzed by BFD and (*R*)-HPP catalysed by BAL are primarily the result of a different size of the binding sites for the acceptor aldehyde. Steric reasons are also responsible for the inability of BFD to cleave (*R*)-2-hydroxy ketones. Further, the detailed structural investigations yield a detailed picture of the interactions relevant for substrate specificity and stereo- and enantioselectivity of this interesting group of enzymes. Based on this work, new variants will be planned in order to increase the substrate range of both enzymes. We expect that with BFD an increase of the substrate range for different acceptor aldehydes to access (*S*)-2-HPP derivatives is more probable than altering the strict (*R*)-stereoselectivity of BAL.

Experimental section

Protein structures

The protein structures of BFD (PDB accession number 1MCZ)^[24] and BAL (PDB accession number 2AG0)^[26] were obtained from the Protein Data Bank (PDB) [<http://www.pdb.org/>]. The structure entry with accession number 1MCZ contains a tetrameric structure of BFD with a resolution of 2.8 Å. The structure of BAL (PDB accession number 2AG0) is available with a resolution of 2.58 Å. 1MCZ contains one inhibitor molecule ((*R*)-mandelic acid) and one molecule of ThDP per monomer. The available structure of BAL contains no substrate or bound inhibitor, but ThDP as cofactor. To investigate the differences of the shape of the binding sites, the structures of BAL and BFD were visualized using the PyMol program (version 0.98).^[30]

Substrate placement

To investigate the differences in stereoselectivities of 2-hydroxy ketone formation, the donor and acceptor aldehydes were placed manually into the binding sites. To explain the enantioselective cleavage of (*R*)-benzoin by BAL, (*R*)- and (*S*)-benzoin were modelled into the binding sites of both enzymes. Molecular models of the molecules were built using the molecule builder implemented in the program SYBYL (Tripos, St. Louis, MO). For placement of the donor aldehyde (benzaldehyde) in a planar orientation to the thiazolium ring, the thiazolium ring was modelled bound to the donor aldehyde. Afterwards the donor

aldehyde was placed into the binding site by superimposing the thiazolium ring linked to the bound donor aldehyde with the thiazolium ring of the cofactor (ThDP) bound to the active site. In case of benzoin forming, the aromatic ring of the acceptor benzaldehyde was placed next to the phenyl ring of the donor benzaldehyde with a distance of one C-C-bond length between the carbonyl C-atoms of which the acyloin bond is formed during the reaction. For 2-HPP formation, the acceptor acetaldehyde was placed with respect of the work of Bürgi and Dunitz^[31] for carbonyl addition reactions next to the donor aldehyde (benzaldehyde). In order to understand the enantioselective preference and the differences between both enzymes concerning the cleavage of benzoin, (*R*)- and (*S*)-benzoin were placed into the binding sites such that the carbonyl group of benzoin approaches the C2-atom of the thiazolium ring of ThDP in a reactive manner, mimicking a situation prior to the transition state formation.

Acknowledgement

We thank M.J. McLeish for providing various BFD and BAL clones and G. Schulz for providing the structural data of BAL in advance. M.P. is grateful to Degussa AG for financial support.

References

- [1] G. D. Hegeman, *Methods Enzymol.* **1970**, *17A*, 674.
- [2] A. Y. Tsou, S. C. Ransom, J. A. Gerlt, D. D. Buechter, P. C. Babbitt, G. L. Kenyon, *Biochemistry* **1990**, *29*, 9856.
- [3] B. Gonzalez, R. Vicuna, *J. Bacteriol.* **1989**, *171*, 2401.
- [4] P. Hinrichsen, I. Gomez, R. Vicuna, *Gene* **1994**, *144*, 137.
- [5] R. Wilcocks, O. P. Ward, S. Collins, N. J. Dewdney, Y. Hong, E. Prosen, *Appl. Environ. Microbiol.* **1992**, *58*, 1699.
- [6] T. Dünwald, M. Müller, *J. Org. Chem.* **2000**, *65*, 8608.
- [7] T. Dünwald, A. S. Demir, P. Siegert, M. Pohl, M. Müller, *Eur. J. Org. Chem.* **2000**, 2161.
- [8] H. Iding, T. Dünwald, L. Greiner, A. Liese, M. Müller, P. Siegert, J. Grötzinger, A. S. Demir, M. Pohl, *Chem. Eur. J.* **2000**, *6*, 1483.
- [9] P. Dünkemann, D. Kolter-Jung, A. Nitsche, A. S. Demir, P. Siegert, B. Lingen, M. Baumann, M. Pohl, M. Müller, *J. Am. Chem. Soc.* **2002**, *124*, 12084.
- [10] B. Lingen, J. Grötzinger, D. Kolter, M. R. Kula, M. Pohl, *Protein Eng.* **2002**, *15*, 585.
- [11] B. Lingen, D. Kolter-Jung, P. Dünkemann, R. Feldmann, J. Grötzinger, M. Pohl, M. Müller, *Chembiochem* **2003**, *4*, 721.

- [12] P. Siegert, M. J. McLeish, M. Baumann, H. Iding, M. M. Kneen, G. L. Kenyon, M. Pohl, *Protein Eng. Des. Sel.* **2005**, *18*, 345.
- [13] A. S. Demir, O. Şeşenoglu, E. Eren, B. Hosrik, M. Pohl, E. Janzen, D. Kolter, R. Feldmann, P. Dünkermann, M. Müller, *Advanced Synthesis & Catalysis* **2002**, *344*, 96.
- [14] M. Sanchez-Gonzalez, J. P. N. Rosazza, *Advanced Synthesis & Catalysis* **2003**, *345*, 819.
- [15] A. S. Demir, O. Şeşenoglu, P. Dünkermann, M. Müller, *Org Lett.* **2003**, *5*, 2047.
- [16] A. S. Demir, M. Pohl, E. Janzen, M. Müller, *J. Chem. Soc. Perkin Trans 1* **2001**, 633.
- [17] A. S. Demir, T. Dünwald, H. Iding, M. Pohl, M. Müller, *Tetrahedron-Asymmetry* **1999**, *10*, 4769.
- [18] P. Domingez de Maria, T. Stillger, M. Pohl, S. Wallert, K.-H. Drauz, H. Gröger, H. Trauthwein, A. Liese, *J. Molec. Cat. B: Enzymatic* **2005**, *38*, 43.
- [19] T. Hischer, D. Gocke, M. Fernandez, P. Hoyos, A. R. Alcantara, J. V. Sinisterra, W. Hartmeier, M. B. Ansorge-Schumacher, *Tetrahedron* **2005**, *61*, 7378.
- [20] E. Janzen, M. Müller, D. Kolter-Jung, M. M. Kneen, M. J. McLeish, M. Pohl, *Bioorg. Chem.*, submitted.
- [21] K. Tittmann, M. Vyazmensky, G. Hübner, Z. Barak, D. M. Chipman, *P. N. A. S.* **2005**, *102*, 553.
- [22] F. Jordan, N. S. Nemeria, *Bioorganic Chemistry* **2005**, *33*, 190.
- [23] K. Tittmann, R. Golbik, K. Uhlemann, L. Khailova, G. Schneider, M. Patel, F. Jordan, D. M. Chipman, R. G. Duggleby, G. Hübner, *Biochemistry* **2003**, *42*, 7885.
- [24] E. S. Polovnikova, M. J. McLeish, E. A. Sergienko, J. T. Burgner, N. L. Anderson, A. K. Bera, F. Jordan, G. L. Kenyon, M. S. Hasson, *Biochemistry* **2003**, *42*, 1820.
- [25] M. S. Hasson, A. Muscate, M. J. McLeish, L. S. Polovnikova, J. A. Gerlt, G. L. Kenyon, G. A. Petsko, D. Ringe, *Biochemistry* **1998**, *37*, 9918.
- [26] T. G. Mosbacher, M. Müller, G. E. Schulz, *FEBS J.* **2005**, *272*, 6067.
- [27] M. M. Kneen, I. D. Pogozheva, G. L. Kenyon, M. J. McLeish, *BBA-Proteins and Proteomics* **2005**, *1753*, 263.
- [28] P. Siegert, doctoral thesis thesis, Heinrich-Heine University (Düsseldorf), **2000**.
- [29] F. Jordan, *Nat. Prod. Rep.* **2003**, *20*, 184.
- [30] W. L. DeLano, *The PyMOL Molecular Graphics System*, <http://www.pymol.org>, San Carlo, CA, USA, **2002**.
- [31] M. B. Bürgi, J. D. Dunitz, *Acc. Chem. Res.* **1983**, *16*, 153.

Table 1: Activity data of variants of BFD and BAL. Kinetic data were measured in triplicates. The error is between 5-10%.

^[a]: data refer to enzyme without His-tag, ^[b]: data refer to enzyme with C-terminal His-tag.

^[c]: different data due to variation of the reaction conditions. ^[e]: crude extract.

n.a.: not assayed; n.d.: no activity determined

BFD variants	Decarboxylase activity (benzoylformate)		Lyase activity (benzoin)	Ligase activity (2-HPP)		Ligase activity (benzoin)		Ref.
	Spec. act. (U/mg)	K _M (mM)	Spec. act. (U/mg)	Spec. act. (U/mg)	ee (%)	Spec. act. (U/mg)	ee (%)	
Wt	270 ^[a] /340 ^[b] /70 ^[c]	0.8	n.d.	6.4-7.1 ^[c]	83-90 ^[c] (<i>S</i>)	0.34	>98% (<i>R</i>)	[8,11,12, 28]
S26A	4.9 ^[a]	4.5	n.d.	n.a.	n.a.	n.a.	n.a	[24, 27]
H70A	0.8 ^[a]	1.9	n.d.	n.a.	57 (<i>S</i>)	n.a.	>98 (<i>R</i>)	[24,27,28]
A460I	12 ^[a]	0.2	n.d.	n.a.	10 (<i>S</i>)	n.a.	>98 (<i>R</i>)	[12]
F464I	28 ^[a]	2.3	n.d.	n.a.	42 (<i>S</i>)	0.55	>98 (<i>R</i>)	[12,28]
A460I/F464I	3.6 ^[a]	1.1	n.d.	n.a.	50 (<i>R</i>)	n.a.	>98 (<i>R</i>)	[12]
H281A	35 ^[a]	0.4	n.d.	n.a.	n.d.	50	>98 (<i>R</i>)	[12,24,28]
G25A	0.3 ^[a,e]	n.a.	n.d.	n.a.	74 (<i>S</i>)	n.a.	n.a.	[28]
L110A	0.2 ^[a,e]	> 35	n.d.	n.a.	n.d.	n.a.	n.a.	[28]
F397A	109 ^[a]	n.a.	n.d.	n.a.	86 (<i>S</i>)	n.a.	>98 (<i>R</i>)	[28]
F397A/F464I	5.7 ^[a]	n.a.	n.d.	n.a.	34 (<i>S</i>)	n.a.	>98 (<i>R</i>)	[28]

Table 1: continued

BAL variants	Decarboxylase activity (benzoylformate)		Lyase activity (benzoin)		Lyase activity (<i>R</i> -2-HPP)		Ligase activity (benzoin)		Ref.
	spec. act. (U/mg)	K _M (mM)	spec. act. (U/mg)	K _M (μM)	spec. act. (U/mg)	K _M (mM)	spec. act. (U/mg)	Ee (%)	
wt	n.d.	n.d.	67 190	48 61	3.4	0.3	336	>98% (<i>R</i>)	[20] [27]
A28S ^[b]	4.7	14.2	9.5	47	1.6	0.6	58	>98% (<i>R</i>)	[20]
	1.33	20	14	45					[27]
A480I ^[b]	n.d.	n.d.	6.6	45	4.9	5	n.a.	n.a.	[20]
F484I ^[b]	n.d.	n.d.	22	67	4.0	1.2	50	>98% (<i>R</i>)	[20]
H286A ^[b]	n.d.	n.d.	56.5	55	2.8	0.9	152	>98% (<i>R</i>)	[20]
H29A ^[b]	n.d.	n.d.	11	50	n.a.	n.a.	n.a.	n.a.	[27]
Q113H ^[b]	n.d.	n.d.	0.54	36	n.a.	n.a.	n.a.	n.a.	[27]
Q113A ^[b]	n.d.	n.d.	0.76	32	n.a.	n.a.	n.a.	n.a.	[27]

Table 2: Amino acid residues mainly forming the binding sites of BFD and BAL. Some corresponding residues are found in similar structural positions. Residues which have already been subjected to site-directed mutagenesis are printed italic (see Tab. 1). bb: back bone.

BAL**BFD****Residues mainly defining the donor-binding sites with corresponding residues**

L 112	L 109
<i>Q 113</i>	<i>L 110</i>
<i>H 286</i> ^[a]	<i>H 281</i> ^[a]
A 394	T 377
C 414 (bb)/Y 397 (spatial)	<i>F 397</i>
H 415	C 398
G 419	G 401
M 421	L 403

Residues mainly defining the acceptor-binding sites with corresponding residues

L 25	N 23
H 26 (bb, C=O)	P 24 (bb, C=O)
G 27	<i>G 25</i>
A 28	<i>S 26</i>
<i>H 29</i>	N 27
<i>A 480</i>	<i>A 460</i>
T 481	L 461
<i>F 484</i>	<i>F 464</i>

Further structurally corresponding residues in BFD and BAL within the binding sites

H 49	Q 46
<i>E 50</i>	<i>E 47</i>
T 73	<i>H 70</i>
G 77	G 74
N 80	N 77
L 398	T 380
S 420	G 402
Y 453	Y 433
H 483	W 463
W 478	Y 458

Structurally non-corresponding residues in BFD and BAL mainly lining the binding sites

Residue in BFD without correspondence in BAL	A 381
T 111	
W 163	
Y 239	Residues in BAL without correspondence in BFD
L 282	
N 283	

^[a]These residues are located in different structural regions but with corresponding relative orientation. Only H281 has significant impact on the active centre (see text).

4.2

Publikation erschienen in *ChemBioChem* **9**: 406-412

Gocke, D., Walter, L., Gauchenova, E., Kolter, G.K., Knoll, M., Berthold, C.L., Schneider, G., Pleiss, J., Müller, M., Pohl, M., 2008. Rational protein design of ThDP-dependent enzymes: engineering stereoselectivity.

Rational Protein Design of ThDP-Dependent Enzymes: Engineering Stereoselectivity

Dörte Gocke,^[a] Lydia Walter,^[b] Ekaterina Gauchenova,^[b] Geraldine Kolter,^[a] Michael Knoll,^[c] Catrine L. Berthold,^[d] Gunter Schneider,^[d] Jürgen Pleiss,^[c] Michael Müller,^[b] and Martina Pohl^{[a]*}

^{[a]*} D. Gocke, G. Kolter, PD Dr. M. Pohl
Institute of Molecular Enzyme Technology
Heinrich-Heine University Düsseldorf
D-52426 Jülich (Germany)
Fax: (+49) 2461-612940
E-mail: ma.pohl@fz-juelich.de
Homepage: www.iet.uni-duesseldorf.de

^[b] L. Walter, Dr. E. Gauchenova, Prof. Dr. M. Müller
Institute of Pharmaceutical Sciences
Albert-Ludwigs University Freiburg
Albertstraße 25, D-79104 Freiburg (Germany)

^[c] M. Knoll, Prof. Dr. J. Pleiss
Institute of Technical Biochemistry
University of Stuttgart
Allmandring 31, D-70569 Stuttgart (Germany)

^[d] C. L. Berthold, Prof. Dr. G. Schneider
Department of Medical Biochemistry and Biophysics
Karolinska Institutet
Tomtebodavägen 6, S-17177 Stockholm (Sweden)

Supporting information for this article is available on the WWW under
<http://www.chembiochem.org>.

Abstract

Benzoylformate decarboxylase (BFD) from *Pseudomonas putida* is an exceptional thiamin diphosphate (ThDP)-dependent enzyme, as it catalyzes the formation of (*S*)-2-hydroxy-1-phenyl-propan-1-one from benzaldehyde and acetaldehyde. This is the only currently known *S*-selective reaction (92% ee) which is catalyzed by this otherwise *R*-selective class of enzymes. Here we describe the molecular basis to introduce *S*-selectivity into ThDP-dependent decarboxylases. By shaping the active site of BFD using rational protein design, structural analysis and molecular modeling, optimal sterical stabilization of the acceptor aldehyde in a structural element called *S*-pocket, as the predominant interaction to adjust stereoselectivity, was identified. Our studies revealed leucine 461 as a hot spot for stereoselectivity in BFD. Exchange for alanine and glycine resulted in variants, which catalyze the *S*-stereoselective addition of larger acceptor aldehydes such as propanal with benzaldehyde and derivatives thereof; a reaction which is not catalyzed by the wild type enzyme. Crystal structure analysis of the variant BFDL461A supports the modeling studies.

Keywords: Biotransformation, carbonylation, 2-hydroxy ketone, site-directed mutagenesis, asymmetric synthesis

Introduction

The potential of thiamin diphosphate (ThDP)-dependent enzymes to catalyze a benzoin condensation-like carbonylation of aldehydes to chiral 2-hydroxy ketones with high stereoselectivity is well established.^[1] Our goal is to generate a toolbox of various ThDP-dependent enzymes to create a platform for diversely substituted and enantio-complementary 2-hydroxy ketones.

With the current set of enzymes including benzoylformate decarboxylase (BFD), benzaldehyde lyase (BAL), branched-chain 2-ketoacid decarboxylase (KdcA), different pyruvate decarboxylases (PDC), and variants thereof, the carbonylation of various aliphatic and aromatic aldehydes is possible yielding symmetrical and mixed (*R*)-2-hydroxy ketones with predominantly high enantioselectivity. However, the corresponding (*S*)-products are hardly accessible by these enzymes. Exceptions are the kinetic resolution of benzoin derivatives by BAL from *Pseudomonas fluorescens*,^[2] and the carbonylation of benzaldehyde derivatives and acetaldehyde yielding (*S*)-2-hydroxypropiophenone derivatives using BFD from *Pseudomonas putida* as a catalyst.^[3]

A molecular explanation for this exceptional behavior of BFD was recently suggested^[4] based on the crystal structure of the enzyme.^[5] A potential *S*-pocket was identified which exactly fits to the size of the small acetaldehyde side chain, when approaching the ThDP-bound aromatic donor aldehyde prior to formation of the new C-C-bond (Fig. 1).^[4] Recent modeling studies showed that larger aldehydes do not fit into this pocket.

Here, we verify the *S*-pocket approach by site-directed mutagenesis of amino acid residues lining this part of the active centre and probing the resulting variants in carbonylation reactions with different aliphatic aldehydes as acyl acceptors. Further, we provide evidences that similar *S*-pockets are also present in other ThDP-dependent decarboxylases, which opens access to a broad range of (*S*)-2-hydroxy ketones as valuable building blocks for compounds such as the taxol side chain and 5'-methoxyhydnocarpin.^[6]

Results and Discussion

The *S*-pocket in BFD from *Pseudomonas putida* is formed by the side chains of P24, A460, and predominantly by L461 (Figure 1). These residues were replaced by smaller amino acid residues in order to dissect the impact of the respective amino acid on an increase of the *S*-pocket. All variants were produced by site-directed mutagenesis. After cloning,

overexpression, and purification the variants were investigated with respect to their decarboxylase and carbonylase activity.

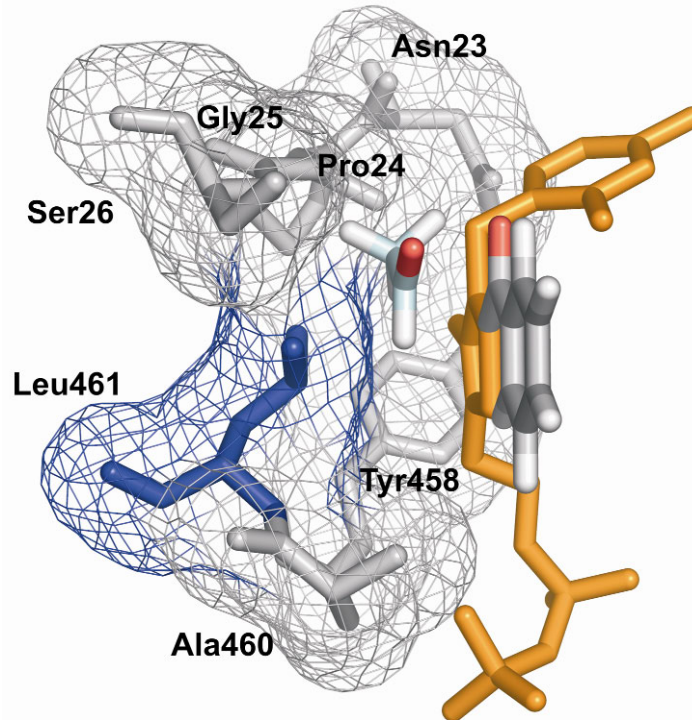


Figure 1: Active site of BFD with the cofactor ThDP (orange). The donor benzaldehyde (grey) and the acceptor acetaldehyde (light blue) are modeled inside. The side chain of acetaldehyde is bound to the *S*-pocket, which is mainly defined by Leu 461 (blue). The diametrically opposed orientation of the side chains of donor and acceptor results in the formation of (*S*)-products.^[6]

Decarboxylase activity of BFD-variants

All variants were still able to catalyze the physiological function of BFD: the decarboxylation of benzoylformate (Tab. 1), showing hyperbolic $v/[S]$ -plots as wild type BFD (BFDwt) but a decreased decarboxylase activity toward benzoylformate. Most pronounced effects were obtained with the mutation in position Leu461 yielding variants with 7 to 8-fold decreased specific decarboxylase activity compared to BFDwt. It is important to note that the K_M -values do not parallel the V_{max} values. Whereas BFDP24A and BFDL461A show K_M -values in the same range as the wild type enzyme, the variant BFDL461V has a 6-fold higher apparent affinity for the substrate, while the K_M -values of the two glycine variants (A460G, L461G) are about 4 to 5-fold higher than that of BFDwt.

The substrate range of the decarboxylase reaction has not been affected by the mutations, with benzoylformate being the main substrate for all variants (see supporting information).

Table 1. Kinetic data determined for the decarboxylation of benzoylformate catalyzed by BFDwt and several BFD-variants.		
enzyme	V_{\max} [U/mg]	K_M [mM]
BFDwt	400 ± 7	0.37 ± 0.03
BFDP24A	367 ± 10	0.51 ± 0.07
BFDA460G	300 ± 9	1.54 ± 0.14
BFDL461V	60 ± 1	0.06 ± 0.01
BFDL461A	53 ± 2	0.25 ± 0.04
BFDL461G	49 ± 2	1.40 ± 0.16

Data were measured in 50 mM potassium phosphate buffer, pH 6.5. Kinetic constants were calculated according to Michaelis-Menten using Origin 7.0 (Origin Lab Corporation, Northampton, MA).

Table 2. Space-time yields and enantioselectivities of different BFD variants concerning the formation of acetoin from acetaldehyde and benzoin from benzaldehyde. n.d.: not determined.				
enzyme	acetoin [g/L/d]	ee [%]	benzoin [g/L/d]	ee [%]
BFDwt	0.11	34 (<i>R</i>)	1.08 ^[b]	n.d.
BFDP24A	0.19	3 (<i>R</i>)	0.58 ^[b]	n.d.
BFDA460G	0.07	38 (<i>R</i>)	0.23 ^[b]	n.d.
BFDL461V	0.1	33 (<i>R</i>)	1.75 ^[a]	n.d.
BFDL461A	0.07	64 (<i>S</i>)	0.14 ^[a]	n.d.
BFDL461G	n.d.	n.d.	n.d.	n.d.

Space-time-yields were calculated within the linear range of ^[a] 3 h, ^[b] 2 h.

Carboligase activity of BFD-variants

All variants were investigated with respect to their carboligase activity toward the self-ligation of benzaldehyde to benzoin and acetaldehyde to acetoin (Tab. 2). Mutations in the putative *S*-pocket effect both, the acetoin and the benzoin synthesis, with the P24A- and the L461V variant being most effective concerning catalysis of the acetoin synthesis and the L461V variant catalyzing also the benzoin synthesis 1.6 times faster than BFDwt. The mutations affect the stereoselectivity of the acetoin formation significantly (Tab. 2): While BFDwt catalyzes predominantly the formation of (*R*)-acetoin (**1**, Tab. 3) (*ee* 34%),^[7] the (*S*)-enantiomer (*ee* 65%) is formed in excess with BFDL461A, supporting the relevance of L461 for the shape of the *S*-pocket.

Further studies with mixed carboligations were focused on the glycine and alanine variants in position of leucine 461. In mixed carboligations with benzaldehyde and acetaldehyde the stereoselectivity of the (*S*)-2-hydroxy-1-phenyl-propan-1-one ((*S*)-HPP, **3**) synthesis was improved by both mutations (*ee* 98%) relative to BFDwt (92% (Table 3, A)).^[3] HPP is formed with benzaldehyde as the donor and acetaldehyde as the acceptor aldehyde (Figure 1). As predicted from modeling studies the most pronounced effect of the increased *S*-pocket should

become apparent, if acetaldehyde is replaced by the larger propanal, which was proved by analytical studies (Table 3, B). Both variants in position 461 catalyze the synthesis of the desired product (*S*)-2-hydroxy-1-phenyl-butan-1-one ((*S*)-**7**) with high stereoselectivity, showing even higher enantioselectivities at pH 7.9 in comparison to standard conditions (pH 7.0), whereas variation of the substrate concentration and reaction temperature had no significant effect on the *ee*-values (data not shown).

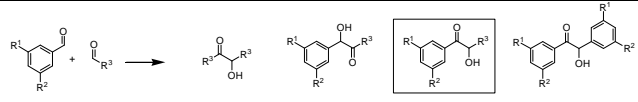
According to the predictions based on the structure of BFDL461A a further increase of the acceptor aldehyde should decrease the stereoselectivity again. This assumption was experimentally confirmed by application of butanal in mixed carboligation reactions with benzaldehyde yielding **10** with decreased stereoselectivity (*ee* 63%, (*S*)) (Table 3, C) and activity. Besides propanal, monomethoxyacetaldehyde is an acceptor aldehyde in presence of benzaldehyde for BFDL461A, resulting in a yield of about 20% in analytical biotransformations (Table 3, D). By contrast, cyclopropylcarbaldehyde results in lower amounts yet high selectivity of the mixed carboligation product (Table 3, E).

Compared to benzaldehyde the stereo control in mixed carboligations with propanal is even better with 3,5-dimethoxybenzaldehyde and 3-cyanobenzaldehyde (Table 3, F, I). Very interesting results have been obtained in carboligations of 3,5-dimethoxybenzaldehyde with butanal and pentanal, respectively. Whereas no product was obtained with the variant BFDL461A, the variant BFDL461G is able to catalyze the mixed carboligation with pentanal even better than with butanal (Table 3, G, H).

The carboligation of benzaldehyde and propanal was investigated in more detail on a preparative scale. After 70 h BFDL461A yielded 35% (w/w) and BFDL461G 31% (w/w) product **7**, respectively, while negligible amounts of **4** and **5** were formed demonstrating very high chemoselectivity for both variants. Confirming the prediction of the modeling studies, the desired product (*S*)-2-hydroxy-1-phenyl-butan-1-one ((*S*)- **7**) was formed with very high stereoselectivity of 93-97% using the variants BFDL461A and BFDL461G, which is in contrast to BFDwt yielding (*R*)-**7** (*ee* 21%) (Table 4). Although carboligation with BFDwt resulted in a total conversion of 37%, a mixture of **4**, **6**, and **7** was formed in almost equal amounts, demonstrating the low chemo- and stereoselectivity of BFDwt for this reaction.

Table 3. Relative product distribution and enantiomeric excesses (*ee*) obtained in analytical scale carbonylation reactions of benzaldehyde derivatives and various aliphatic aldehydes catalyzed by BFD variants. Relative product distribution is given in mol% (NMR); *ee*-values were determined by HPLC-analysis.

All studies were performed with equimolar concentrations of both aldehydes (18 mM) in potassium phosphate buffer (50 mM, pH 7, 2.5 mM MgSO₄, 0.1 mM ThDP, 20 vol% DMSO) 0.3 mg/mL purified enzyme, 30 °C, if not otherwise indicated.



A $R^1 = R^2 = H$ $R^3 = CH_3$	products			
	1	2	3	4
	BFDwt			
	-	-	81% 92% (S)	4% 99% (R)
	BFDL461A			
	-	-	79% 98% (S)	-
BFDL461G				
-	-	74.5% 98% (S)	-	
B $R^1 = R^2 = H$ $R^3 = C_2H_5$	Products			
	5	6	7	4
	BFDwt			
	8% <i>ee</i> n.d.	12% 98% (R)	6% 21% (R)	8% 99% (R)
	BFDL461A			
	6% <i>ee</i> n.d.	-	21.5% 88%/93% ^a (S)	1.5% <i>ee</i> n.d.
BFDL461G				
-	-	23% 93%/97% ^a (S)	0.5% <i>ee</i> n.d.	
C $R^1 = R^2 = H$ $R^3 = C_3H_7$	Products			
	8	9	10	4
	BFDwt			
	34% <i>ee</i> n.d.	32% 99% (R)	2% 66% (R)	3% 99% (R)
	BFDL461A			
	8% <i>ee</i> n.d.	3% >99% (R)	1% 63% (S)	1% 97% (R)
BFDL461G				
6% <i>ee</i> n.d.	2% <i>ee</i> n.d.	1% <i>ee</i> n.d.	-	
D $R^1 = R^2 = H$ $R^3 = OCH_3$	products			
	11	12	13	4
	BFDL461A			
-	-	9.5%, 21% ^[b] 93% (S)	<1% <i>ee</i> n.d.	

Table 3: continued

E	products				
	15	16	17	4	
	BFDL461A				
$R^1 = R^2 = H$ $R^3 =$ cyclopropyl	-	-	5% 97.5% (S)	-	
F	products				
	1	18	19	20	
	BFDwt				
	$R^1 = R^2 =$ OCH_3 $R^3 = C_2H_5$	-	-	<1% ee n.d.	-
	BFDL461A				
		-	-	7.5%; 9% ^[b] >99% (S)	-
BFDL461G					
	-	-	31% >99% (S)	-	
G	products				
	21	22	23	20	
	BFDwt				
	$R^1 = R^2 =$ OCH_3 $R^3 = C_3H_7$	19% ee n.d.	26% ee n.d.	-	-
	BFDL461A				
		-	-	-	-
BFDL461G					
	2% ee n.d.	26% ee n.d.	4% ee n.d.	-	
H	products				
	24	25	26	20	
	BFDL461A				
$R^1 = R^2 =$ OCH_3 $R^3 = C_4H_{11}$	6% ee n.d.	4% ee n.d.	11% > 90% (S)	<1% ee n.d.	
I	products				
	1	27	28	29	
	BFDL461A				
$R^1 = CN$ $R^2 = H$ $R^3 = C_2H_5$	-	-	22.5%, 60% ^[b] 98% (S)	-	
[a] Carboligations have been performed at pH 7.9.					
[b] Carboligations have been performed with a 3-fold excess of the aliphatic aldehyde (54 mM).					

Structural investigation of BFDL461A

In order to prove that the site-specific mutagenesis did not alter the 3D structure of the enzyme besides the exchanged amino acid, the crystal structure of the BFDL461A variant was solved with a resolution of 2.2 Å. Despite the desired increase of the *S*-pocket, no significant structural modifications have been detected. In contrast to BFDwt the *S*-pocket in the variant BFDL461A offers optimal space for the ethyl group of propanal providing higher stereoselectivity (Figure 2, B, D).

Our data show that stereoselectivity is predominantly a consequence of the optimal stabilization of the acceptor aldehyde side chain in the *S*-pocket. If this fit is not optimal, as observed for BFDwt already with propanal (Fig. 2, A, C) and for BFDL461A with butanal, the *S*-selectivity is reduced (Tab. 3, C).

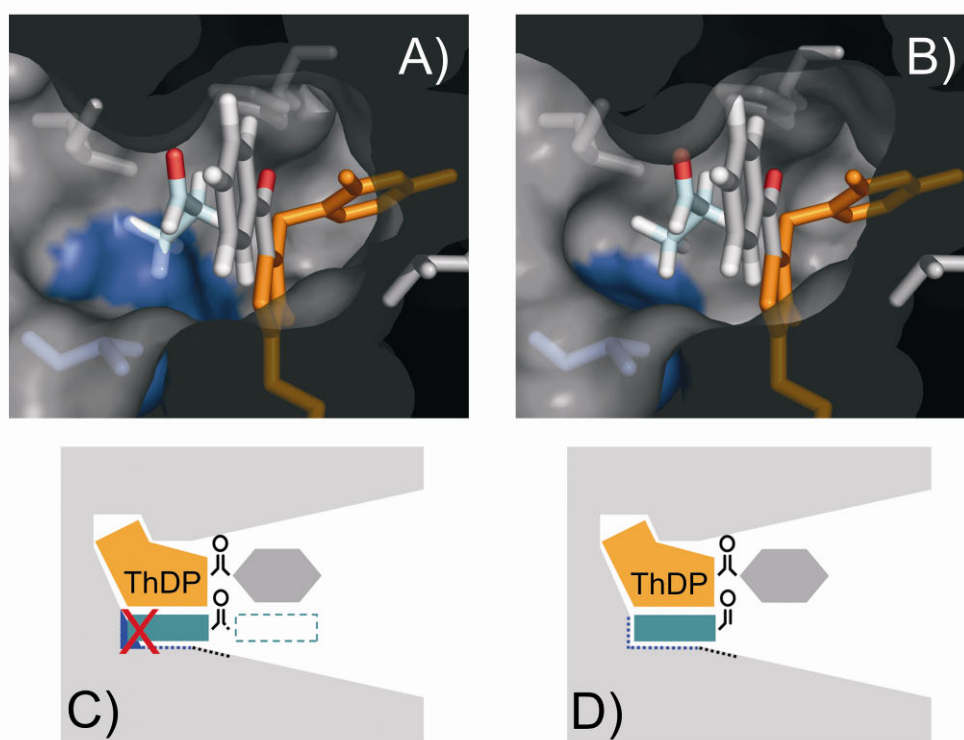


Figure 2: Crystal structure of the active site of BFDwt (A) and BFDL461A (B) with benzaldehyde and propanal modeled inside. The side chain of the amino acid residue in position 461 is marked in blue. Benzaldehyde (grey) is bound to the C2-atom of the thiazolium ring (orange) and is coplanarly arranged due to steric and electronic demands. Propanal (light blue) is located in the *S*-pocket. Models show a perfect stabilization of the acceptor aldehyde in the *S*-pocket of the variant (B, D), while L461 causes steric hindrance with propanal in BFDwt (A, C). Consequently, in BFDwt the propanal approaches predominantly parallel to benzaldehyde (dotted square, C) yielding mainly the (*R*)-enantiomer while in BFDL461A the perfect fitting allows an antiparallel arrangement (D) leading to an excess of the (*S*)-product.

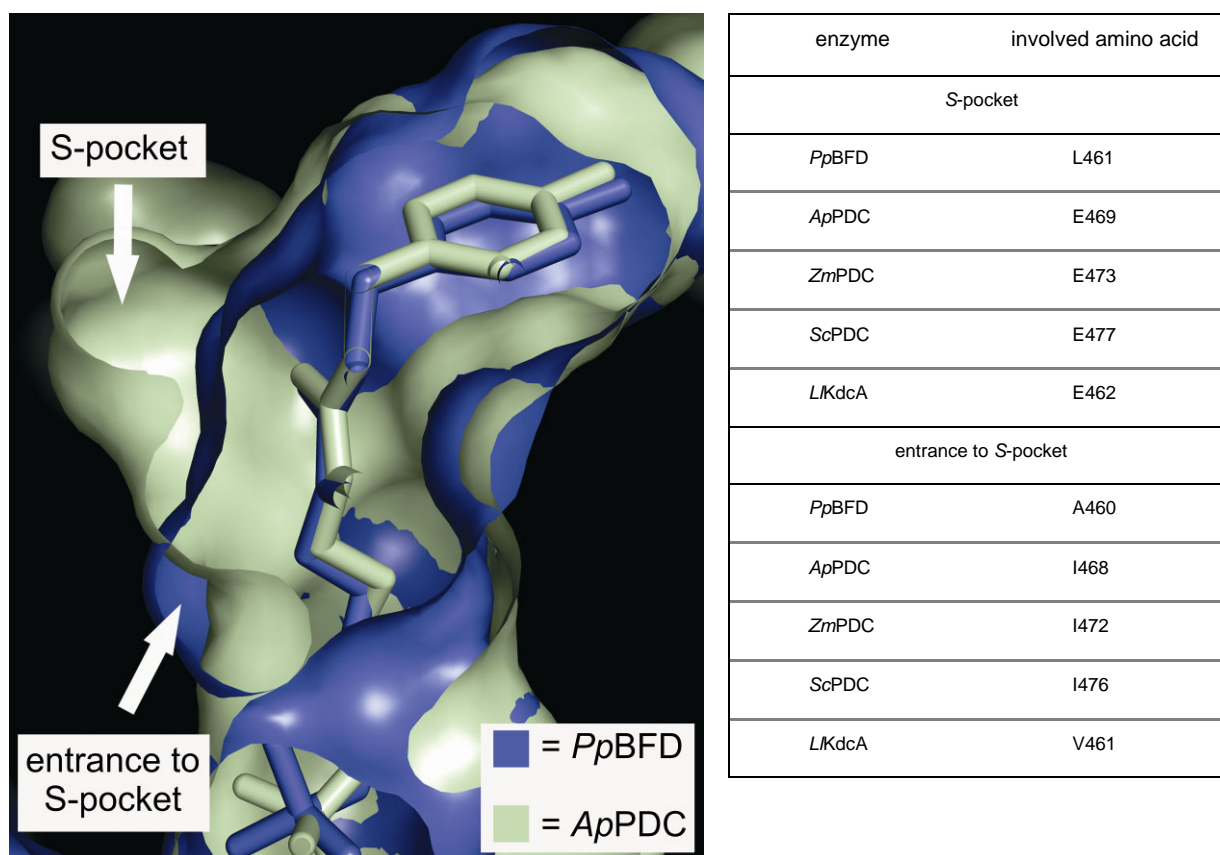


Figure 3: Superimposition of the *S*-pockets of BFDwt and *ApPDC*. Compared to BFD *ApPDC* shows an enlarged *S*-pocket but the entrance is blocked by residue I468 (left). The amino acids mainly bordering the *S*-pocket as well as those defining the entrance to the pocket are given for BFD, *ApPDC* and additionally for *ZmPDC*, *ScPDC* and *KdcA* (right).

Table 4. Preparative scale carbonylation of benzaldehyde and propanal catalyzed by different BFD variants. Products were isolated by flash column chromatography.

substrate/ product ^a	BFDwt (ee)	BFDL461A (ee)	BFDL461G (ee)
benzaldehyde	1.1%	-	-
4	35.6% (96% <i>R</i>)	-	-
6	36.7% (98% <i>R</i>)	-	-
7	26.7% (21% <i>R</i>)	> 99% (93% <i>S</i>)	> 99% (97% <i>S</i>)
isolated yield	16.2 mg 37 mol% (w/w)	15 mg 35 mol% (w/w)	13.7 mg 31 mol% (w/w)

[a] Numbers refer to Table 3. Product composition is given in mol% (NMR); ee-values were determined by chiral HPLC.

Conclusion

We have successfully engineered *S*-specific BFD variants using a structure-guided approach. By investigating the origin of *S*-selectivity we demonstrate the possibility to shape this part of the active site selectively for longer chain aliphatic acceptor aldehydes. The experimental data are very well predictable by modeling studies, which allows the *in silico* design of *S*-specific biocatalysts for special requirements. In order to generalize this strategy the 3D structures of other ThDP-dependent enzymes related to BFD have been compared.

A superimposition of the crystal structures of BFDwt^[5], BAL from *Pseudomonas fluorescens*,^[8] PDC from *Zymomonas mobilis* (ZmPDC),^[9a] *Saccharomyces cerevisiae* (ScPDC)^[9b] as well as the recently solved structures of PDCs from *Acetobacter pasteurianus* (ApPDC)^[10] and KdcA from *Lactococcus lactis* ^[11, 12, 13] gave profound insights. While there is no *S*-pocket visible in BAL, the *S*-pockets of the other enzymes increase in the series *Pp*BFD < *Ll*KdcA < ZmPDC / ScPDC < ApPDC (Figure 3).

However, the entrance of the *S*-pocket in KdcA and in both PDCs is restricted by bulky residues, such as isoleucine or valine, which could explain, why all these enzymes are strictly *R*-selective. Consequently, (*S*)-2-hydroxy ketones could be formed by improving the access to this pocket. This has successfully been shown with the variant ZmPDCI472A which catalyzes the formation of (*S*)-HPP (**3**) (*ee* 70%) while exclusively (*R*)-phenylacetylcarbinol (**2**) (*ee* > 98%) is formed with the wild type enzyme using benzaldehyde and acetaldehyde as substrates.^[14]

The predominant *R*-selectivity of ThDP-dependent enzymes is therefore soundly explained mainly from their structures. However, many of these enzymes have the latent inherent property of *S*-selectivity, since such *S*-pockets are visible in almost all 3D structures mentioned above, although they are not accessible in many cases. Our results pave the way to expand the shaping strategy of the *S*-pockets to a broad range of other 2-ketoacid decarboxylases. This is a powerful tool to enlarge the toolbox of enzymatically accessible 2-hydroxy ketones by *S*-enantiomers, therewith providing a valuable platform for chemo-enzymatic synthesis.

Experimental Section

Site-directed mutagenesis

The 1611 bp gene of benzoylformate decarboxylase (BFD, E.C. 4.1.1.7) from *Pseudomonas putida* was ligated into a pKK233-2 plasmid (Pharmacia),^[3,15] which contained the information for a C-terminal His₆-tag. For mini- and midi preparations *E. coli* XL1-blue (Stratagene) was transformed with the construct by electroporation. For overexpression *E. coli* SG13009/pRep4 (Qiagen) was used as a host. Site-directed mutagenesis was performed using the QuikChange® Site-directed Mutagenesis Kit (Stratagene). Mutagenesis primers are given in the supplementary material. Gene sequences were confirmed by DNA sequencing (Sequiserve).

Expression and purification

Cultivation of the variants was done in shaking cultures (1 L LB-medium, pH 7.5, 5 L flasks). Overexpression was induced by addition of 1 mM IPTG at OD₆₀₀ ≤ 0.45. For biotransformations the variants were purified according to a protocol previously developed for the wild type enzyme (BFDwt)^[3,15] (Ni-NTA-chromatography: disintegration buffer (50 mM potassium phosphate, pH 7.0, 2.5 mM MgSO₄, 0.1 mM ThDP), washing buffer (50 mM potassium phosphate, pH 7.0, 20 mM imidazole), elution buffer (50 mM potassium phosphate, pH 7.8, 250 mM imidazole); G25-chromatography (10 mM potassium phosphate pH 7.0, 2.5 mM MgSO₄, 0.1 mM ThDP)). After purification the enzyme variants were either freeze dried or diluted with 50% (v/v) glycerol and stored at -20 °C.

For enzyme crystallization of BFDL461A the purification protocol was kept but buffers were changed (Ni-NTA-chromatography: disintegration buffer (Mes/NaOH, 50 mM, pH 7.0, 2.5 mM MgSO₄, 0.1 mM ThDP), washing buffer (Mes/HCl, 50 mM, pH 7.0, 50 mM imidazole), elution buffer (Mes/HCl, 50 mM, pH 7.0, 250 mM imidazole); G25-chromatography (Mes/NaOH, 20 mM, pH 7.0, 2.5 mM MgSO₄, 0.1 mM ThDP). Concentration of enzyme solutions was performed in vivaspin 20 centrifuge columns (Sartorius, cut off 10 kDa) up to 130 mg/mL. Superdex G200-size (GE Healthcare) exclusion chromatography showed a purity of 97.5% of the tetrameric BFDL461A (2% dimeric BFDL461A, 0.5% impurity). For storage the enzyme solution was shock-frozen in liquid nitrogen and kept at -20 °C.

Decarboxylase activity assay

One unit of decarboxylase activity is defined as the amount of enzyme which catalyzes the decarboxylation of 1 μmol benzoylformate per minute under standard conditions (pH 6.5, 30°C). Activity was measured using a coupled photometric assay as previously described.^[3] For determination of the substrate range different 2-keto acids were applied in a final concentration of 30 mM in this assay; except indole-3-pyruvate (1 mM) (see supplementary material).

Protein determination was performed according to Bradford^[16] using bovine serum albumin (BSA) as the standard.

Benzoin syntheses

Reaction conditions: benzaldehyde (20 mM), DMSO (20 vol%), BFD variant (0.3 mg/mL), potassium phosphate buffer (50 mM, pH 7.5, 2.5 mM MgSO_4 and 0.1 mM ThDP), 30°C, 100 rpm. To avoid evaporation of the aldehydes the reaction batch was divided into GC-vials with a volume of 400 μL , after taking the starting sample. The reaction was stopped by addition of acetonitrile (400 μL) followed by intense vortexing and centrifugation of the precipitate. Calibration curves with benzoin were prepared in the same way. Conversions were determined by HPLC, employing a Dionex HPLC (Germering), equipped with a 250 x 4,6 Multohyp ODS-5 μ (CS-Chromatography) and an UV-detector (mobile phase 60% (v/v) H_2O : 40% (v/v) acetonitrile, flow 1.1 mL/min, pressure 130 bar, 20 μL injection volume, detection $\lambda = 250$ nm), R_t (benzoin) = 32.2 min.

Acetoin syntheses

Reaction conditions: acetaldehyde (40 mM), DMSO (20 vol%), BFD variant (0.3 mg/mL), potassium phosphate buffer (50 mM, pH 7.5, 2.5 mM MgSO_4 , 0.1 mM ThDP), 30°C, 100 rpm. As described for the benzoin synthesis the reaction batch was divided into GC-vials with a volume of 400 μL each. For enzyme inactivation the vial was heated for 60 sec at 90°C followed by centrifugation of the precipitate. Conversion and enantiomeric excess were determined by chiral GC, employing a 6890 N Agilent GC (Palo Alto), equipped with a Cyclodex b-1/P column (50 m x 320 μm) and a FID detector (flow 3.4 mL/min, pressure 0.8 bar, split 5:1, 1 μL injection volume, temperature gradient: 50 °C for 5 min, 40 °C/min to 190 °C), R_t ((R)-acetoin) = 6.98 min, R_t ((S)-acetoin) = 7.11 min.

Mixed carboligations of benzaldehyde and different aliphatic aldehydes

Analytical scale: Reaction conditions: 1.5 mL-scale, benzaldehyde (0.027 mmol, 2.9 mg) was dissolved in a mixture of DMSO (0.3 mL) and potassium phosphate buffer (50 mM, 1.2 mL, pH 7, 2.5 mM MgSO₄, 0.1 mM ThDP). To this solution acetaldehyde, propanal, or butanal (0.027 mmol) was added. After addition of purified enzyme (0.45 mg) the reaction was stirred slowly at 30°C for 72 h. The reaction mixture was extracted with CDCl₃.

Preparative scale synthesis: Reaction conditions: 15 mL scale: benzaldehyde (29 mg, 0.27 mmol) and propanal (16 mg, 0.27 mmol) were dissolved in DMSO (3 mL). After addition of potassium phosphate buffer (50 mM, 12 mL, pH 7.9, 2.5 mM MgSO₄, 0.1 mM ThDP) the reaction was started with the purified BFD variant (4.5 mg) and stirred slowly at 30°C. After 26.5 h further of BFDwt or the variants (4.5 mg), respectively, were added. The reaction was stopped after 70 h (BFDwt, BFDL461A) or 50 h (BFDL461G) by extracting three times with ethyl acetate (25 mL) and the organic layer was dried over Na₂SO₄. The solvent was evaporated and the crude product was dissolved in ether (5 mL). The ether extract was washed with brine and dried over Na₂SO₄ followed by evaporation of the solvent.

Analysis of (S)-2-hydroxy-1-phenylbutan-1-one ((S)-7)

HPLC: (chiral OD-H, n-hexane: 2-propanol, 95:5, 0.5 mL min⁻¹, 40°C): R_t (S) = 11.8 min, R_t (R) = 13.9 min; [α]_D²⁵ = -11.52 (c = 0.4, CHCl₃); CD (acetonitrile): λ [nm] (mol. CD) = 298 (-0.3250), 281 (-1.7045), 239 (6.2025), 206 (-7.6004); ¹H NMR (400 MHz, CDCl₃, 300 K, ppm): δ = 0.96 (t, 3H, ³J (H, H) = 7.4 Hz, CH₃), 1.63 (dq, 1H, ²J (H, H) = 14.3 Hz, ³J (H, H) = 7.4 Hz, 7.3 Hz, CH₂), 1.98 (dq, 1H, ²J (H, H) = 14.3 Hz, ³J (H, H) = 7.4 Hz, 3.8 Hz, CH₂), 3.73 (d, 1H, ³J (H, H) = 6.4 Hz, OH), 5.08 (ddd, 1H, ³J (H, H) = 7.3 Hz, 6.4 Hz, 3.8 Hz, CHOH), 7.52 (ddm, 2H, ³J (H, H) = 7.4 Hz, Ar-H), 7.64 (ddm, 1H, ³J (H, H) = 7.4 Hz, Ar-H), 7.93 (dm, 2H, ³J (H, H) = 7.4 Hz, Ar-H); ¹³C NMR (100 MHz, CDCl₃, 300 K, ppm): δ = 8.8 (CH₃), 28.8 (CH₂), 73.9 (CHOH), 128.4 (2 x CH_{Ar}), 128.8 (2 x CH_{Ar}), 133.3 (CH_{Ar}), 202.1 (CO); GCMS R_t = 8.7 min; MS (70 eV, EI): m/z (%): 164 (0.1%) [M⁺], 105 (100%), 77 (46%).

Crystallization

BFDL461A was crystallized using the hanging drop vapor diffusion method. Droplets were set up for crystallization by mixing of protein solution (2 μ L) containing protein (13 g/mL) in Mes/NaOH (20 mM, pH 7.0, 2.5 mM MgSO₄, 0.1 mM ThDP) and 2 μ L of the reservoir solution. Screening and optimization revealed a reservoir solution of 18-24% (w/v) PEG 2000

MME, 0.1 M sodium citrate, pH 5.2-5.8, 100-150 mM $(\text{NH}_4)_2\text{SO}_4$ to be best. After equilibration for 3 days diffraction-quality crystals appeared.

Data collection and processing

For cryoprotection the crystals were quickly dipped into the well solution supplemented with 25% ethylene glycol before being frozen in a cryogenic nitrogen gas stream at 110 K. Data were collected to a resolution of 2.2 Å at beamline I911-3 at Max-lab, Lund, Sweden. Images were processed using MOSFLM,^[17] where unit-cell parameters were determined by the autoindexing option. The data set was scaled with the program SCALA implemented in the CCP4 program suite.^[18] The crystal belongs to the space group $P2_12_12_1$ with the cell dimensions $a = 96$ Å, $b = 140$ Å and $c = 169$ Å. Four monomers were packed in one asymmetric unit. Data collection statistics are given in Tab. 5.

Table 5. Data collection statistics of BFDL461A. Values in parentheses are given for the highest resolution interval.	
Resolution [Å]	2.2 (2.32-2.2)*
No. of observations	321349 (46653)
No. of unique reflections	111569 (16561)
Completeness [%]	97.0 (99.2)
Multiplicity [%]	2.9 (2.8)
Mean $\{I/\sigma(I)\}$	8.8 (2.2)
Wilson B factor [Å ²]	27.8
R_{merge} [%]	12.4 (45.8)

Structure solution and crystallographic refinement

The structure of BFDL461A was determined by molecular replacement using the program MOLREP.^[18,19] The BFDwt structure (pdb accession code: 1BFD)^[5] was used as search model to place the four monomers into the asymmetric unit. Atomic positions and B-factors of the model were refined by the maximum likelihood method in REFMAC5^[18,20] which was interspersed with rounds of manual model building in COOT.^[21] Non-crystallographic symmetry restraints were initially applied but released toward the end of refinement. Water assignment was performed in COOT where the model was validated as well. The quality of

the final structure was examined using PROCHECK.^[18, 22] Statistics of the refinement and final model are given in the supplementary material. The coordinates of BFDL461A, in addition to the structure factors, have been deposited in the Research Collaboratory for Structural Bioinformatics Protein Databank PDB with the accession code 2V3W.

Structural analysis and substrate placement

The modeling studies were carried out applying the programs PyMol^[23] and Swiss-Pdb Viewer.^[24] To investigate the differences in the stereoselectivities of 2-hydroxy ketone formation of BFDwt and BFDL461A and to predict the optimal substrate size of the acceptor aldehyde fitting in the *S*-pocket, the acceptor and donor aldehydes were placed into the active sites as described previously.^[6] Models of the molecules were created using the molecular builder in the program SYBYL (Tripos, St. Louis, MO).

Acknowledgements

This research was kindly supported by Degussa. Geraldine Kolter thanks the Deutsche Forschungsgemeinschaft for financial support in frame of the Graduiertenkolleg "BioNoco". We gratefully thank Ilona Frindi-Wosch for skilful technical assistance.

References

- [1] M. Pohl, G. A. Sprenger, M. Müller, *Curr. Opin. Biotechnol.* **2004**, *15*, 335-342.
- [2] A. S. Demir, M. Pohl, E. Janzen, M. Müller, *J. Chem. Soc. Perkin Trans. 1* **2001**, 633-635.
- [3] a) H. Iding, T. Dünwald, L. Greiner, A. Liese, M. Müller, P. Siegert, J. Grötzinger, A. S. Demir, M. Pohl, *Chem. Eur. J.* **2000**, *6*, 1483-1495.; b) B. Lingen, D. Kolter-Jung, P. Dünkermann, R. Feldmann, J. Grötzinger, M. Pohl, and M. Müller **2003** *Chembiochem* *4*, 721-726
- [4] M. Knoll, M. Müller, J. Pleiss, M. Pohl, *Chembiochem* **2006**, *7*, 1928-1934.
- [5] a) M. S. Hasson, A. Muscate, M. J. McLeish, L. S. Polovnikova, J. A. Gerlt, G. L. Kenyon, G. A. Petsko, D. Ringe, *Biochemistry* **1998**, *37*, 9918-9930; b) E. S. Polovnikova, M. J. McLeish, E. A. Sergienko, J. T. Burgner, N. L. Anderson, A. K. Bera, F. Jordan, G. L. Kenyon, M. S. Hasson, *Biochemistry* **2003**, *42*, 1820-1830.
- [6] a) K. C. Nicolaou, Z. Yang, J. J. Liu, H. Ueno, P. G. Nantermet, R. K. Guy, C. F. Clalborne, J. Renaud, E. A. Couladouros, K. Paulvannan, E. J. Sorensen, *Nature* **1994**, *367*, 630-634; b) F. R. Stermitz, P. Lorenz, J. N. Tawara, L. A. Zenewicz, K. Lewis, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 1433-1437.

- [7] P. Domínguez de María, M. Pohl, D. Gocke, H. Gröger, H. Trauthwein, T. Stillger, L. Walter, M. Müller, *Eur. J. Org. Chem.* **2007**, 2940-2944.
- [8] T. G. Mosbacher, M. Müller, G. E. Schulz, *FEBS J.* **2005**, 272, 6067-6076.
- [9] a) D. Dobritzsch, S. König, G. Schneider, G. Lu, *J. Biol. Chem.* **1998**, 273, 20196-20204; b) P. Arjunan, T. Umland, F. Dyda, S. Swaminathan, W. Furey, M. Sax, B. Farrenkopf, Y. Gao, D. Zhang, F. Jordan, *J. Mol. Biol.* **1996**, 256, 590-600.
- [10] D. Gocke, C. Berthold, T. Graf, H. Brosi, I. Frindi-Wosch, M. Knoll, T. Stillger, L. Walter, M. Müller, J. Pleiss, G. Schneider, M. Pohl, *submitted*.
- [11] A. Yep, G. L. Kenyon, M. J. McLeish, *Bioorg. Chem.* **2006**, 34, 325-336.
- [12] D. Gocke, C. L. Nguyen, M. Pohl, T. Stillger, L. Walter, M. Müller, *Adv. Synth. Catal.* **2007**, 349, 1425-1435.
- [13] C. Berthold, D. Gocke, M. D. Wood, F. Leeper, M. Pohl, G. Schneider, *Acta Crystallogr. Sec. D: Biol. Crystallogr.*, *accepted*.
- [14] P. Siegert, M. J. McLeish, M. Baumann, H. Iding, M. M. Kneen, G. L. Kenyon, M. Pohl, *Protein Eng. Des. Sel.* **2005**, 18, 345-357.
- [15] A. Y. Tsou, S. C. Ransom, J. A. Gerlt, D. D. Buechter, P. C. Babbitt, G. L. Kenyon, *Biochemistry* **1990**, 29, 9856-9862.
- [16] M. M. Bradford, *Anal. Biochem.* **1976**, 72, 248-254.
- [17] A. G. W. Leslie, *Joint CCP4 + ESF-EAMCB Newsl. Protein Crystallogr.* **1992**, 26.
- [18] Collaborative Computational Project Number 4, *Acta Crystallogr. Sec. D: Biol. Crystallogr.* **1994**, 50, 760-763.
- [19] A. Vagin, A. Teplyakov, *J. Appl. Crystallogr.* **1997**, 30, 1022-1025.
- [20] G. N. Murshudov, A. A. Vagin, E. J. Dodson, *Acta Crystallogr. Sec. D: Biol. Crystallogr.* **1997**, 53, 240-255.
- [21] P. Emsley, K. Cowtan, *Acta Crystallogr. Sec. D: Biol. Crystallogr.* **2004**, 60, 2126-2132.
- [22] R. A. Laskowsky, M. W. McArthur, D. S. Moss, J. M. Thornton, *J. Appl. Crystallogr.* **1993**, 26, 283-291.
- [23] W. L. DeLano, The PyMOL Molecular Graphics System, <http://www.pymol.org>, San Carlo, CA, USA, **2002**.
- [24] N. Guex, C. Peitsch, *Electrophoresis* **1997**, 18, 2714-2723. <http://www.expasy.org/spdbv/>.

4.3

Publikation eingereicht bei *BBA: Proteins & Proteomics*

Gocke, D., Berthold C. L., Graf, T., Brosi, B., Frindi-Wosch, I., Knoll, M., Stillger, T., Walter, L., Müller, M., Pleiss, J., Schneider, G., Pohl, M., Pyruvate Decarboxylase from *Acetobacter pasteurianus*: BIOCHEMICAL AND STRUCTURAL CHARACTERISATION.

**Pyruvate Decarboxylase from *Acetobacter pasteurianus*:
BIOCHEMICAL AND STRUCTURAL CHARACTERISATION**

**Dörte Gocke¹, Catrine L. Berthold², Thorsten Graf¹, Helen Brosi¹, Ilona Frindi-Wosch¹,
Michael Knoll³, Thomas Stillger⁴, Lydia Walter⁴, Michael Müller⁴, Jürgen Pleiss³,
Gunter Schneider², Martina Pohl^{1*}**

¹*Institute of Molecular Enzyme Technology, Heinrich-Heine University Düsseldorf, Research Centre Jülich, D-52426 Jülich, Germany;* ²*Department of Medical Biochemistry and Biophysics, Karolinska Institutet, S-17177 Stockholm, Sweden;* ³*Institute of Technical Biochemistry, University of Stuttgart, D-70569 Stuttgart, Germany;* ⁴*Institute of Pharmaceutical Sciences, Albert-Ludwigs-University Freiburg, 79104 Freiburg, Germany*

*Corresponding author.

Tel.: +49-(0)2461-613704; Fax: +49-(0)2461-612490; E-mail: ma.pohl@fz-juelich.de

Keywords: *Acetobacter pasteurianus*; pyruvate decarboxylase; enzymatic carbonylation; thiamin diphosphate; X-ray structure; 2-hydroxyketones

Summary

Pyruvate decarboxylase from *Acetobacter pasteurianus* (*ApPDC*), a thiamin diphosphate (ThDP)-dependent enzyme, has been investigated concerning its decarboxylase activity as well as the potential to catalyse the carbonylation of aldehydes to chiral 2-hydroxyketones. Beside a detailed characterisation of the substrate range, the stability of *ApPDC* was studied with regard to the influences of different pH-values and temperatures. Data were compared to those obtained with similar PDCs from *Zymomonas mobilis*, *Zymobacter palmae* and *Saccharomyces cerevisiae*. Although *ApPDC* resembles to a large extent the well characterised enzyme from *Zymomonas mobilis* with respect to sequence, kinetic behaviour and stability, there are some remarkable differences concerning the decarboxylation of aromatic 2-ketoacids and the stereoselectivity of the carbonylation reaction, which hint to a larger active site in *ApPDC*. These results can be explained based on the X-ray structure of the enzyme, which was determined to a resolution of 2.75 Å.

1. Introduction

Pyruvate decarboxylases (PDC, E.C. 4.1.1.1) catalyse the non-oxidative decarboxylation of pyruvate to acetaldehyde requiring the cofactors thiamin diphosphate (ThDP) and divalent cations like Mg^{2+} . They are the key enzymes in homo-fermentative pathways cleaving the central intermediate pyruvate into acetaldehyde and CO_2 . PDCs are commonly found in plants, yeasts and fungi, but absent in mammals [1]. Currently four PDCs from prokaryotes have been cloned and characterised, encompassing three PDCs from the gram-negative bacteria *Zymomonas mobilis* (ZmPDC) [2], *Zymobacter palmae* (ZpPDC) [3], and *Acetobacter pasteurianus* (ApPDC) [4], as well as the PDC from the gram-positive bacterium *Sarcina ventriculi* (SvPDC) [5].

A. pasteurianus, a common spoilage organism of fermented juices, is an obligate oxidative bacterium and in contrast to the fermentative ethanol pathways in plants, yeast and bacterial PDCs, ApPDC is a central enzyme of the oxidative metabolism [4].

Pyruvate decarboxylases are key enzymes in the production of biofuels from bulk plant materials like sugars or cellulose. Currently bacterial strains with engineered ethanol pathways, e.g. *Escherichia coli* strains with an inserted ZmPDC, alcohol dehydrogenase (ADH) and acyltransferase are investigated for industrial applications [6]. Moreover, PDCs are able to catalyse an acyloin-condensation like carboligation of aldehydes. Since the 1930s this activity of PDC from yeast is used in an industrial fermentative process for the production of (*R*)-phenylacetylcarbinol (**20**), a pre-step in the synthesis of ephedrine [7, 8]. A similar carboligase activity yielding chiral 2-hydroxy ketones was detected and intensively studied for other PDCs and further ThDP-dependent enzymes like benzoylformate decarboxylase (BFD, E.C. 4.1.1.7), branched-chain 2-keto acid decarboxylase (KdcA, E.C. 4.1.1.72) and benzaldehyde lyase (BAL, E.C. 4.1.2.38) [9-18]. Different substrate preferences of these enzymes open the way to a broad range of 2-hydroxy ketones. Our intention is to create a 2-hydroxy ketone platform encompassing a high diversity of enantiocomplementary molecules. To achieve this goal a detailed investigation of structure-function relationships of ThDP-dependent enzymes is necessary in order to elucidate the mechanisms regulating chemo- and enantioselectivity and to enable the design of biocatalysts with tailor-made carboligase activity. Using a structure guided approach we have recently opened the way to access (*S*)-hydroxy ketones by site-directed mutagenesis of the predominantly strictly *R*-selective ThDP-dependent enzymes [19, 20].

Here we report the characterisation of the pyruvate decarboxylase from *Acetobacter pasteurianus* (*ApPDC*) with regard to the substrate range of the decarboxylase, the carboligase reaction and the pH- and temperature dependent stability and activity in comparison to other PDCs from *Zymomonas mobilis* (*ZmPDC*), *Zymobacter palmae* (*ZpPDC*) and *Saccharomyces cerevisiae* (*ScPDC*). Furthermore the three-dimensional (3D) structure of *ApPDC* was determined by X-ray crystallography demonstrating a significant larger active site compared to other PDCs of known structure, which explains the observed differences in the substrate range and the stereoselectivity of *ApPDC*.

2. Materials and methods

2.1. Cloning of *ApPDC*

Genomic DNA of *Acetobacter pasteurianus* subspec. *ascendens* (DSM-No. 2347 (= ATCC 12874), DSMZ) was isolated with the DNeasy Tissue Kit (Qiagen) using the manufacturers protocol for gram-negative bacteria (DNeasy Tissue Kit Handbook 05/2002). The gene was amplified by PCR using the following primers:

*ApPDC*_hisC-up:

5'-ATAT**CATATG**ACCTATACTGTTGGCAT-3'

*NdeI*start

*ApPDC*_hisC-down:

5'-ATAT**CTCGAG**GGCCAGAGTGGTCTTGC-3'

XhoI

After a single digest of the amplified gene with *XhoI* the gene was first ligated into the vector pBluescriptII (Fermentas) in order to allow restriction by *NdeI* and *XhoI*. The 1674 bp long gene was digested with both restriction endonucleases and ligated into the vector pET22b (Novagen) carrying an ampicillin-resistance and the information for a C-terminal His₆-tag. Transformation of the expression host *E. coli* BL21(DE3) (Novagen) was performed by electroporation. The DNA sequence was verified by sequencing (Sequiserie) and compared to the published sequence [4] (see supplementary material). *ZpPDC* was cloned into pET28a(+) carrying a C-terminal His₆-tag.

2.2. Cultivation and expression

Cultivation, expression, purification and storage of *ZpPDC* and *ScPDC* were performed as described for *ApPDC* in the following. High cell density cultivation according to the method of Korz et al. [21] was performed in a 40 L Techfors reactor (Infors AG, CH) at 30°C, pH 7.0.

Starting on minimal media the fed-batch process was run by a continuous feed of glucose. pH 7.0 was maintained by NH₃-addition. Dissolved oxygen saturation was regulated between 30-40% saturation by increasing the stirrer speed and the air flow rate and by a stepwise pressure increase during the high oxygen demanding expression. After a preliminary growth phase protein production was induced by the addition of 1.5 mM isopropyl β -D-1-thiogalactopyranoside (IPTG) at an OD₆₀₀~60. Cells were harvested by a separator (Westfalia Separator AG) and stored at -20°C.

2.3. Purification and storage

Harvested *E. coli* cells were suspended by addition of 1:10 (w/v) buffer A (50 mM potassium phosphate buffer, pH 6.5, including 2.5 mM MgSO₄ and 0.1 mM ThDP) and 10 mM. After addition of 0.33 mg/mL lysozyme and incubation for 30 min on ice, cells were further disrupted by ultra-sonification (10 times 30 sec) followed by centrifugation.

Purification of ApPDC for biochemical studies was performed by immobilised nickel chelate chromatography and additional size exclusion chromatography, using a purification protocol previously developed for ZmPDC [22] with the following alterations: The pH of all 50 mM potassium phosphate buffers was set to 6.5; all buffers contained 2.5 mM MgSO₄ and 0.1 mM ThDP; non-specifically bound proteins were eluted with 50 mM imidazole; ApPDC was eluted with 250 mM imidazole. The enzyme was either freeze dried or diluted with 50% (v/v) glycerol and stored at -20 °C.

For crystallisation of ApPDC the purification protocol was kept but potassium buffer was exchanged by Mes/NaOH-buffer as following: Ni-NTA-chromatography: disintegration buffer (50 mM Mes/NaOH-buffer, pH 6.5, 2.5 mM MgSO₄, 0.1 mM ThDP), washing buffer (50 mM Mes/HCl-buffer, pH 6.5, 50 mM imidazole), elution buffer (50 mM Mes/HCl-buffer, pH 6.5, 250 mM imidazole); G25-chromatography: 20 mM Mes/NaOH-buffer, pH 7.0, 2.5 mM MgSO₄, 0.1 mM ThDP. The enzyme solution was concentrated up to 35 mg/mL using vivaspin 20 centrifuge columns (Sartorius, cut off 10 kDa). For storage the enzyme solution was shock-frozen in liquid nitrogen and kept at -20 °C.

2.4. Determination of the molecular mass

Size-exclusion chromatography was performed using a Superdex G200 16/60 prep grade column (Amersham) and buffer A (pH 6.5) including 150 mM KCl. The coefficients of available volume ($K_{av} = (V_e - V_o)/(V_t - V_o)$, with V_e = elution volume of the respective protein

and V_t = total volume, elution volume of blue dextran) for *ApPDC* and the standard proteins were determined twice with a standard deviation of 0.2% (see supplementary material).

Calibration was performed using ribonuclease A (13.7 kD, $K_{av} = 0.63$), chymotrypsinogen A (25 kD, $K_{av} = 0.59$), ovalbumin (43 kD; $K_{av} = 0.46$), BSA (67 kD, $K_{av} = 0.37$), aldolase (158 kD, $K_{av} = 0.27$), catalase (232 kD, $K_{av} = 0.23$), ferritin (440 kD, $K_{av} = 0.12$) and thyroglobulin (669 kD, $K_{av} = 0.05$). The lyophilisate (56 mg) was solved in 600 μL buffer A (pH 6.5) including 60 μL 1.5 M KCl; flow 1 mL/min. The K_{av} -value of *ApPDC* was determined to 0.125. Data were plotted as $\log M_r$ over K_{av} resulting in a linear correlation with $R^2 = 0.993$.

2.5. Decarboxylase activity

One unit of decarboxylase activity is defined as the amount of *ApPDC* which catalyses the decarboxylation of 1 μmol pyruvate (**1**) per minute under standard conditions (pH 6.5, 30°C). Protein determination was performed according to Bradford [23] using BSA as the standard.

Two continuous decarboxylase assays were used. In the *coupled decarboxylase assay* yeast alcohol dehydrogenase (yeast ADH) reduces the aldehydes obtained after decarboxylation of 2-keto acids by *ApPDC*. The simultaneous consumption of NADH was followed spectrophotometrically for 90 sec at 340 nm (molar extinction coefficient ϵ of NADH = 6.22 L $\text{mmol}^{-1} \text{cm}^{-1}$). Assay composition: 890 μL buffer A (pH 6.5), 50 μL sodium pyruvate or another 2-keto acid (350 mM) in buffer A (final concentration 17.5 mM), 50 μL NADH 4.6 mM in buffer A (final concentration 0.23 mM), 10 μL yeast ADH (Roche, EC 1.1.1.1, 1 g in 34 mL 3.2 M ammonium sulphate, pH 6.0, 300 U/mg at 25°C). The assay was started by addition of 50 μL *ApPDC*. The concentration of *ApPDC* was chosen in such a way that a linear decay of NADH was observed over 90 sec (ca. 0.2 mg/mL *ApPDC* with pyruvate as the substrate). The concentration of yeast ADH was sufficient to make certain that the coupling reaction was not rate limiting. The 3.2 M ammonium sulphate storage buffer of yeast ADH did not influence the results.

Further, a *direct decarboxylase assay* following the direct decay of pyruvate was developed to measure *ApPDC* activity under NADH degrading or ADH inactivating conditions. Assay composition: 950 μL sodium pyruvate (15 mM in buffer A, pH 6.5). The reaction was started by addition of 50 μL *ApPDC* solution (ca 0.8 mg/mL). The decay of pyruvate was followed at 320 nm ($\epsilon_{\text{pyruvate}} = 0.022 \text{ L mmol}^{-1} \text{cm}^{-1}$) in quartz cuvettes.

Kinetic constants were calculated by non-linear regression using the Michaelis-Menten equation in Origin 7G SR4 (OriginLab Coop., Northampton) except for *ScPDC*, where the Hill equation including substrate inhibition was used (see supplementary material). *ScPDC*

decarboxylase activity was measured using the same assay as described for *ApPDC*. The activity assay for *ZpPDC* was performed at pH 7.0, which is the optimum for initial rate activity of *ZpPDC* (Fig. 2). Assays with *ZmPDC* were performed in 50 mM Mes/KOH-buffer, pH 6.5, including 2 mM MgSO₄ and 0.1 mM ThDP [24].

Both decarboxylase assays show a standard deviation of about 10%. Therefore all decarboxylase activities in this paper are given as the average of minimum 3-6 measurements.

Substrate range of decarboxylase activity- Activity toward different 2-keto acids was measured with the coupled decarboxylase assay using horse liver ADH as coupling enzyme. It was verified that the respective aldehydes, which are obtained by decarboxylation, are substrates for the horse liver ADH. *ZmPDC* activity was detected in Mes/KOH-buffer, pH 6.5 [24] while the presented data were all measured in potassium phosphate buffer, pH 6.5. All substrates were applied at a concentration of 30 mM, except for indole-3-pyruvate (**14**), being applied at 1 mM due to low solubility and strong absorbance.

Optima and stability investigations- For *ApPDC*, *ScPDC* and *ZpPDC* the pH- and temperature optima were measured using the direct decarboxylase assay. To investigate the stability towards pH, temperature and organic solvents, *ApPDC* was incubated under the conditions given in the figure legends. Residual activity was assayed with the coupled decarboxylase assay. It was verified, that the pH in the cuvettes was stable during the time of detection.

2.6. Carboligation

Benzoin formation- To avoid evaporation of the aldehydes the reaction batch was divided into 200 μ L tubes. Reaction conditions (50 μ L for each tube): 0-89 mM benzaldehyde in 15 μ L dimethylsulfoxide (DMSO) (final concentration: 0-38 mM benzaldehyde, 30% (v/v) DMSO). The reaction was started by addition of 35 μ L *ApPDC* (4.2 mg/mL) in buffer A (pH 6.5), 30 °C, 100 rpm. The reaction was stopped by addition of 50 μ L acetonitrile followed by vortexing and centrifugation of the precipitate. Calibration curves with benzoin were handled like the reaction samples.

The conversion was determined by HPLC, employing a Dionex HPLC (Germering), equipped with a 250 x 4.6 Multohyp ODS-5 μ (CS-Chromatography) and a UV-detector (mobile phase 59.5% H₂O : 40% acetonitrile: 0.5% acetic acid, flow 1.1 mL/min, pressure 130 bar, 20 μ L injection volume, detection at $\lambda = 250$ nm), R_t (benzoin) = 32.2 min.

Mixed ligations of aromatic and aliphatic aldehyde- The conversion was followed by GCMS, employing a HP 6890 series GC-system fitted with a HP 5973 mass selective detector (Hewlett Packard; column HP-5MS, 30 m x 250 μ m; T_{GC} (injector) = 250°C, T_{MS} (ion

source) = 200°C, time program (oven): $T_{0 \text{ min}} = 60^\circ\text{C}$, $T_{3 \text{ min}} = 60^\circ\text{C}$, $T_{14 \text{ min}} = 280^\circ\text{C}$ (heating rate $20^\circ\text{C}\cdot\text{min}^{-1}$), $T_{19 \text{ min}} = 280^\circ\text{C}$). The enantiomeric excess (*ee*) was determined by chiral phase HPLC employing a HP 1100 HPLC system (Agilent) fitted with a diode-array detector. Yields were calculated in mol% after NMR detection. NMR spectra were recorded on a Bruker DPX-400. Chemical shifts are reported in ppm relative to CHCl_3 (^1H NMR: $\delta = 7.27$) and CDCl_3 (^{13}C NMR: $\delta = 77.0$) as internal standards. CD-spectra were recorded on a JASCO J-810 spectropolarimeter using acetonitrile as solvent.

(R)-1-Hydroxy-1-phenylpropan-2-one (**20**)-

Reactions conditions: benzaldehyde (150 mg, 30 mM) and acetaldehyde (110 mg, 50 mM) were dissolved in 40 mL buffer A (pH 7.0) and 10 mL DMSO. After addition of purified ApPDC dissolved in the same buffer (63 U/mL decarboxylase activity) the reaction was stirred slowly at 30°C. After 24 h additional acetaldehyde (50 mM) and ApPDC (42 U/mL) were added. After 72 h the reaction mixture was extracted three times with ethyl acetate (25 mL) and the organic layer was dried over Na_2SO_4 . The solvent was evaporated and the crude product was dissolved in diethylether (5 mL). The solution was washed with brine and dried over Na_2SO_4 followed by evaporation of the solvent. Purification of the crude product by flash chromatography revealed the designated product. *Analytical performance:* flash chromatography: cyclohexane/ethyl acetate 10:1; yield 30%; HPLC: Chiralcel-OD-H, n-hexane/2-propanol 90:10, 0.75 mL min^{-1} , 25°C , R_t [(*S*)-**20**] = 11.25 min, R_t [(*R*)-**20**] = 12.56 min; ^1H NMR (400 MHz, CDCl_3 , 300 K, ppm), $\delta = 2.10$ (s, 3 H, CH_3), 4.40 (bs, 1 H, OH), 5.10 (s, 1 H, CHOH), 7.30-7.41 (m, 5 H, Ar-H); ^{13}C NMR (100 MHz, CDCl_3 , 300 K, ppm), $\delta = 25.3$ (CH_3), 80.2 (CHOH), 127.4 (CH), 128.8 (CH), 129.1 (CH), 138.1 (Cq), 207.2 (C=O); GCMS $R_t = 7.83$ min; MS (70 eV, EI): m/z (%): 150 (3%) [M^+], 107 (100%), 79 (68%).

(R)-1-Hydroxy-1-phenylbutan-2-one (**21**)-

Reactions conditions: benzaldehyde (32 mg, 20 mM) and propanal (35 mg, 40 mM) were dissolved in 12 mL buffer A (pH 7.0) and 3 mL DMSO. After addition of ApPDC (72 U/mL) the reaction was stirred slowly at 30°C for 44 h before additional propanal (35 mg, 40 mM) and ApPDC (72 U/mL) were added. After 72 h product extraction was performed as described for **20**. After purification of the crude product by thin layer chromatography 84 mol% **21** and 16 mol% propioid were obtained. *Analytical performance:* thin layer chromatography: cyclohexane/ethyl acetate 5:1, yield = 45%; HPLC: Chiralcel OD-H, n-hexane/2-propanol 95:5, 0.5 mL min^{-1} , 40°C , R_t [(*R*)-**21**] = 19.69 min, R_t [(*S*)-**21**] = 16.83 min; ^1H NMR (400 MHz, CDCl_3 , 300 K, ppm), $\delta = 1.01$ (t, $J = 7.4 \text{ Hz}$, 3 H, CH_3), 2.29-2.46 (m, 2 H, CH_2), 4.30-

4.4 (bs, 1 H, OH), 5.10 (s, 1 H, CHOH), 7.30-7.41 (m, 5 H, Ar-H); ^{13}C NMR (100 MHz, CDCl_3 , 300 K, ppm), δ = 7.6 (CH_3), 31.2 (CH_2), 79.5 (CHOH), 127.3 (CH), 128.6 (CH), 128.9 (CH), 138.3 (Cq), 210.1 (C=O); GCMS R_t = 8.51 min, MS (70 eV, EI): m/z (%): 164 (0.1%) [M^+], 107 (100%), 79 (81%).

(R)-1-Cyclopropyl-2-hydroxy-2-phenylethanone (**22**)-

Reactions conditions: Reaction was performed as described for **21** but with cyclopropyl-carbaldehyde (40 mM). After purification of the crude product by thin layer chromatography 76 mol% **22** and 24 mol% 2-hydroxy-1,2-dicyclopropylethanone were obtained. *Analytical performance:* thin layer chromatography: cyclohexane/ethyl acetate 5:1, yield = 6%; HPLC: Chiralcel OD-H, n-hexane/2-propanol 95:5, 0.5 mL min $^{-1}$, 40°C: R_t [(*S*)-**22**] = 16.6 min, R_t [(*R*)-**22**] = 21.15 min; CD (acetonitrile): λ ($\Delta\epsilon$) [nm] (mol. CD) = 276 (-7.19), 219 (+8.36), 209 (+8.30); ^1H NMR (400 MHz, CDCl_3 , 300 K, ppm), δ = 0.78-0.86 (m, 1 H), 0.94-1.10 (m, 1 H), 1.16-1.20 (m, 2 H), 1.83-1.90 (m, 1 H, CH), 4.39 (bs, 1 H, OH), 5.27 (s, 1 H, CHOH), 7.30-7.45 (m, 5 H, Ar-H); ^{13}C NMR (100 MHz, CDCl_3 , 300 K, ppm), δ = 12.1 (CH_2), 12.7 (CH_2), 17.6 (CH), 80.1 (CHOH), 127.7 (CH), 128.6 (CH), 128.9 (CH), 138.1 (Cq), 209.6 (C=O); GCMS R_t = 9.45 min, MS (70 eV, EI): m/z (%): 176 [M^+], 107 (100%), 79 (71%).

2.7. Crystallisation

ApPDC was crystallised using the hanging drop vapor diffusion method. Droplets were set up for crystallisation by mixing 2 μL of purified protein solution containing 3 mg/mL protein in 20 mM Mes/NaOH-buffer, pH 6.5, 2.5 mM MgSO_4 , 0.1 mM ThDP with 2 μL of reservoir solution. Screening and optimisation revealed a reservoir solution of 16-21% (w/v) PEG 1500, 0.1 M Mes/malic acid/Tris-buffer, pH 7-7.2 and 5-10 mM dithiothreitol (DTT) to be optimal. Diffraction-quality crystals appeared after three days of equilibration.

2.8. Data collection and processing

For cryoprotection the crystals were quickly dipped into the well solution supplemented with 25% ethylene glycol before being frozen in a cryogenic nitrogen gas stream at 110 K. Data were collected to a resolution of 2.75 Å at beamline I911-2 at Max-lab, Lund, Sweden. Images were processed using MOSFLM [25], where unit-cell parameters were determined by the auto indexing option. The data set was scaled with the program SCALA implemented in the CCP4 program suite [26].

The crystal belongs to the space group C2 with eight monomers packed in the asymmetric unit. Data collection statistics are given in Tab. 1.

2.9. Structure solution and crystallographic refinement

The structure of *Ap*PDC was determined by molecular replacement using the program MOLREP [26, 27]. As the sequence similarity between *Ap*PDC and *Zm*PDC is 62% (see supplementary material), the structure of *Zm*PDC (pdb accession code: 1zpd) [28] was used as the search model to place the eight monomers in the asymmetric unit. Atomic positions and B-factors of the model were refined by the maximum likelihood method in REFMAC5 [26, 29] which was interspersed with rounds of manual model building in COOT [30]. Non-crystallographic symmetry restraints were applied using tight main chain and side chain restraints. Water assignment was performed in COOT. The quality of the final structure was examined using PROCHECK [26, 31]. Statistics of the refinement and final model are given in Tab. 2. The coordinates of *Ap*PDC, in addition to the structure factors, have been deposited at the Research Collaboratory for Structural Bioinformatics Protein Databank PDB with accession code **2vbi**.

2.10. Structural analysis and models

Structural comparison of *Ap*PDC with *Zm*PDC (Fig. 5) was done with COOT [32]. Comparison of *Ap*PDC with other ThDP-dependent enzymes (Fig. 7+8) was performed using the PyMOL program for visualisation [32] and the SwissPDB Viewer for superimposition [33].

3. Results and Discussion

3.1. Cloning, overexpression, purification and storage

The coding gene (*pdc*, gene bank AAM21208) was cloned and expressed in *E. coli* as a C-terminal hexahistidine fusion protein. Comparison of the amino acid sequences of *ApPDC* with the one published by Raj et al. [4] revealed only 98% identity. Beside 9 amino acids exchanges in different positions of the protein, the *ApPDC* described in this work contains 558 amino acids and is thus one amino acid longer than the published sequence [4] (for sequence alignment see supplementary material). A further C-terminal elongation by at least eight amino acids is due to the hexahistidine tag. Although the strains used in this work (DSM-No. 2347, DSMZ) and by Raj et al. [4] (ATCC 12874, American Type Culture Collection) were described to be the same the differences might be explained by varying sources of the *Acetobacter pasteurianus* strain sold by the companies. Because of the type of mutations it is unlikely that the differences appeared due to mistakes in the polymerase chain reaction. The sequence was confirmed twice in different clones.

A 15 L fed-batch fermentation of the recombinant *E. coli* strain yielded 1.42 kg cells with a cell specific activity of 750 U/g_{cells} (decarboxylase activity) in the crude extract and an overall activity of 1.07×10^6 U. The enzyme was purified by immobilised metal chelate chromatography yielding 48 g lyophilisate comprising, beside freeze-dried salts from the buffer solution, 14 g pure *ApPDC* (protein purity according to SDS-gel > 95%). The specific activity measured directly after desalting was 110 U/mg. For the detection of the best storage conditions the specific activity of *ApPDC* was followed for 4 weeks. Lyophilised *ApPDC* as well as frozen fractions, both stored at -20 °C, kept an activity of 100 U/mg. However, storage under sterile conditions at 4 °C in 10 mM potassium phosphate buffer, pH 6.5, containing 2.5 mM MgSO₄ and 0.1 mM ThDP resulted in a drastically reduced catalytic activity (46 U/mg after 4 weeks).

3.2. Determination of the native molecular weight

The native molecular weight of the recombinant *ApPDC* was determined by size-exclusion chromatography. The main peak (85.4%) eluted at an apparent molecular weight of 226 ± 1 kDa. As the calculated size of the monomeric subunit is 59.9 kDa, the observed data correlate best with a tetrameric structure of *ApPDC* in the native state under the applied conditions. This is in accordance to most other tetrameric 2-keto acid decarboxylases, such as *ZmPDC* [28], *ScPDC* [34, 35] and *BFD* [36], except for *KdcA* [37] showing a dimeric

quaternary structure. A second small peak (14.6%) was detected corresponding to a molecular weight of 476 kDa, correlating well with an octamer and showing a specific activity comparable to the tetramer. Similar findings have also been described by Raj et al. [38]. In plant PDCs oligomers up to 300-500 kDa are rather common [39]. Small amounts of octamers were also found in *Zp*PDC (data not shown) and *Zm*PDC [40].

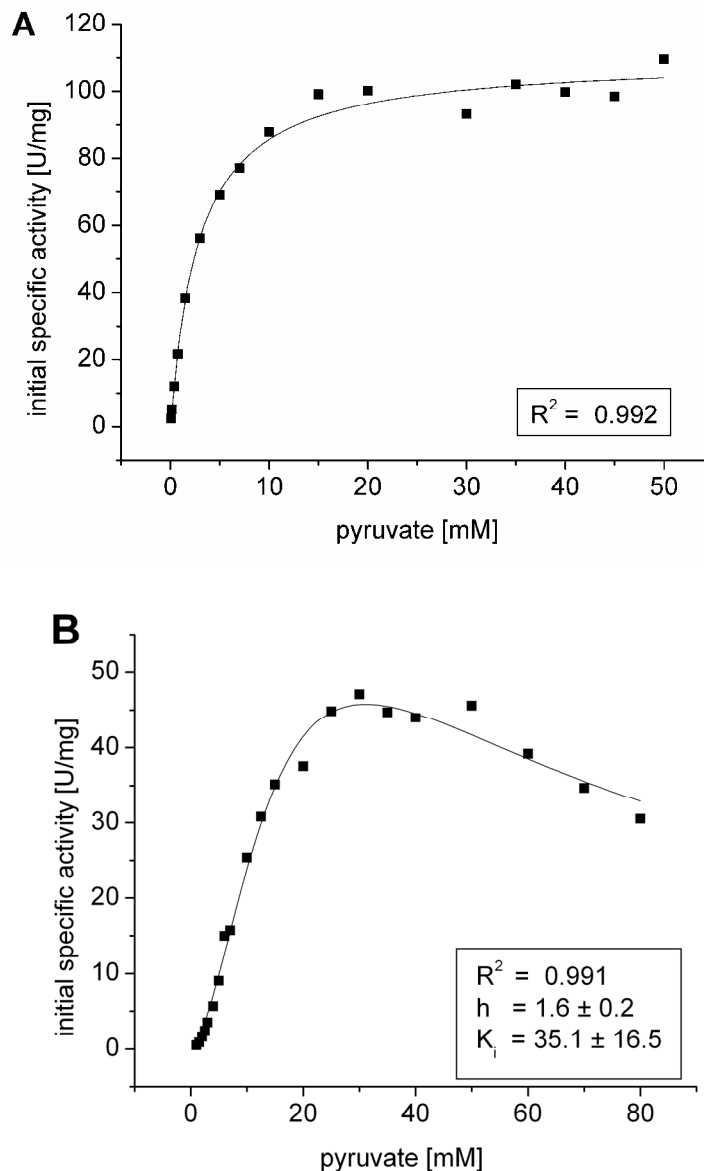


Figure 1: Kinetics of *Ap*PDC (A) and *Sc*PDC (B) for the decarboxylation of pyruvate. h = Hill coefficient (if > 1 , cooperativity confirmed), K_s = inhibition constant (B). Data were measured in 50 mM potassium phosphate buffer, pH 6.5, 2.5 mM $MgSO_4$, 0.1 mM ThDP.

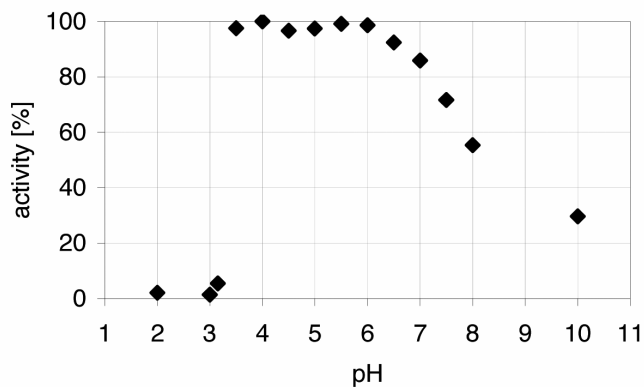


Figure 2: Determination of the pH-optimum of *ApPDC* for the decarboxylation of pyruvate. Data were obtained using the direct decarboxylase assay under initial rate conditions.

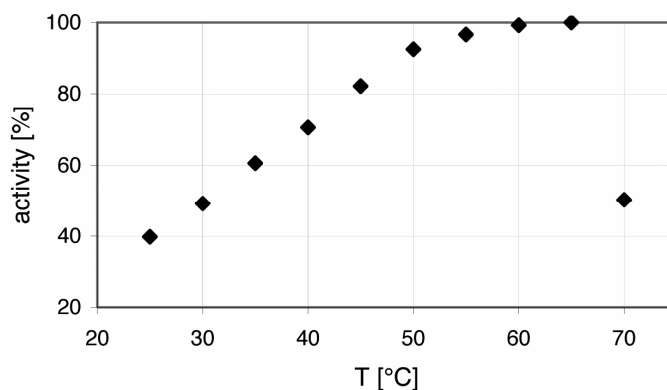


Figure 3: Determination of the temperature optimum of *ApPDC* for the decarboxylation of pyruvate. Data were obtained using the direct decarboxylase assay under initial rate conditions.

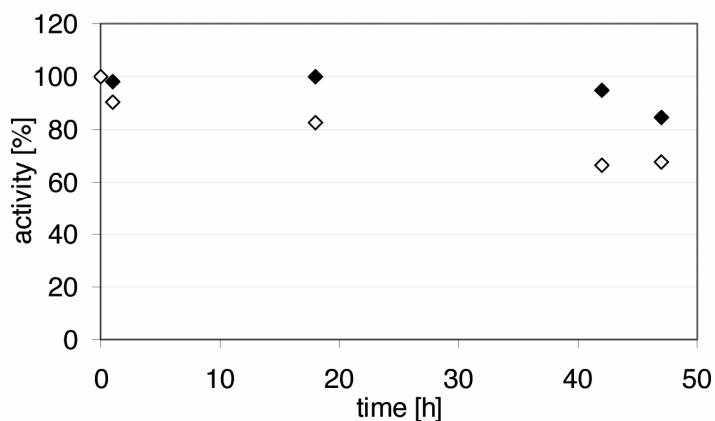


Figure 4: Stability of *ApPDC* in the presence and absence of DMSO. Incubation in the presence \blacklozenge and absence \diamond of 30% (v/v) DMSO in 50 mM potassium phosphate puffer, pH 6.5 (pH was adjusted before DMSO addition), 2.5 mM MgSO_4 , 0.1 mM ThDP at 30 °C. Residual activities were measured using the coupled decarboxylase assay.

3.3. Substrate range and kinetics of decarboxylase activity

The investigation of the substrate range of the decarboxylase reaction is of interest in order to deduce information about the acyl donor spectrum for carboligase activity of *Ap*PDC. If a respective 2-keto acid is a substrate for decarboxylation, the binding of the corresponding aldehyde to the C2-atom of ThDP located in the active centre is most likely, meaning that this aldehyde may be a possible donor aldehyde in enzyme catalysed carboligation reactions.

As demonstrated in Tab. 3, *Ap*PDC prefers aliphatic non-branched 2-keto acids. The longer the side chain, the less activity is observed. Nevertheless, activity is also evident with branched-chain (**7-9**) and aromatic substrates (**11-13**), with an exceptionally high activity for benzoylformate compared to other PDCs (**11**). Thus, the substrate range of *Ap*PDC is similar to *Zp*PDC and broader compared to *Zm*PDC [24] and *Sc*PDC (Tab. 3). It should be mentioned that the substrate range was investigated with a concentration of 30 mM for each substrate. It is not very likely but still possible that the maximum rate of each reaction could not be detected correctly due to high K_M values or substrate surplus effects.

Kinetic constants were determined for the physiological substrate pyruvate (**1**). *Ap*PDC shows a hyperbolic $v/[S]$ -plot up to 75 mM with a maximal velocity (V_{max}) of 110 U/mg and a K_M of 2.7 mM (Tab. 4, Fig. 1A). These values agree well with the published data obtained in citrate/phosphate buffer ([38], Tab. 4). Whereas *Zm*PDC [41] and *Zp*PDC show similar hyperbolic $v/[S]$ -plots, substrate activation in *Sc*PDC gives rise to a sigmoidal $v/[S]$ -plots. Additionally the plot shows a decrease of activity at concentrations > 30 mM pyruvate (Fig. 1B). The latter might be a result of substrate surplus inhibition or increased ionic strength. The sigmoidal $v/[S]$ -plot of brewers yeast PDC [42] and *Sc*PDC [39] have already been described and similar results have been obtained with plant PDCs [43] as well as with the prokaryotic PDC from *Sarcina ventriculi* (*Sv*PDC) [44]. Just recently the mechanism for the allosteric activation of yeast PDC could be elucidated [45, 46].

3.4. Optimal reaction conditions

Optimal cofactor concentration- As in all ThDP-dependent enzymes the cofactors in *Ap*PDC are bound non-covalently to the active site. The main contributions to the binding arise from the coordinative interaction of the diphosphate moiety via Mg^{2+} to the protein as well as from hydrophobic and ionic interactions between protein side chains and the thiazol- and pyrimidine ring of ThDP [47]. For stability of the holoenzyme most ThDP-dependent enzymes require the addition of cofactors to the buffer. Under the applied *Ap*PDC

concentrations the addition of 2.5 mM MgSO₄ and 0.1 mM ThDP is sufficient to keep the enzyme stable and active in buffer for several days.

pH-dependent activity and stability- For *ApPDC* maximal initial rate activity for the decarboxylation of pyruvate (**1**) was observed in an exceptional broad pH-range of pH 3.5-6.5 (Fig. 2). Similar results were observed by Raj et al. in sodium citrate buffer at 25°C [38]. However, the stability of *ApPDC* decreases significantly at pH 4 (half-life time: 2.3 h), while it is completely stable in the range of pH 5-7 within 120 min (data not shown). In comparison the pH-optima of *ScPDC* (pH 5-7), *ZpPDC* (pH 7) (data not shown) and *ZmPDC* (pH 6-6.5) [48] are significantly less broad and overlap well with the stability optima.

Temperature-dependent activity and stability- Under initial rate conditions (90 sec) the temperature optimum of *ApPDC* was observed at 65°C with a specific activity of 188 U/mg. A further temperature increase resulted in a fast decay of activity with a midpoint of thermal inactivation (*T_m*) at about 70°C (Fig. 3). This optimum is considerably higher than that of *ScPDC* (43°C) and *ZpPDC* (55°C) (data not shown) and comparable to data obtained with *ZmPDC* (60°C) [49].

From these data the activation energy of 27.1 kJ/mol was calculated for the decarboxylase reaction from an $\ln V_{\max}/[1/T]$ -plot in the linear range of 20-50°C (see supplementary material), which is lower compared to other pyruvate decarboxylases. For *ZpPDC* an activation energy of 41 kJ/mol and for *ZmPDC* 43 kJ/mol were determined.

In order to identify optimal conditions for the application of *ApPDC* in enzymatic syntheses, the temperature stability of the enzyme was determined (Tab. 5). While the enzyme is sufficiently stable up to 30°C, it rapidly loses activity at higher temperatures. The same is valid for *ZpPDC*, whereas the stability optimum of *ScPDC* is below 30°C (Tab. 5). In comparison *ZmPDC* is more thermostable with a half-life time of 24 h at 50°C [49].

Stability towards dimethylsulfoxide (DMSO)- The biotransformation of aromatic aldehydes is often hampered by their low solubility in aqueous systems. Since the addition of 20% (v/v) DMSO has been successfully applied for BAL, BFD and KdcA catalysed carbonylase reactions [9, 11, 12, 14, 18, 50] addition of DMSO was tested also with *ApPDC*. As demonstrated in Fig. 4, *ApPDC* is not just stable in the presence of up to 30% (v/v) DMSO (half-life: ~430 h), the stability is even considerably enhanced compared to the buffer without DMSO addition.

3.5. Carboligase activity

The substrate range of the decarboxylase reaction suggests *ApPDC* to prefer short non-branched aliphatic donor aldehydes in carboligase reactions. We investigated the carboligase activity in mixed carboligation reactions with benzaldehyde and various aliphatic aldehydes. The results in Tab. 6 show that the main product in such biotransformations is indeed always the phenylacetylcarbinol derivative (**20-22**), which results from the reaction of an aliphatic donor aldehyde (**17-19**) and benzaldehyde (**16**) as the acceptor aldehyde. As the aliphatic aldehyde was applied in excess, its self-ligation product is obtained as a side product (~16 mol% 4-hydroxyhexan-3-one (entry **2**, Tab. 6) and 24 mol% 1,2-dicyclopropyl-2-hydroxyethanone (entry **3**, Tab. 6), respectively). However, neither benzoin nor mixed carboligation products derived from benzaldehyde as a donor aldehyde have been observed. Even with benzaldehyde as the only substrate just a very low benzoin forming activity was determined ($2.4 \cdot 10^{-3}$ U/mg), which is typical for pyruvate decarboxylases [19].

Concerning the carboligation *ApPDC* is less enantioselective (*ee* 81-91%, Tab. 6) than *ZmPDC*, which is strictly *R*-selectivity (98%) with respect to the formation of **20** [17].

The reduced stereo control of *ApPDC* suggests that its active site might be different to *ZmPDC*. In order to rationalise these differences, the enzyme was crystallised to determine the structure.

3.6 Crystal structure

Quality of the electron density map and the model- The crystal structure of *ApPDC*, determined to 2.75 Å resolution, was solved by molecular replacement. The asymmetric unit contains eight monomers, arranged as two tetrameric *ApPDC* units. The structure was refined to an R-factor of 21.9% and an R_{free} of 24.0%.

8 x 554 amino acids (comprising residue 2-555) could be placed in the electron density map, with each subunit binding one ThDP molecule and one Mg^{2+} -ion anchoring the cofactor. 91.3% of the amino acid residues in the final model are placed in the most favored regions of the Ramachandran plot. One residue, S74, has a strained conformation in the disallowed region of the plot, but is well defined in the electron density (Tab. 1).

Overall and quaternary structure- *ApPDC* shows the general fold of ThDP-dependent enzymes consisting of three domains with α/β -topology [35, 39] (Fig. 5). Each domain is comprised of a central β -sheet surrounded by α -helices. The N-terminal PYR (pyrimidine)- and C-terminal PP (pyrophosphate)-domains are topologically equivalent (residues 1-188 and 349-555) while the R (regulatory)-domain has a different fold (residues 189-348).

Two monomers form the minimal functional unit, with the coenzyme bound in the cleft between the PYR-domain of one monomer and the R-domain of the second monomer. ThDP is bound in the conserved V-conformation, forced by a large hydrophobic residue next to the thiazolium ring (Ile-411), positioning the N4'-imino group of the pyrimidine ring in a reactive distance to the C2 of the thiazolium ring where the catalytic reaction takes place. The packing of tetrameric units in the crystal supports the size exclusion chromatography results indicating that *ApPDC* is a tetramer in solution. The oligomeric form can be best described as a dimer of dimers. Superimposition of *ZmPDC* and *ApPDC* (Fig. 5) reveals the backbones of all three domains to be very similar. There are only minor differences in the loops connecting the domains (encompassing the residues 181-190 and 345-356 in *ApPDC*) and the C-terminus, which is elongated in *ZmPDC* by four extra residues inserted in a loop relative to *ApPDC*. Like *ZmPDC*, *ApPDC* shows a similar compact arrangement of the tetrameric structure.

ThDP-binding site- A superimposition of the crystal structure of *ApPDC* with the known structures of *ZmPDC* [28], *ScPDC* [34], *KdcA* [37], *BFD* [36] and *BAL* [51] combined with the biochemical data concerning substrate specificities of these enzymes and variants thereof provides insights into several aspects of the active site architecture in these enzymes. In Tab. 7 important residues lining the substrate binding site including the *S*-pocket [19], catalytically important residues as well as amino acids forming the substrate channel are given. Their relative positions are visualised in Fig. 6.

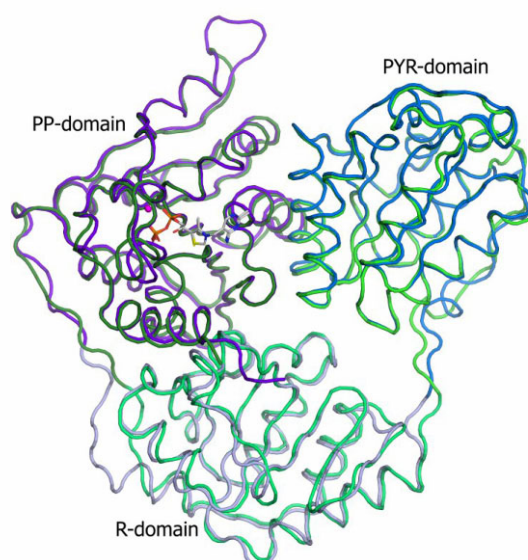


Figure 5: Ribbon representation of the *ApPDC* subunit (green) superimposed with the *ZmPDC* subunit (blue) [28]. Superposition of the two structures results in an rms deviation of 0.81 Å over 549 Ca-atoms. The three domains are coloured in different shades. ThDP is shown as sticks and the Mg²⁺ ion is shown as a sphere coloured in magenta.

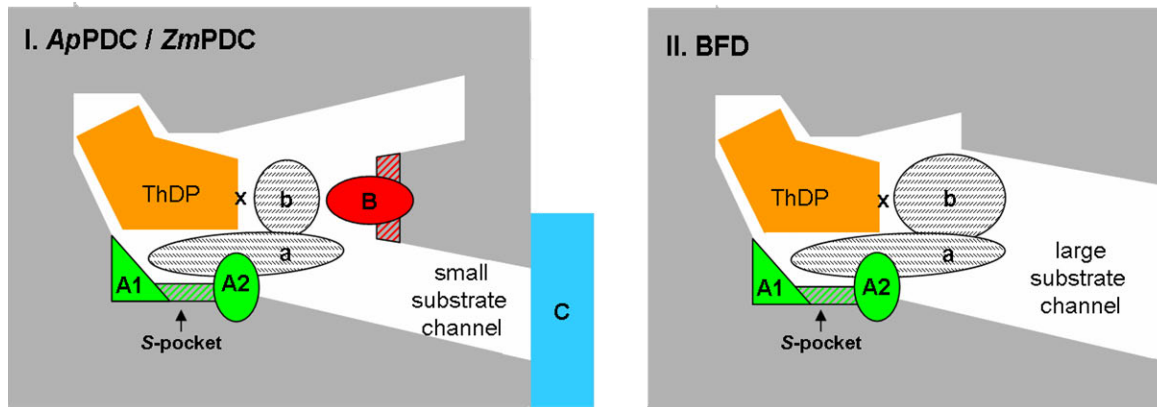


Figure 6: Schematic presentation of the substrate channel and the active site of *ApPDC/ZmPDC* (I) and *BFD* (II). ThDP is bound in the V-conformation in the active site. x = C2-atom of the thiazolium ring, A = residues lining the S-pocket, A1 = most prominent residues determining the size of the S-pocket, A2 = residues defining the entrance to the S-pocket, a = acceptor binding site; B = bulky residue in *ApPDC/ZmPDC* situated opposite to the substrate binding site where the donor binds (I), b = binding site for 2-keto acid and donor aldehyde, x = C2-atom of the thiazolium ring; C = C-terminal α -helix covering the entrance of the substrate channel in *ApPDC/ZmPDC* (I).

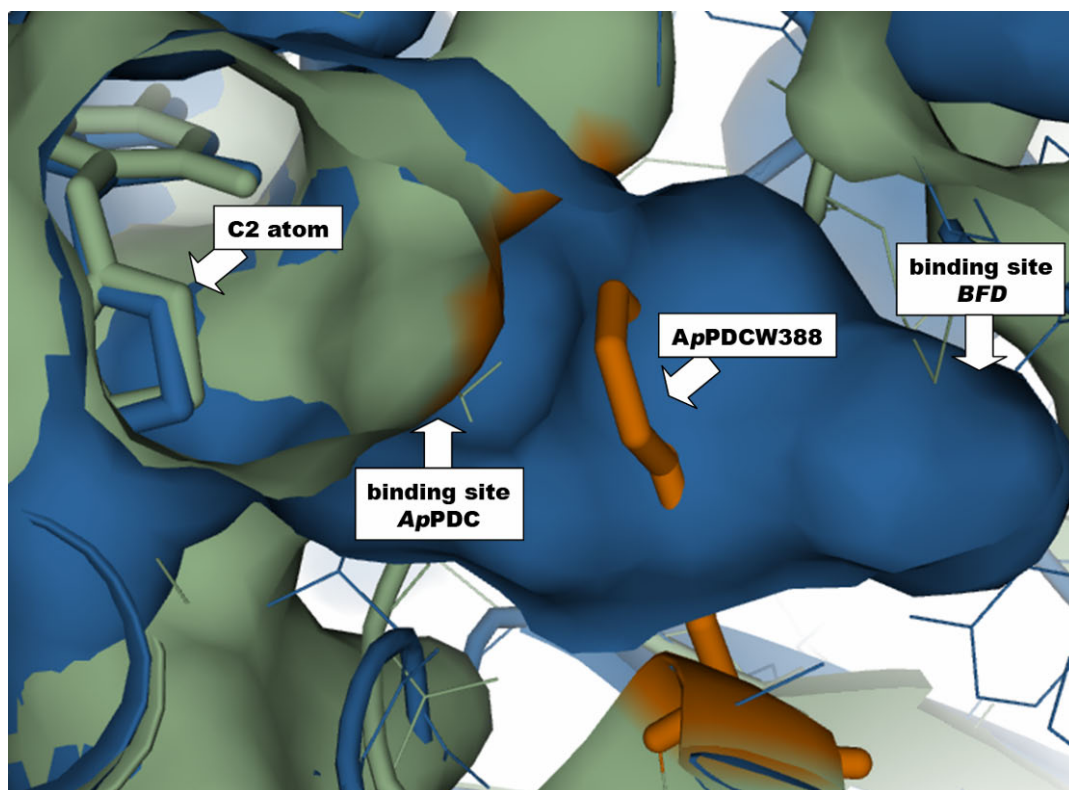


Figure 7: Comparison of the substrate binding sites of *ApPDC* and *BFD*. Superimposition of the crystal structures of the binding site of *ApPDC* (green) and *BFD* (blue). In *ApPDC* residue W388 (marked in orange) limits the pockets size.

Substrate channel and entrance- In contrast to BFD, which has no C-terminal helix [20], the entrance to the substrate channel in *ApPDC* is not directly visible on the surface of the 3D structure due to a C-terminal helix covering this region (area C-Fig. 6, I). A similar 4-turn helix is found in *ZmPDC* while *ScPDC* has a helix with 2-3 turns and BAL a shorter one with 1.5 turns (Tab. 7). It is assumed that the movement of the C-terminal helix is important for the catalytic activity [28]. This is supported by results published by Chang et al. [52] who investigated variants of *ZmPDC* with a C-terminal truncation. Whereas the deletion of the seven very last C-terminal amino acids has no effect on the decarboxylase activity, the next few, R561 and S560, are critical not only for the decarboxylase activity but also for the cofactor binding.

A comparison of the substrate channels leading to the active centres shows a very wide and straight substrate channel for BFD (Fig. 6, II), whereas the channels in *ApPDC* and *ZmPDC* are less wide and curved (Fig. 6, I). The dimensions of the channels in BAL, KdcA and *ScPDC* range between those of BFD and *ApPDC/ZmPDC*.

Donor binding site- The reaction mechanism of ThDP-dependent decarboxylation has been intensively studied for PDCs [35, 53, 54] and BFD [55, 56]. The substrate is attacked by the activated ylid of ThDP followed by the release of CO₂ and the formation of the covalent α -carbanion/enamine intermediate, the so-called “active aldehyde”. In the decarboxylation reaction an aldehyde is released after protonation of the α -carbon, while during carboligation it reacts with a second aldehyde as an acyl acceptor [57]. Whereas the decarboxylation requires the binding of only one substrate (a 2-keto acid), the arrangement of two substrates (aldehydes) in close proximity in the active site is necessary for carboligation. The binding site and orientation of the second (acceptor) aldehyde is defined by the space which is left after the first (donor) aldehyde is bound to the C2 atom of the ThDP in V-conformation, and the necessity to arrange the carbonyl groups of the reacting aldehydes close to a histidine residue operating as a proton relay system during the catalysis [16] (Fig. 6). As a consequence overlapping but clearly distinct binding regions can be assumed for the keto-acid or respectively the donor and the acceptor aldehyde (Fig. 6).

Differences in the substrate range of the ThDP-dependent lyases are rationalised by different sizes of the respective binding sites. The preference of PDCs for short aliphatic 2-keto acids as well as short aliphatic donor aldehydes is mainly a result of the amino acids in position B (Fig. 6), which is occupied by a bulky tryptophane in bacterial PDCs (Tab. 7). Compared to *ZmPDC* the binding site for 2-keto acids and donor aldehydes is slightly larger in *ApPDC*. This might explain why *ApPDC* is able to decarboxylate benzoylformate at least with low

activity in contrast to *ZmPDC*, not showing any product formation (Tab. 3). Modeling studies demonstrate that the space in the substrate binding site, which is mainly restricted by W388, is just large enough for a ThDP-bound benzaldehyde (Fig. 7). Nevertheless, aliphatic aldehydes are preferred as donors in mixed carboligations with aliphatic and aromatic aldehydes.

In BFD as well as in BAL and KdcA this part of the active site is wide enough to bind large aromatic substrates. However, the size of the binding site is not the only criterion to explain substrate transformation. Optimal stabilisation of a molecule in a binding site is also important to allow the rapid formation of an enzyme-substrate complex. This may explain why BFD shows only very little activity towards aliphatic 2-keto acids and aliphatic donor aldehydes although there is enough space available in the substrate binding site.

The aspect of optimal substrate stabilisation in the binding site is extremely important in mixed carboligations of two different aldehydes. The product range accessible by such biotransformations can yield a maximum of four different products each as pair of enantiomers [57]. Fortunately, the various constrains in the different enzymes result in an almost selective formation of one predominant ligation product, if the two aldehydes to be ligated are sufficiently different. Thus, in the case of an aliphatic and an aromatic aldehyde, PDCs prefer the binding of the aliphatic aldehyde as the donor, whereas BFD prefers the aromatic aldehyde. There are two possibilities for the acceptor aldehyde to approach the ThDP-bound donor: in a parallel mode, where both side chains of the donor- and the acceptor aldehyde point into the substrate channel or in an antiparallel mode, with the acceptor side chain pointing into the so-called *S*-pocket [20]. This step is decisive for the stereoselectivity of the ligation step and the chemoselectivity of the reaction, since both aldehydes present in the reaction mixture can also react as acceptors. Again selectivity is decided by steric requirements and optimal stabilisation as was recently described for BFD and BAL [19,20]. Currently BFD is the only known enzyme, where such an *S*-pocket is accessible in the wild-type enzyme making (*S*)-2-hydroxy propiophenones available.

S-pocket- This structural feature of the substrate binding site enables BFD to catalyse *S*-selective carboligation (Fig. 6). Recently we described similar but due to sterically reasons not accessible pockets in the 3D structures of other ThDP-dependent decarboxylases, such as KdcA, *ScPDC* and *ZmPDC* [19]. In *ApPDC* this pocket is by far the largest pocket observed so far (Fig. 8). However, the *S*-pockets in PDCs are blocked by bulky isoleucine residues (Tab. 7), which is the main reason for the usually high *R*-selectivity of the PDCs [19]. For *ZmPDC* it was demonstrated that chemo- and stereoselectivity of the carboligase reaction can

be influenced by point mutation of this position (*ZmPDCI472A*) [17]. Since *ApPDC* (91% **20**) is less enantioselective compared to *ZmPDC* (98% **20**) [17] concerning the carboligation of acetaldehyde and benzaldehyde, it can be assumed that the large *S*-pocket in *ApPDC* is not completely blocked by isoleucine 468, which would explain the relative large amount of (*S*)-products formed, lowering the enantiomeric excess of the predominant (*R*)-product (Tab. 6). This hypothesis is supported by modelling studies which show that the *S*-pocket in *ApPDC* is large enough to accept the bulky benzaldehyde substrate, if flexible side chains that adopt to the bound substrate (especially in the entrance region to the large *S*-pocket) are assumed.

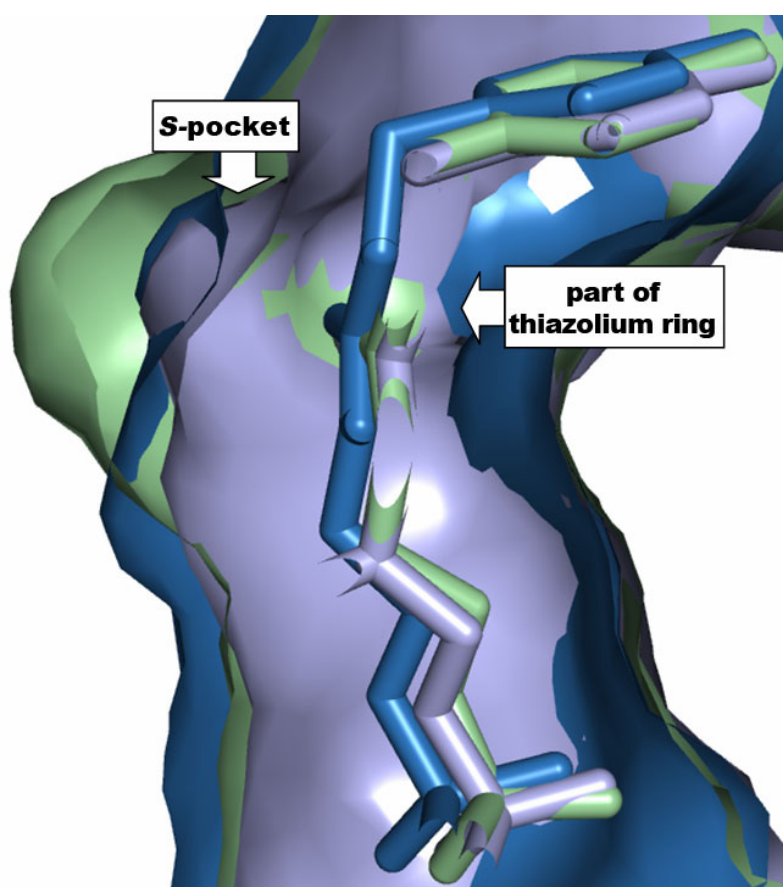


Figure 8: Superimposition of the active site of *ApPDC* (green), *ZmPDC* (light blue) and *ScPDC* (dark blue). The pocket is mainly defined by the side chain of a conserved glutamate residue (E469 in *ApPDC*, E473 in *ZmPDC* and E477 in *ScPDC*). Differences in the size of the *S*-pocket result from differences in the orientation of this glutamate residue and are not due to alternate backbone conformations.

4. Conclusions

ApPDC resembles very much the well characterised enzyme from *Zymomonas mobilis* with respect to sequence, kinetic behaviour and stability. However, there are some remarkable differences concerning the catalytic activity: the decarboxylation of aromatic 2-ketoacids, especially benzoylformate and a significantly lower stereoselectivity during carboligation. Both observations can be explained based on the 3D structure, showing a larger active site compared to *ZmPDC*. *ApPDC* prefers aliphatic donor aldehydes, as these fit best into the binding site. But the donor binding site is even large enough to accept benzaldehyde, which is not possible with *ZmPDC*. There are also differences in the acceptor binding site. In *ZmPDC* the acceptor binding site (Fig. 6) is wide enough to bind aromatic aldehydes, such as benzaldehyde, in one preferred orientation: with the aromatic side chain being directed into the substrate channel. Thus *ZmPDC* catalyses the synthesis of enantiomerically pure (*R*)-phenylacetylcarbinol from acetaldehyde as the donor and benzaldehyde as the acceptor. In *ApPDC* the situation is different. The *S*-pocket, as part of the acceptor binding site, is also accessible. As the *S*-pocket in *ApPDC* is large enough to bind even aromatic side chains, the stereocontrol of carboligase reactions with benzaldehyde is less strict than for *ZmPDC*. The amount of (*S*)-product formed is directly correlated with the size of the donor aldehyde. Thus if acetaldehyde is replaced by propanal or cyclopropylcarbaldehyde the amount of *S*-product increases resulting in a decrease of the enantioselectivity of the (*R*)-product (Tab. 6). Here the larger donor aldehydes occupy more space in the substrate binding site, thereby making the antiparallel arrangement of benzaldehyde in the *S*-pocket more likely. Together with our previous studies on BFD and BAL [19, 20], these structure-function studies pave the way for the design of *S*-specific ThDP-dependent enzymes.

Acknowledgements

The skilful technical assistance of Katharina Range is gratefully acknowledged. The authors thank Degussa AG and the Swedish Research Council for financial support. The *ScPDC* gene was kindly provided by PD Dr. Stephan König from the Martin-Luther University Halle-Wittenberg

References

- [1] M. Pohl, Protein design on pyruvate decarboxylase (PDC) by site-directed mutagenesis, *Adv. Biochem. Eng./Biotechnol.* 58 (1997) 15-43.
- [2] E.A. Dawes, D.W. Ribbons, P.J. Large, The Route of Ethanol Formation in *Zymomonas mobilis*, *Biochem. J.* 98 (1966) 795-803.
- [3] T. Okamoto, K. Taguchi, H. Nakamura, H. Ikenaga, H. Kuraishi, K. Yamasato, *Zymobacter palmae* gen. nov., sp. nov., a new ethanol-fermentating peritrichous bacterium isolated from palm sap, *Arch. Microbiol.* 160 (1993) 333-337.
- [4] K.C. Raj, L.O. Ingram, J.A. Maupin-Furlow, Pyruvate decarboxylase: a key enzyme for the oxidative metabolism of lactic acid by *Acetobacter pasteurianus*, *Arch. Microbiol.* 176 (2001) 443-451.
- [5] S.E. Lowe, J.G. Zeikus, Purification and characterization of pyruvate decarboxylase from *Sarcina ventriculi*, *J. Gen. Microbiol.* 138 (1992) 803-807.
- [6] R. Kalscheuer, T. Stolting, A. Steinbüchel, Microdiesel: *Escherichia coli* engineered for fuel production, *Microbiology* 152 (2006) 2529-2536.
- [7] C. Neuberg, J. Hirsch, Über ein Kohlenstoffketten knüpfendes Ferment (Carboligase), *Biochem. Zeitschr.* 115 (1921) 282.
- [8] V.B. Shukla, P.R. Kulkarni, L-Phenylacetylcarbinol (L-PAC): biosynthesis and industrial applications, *World J. Microbiol. Biotechnol.* 16 (2000) 499-506.
- [9] A.S. Demir, O. Şeşenoglu, P. Dünkermann, M. Müller, Benzaldehyde lyase-catalyzed enantioselective carbonylation of aromatic aldehydes with mono- and dimethoxy acetaldehyde, *Org. Lett.* 5 (2003) 2047-2050.
- [10] A.S. Demir, O. Şeşenoglu, E. Eren, B. Hosrik, M. Pohl, E. Janzen, D. Kolter, R. Feldmann, P. Dünkermann, M. Müller, Enantioselective synthesis of alpha-hydroxy ketones via benzaldehyde lyase-catalyzed C-C bond formation reaction, *Adv. Synth. Catal.* 344 (2002) 96-103.
- [11] P. Domínguez de María, M. Pohl, D. Gocke, H. Gröger, H. Trauthwein, L. Walter, M. Müller, Asymmetric synthesis of aliphatic 2-hydroxy ketones by enzymatic carbonylation of aldehydes, *Eur. J. Org. Chem.* (2007) 2940-2944.
- [12] P. Domínguez de María, T. Stillger, M. Pohl, S. Wallert, K. Drauz, H. Gröger, H. Trauthwein, A. Liese, Preparative enantioselective synthesis of benzoin and (*R*)-2-hydroxy-1-phenylpropane using benzaldehyde lyase, *J. Mol. Catal. B- Enzym.* 38 (2006) 43-47.
- [13] P. Dünkermann, M. Pohl, M. Müller, Enantiomerically pure 2-hydroxy carbonyl compounds through enzymatic C-C bond formation, *Chim. Oggi/Chemistry Today supplement Chiral Catalysis* 22 (2004) 24-28.
- [14] D. Gocke, C.L. Nguyen, M. Pohl, T. Stillger, L. Walter, M. Müller, Branched-Chain Keto Acid Decarboxylase from *Lactococcus lactis* (KdcA), a Valuable Thiamine Diphosphate-Dependent Enzyme for Asymmetric C-C Bond Formation, *Adv. Synth. Catal.* 349 (2007) 1425-1435.
- [15] M. Pohl, B. Lingen, M. Müller, Thiamin-diphosphate-dependent enzymes: new aspects of asymmetric C-C bond formation, *Chem. Eur. J.* 8 (2002) 5288-5295.
- [16] M. Pohl, G.A. Sprenger, M. Müller, A new perspective on thiamine catalysis, *Curr. Opin. Biotech.* 15 (2004) 335-342.
- [17] P. Siegert, M.J. McLeish, M. Baumann, H. Iding, M.M. Kneen, G.L. Kenyon, M. Pohl, Exchanging the substrate specificities of pyruvate decarboxylase from *Zymomonas mobilis* and benzoylformate decarboxylase from *Pseudomonas putida*, *Protein Eng. Des. Sel.* 18 (2005) 345-357.

- [18] T. Stillger, M. Pohl, C. Wandrey, A. Liese, Reaction engineering of benzaldehyde lyase catalyzing enantioselective C-C bond formation, *Org. Process Res. Dev.* 10 (2006) 1172-1177.
- [19] D. Gocke, L. Walter, E. Gauchenova, G. Kolter, M. Knoll, C.L. Berthold, G. Schneider, J. Pleiss, M. Müller, M. Pohl, Rational Protein Design of ThDP-dependent Enzymes: Engineering Stereoselectivity, *ChemBioChem* (2008) in press.
- [20] M. Knoll, M. Müller, J. Pleiss, M. Pohl, Factors Mediating Activity, Selectivity, and Substrate Specificity for the Thiamin Diphosphate-Dependent Enzymes Benzaldehyde Lyase and Benzoylformate Decarboxylase, *ChemBioChem* 7 (2006) 1928-1934.
- [21] D.J. Korz, U. Rinas, K. Hellmuth, E.A. Sanders, W.-D. Deckwer, Simple fed-batch technique for high cell density cultivation of *Escherichia coli*, *J. Biotechnol.* 39 (1995) 59-65.
- [22] M. Pohl, P. Siegert, K. Mesch, H. Bruhn, J. Grötzinger, Active site mutants of pyruvate decarboxylase from *Zymomonas mobilis* - a site-directed mutagenesis study of L112, I472, I476, E473, and N482, *Eur. J. Biochem.* 257 (1998) 538-546.
- [23] M.M. Bradford, A rapid and sensitive method for the quantification of microgram quantities of protein using the principle of protein-binding dye, *Anal. Biochem.* 72 (1976) 248-254.
- [24] P. Siegert, Vergleichende Charakterisierung der Decarboxylase- und Carboligasereaktion der Benzoylformiatdecarboxylase aus *Pseudomonas putida* und der Pyruvatdecarboxylase aus *Zymomonas mobilis* mittels gerichteter Mutagenese, doctoral thesis, Heinrich-Heine University Duesseldorf (2000).
- [25] A.G.W. Leslie, Recent changes to the MOSFLM package for processing film and image plate data, *Joint CCP4 + ESF-EAMCB Newsl. Protein Crystallogr.* 26 (1992).
- [26] Collaborative Computational Project Number 4, The CCP4 suite: programs for protein crystallography, *Acta Crystallogr. Sec. D: Biol. Crystallogr.* 50 (1994) 760-763.
- [27] A. Vagin, A. Teplyakov, MOLREP: an automated program for molecular replacement, *J. Appl. Crystallogr.* 30 (1997) 1022-1025.
- [28] D. Dobritzsch, S. König, G. Schneider, G. Lu, High resolution crystal structure of pyruvate decarboxylase from *Zymomonas mobilis*. Implications for substrate activation in pyruvate decarboxylases, *J. Biol. Chem.* 273 (1998) 20196-20204.
- [29] G.N. Murshudov, A. Vagin, E.J. Dodson, Refinement of macromolecular structures by the maximum-likelihood method, *Acta Crystallogr. Sect. D: Biol. Crystallogr.* 53 (1997) 240-255.
- [30] P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics, *Acta Crystallogr. Sec. D: Biol. Crystallogr.* 60 (2004) 2126-2132.
- [31] R.A. Laskowsky, M.W. McArthur, D.S. Moss, J.M. Thornton, PROCHECK, *J. Appl. Crystallogr.* 26 (1993) 282-291.
- [32] W.L. DeLano, The PyMOL Molecular Graphics System, DeLano Scientific, San Carlo, CA, USA, 2002.
- [33] N. Guex, M.C. Peitsch, SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling, *Electrophoresis* 18 (1997) 2714-2723.
- [34] P. Arjunan, T. Umland, F. Dyda, S. Swaminathan, W. Furey, M. Sax, B. Farrenkopf, Y. Gao, D. Zhang, F. Jordan, Crystal structure of the thiamin diphosphate-dependent enzyme pyruvate decarboxylase from the yeast *Saccharomyces cerevisiae* at 2.3 Å resolution, *J. Mol. Biol.* 256 (1996) 590-600.
- [35] F. Jordan, M. Liu, E. Sergienko, Z. Zhang, A. Brunskill, P. Arjunan, W. Furey, Yeast pyruvate decarboxylase: New features of the structure and mechanism, in: Jordan, F. and Patel, R.N. (Eds.), *Thiamine-Catalytic Mechanisms in Normal and Disease States*, Marcel Dekker Inc., New York / Basel, 2004.

- [36] M.S. Hasson, A. Muscate, M.J. McLeish, L.S. Polovnikova, J.A. Gerlt, G.L. Kenyon, G.A. Petsko, D. Ringe, The crystal structure of benzoylformate decarboxylase at 1.6 Å resolution: diversity of catalytic residues in thiamin diphosphate-dependent enzymes, *Biochemistry* 37 (1998) 9918-9930.
- [37] C.L. Berthold, D. Gocke, M.D. Wood, F.J. Leeper, M. Pohl, G. Schneider, Crystal structure of the branched-chain keto acid decarboxylase (KdcA) from *Lactococcus lactis* provides insights into the structural basis for the chemo- and enantioselective carboligation reaction, *Acta Crystallogr. Sec. D: Biol. Crystallogr.* 63 (2007) 1217-1224.
- [38] K.C. Raj, L.A. Talarico, L.O. Ingram, J.A. Maupin-Furlow, Cloning and characterization of the *Zymobacter palmae* pyruvate decarboxylase gene (pdc) and comparison to bacterial homologues, *Appl. Environ. Microbiol.* 68 (2002) 2869-2876.
- [39] S. König, Subunit structure, function and organisation of pyruvate decarboxylases from various organisms, *Biochim. Biophys. Acta* 1385 (1998) 271-286.
- [40] M. Pohl, J. Grötzinger, A. Wollmer, M.R. Kula, Reversible dissociation and unfolding of pyruvate decarboxylase from *Zymomonas mobilis*, *Eur. J. Biochem.* 224 (1994) 651-661.
- [41] S. Bringer-Meyer, K.L. Schimz, H. Sahm, Pyruvate decarboxylase from *Zymomonas mobilis*. Isolation and partial characterization, *Arch. Microbiol.* 146 (1986) 105-110.
- [42] G. Hübner, R. Weidhase, A. Schellenberger, The mechanism of substrate activation of pyruvate decarboxylase. A first approach, *Eur. J. Biochem.* 92 (1978) 175-181.
- [43] A. Dietrich, S. König, Substrate activation behaviour of pyruvate decarboxylase from *Pisum sativum* cv. Miko, *FEBS Lett.* 400 (1997) 42-44.
- [44] L.A. Talarico, L.O. Ingram, J.A. Maupin-Furlow, Production of the Gram-positive *Sarcina ventriculi* pyruvate decarboxylase in *Escherichia coli*, *Microbiology* 147 (2001) 2425-2435.
- [45] G. Lu, D. Dobritzsch, S. Baumann, G. Schneider, S. König, The structural basis of substrate activation in yeast pyruvate decarboxylase. A crystallographic and kinetic study, *Eur. J. Biochem.* 267 (2000) 861-868.
- [46] G. Lu, D. Dobritzsch, S. König, G. Schneider, Novel tetramer assembly of pyruvate decarboxylase from brewer's yeast observed in a new crystal form, *FEBS Lett.* 403 (1997) 249-253.
- [47] R.A.W. Frank, F.J. Leeper, B.F. Luisi, Structure, mechanism and catalytic duality of thiamine-dependent enzymes, *Cell. Mol. Life S.* 64 (2007) 829-905.
- [48] A.D. Neale, R.K. Scopes, R.E. Wettenhall, N.J. Hoogenraad, Pyruvate decarboxylase of *Zymomonas mobilis*: isolation, properties, and genetic expression in *Escherichia coli*, *J. Bacteriol.* 169 (1987) 1024-1028.
- [49] M. Pohl, K. Mesch, A. Rodenbrock, M.R. Kula, Stability investigations on the pyruvate decarboxylase from *Zymomonas mobilis*, *Biotechnol. Appl. Biochem.* 22 (1995) 95-105.
- [50] F. Hildebrand, S. Kühl, M. Pohl, D. Vasic-Racki, M. Müller, C. Wandrey, S. Lütz, The production of (*R*)-2-hydroxy-1-phenyl-propan-1-one derivatives by benzaldehyde lyase from *Pseudomonas fluorescens* in a continuously operated membrane reactor, *Biotechnol. Bioeng.* 96 (2007) 835-843.
- [51] T.G. Mosbacher, M. Müller, G.E. Schulz, Structure and mechanism of the ThDP-dependent benzaldehyde lyase from *Pseudomonas fluorescens*, *FEBS J.* 272 (2005) 6067-6076.
- [52] A.K. Chang, P.F. Nixon, R.G. Duggleby, Effects of deletions at the carboxyl terminus of *Zymomonas mobilis* pyruvate decarboxylase on the kinetic properties and substrate specificity, *Biochemistry* 39 (2000) 9430-9437.

- [53] D. Kern, G. Kern, H. Neef, K. Tittmann, M. Killenberg-Jabs, C. Wikner, G. Schneider, G. Hübner, How thiamine diphosphate is activated in enzymes, *Science* 275 (1997) 67-70.
- [54] A. Schellenberger, Sixty years of thiamin diphosphate biochemistry, *Biochim. Biophys. Acta* 1385 (1998) 177-186.
- [55] E.S. Polovnikova, M.J. McLeish, E.A. Sergienko, J.T. Burgner, N.L. Anderson, A.K. Bera, F. Jordan, G.L. Kenyon, M.S. Hasson, Structural and kinetic analysis of catalysis by a thiamin diphosphate-dependent enzyme, benzoylformate decarboxylase, *Biochemistry* 42 (2003) 1820-1830.
- [56] P.M. Weiss, G.A. Garcia, G.L. Kenyon, W.W. Cleland, P.F. Cook, Kinetics and mechanism of benzoylformate decarboxylase using ^{13}C and solvent deuterium isotope effects on benzoylformate and benzoylformate analogues, *Biochemistry* 27 (1988) 2197-2205.
- [57] H. Iding, P. Siegert, K. Mesch, M. Pohl, Application of alpha-keto acid decarboxylases in biotransformations, *Biochim. Biophys. Acta* 1385 (1998) 307-322.

Tables

Tab. 1. Data collection statistics of *ApPDC*. Values in parentheses are given for the highest resolution interval.

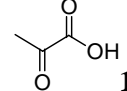
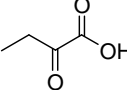
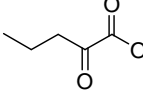
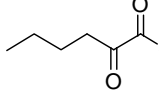
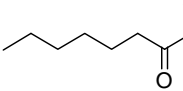
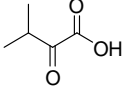
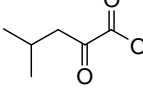
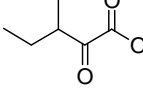
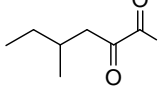
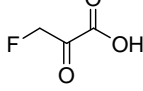
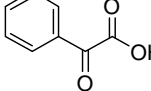
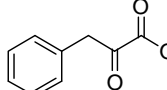
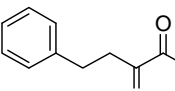
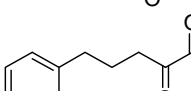
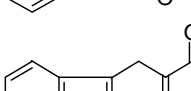
Space group	C2
Molecules in asymmetric unit	8
Unit cell dimensions	$a = 179.62 \text{ \AA}$, $b = 162.43 \text{ \AA}$, $c = 169.33 \text{ \AA}$
No. of observations	289630 (42113)
No. of unique reflections	119238 (17599)
Resolution	2.75 (2.9-2.75) \AA
R-merge	0.083 (0.433)
Mn ($I/\sigma(I)$)	10.8 (2.0)
Completeness	97.2 (98.5)%
Multiplicity	2.4 (2.4)%
Wilson B-factor	69.7 \AA^2

Tab. 2. Structure refinement and final model statistics. TLS model = twin lattice symmetry model; rms deviation = root means square deviation

Refinement program	REFMAC5
TLS model	8 TLS groups
Reflections in working set	107178
Reflections in test set	5995
R-factor	21.9%
R-free	24.0%
Atoms modeled	33860
Number of amino acids	8 x 554
Number of ThDPs	8
Number of magnesium ions	8
Number of waters	92
Average B-factor protein*	42.4 Å ²
Average B-factor ligands	38.3 Å ²
Average B-factor magnesium ions	34.7 Å ²
Average B-factor waters	29.0 Å ²
RMS deviations from ideals	
Bonds	0.009 Å
Angles	1.1044 °
Ramachandran distribution	
Most favored	91.3%
Additionally allowed	8.5%
Generously allowed	0.0%
Disallowed	0.2%

* Contains residual B-factors only, does not include TLS contribution

Tab. 3. Substrate range of the decarboxylation reaction of *Ap*PDC compared to *Zm*PDC [24], *Sc*PDC and *Zp*PDC. Activity towards different 2-keto acids was measured using the coupled decarboxylase assay. All substrates were applied in a concentration of 30 mM, except **15** (1 mM).

2-keto acid	specific activity [U/mg]			
	<i>Ap</i> PDC	<i>Zm</i> PDC	<i>Sc</i> PDC	<i>Zp</i> PDC
 1	89.3	120	43.4	147.0
 2	60	79	16.9	85.1
 3	12.9	13	18.8	20.7
 4	4.2	0.2	5.3	6.2
 5	1.1	-	0.0	0.6
 6	-	0.0	6.9	10.4
 7	1.1	0.3	0.3	2.8
 8	2.2	0.0	0.0	2.8
 9	0.6	0.0	0.0	0.4
 10	0.3	-	0.0	0.6
 11	1.1	0.0	0.0	0.3
 12	0.8	0.0	0.0	1.8
 13	0.3	0.0	1.7	0.3
 14	0.0	0.0	0.2	0.2
 15	0.0	-	0.0	0.5

- = not determined, 0.0 = activity < 0.05 U/mg


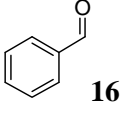
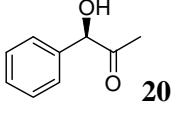
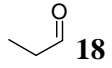
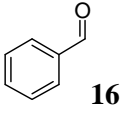
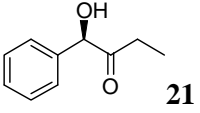
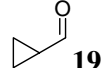
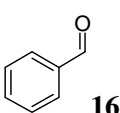
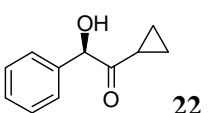
Tab. 4. Kinetic parameters for the decarboxylation of pyruvate by *ApPDC* compared to *ZmPDC*, *ScPDC* and *ZpPDC*. Data from Raj et al. [38] were measured in sodium citrate buffer at 25°C; *ZmPDC* data in Mes/KOH-buffer, pH 6.5, at 30°C [17], while data obtained in this work were measured in potassium phosphate buffer, pH 6.5, (*ZpPDC*: pH 7.0) at 30°C.

	<i>ApPDC</i>	<i>ZmPDC</i>	<i>ScPDC</i>	<i>ZpPDC</i>
V_{\max} [U/mg]	110 ± 1.9 97 (pH 5.0) [38] 79 (pH 7.0) [38]	120-150 [17]	112.0 ± 35.6	116 ± 2.0 130 (pH 6.0) [38] 140 (pH 7.0) [38]
K_M [mM]	2.8 ± 0.2 0.35 (pH 5.0) [38] 5.1 (pH 7.0) [38]	1.1 ± 0.1 [17]	$S_{0.5} = 21.6 \pm 7.4$	2.5 ± 0.2 0.24 (pH 6.0) [38] 0.71 (pH 7.0) [38]

Tab. 5. Temperature stability of *ApPDC* compared to *ScPDC* and *ZpPDC*. Residual activity was measured using the coupled decarboxylase assay with pyruvate; - = not determined.

T (°C)	half-life $t_{1/2}$	half-life $t_{1/2}$	half-life $t_{1/2}$
	[h]	[h]	[h]
	<i>ApPDC</i>	<i>ScPDC</i>	<i>ZpPDC</i>
	pH 6.5	pH 6.5	pH 7
20	193	235	-
25	-	243	-
30	144	78	150
35	-	62	-
40	34	-	40
50	12	-	10
60	2	-	0.4
70	0.4	-	-

Tab. 6. Mixed carboligations of different aliphatic aldehydes with benzaldehyde catalysed by *Ap*PDC. Reaction conditions: 50 mM potassium phosphate buffer, pH 7.0, containing 2.5 mM MgSO₄, 0.1 mM ThDP and 20% DMSO, 30 °C. Product detection was done by NMR after 3 d, the *ee* was determined by chiral phase HPLC.

no.	donor	acceptor	product	ee
1	 17	 16	 20	91% (R)
2	 18	 16	 21	82% (R)
3	 19	 16	 22	81% (R)

Tab. 7. Residues lining the substrate channel and the active site of various ThDP-dependent enzymes. Corresponding residues are found in similar structural positions, if not otherwise indicated. For the positions of the respective areas see the schematic presentation (Fig. 6). Residues in parenthesis are situated in a similar position but are only partially involved in lining the respective site.

BAL	BFD	<i>Ap</i> PDC	<i>Zm</i> PDC	<i>KdcA</i>	<i>Sc</i> PDC	remarks
pdb codes						
2ag0	1mcz, 1bfd	2vbi	1zpd	2vbf	1pyd	
Residues lining the S-pocket (area A1-Fig. 6 I+II)						
L25	(N23)	V24	V24	V23	L25	
H26	P24	A25	A25	P24	P26	
G27	G25	G26	G26	G25	G27	
A28	S26	D27	D27	D26	D28	
F484	F464	I472	I476	I465	I480	
(W487)	(Y458)	Y466	Y470	Y459	Y474	
-	L461	E469	E473	E462	E477	
Entrance to the S-pocket (area A2-Fig. 6 I + II)						
A480	A460	I468	I472	V461	I476	
Further important residues lining the donor binding site (area B-Fig. 6 I)						
L112 Q113	L109 L110	H113 H114	H113 H114	H112 H113	H114 H115	backbone-displacement, BFDL109/L110 correspond to H113/H114/Y290 in <i>Ap</i> PDC
H29 Q113	H70					proton relay system; in PDCs H113, H114, respectively H115 function as proton relay system
(L398) H415 (C414) Y397	(T380) C398 F397	W388	W392	F381	A392	residue defining the size of the binding pocket for 2-keto acid and donor aldehyde
M421	L403	I411	I415	I404	I415	stabilising the V-conformation of ThDP
C-terminal α-helix covering the entrance to the substrate channel (area C-Fig. 6 I)						
aa 550-555 α -helix, 1.5 turns	no helix	aa 538-555 α -helix, 4 turns	aa 544-566 α -helix, > 4 turns	aa 532-547 α -helix, 4 turns	aa 547-556 α -helix, 2-3 turns	

Supplementary Material**Pyruvate Decarboxylase from *Acetobacter pasteurianus*: BIOCHEMICAL AND STRUCTURAL CHARACTERISATION**

Dörte Gocke, Catrine L. Berthold, Thorsten Graf, Helen Brosi, Ilona Frindi-Wosch, Michael Knoll, Thomas Stillger, Lydia Walter, Michael Müller, Jürgen Pleiss, Gunter Schneider, Martina Pohl

1. Alignment of the *Ap*PDC protein sequence in this work and the published sequence at NCBI.

Differences are marked in yellow.

*Ap*PDC_Pohl = *Ap*PDC sequence cloned by Gocke *et al.*

*Ap*PDC_NCBI = *Ap*PDC sequence cloned by Raj *et al.* (1), accession code at NCBI:

AAM21208

→ 98 % identity (9 different amino acids, 1 gap)

```

ApPDC_Pohl      MTYTVGMYLAERLVQIGLKHFFAVAGDYDNLVLLDQQLLNKDMKQIYCCNELNCGFSAEGY 60
ApPDC_NCBI      MTYTVGMYLAERLVQIGLKHFFAVGGDYDNLVLLDQQLLNKDMKQIYCCNELNCGFSAEGY 60
*****
ApPDC_Pohl      ARSNGAAAAVVTFVSGAISAMNALGGAYAENLPVILISGAPNSNDQGTGHILHHTIGKTD 120
ApPDC_NCBI      ARSNGAAAAVVTFVSGAISAMNALGGAYAENLPVILISGAPNSNDQGTGHILHHTIGKTD 120
*****
ApPDC_Pohl      YSYQLEMARQVTCAAESITDAHSAPAKIDHVIRTALRERKPAYLDIACNIASEPCVRPGP 180
ApPDC_NCBI      YSYQLEMARQVTCAAESITDAHSAPAKIDHVIRTALRERKPAYLDIACNIASEPCVRPGP 180
*****
ApPDC_Pohl      VSSLLSEPEIDHTSLKAAVDATVALLKKSASPVMLLGSKLRAANALAAATETLADKLQCAV 240
ApPDC_NCBI      VSSLLSEPEIDHTSLKAAVDATVALLKNRPA PVMLLGSKLRAANALAAATETLADKLQCAV 240
*****
ApPDC_Pohl      TIMAAAKGFFPEDHAGFRGLYWGEVSNPGVQELVETSDALLCIAPVFN DYSTVGWSAWPK 300
ApPDC_NCBI      TIMAAAKGFFPEDHAGFRGLYWGEVSNPGVQELVETSDALLCIAPVFN DYSTVGWSGMPK 300
*****
ApPDC_Pohl      GPNVILAEPDRVTVDGRAYDGFTLRAFLQALAEKAPARPASAQKSSVPTCSLTATSDEAG 360
ApPDC_NCBI      GPNVILAEPDRVTVDGRAYDGFTLRAFLQALAEKAPARPASAQKSSVPTCSLTATSDEAG 360
*****
ApPDC_Pohl      LTNDEIVRHINALLTSNTTLVAETGDSWFNAMRMTLPRGARVELEMQWGHIGWSVPSAFG 420
ApPDC_NCBI      LTNDEIVRHINALLTSNTTLVAETGDSWFNAMRMTLA-GARVELEMQWGHIGWSVPSAFG 419
*****
ApPDC_Pohl      NAMGSQDRQHVVMVGDGSFQLTAQEVAQMVRYELPVIIIFLINNRGYVIEIAIHDGPYNYI 480
ApPDC_NCBI      NAMGSQDRQHVVMVGDGSFQLTAQEVAQMVRYELPVIIIFLINNRGYVIEIAIHDGPYNYI 479
*****
ApPDC_Pohl      KNWDYAGLMEVFNAGEGHGLGKATTPKELTEAIARAKANTRGPTLIECQIDRTDCTDML 540
ApPDC_NCBI      KNWDYAGLMEVFNAGEGHGLGKATTPKELTEAIARAKANTRGPTLIECQIDRTDCTDML 539
*****
ApPDC_Pohl      VQWGRKVASTNARKTTLA 558
ApPDC_NCBI      VQWGRKVASTNARKTTLA 557
*****

```

2. Alignment of ApPDC and pyruvate decarboxylase from *Zymomonas mobilis* (ZmPDC)

ZmPDC_NCBI: ZmPDC sequence cloned by Pohl *et al.* (2)

ApPDC_Pohl: ApPDC sequence cloned by Gocke *et al.*

→ ZmPDC used for molecular replacement, homology 62 %

```

ZmPDC_NCBI      MSYTVGTYLAERLVQIGLKHHFAVAGDYNLVLDDNLLLNKNMEQVYCCNELNCGFSAEGY 60
ApPDC_Pohl      MTYTVGMYLAERLVQIGLKHHFAVAGDYNLVLDDQLLLNKMKQIYCCNELNCGFSAEGY 60
*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*

ZmPDC_NCBI      ARAKGAAAAVVTYSVVGALSFAFDAIGGAYAENLPVILISGAPNNNDHAAGHVLHHALGKTD 120
ApPDC_Pohl      ARSNGAAAAVVTFSVGAISAMNALGGAYAENLPVILISGAPNSNDQGTGHILHHTIGKTD 120
*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*

ZmPDC_NCBI      YHQYLEMAKNITAAAEAIYTPPEAPAKIDHVIKTALREKKPVYLEIACNIASMPCAAPGP 180
ApPDC_Pohl      YSYQLEMARQVTCAAESITDAHSAPAKIDHVIRTALRERKPAYLDIACNIASEPCVRPGP 180
* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

ZmPDC_NCBI      ASALFNDEASDEASLNAAVEETLKFIANRDKVAVLVGSKLRAAGAEAAVKFADALGGAV 240
ApPDC_Pohl      VSSLSEPEIDHTSLKAAVDATVALLEKSASPVMLLGSKLRAANALAAATETLADKLQCAV 240
.*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*

ZmPDC_NCBI      ATMAAAKSFFPEENPHYIGTSWGEVSYPGVEKTMKEADAVIALAPVFNDDYSTTGWTDIPD 300
ApPDC_Pohl      TIMAAAKGFFPEDHAGFRGLYWGEVSNPGVQELVETSDALLCIAPVFNDDYSTVGWSAWPK 300
: * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

ZmPDC_NCBI      PKKLVLAEPERSVVVNGIRFSPVHLKDYLTRLAQKVSCKKTGALDFFKSLNAGELKKAAPAD 360
ApPDC_Pohl      GPNVILAEPDRVTVDGRAYDGFTRLAFLQALAEKAPARP-ASAQKSSVPTCSLTATS--- 356
:*:*:*:* *.*:* * : .. * : * * * : * .. : . * . * : : * . : :

ZmPDC_NCBI      PSAPLVNAE IARQVEALLTPNTTVIAETGDSWFNAQRMKLPNGARVEYEMQWGHIGWSVP 420
ApPDC_Pohl      DEAGLTNDEIVRHINALLTSNTTLVAETGDSWFNAMRMTLPRGARVELEMQWGHIGWSVP 416
. * * . * * . * : : * * * * * * * * * * * * * * * * * * * * * * * *

ZmPDC_NCBI      AAFGYAVGAPERRNILMVGDSGFSQLTAQEVAQMVRLLKLPVIFLINNYGYTIEVMIHDGP 480
ApPDC_Pohl      SAFGNAMGSQDRQHVVVMVGDSGFSQLTAQEVAQMVRVYELPVIIFLINNRGYVIEIAIHDGP 476
: * * * * * * * * : * : : : * * * * * * * * * * * * * * * * * * * * *

ZmPDC_NCBI      YNNIKNWDYAGLMEVFNNGGYDSGAGKGLKAKTGGELAEAIKVALANTDGP TLIECFIG 540
ApPDC_Pohl      YNYIKNWDYAGLMEVFNAG---EGHGLGLKATTPKELTEAIARAKANTRGPTLIECQID 532
** * * * * * * * * * * . * * * * * * * * * * * * * * * * * * * * *

ZmPDC_NCBI      REDCTEELVKWGRVAAANSRKPVNK 566
ApPDC_Pohl      RTDCTDMLVQWGRKVASTNARKTTLA 558
* * * * * * * * * * * * * * * * * * * *
    
```

3. Alignment of our *ZpPDC* protein sequence and the published sequence at NCBI

ZpPDC_Pohl = *ZpPDC* sequence cloned by Gocke *et al.*

ZpPDC_NCBI = *ZpPDC* sequence cloned by Raj *et al.* (3)

→ 99 % identity (2 different amino acids, 1 insertion)

```

ZpPDC_Pohl      MGYTVGMYLAERLAQIGLKHHFAVAGDYNLVLLDQQLLNKDMEQVYCCNELNCGFSAEGY 60
ZpPDC_NCBI     -MYTVGMYLAERLAQIGLKHHFAVAGDYNLVLLDQQLLNKDMEQVYCCNELNCGFSAEGY 59
                *****

ZpPDC_Pohl      ARARGAAAAIVTFSVGAISAMNAIGGAYAENLPVILISGSPNTNDYGTGHILHHTIGTTD 120
ZpPDC_NCBI     ARARGAAAAIVTFSVGAISAMNAIGGAYAENLPVILISGSPNTNDYGTGHILHHTIGTTD 119
                *****

ZpPDC_Pohl      YNYQLEMVKHVTCAAESIVSAEEAPAKIDHVIRTALRERKPAYLEIACNVAGAECVRPGP 180
ZpPDC_NCBI     YNYQLEMVKHVTCAARESIVSAEEAPAKIDHVIRTALRERKPAYLEIACNVAGAECVRPGP 179
                *****

ZpPDC_Pohl      INSLLEREVDQTSVTAAVDAAVEWLQDRQNVVMLVGSKLRAAAAQKQAVLADRLGCAV 240
ZpPDC_NCBI     INSLLEREVDQTSVTAAVDAAVEWLQDRQNVVMLVGSKLRAAAAQKQAVLADRLGCAV 239
                *****

ZpPDC_Pohl      TIMAAAKGFFPEDHDPNFRGLYWGEVSSEGAQELVENADAILCLAPVFNDYATVGWNSWPK 300
ZpPDC_NCBI     TIMAAEKGFFPEDHDPNFRGLYWGEVSSEGAQELVENADAILCLAPVFNDYATVGWNSWPK 299
                *****

ZpPDC_Pohl      GDNVMVMDTDRVTFAGQSFEGLSLSTFAAALAEKAPSRPATQGTQAPVLGIEAAEPNAP 360
ZpPDC_NCBI     GDNVMVMDTDRVTFAGQSFEGLSLSTFAAALAEKAPSRPATQGTQAPVLGIEAAEPNAP 359
                *****

ZpPDC_Pohl      LTNDEMTRIQSLITSDTTLTAETGDSWFNASRMPIPGGARVELEMQWGHIGWSVPSAFG 420
ZpPDC_NCBI     LTNDEMTRIQSLITSDTTLTAETGDSWFNASRMPIPGGARVELEMQWGHIGWSVPSAFG 419
                *****

ZpPDC_Pohl      NAVGSPERRHIMMVGDSFQLTAQEVAQMIRYEIPVIIIFLINNRGYVIEIAIHDGPYNYI 480
ZpPDC_NCBI     NAVGSPERRHIMMVGDSFQLTAQEVAQMIRYEIPVIIIFLINNRGYVIEIAIHDGPYNYI 479
                *****

ZpPDC_Pohl      KNWNYAGLIDVFNDEDGHGLGLKASTGAELEGAIKKALDNRRGPTLIECNIAQDDCTETL 540
ZpPDC_NCBI     KNWNYAGLIDVFNDEDGHGLGLKASTGAELEGAIKKALDNRRGPTLIECNIAQDDCTETL 539
                *****

ZpPDC_Pohl      IAWGKRVAATNSRKPQALEHHHHHHH 565
ZpPDC_NCBI     IAWGKRVAATNSRKPQA----- 566
                *****
    
```

4. Calibration G200 size exclusion chromatography column

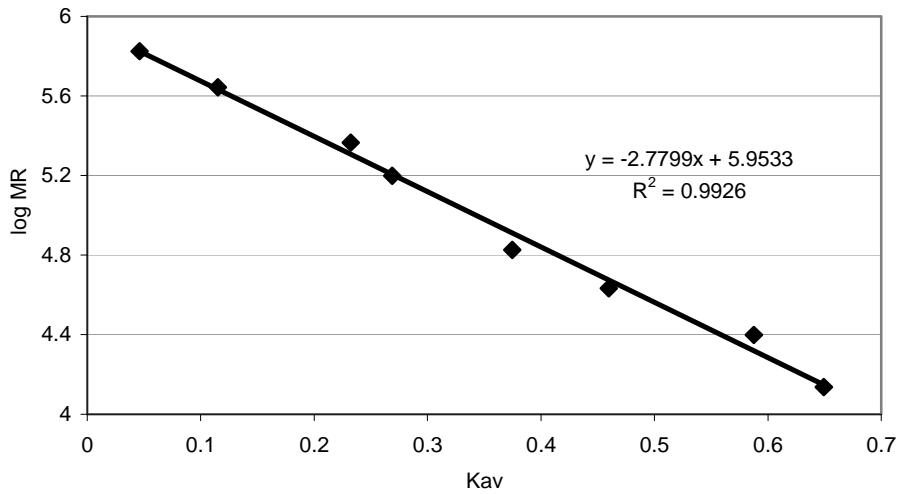


FIGURE 1. Calibration G200 size exclusion chromatography column.

Buffer: 50 mM potassium phosphate buffer, 0.1 mM ThDP, 2.5 mM MgSO₄, 150 mM KCl, pH 6.5.

For ApPDC log M_R = 5.35.

5. Hill-equation for determination of kinetic constants of ScPDC

$$V = \frac{V_{\max} \cdot [S]^h}{S_{0.5}^h + S^h \left(1 + \frac{S}{K_i}\right)}$$

6. Activation energy for the ApPDC catalysed decarboxylation of pyruvate

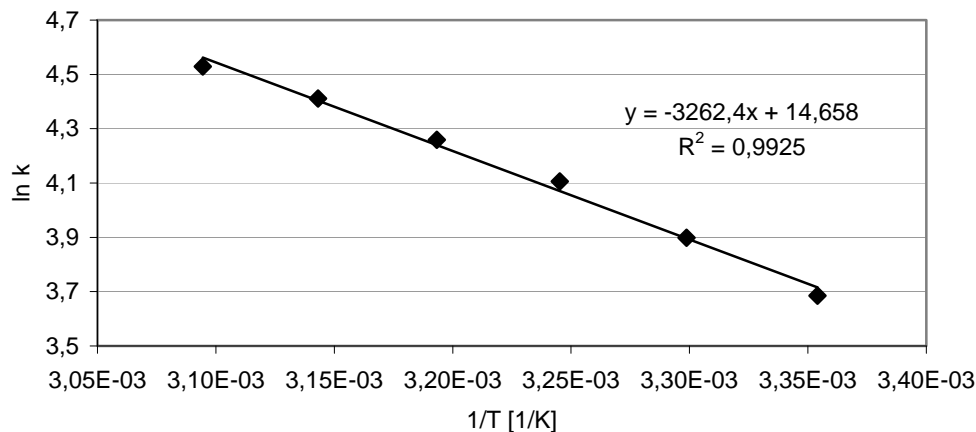


FIGURE 2. lnV_{max}/[1/T]-plot to determine the activation energy of the decarboxylase reaction. Values used in the linear area of 20-50°C only.

REFERENCES

1. Raj, K. C., Ingram, L. O., Maupin-Furlow, J. A. (2001) *Arch. Microbiol.* **176**, 443-451
2. Pohl, M., Mesch, K., Rodenbrock, A., Kula, M.R. (1995) *Biotechnol. Appl. Biochem.* **22**, 95-105
3. Raj, K. C., Talarico, L. A., Ingram, L. O., Maupin-Furlow, J. A. (2002) *Appl. Environ. Microbiol.* **68**, 2869-2876

4.4

Publikation erschienen in *Protein Science* **17**: 1689-1697

Knoll, M., Pleiss, J., 2008. The Medium-Chain Dehydrogenase/Reductase Engineering Database: A systematic analysis of a diverse protein family to understand sequence-structure-function relationship.

Copyright © The Protein Society. Published by Cold Spring Harbor Laboratory Press, 2008. Reprinted with kind permission

The Medium-Chain Dehydrogenase/Reductase Engineering Database: A systematic analysis of a diverse protein family to understand sequence-structure-function relationship

MICHAEL KNOLL AND JÜRGEN PLEISS

Institute of Technical Biochemistry, University of Stuttgart, D-70569 Stuttgart, Germany

(RECEIVED March 18, 2008; FINAL REVISION June 30, 2008; ACCEPTED July 1, 2008)

Abstract

The Medium-Chain Dehydrogenase/Reductase Engineering Database (MDRED, <http://www.mdred.uni-stuttgart.de>) has been established to serve as an analysis tool for a systematic investigation of sequence-structure-function relationships. It includes sequence and structure information of 2684 and 42 medium-chain dehydrogenases/reductases (MDRs), respectively. Although MDRs are very diverse in sequence, they have a conserved tertiary structure. MDRs are assigned to 199 homologous families and 29 superfamilies. For each family, annotated multiple sequence alignments are provided, and functionally relevant residues are annotated. Twenty-five superfamilies were classified as zinc-containing MDRs, four as non-zinc-containing MDRs. For the zinc-containing MDRs, three subclasses were identified by systematic analysis of a variable loop region, the quaternary structure determining loop (QSDL): the class of short, medium, and long QSDL, which include 11, 3, and 5 superfamilies, respectively. The length of the QSDL is predictive for tetramer (short QSDL) and dimer (long QSDL) formation. The class of medium QSDL includes both tetrameric and dimeric MDRs. The shape of the substrate-binding site is highly conserved in all zinc-containing MDRs with the exception of two variable regions, the substrate recognition sites (SRS): two residues located on the QSDL (SRS1) and, for the class of long QSDL, one residue located in the catalytic domain (SRS2). The MDRED is the first online-accessible resource of MDRs that integrates information on sequence, structure, and function. Annotation of functionally relevant residues assist the understanding of sequence-structure-function relationships. Thus, the MDRED serves as a valuable tool to identify potential hotspots for engineering properties such as substrate specificity.

Keywords: sequence-structure-function relationship; protein family database; protein family classification; structure analysis

Reprint requests to: Jürgen Pleiss, Institute of Technical Biochemistry, University of Stuttgart, Allmandring 31, D-70569 Stuttgart, Germany; e-mail: Juergen.Pleiss@itb.uni-stuttgart.de; fax: 49-711-68563196.

Abbreviations: ACR, acyl-CoA reductase; ADH, alcohol dehydrogenase; ApADH, alcohol dehydrogenase from *Aeropyrum pernix*; CAD, cinnamyl alcohol dehydrogenase; HLADH, horse liver alcohol dehydrogenase; HMM, Hidden Markov model; LTD, leukotriene B4 dehydrogenase; MI, macrofamily I; MII, macrofamily II; MDR,

medium-chain dehydrogenase/reductase; MDRED, Medium-Chain Dehydrogenase/Reductase Engineering Database; MRF, mitochondrial response protein; PDH, polyoldehydrogenase; QOR, quinone oxidoreductase; QSDL, quaternary structure determining loop; SRS, substrate recognition site; YADH, yeast alcohol dehydrogenase/tetrameric alcohol dehydrogenase.

Article and publication are at <http://www.proteinscience.org/cgi/doi/10.1110/ps.035428.108>.

The protein family of medium-chain dehydrogenases/reductases (MDRs) comprises a large enzyme family with a broad range of enzymatic activities. They are found in all kingdoms of life and are involved in metabolism, regulatory processes, and protection against cell damage (Jörnvall et al. 1999; Nordling et al. 2002). Despite their low sequence similarity, they have a similar size of 350 to 400 residues and a conserved overall structure formed by two domains, a cofactor binding domain and a catalytic domain (Jörnvall et al. 1978). While all MDRs use NAD(H) or NADP(H) as cofactor, they can be divided into two classes with a different reaction mechanism: zinc-containing and non-zinc-containing MDRs (Nordling et al. 2002). In addition, many MDRs bind a second, non-catalytic, structural zinc ion (Eklund et al. 1976; Chase Jr. 1999).

Most MDRs are active as dimers or tetramers. Previously, a loop segment in the catalytic domain subsequent to the structural zinc binding site has been suggested to mediate quaternary structure formation (Jörnvall 1977; Persson et al. 1994; Norin et al. 1997). In addition, this loop segment is part of the substrate-binding site (Shafqat et al. 1999), which consists of conserved and variable regions. The conserved regions are formed by the cofactor NAD(H) or NADP(H), which is located at the bottom of the substrate-binding site, and the catalytic zinc ion, which is located at the back wall of the substrate-binding site. The left wall is formed by the highly conserved cofactor binding residues. The substrate enters from the front through the substrate access channel. Thus, there are only two variable regions, the ceiling and the right wall of the binding site. Previously, it has been shown that mutating residues located in these two regions changed substrate specificity (Hurley and Bosron 1992; Xie and Hurley 1999; Ziegelmann-Fjeld et al. 2007).

Two classification schemes exist for MDRs. One classification (Nordling et al. 2002; Jörnvall et al. 2003) is based on a consensus evolutionary tree constructed from an alignment of about 100 MDR sequences of six genomes. The MDRs were assigned to eight functional families which belong to two classes: The zinc-containing MDRs include cinnamyl alcohol dehydrogenases (CADs), polyoldehydrogenases (PDHs), dimeric alcohol dehydrogenases (ADHs), and yeast alcohol dehydrogenases/tetrameric alcohol dehydrogenases (YADHs), while the non-zinc-containing MDRs include quinone oxidoreductases (QORs), mitochondrial response proteins (MRF), leukotriene B4 dehydrogenases (LTDs), and acyl-CoA reductases (ACRs) (Nordling et al. 2002; Jörnvall et al. 2003). A second classification (Riveros-Rosas et al. 2003) extended this concept based on a much larger data set of 583 MDRs. They were grouped into three macrofamilies: Macrofamilies I and II include the zinc-containing MDRs, macrofamily III the non-zinc-containing MDRs. Each macrofamily was divided into further subfamilies, which were not consistent with the eight functional families introduced by Nordling et al. (2002).

In the meantime, the number of MDR sequences in public databases has increased more than fivefold, and structures of 42 proteins became available. We took advantage of this wealth of data and established the Medium-Chain Dehydrogenase/Reductase Engineering Database (MDRED), applying the extensible database system DWARF (Fischer et al. 2006). The MDRED provides a predictive classification scheme based on sequence and structure. By a systematic analysis of sequence and structure, conserved motifs and functionally relevant residues were identified.

Results

Database and data content

For a systematic analysis of sequence, structure, and function of the huge and diverse protein family of medium-chain dehydrogenases/reductases, the Medium-Chain Dehydrogenase/Reductase Engineering Database (MDRED) has been established. The MDRED contains 6420 sequence entries for 2684 proteins and 257 structure entries for 42 proteins. The MDRs were assigned to 29 superfamilies based on sequence similarity (Table 1). The superfamilies were further divided into 199 homologous families based on multiple sequence alignments of the superfamilies. For 24 homologous families (13 superfamilies), at least one family member with experimentally determined structure is available. A systematic nomenclature *mdrx.y* was introduced, where *x* describes the superfamily number and *y* the homologous family number. For each family, a multiple sequence alignment was performed and a phylogenetic tree was calculated. Functionally relevant residues were annotated. Annotation information was extracted from GenBank and transferred to all family members with a conserved residue at the respective position. All alignments and phylogenetic trees are accessible at <http://www.mdred.uni-stuttgart.de>. The MDRED can be browsed on the level of family classification, structure, and organisms. Sequence and protein entries can be searched by providing their GI number (general identifier) from NCBI. Protein name, source organism, and links to the respective GenBank entry are provided for each protein. The MDRED supports classification of new sequences by providing a BLAST interface and by pre-calculated HMM profiles for each superfamily and homologous family. The complete data are available via a tar archive.

The largest superfamilies are *mdr2*, *mdr3*, and *mdr4* with 335, 292, and 218 protein entries, respectively, which account for 32% of all protein entries (Fig. 1). Sequence lengths vary from 272 to 437 residues, with an average length of 357. The average sequence identity within each superfamily varies from 30% (superfamily *mdr8*) to 71%

Table 1. Superfamilies of the Medium-Chain Dehydrogenase/Reductase Engineering Database and subclassification of zinc-containing MDRs according to the QSDL

MDRED superfamily	MDRED superfamily name	Functional family ^a	MDR subclass	Catalytic zinc	Number of homologous families	Proteins in superfamily	Sequences in superfamily	Proteins with structure
mdr2	YADH	YADH-MII	Short QSDL	Zinc-containing	8	335	783	7
mdr4	Sugar alcohol DH	PDH-MI			5	218	474	3
mdr6	Threonine DH	PDH-MI			18	126	368	3
mdr8	PDH- and CAD-like				13	124	306	
mdr11	PDH- and CAD-like				12	122	273	
mdr12	2,3-Butanediol DH	PDH-MI			8	127	271	
mdr18	Sugar alcohol DH-like	PDH-MI			7	36	120	
mdr22	PDH-like	PDH-MI			2	14	49	
mdr23	Secondary ADH	PDH-MI			3	15	57	3
mdr24	PDH-like	PDH-MI			1	6	34	
mdr29	Glucose DH-like	PDH-MI			1	3	5	1
mdr7	CAD-like	CAD-MII	Medium QSDL	Zinc-containing	9	149	352	4
mdr17	Glutathione-independent FDH	PDH-MI			6	52	124	2
mdr26	Putative ADH	ADH-MI			3	5	10	
mdr1	ADH-like	ADH-MI	Long QSDL	Zinc-containing	3	151	417	
mdr3	Glutathione-dependent FDH	ADH-MI			10	292	677	2
mdr9	ADH-like	ADH-MI			5	79	323	9
mdr19	Benzyl-/Aryl ADH	ADH-MI			8	56	93	1
mdr21	ADH-like	ADH-MI			5	32	72	
mdr5	Glutathione-dependent FDH	ADH-MI	Not assigned	Zinc-containing	7	146	346	
mdr14	Threonine-/Sorbitol DH	PDH-MI			11	78	207	
mdr16	CAD like	CAD-MII			9	86	167	
mdr25	5-Exo-hydroxycamphor DH	PDH-MI			4	10	22	1
mdr27	Putative ADH	ADH-MI			3	4	9	
mdr28	Putative ADH	ADH-MI			2	3	8	
mdr10	QOR-like (VAT-1 protein, ζ -crystallin, tumor protein p53 inducible)			Non-zinc-containing	16	183	347	4
mdr13	QOR-like (ζ -crystallin-like)				6	99	223	2
mdr15	QOR-like				7	86	191	
mdr20	QOR-like				3	8	8	

^a "Functional family" refers to previously introduced classifications of MDRs by functional families (PDH, CAD, YADH, ...) (Nordling et al. 2002) and macrofamily (MI, MII) (Riveros-Rosas et al. 2003).

(superfamily mdr24). The superfamilies were grouped into two classes, zinc-containing and non-zinc-containing MDRs, dependent on the presence of a strictly conserved and annotated sequence motif G-H-E of the catalytic zinc binding site. In the non-zinc-containing MDRs, the highly conserved active-site residues asparagine, aspartic acid/ glutamic acid, and threonine were annotated. MDRs of superfamily mdr28 are an exception, as they lack both motifs. They were classified as zinc-containing MDRs owing to their global sequence similarity.

Non-zinc-containing MDRs

Superfamilies mdr10, mdr13, mdr15, and mdr20 comprise family members of QOR, LTD, MRF, and ACR

(Nordling et al. 2002). These superfamilies were assigned to the class of non-zinc-containing MDRs. All non-zinc-containing MDRs are lacking the catalytic zinc binding motif. Most of the non-zinc-containing MDRs are also lacking a sequence motif capable for coordination of a structural zinc atom. However, in some family members of superfamilies mdr15 and mdr20, such a sequence motif was found.

Zinc-containing MDRs

Proteins belonging to the functional family of ADHs (Nordling et al. 2002) and homologs were found in superfamilies mdr1, mdr3, mdr5, mdr9, mdr19, and mdr21, glutathione-dependent formaldehyde dehydrogenases in

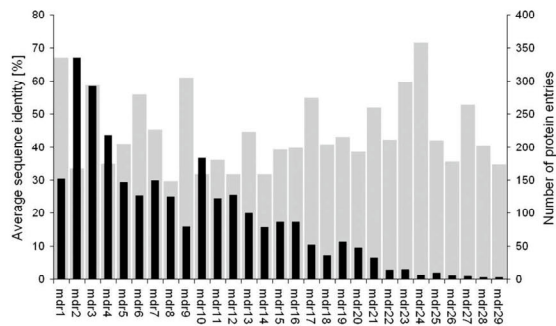


Figure 1. Average sequence identity (gray) and number of protein entries (black) per superfamily.

mdr3 and mdr5, benzyl/aryl ADHs in mdr19, and the majority of YADHs in mdr2. Members of PDHs and CADs were distributed over superfamilies mdr4, mdr6, mdr7, mdr8, mdr11, mdr12, mdr14, mdr17, mdr18, mdr22, mdr23, mdr24, mdr25, and mdr29, with the majority of sugar alcohol dehydrogenases (sorbitol, arabinitol, iditol, idionate, and xylitol dehydrogenases) found in superfamily mdr4. Secondary ADHs were found in superfamily mdr23, glucose dehydrogenases in mdr29, glutathione-independent formaldehyde dehydrogenases in mdr17, 5-exo-hydroxycamphor dehydrogenases in mdr25, and 2,3-butanediol dehydrogenases in mdr12. While threonine dehydrogenases were found in almost all superfamilies of zinc-containing MDRs, most of them as well as sorbitol dehydrogenases were found in superfamilies mdr6 and mdr14.

Subclassification of zinc-containing MDRs

Comparison of 37 superimposed structures of zinc-containing MDRs revealed a highly conserved overall tertiary structure, although the proteins are diverse in sequence. Horse liver ADH (HLADH, PDB entry 1HEU) (Meijers et al. 2001) and the ADH from *Aeropyrum permix* (ApADH, PDB entry 1H2B) (Guy et al. 2003) have only a small C_{α} root mean square deviation of 1.6 Å for 235 out of 343 residues, although these enzymes considerably differ in sequence (only 28% sequence identity). The most variable region is a loop segment located subsequent to the structural zinc binding site. This loop segment varies in sequence, length, and conformation. Because this loop segment was previously postulated to mediate quaternary structure formation of MDRs, we named it quaternary structure determining loop (QSDL). As seen from multiple sequence alignments and superimposition of all available structures of zinc-containing MDRs, the QSDL is flanked by residues that are highly conserved in sequence and structure. These residues were annotated as *QSDL-start* and *QSDL-end* (Fig. 2). In HLADH and ApADH, the residue at

QSDL-start was a cysteine (position 111 in HLADH and position 123 in ApADH). At *QSDL-end*, a serine was found in the structure of HLADH at position 144 and a glycine in ApADH at position 135. In all annotated sequences of zinc-containing MDRs, only cysteine and serine were found at position *QSDL-start* (98% and 2%, respectively). Because it represents the fourth zinc-coordinating residue of the structural zinc binding motif, it could be easily identified on sequence level. *QSDL-end* is more difficult to find. However, it is embedded in a conserved sequence motif: At position *QSDL-end*, glycine was found in 75%, and serine and threonine were found in 22% of all sequences. A high frequency of occurrence was found for phenylalanine (52%) and leucine, tyrosine, or glutamine (12% each) at position *QSDL-end+2*. At position *QSDL-end+3*, alanine and serine (50% and 30%, respectively) were found; at position *QSDL-end+4*, glutamic acid and glutamine (62% and 11%, respectively) were found. At position *QSDL-end+5*, tyrosine was found in 67% of all cases. Using this motif, for 2131 zinc-containing MDRs (92%), *QSDL-start* and *QSDL-end* could be annotated. Counting the number of amino acids of each annotated QSDL revealed two peaks in the QSDL length distribution (Fig. 3). Thus, three classes were defined. The class of short QSDL has a QSDL of less than 19 amino acids and comprises 52% of all proteins with annotated QSDL. Only 15% of proteins with annotated QSDLs have a QSDL of medium length of 19 to 31 residues. The class of long QSDL has a QSDL of more than 31 amino acids and comprises 33% of all zinc-containing MDRs with annotated QSDL. An MDRED superfamily was assigned to the class of short QSDL if >90% of its members have a QSDL of less than 19 residues in length. Thus, superfamilies mdr2, mdr4, mdr6, mdr8, mdr11, mdr12, mdr18, mdr22, mdr23, mdr24, and mdr29 were assigned to the class of short QSDL. Likewise,

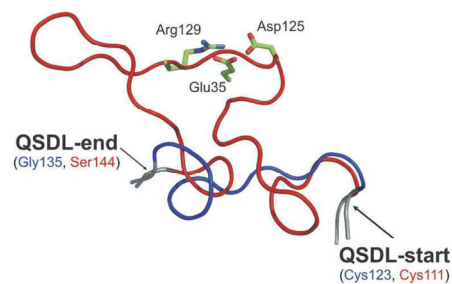


Figure 2. Structurally conserved loop segments (QSDL) of the class of long QSDL (red) and short QSDL (blue). The representative QSDL for each family is shown. (Red) Horse liver ADH (PDB entry: 1HEU); (blue) *Aeropyrum permix* ADH (PDB entry: 1H2B). *QSDL-start* and *QSDL-end* are labeled. The conserved salt bridge network in long QSDL MDRs (side chains of Asp125, Arg129, Glu35) are shown as sticks and colored by atom type.

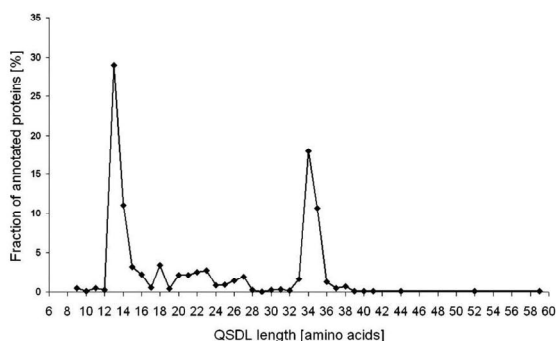


Figure 3. Fraction of proteins with annotated QSDL for each QSDL length.

superfamilies mdr7, mdr17, and mdr26 were assigned to the class of medium QSDL, because >90% of the family members have a QSDL length between 19 and 31 residues. Superfamilies mdr1, mdr3, mdr9, mdr19, and mdr21 were assigned to the class of long QSDL, because >90% of proteins possess a QSDL of more than 31 residues. All remaining superfamilies of zinc-containing MDRs that were not assigned to either of these classes were classified as “not assigned,” because their QSDLs could not be annotated for the superfamily members (mdr27, mdr28) or because family members have QSDLs of mixed lengths (mdr5, mdr14, mdr16, and mdr25).

A sequence conservation analysis of the QSDL was performed for all proteins of each class of short and long QSDL to identify highly conserved and thus potentially structurally relevant residues located in the QSDL. The QSDL sequences of the class of long QSDL are slightly more conserved than those of the class of short QSDL and contain two highly conserved residues, Asp125 and Arg129 (numbering refers to PDB entry 1HEU). A third highly conserved residue (Glu35) is located in the cata-

lytic domain, pointing toward the QSDL. Owing to their side chain distances, they form a strong, highly conserved salt bridge network, which probably stabilizes the conformation of the QSDL. Such highly conserved and potentially stabilizing residues were not found for the QSDLs of the class of short and medium QSDL.

Shape of the binding site

A comparison of 37 structures of zinc-containing MDRs revealed positions that are variable and contribute to the shape of the binding site. The residues at these positions were annotated in all superfamilies where structure information was available. For the class of long QSDL, two neighboring hydrophobic residues in the QSDL are forming the variable ceiling region above the cofactor NAD(P), which was named substrate recognition site 1 (SRS1). In the majority of proteins, a phenylalanine or tyrosine residue followed by leucine, valine, methionine, isoleucine, or phenylalanine was found. In HLADH (PDB entry 1HEU), these residues are Phe140 and Leu141. A third residue (Phe93 in PDB entry 1HEU of HLADH) forms the right wall of the binding site, which was named the substrate recognition site 2 (SRS2) (Fig. 4A). It is located two residues subsequent to a highly conserved proline (Pro91 in PDB entry 1HEU of HLADH) outside the QSDL. Mainly phenylalanine or tyrosine was found at SRS2. In the class of short QSDL, two hydrophobic residues form the binding site at SRS1 (Fig. 4B). These two residues are not adjacent as observed for the class of long QSDL, but separated by two residues. In the structure of *A*pADH, these residues are Phe128 and Leu131 (PDB entry 1H2B). No conserved residue at SRS2 was found for proteins in the class of short QSDL. However, this region contributes to substrate specificity, which has been demonstrated for yeast ADH, where exchanging tryptophan at position 93 by alanine led to an increased oxidation rate of 2-propanol (Creaser et al. 1990).

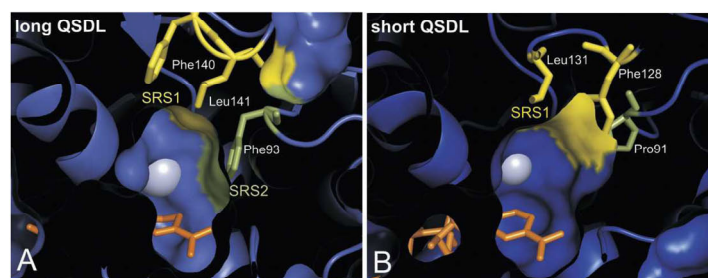


Figure 4. Top view of the MDR binding-site. Representative structures of (A) long QSDL (PDB entry 1HEU) and (B) short QSDL (PDB entry 1H2B). The variable regions SRS1 (yellow) and SRS2 (green) are marked. The (blue) zinc atom is shown as a sphere; the (orange) nicotinamide part of NAD(P) as sticks.

Generally, the more bulky the residues at positions SRS1 and SRS2 are, the smaller the available space for the substrate in the binding site is expected to be. In a few MDRs, a tryptophan residue is found at either SRS1 or SRS2, which restricts the space of the binding site considerably.

Discussion

Classification

The Medium-Chain Dehydrogenase/Reductase Engineering Database (MDRED) has been designed to serve as a navigation and analysis tool of MDRs. The MDRED includes 2684 MDRs in a consistent data structure. It is based on the extensible database system DWARF (Fischer et al. 2006), which integrates information on sequence, structure, and function. The DWARF system has already been applied successfully to build up the Lipase Engineering Database (<http://www.led.uni-stuttgart.de>) (Pleiss et al. 2000; Fischer and Pleiss 2003) and the Cytochrome P450 Engineering Database (<http://www.cyped.uni-stuttgart.de>) (Fischer et al. 2007). The MDRED is a platform to analyze sequence–structure–function relationships and to classify new sequences by providing multiple sequence alignments, phylogenetic trees, and family-specific HMM profiles. Multiple sequence alignments with annotated functionally relevant residues are provided for each superfamily and homologous family.

In this study, 2684 MDRs were assigned to 29 superfamilies using a Markov cluster algorithm based on a sequence similarity graph, as implemented in the program TRIBE-MCL (Enright et al. 2002). Owing to the large number of sequences analyzed here, we suggest a subsequent classification of superfamilies into homologous families. The assignment of proteins to homologous families is based on multiple sequence alignments of the superfamilies, whereby a homologous family is defined as a cluster of protein sequences that are more similar to each other than to other protein sequences within the same superfamily. In accordance with previous classifications based on smaller numbers of about 100 MDRs (Nordling et al. 2002; Jörnvall et al. 2003) and of about 500 MDRs (Riveros-Rosas et al. 2003), the MDRs were assigned to two groups of zinc-containing and non-zinc-containing MDRs. The group of non-zinc-containing MDRs corresponds to macrofamily III, as introduced by Riveros-Rosas et al. (2003). The group of zinc-containing MDRs includes functional families of ADHs, PDHs, CADs, and YADHs as defined by Nordling et al. (2002) and corresponds to macrofamilies I and II by Riveros-Rosas et al. (2003). Our classification into zinc-containing and non-zinc-containing MDRs and the assignment of MDRs into superfamilies are consistent with previous

classifications. The assignment of MDRs to superfamilies generally corresponds to the classification of Nordling et al. (2002), although more than one superfamily might exist for a single functional family as defined by Nordling et al. (2002) because of the large number of sequences investigated in this study (Table 1). Although the assignment of MDRs into superfamilies globally corresponds to the functional families, some exceptions were found. In two superfamilies (mdr8 and mdr11), proteins from several functional families (PDH and CAD) were found. Therefore, it seems that the specific patterns of PDH and CAD introduced previously to identify functional families (Nordling et al. 2002) are not yet specific enough to distinguish between the two functional families. Interestingly, the functional families PDH and CAD were grouped into different macrofamilies (PDH into macrofamily I and CAD into macrofamily II) by Riveros-Rosas et al. (2003). Thus, the currently defined sequence-specific patterns for PDH and CAD are not consistent with their global sequence similarity based on the large number of MDRs analyzed here.

Functional families according to Nordling et al. (2002) were further subclassified into subfamilies by Riveros-Rosas et al. (2003). This subclassification is generally in accordance with our sequence-based classification into superfamilies. However, owing to the large number of sequences analyzed here, functionally different MDRs were not necessarily found separated into different superfamilies of the MDRED: Threonine dehydrogenases (belonging to the functional family of PDHs) were found in different superfamilies of the MDRED in which also other PDHs were found. Thus, substrate specificity is not always linked to sequence similarity. Our classification by sequence similarity therefore enables the correlation of sequence and function, and helps us to understand the sequence–structure–function relationships of MDRs.

As a new classification criterion, we grouped superfamilies of zinc-containing MDRs into three structural classes based on the QSDL length (short, medium, and long QSDL). This assignment was successful for 86% of all zinc-containing MDRs. Superfamilies of the class of short QSDL comprise the functional families of YADHs, PDHs, and CADs; the class of long QSDL, the functional family of ADHs (Nordling et al. 2002). This assignment of superfamilies to classes based on structural properties further assists the understanding of sequence–structure–function relationships of MDRs.

Application

By systematically comparing sequence and structure of zinc-containing MDRs, a variable loop region was identified that varies in sequence and conformation. This loop (QSDL) mediates quaternary structure and possesses

residues relevant for the shape of the substrate-binding site. Depending on the length of QSDL, three classes (class of short, medium, and long QSDL) were introduced. Superfamilies whose members possess a QSDL of a special length are assigned to one of the three classes. A correlation of the quaternary structure of MDRs with the length of their QSDLs was observed. With only one exception, the benzyl alcohol dehydrogenase from *Acinetobacter calcoaceticus* (PDB entry: 1F8F; <http://dx.doi.org/10.2210/pdb1f8f/pdb>) (MacKintosh and Fewson 1988), all proteins of the class of long QSDL with known three-dimensional structure are active as dimers, whereas all members of the class of short QSDL are tetramers. This is in accordance with previous observations, where this loop segment was suggested to mediate quaternary structure (Jörnvall 1977; Persson et al. 1994; Norin et al. 1997). The QSDL was successfully annotated in most of the superfamilies of zinc-containing MDRs, enabling the prediction of quaternary structure of more than 2000 MDRs. This prediction is in accordance with experimental data on quaternary structure for secondary ADHs (*mdr23*), threonine and sorbitol dehydrogenases, glutathione-dependent formaldehyde dehydrogenases (*mdr3*), hydroxynitrile lyases (*mdr3*), and the YADH functional families (Riveros-Rosas et al. 2003). However, there are a small number of exceptions: The (*R,R*)-butanediol dehydrogenase from *Saccharomyces cerevisiae* (*mdr12.1*) has been reported to be active as the dimer (Gonzalez et al. 2000), although another homologous (*R,R*)-butanediol dehydrogenase from *Saccharomyces cerevisiae* is known to be tetrameric (Heidlas and Tressl 1990). Furthermore, proteins of the family of galactitol 1-phosphate dehydrogenases (*mdr18*) have been classified in the class of short QSDL according to their QSDL length, although forming dimers (Riveros-Rosas et al. 2003).

MDRs of the class of medium QSDL whose structures are known show both, dimeric and tetrameric quaternary structures, such as the alcohol dehydrogenase from *Escherichia coli* (PDB entry 1UUF; <http://dx.doi.org/10.2210/pdb1uuf/pdb>), which has a QSDL length of 21 residues and is active as the dimer, while the formaldehyde dismutase from *Pseudomonas putida* (PDB entry 2DPH; <http://dx.doi.org/10.2210/pdb2dph/pdb>) has a QSDL length of 24 residues and is active as the tetramer. Even an active trimeric structure was found within the class of medium QSDL, the mycothiol-dependent formaldehyde dehydrogenase (Norin et al. 1997) with a QSDL of 26 residues in length.

The QSDL is not only involved in quaternary structure formation, but is also part of the substrate-binding site (Shafqat et al. 1999). Residues located within the QSDL contribute to the shape of the binding site at the variable ceiling region above the NAD(P) cofactor (SRS1). Since these residues are located at different positions within the

classes of short QSDL and long QSDL, they are generally difficult to identify. By grouping zinc-containing MDRs into three classes, these residues are predictable for most MDRs within the classes of short and long QSDL. For most of the proteins of the class of short and long QSDL, annotation of these relevant residues was possible, which accounts for about 1500 proteins. The majorities of residues found at these positions are hydrophobic and point toward the substrate. Thus, the bulkiness of their side chain is expected to mediate substrate specificity. This was demonstrated for members of the class of short QSDL. Exchanging Trp110 by alanine significantly broadened substrate specificity toward phenyl-substituted alcohols and ketones of the alcohol dehydrogenase of *Thermoanaerobacter ethanolicus* (Ziegelmann-Fjeld et al. 2007). For members of the class of long QSDL, an increased affinity for the 4-methylpyrazole inhibitor was attributed to the exchange of Met141 by leucine in human $\sigma\sigma$ alcohol dehydrogenase (Xie and Hurley 1999). Additionally, exchange of hydrophobic residues at the second substrate recognition site (SRS2), which contribute to the shape of the right wall of the binding site, have led to changes in substrate specificity. A significant increase of activity toward bulky secondary alcohols of a double mutant, Phe93Ala/Thr94Ile, of human liver alcohol dehydrogenase as compared to the wild type was explained by removal of steric hindrance mainly caused by Phe93 (Hurley and Bosron 1992). Thus, residues located at positions SRS1 and SRS2 determine the available space within the binding site, and annotation of these residues enables a reliable prediction for a large number of MDRs.

For the first time, a classification of MDRs based on structural differences (length of QSDL) is proposed that is predictive for the quaternary structure of MDRs. In addition, the QSDL contributes amino acids that are relevant for the shape of the binding site (SRS1). These functionally relevant residues are annotated within the database. The MDRED provides a comprehensive resource of information on the MDR family in a consistent format, and thus is a valuable tool for a deeper understanding of biochemical properties of MDRs. By a systematic classification and annotation of MDRs, it facilitates extracting rules for substrate specificity and identifying promising mutation sites in order to engineer proteins with improved properties.

Materials and Methods

Database setup

The MDRED was established by applying the data warehouse system DWARF (Fischer et al. 2006). The DWARF system integrates data on sequence, structure, and functional annotation for protein fold families and provides tools for extracting, transforming, and loading data from public resources to

populate a local protein family database. Data and annotation information were extracted from GenBank (Benson et al. 2007) and the Protein Data Bank (Berman et al. 2000). Additional annotation information (zinc-binding site, active-site residues, *QSDL-start*, *QSDL-end*, SRS1, and SRS2) was added manually. Proteins were assigned to superfamilies and homologous families based on sequence similarity. The protein entries were classified into superfamilies applying the TribeMCL method (Enright et al. 2002) with an inflation value $I = 2$. The subsequent classification of each superfamily into homologous families was achieved by multiple sequence alignments and phylogenetic trees, as calculated by CLUSTALW (v1.83) (Thompson et al. 1994) with default parameters.

All sequence entries that share >98% sequence identity and descent from the same organism are considered as a single protein entry in the database. In case of multiple sequence entries for each protein, the longest sequence for one protein entry was assigned as the reference sequence for analysis. All sequences shorter than a minimal length of 272 amino acids and longer than a maximum length of 439 amino acids were discarded. These minimal and maximal values were derived based on the shortest and longest sequence within the database with known 3D structure $\pm 10\%$ (quinone oxidoreductase from *Thermus thermophilus*, 302 amino acids; formaldehyde dehydrogenase from *Pseudomonas putida*, 399 amino acids). For sequence entries where structure information was available, structure data were stored as a structure entry. Secondary structure information was calculated using DSSP (Kabsch and Sander 1983) and displayed within one set of the annotated multiple sequence alignments. Multiple sequence alignments of each family were used for improvement of classification, as well as for enrichment of annotation information.

Web accessibility

Annotated multiple sequence alignments and phylogenetic trees are provided via the online accessible version of the MDRED at <http://www.mdred.uni-stuttgart.de>. For each alignment, the information for amino acid conservation is given as calculated by PLOTCON (Rice et al. 2000). For each homologous family and superfamily, family-specific HMM profiles, calculated with the HMMER program (<http://hmmer.janelia.org/>), and phylogenetic trees that are visualized applying the program PHYLONDENDRON (<http://iubio.bio.indiana.edu/soft/molbio/java/apps>) are supplied. All protein entries are linked to the respective GenBank entries, and all annotated multiple sequence alignments, phylogenetic trees, structural monomers, and HMM profiles can be visualized and accessed via the website. Additionally, an archive, comprising sequences, structures, alignments, and phylogenetic trees grouped by families, and a formatted text file listing all protein information can be downloaded.

Data analysis

For structural analysis and visualization, the PyMOL program (DeLano Scientific) and Swiss-PdbViewer (Guex and Peitsch 1997) were used. In case of multiple structure entries for one protein, the structural monomer originating from the wild-type enzyme with the best resolution was used for analysis. All structures were superimposed onto the structure of HLADH with PDB entry 1HEU (Meijers et al. 2001) for analysis. Conservation analysis was performed for each class of short,

medium, and long QSDL, using the program A12Co (Pei and Grishin 2001) based on a multiple sequence alignment including all proteins of each class. Scripts for analysis were written in Perl.

Acknowledgments

Valuable contributions by Ima Avalos Vizcarra and Florian Wagner are acknowledged. This work was financially supported by the Deutsche Bundesstiftung Umwelt.

References

- Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Wheeler, D.L. 2007. GenBank. *Nucleic Acids Res.* **35**: D21–D25.
- Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. 2000. The Protein Data Bank. *Nucleic Acids Res.* **28**: 235–242.
- Chase Jr., T. 1999. Alcohol dehydrogenases: Identification and names for gene families. *Plant Mol. Biol. Rep.* **17**: 333–350.
- Creaser, E.H., Murali, C., and Britt, K.A. 1990. Protein engineering of alcohol dehydrogenases: Effects of amino acid changes at positions 93 and 48 of yeast ADH1. *Protein Eng.* **3**: 523–526.
- Eklund, H., Nordström, B., Zeppezauer, E., Söderlund, G., Ohlsson, I., Boiwe, T., Söderberg, B.-O., Tapia, O., Brändén, C.-I., and Åkeson, Å. 1976. Three-dimensional structure of horse liver alcohol dehydrogenase at 2.4 Å resolution. *J. Mol. Biol.* **102**: 27–59.
- Enright, A.J., Van Dongen, S., and Ouzounis, C.A. 2002. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* **30**: 1575–1584.
- Fischer, M. and Pleiss, J. 2003. The Lipase Engineering Database: A navigation and analysis tool for protein families. *Nucleic Acids Res.* **31**: 319–321.
- Fischer, M., Thai, Q.K., Grieb, M., and Pleiss, J. 2006. DWARF—a data warehouse system for analyzing protein families. *BMC Bioinformatics* **7**: 495. doi: 10.1186/1471-2105-7-495.
- Fischer, M., Knoll, M., Sirim, D., Wagner, F., Funke, S., and Pleiss, J. 2007. The Cytochrome P450 Engineering Database: A navigation and prediction tool for the cytochrome P450 protein family. *Bioinformatics* **23**: 2015–2017.
- Gonzalez, E., Fernandez, M.R., Larroy, C., Sola, L., Pericas, M.A., Pares, X., and Biosca, J.A. 2000. Characterization of a (2R,3R)-2,3-butanediol dehydrogenase as the *Saccharomyces cerevisiae* YAL060W gene product. Disruption and induction of the gene. *J. Biol. Chem.* **275**: 35876–35885.
- Guex, N. and Peitsch, M.C. 1997. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* **18**: 2714–2723.
- Guy, J.E., Isupov, M.N., and Littlechild, J.A. 2003. The structure of an alcohol dehydrogenase from the hyperthermophilic archaeon *Aeropyrum pernix*. *J. Mol. Biol.* **331**: 1041–1051.
- Heidlas, J. and Tressl, R. 1990. Purification and characterization of a (R)-2,3-butanediol dehydrogenase from *Saccharomyces cerevisiae*. *Arch. Microbiol.* **154**: 267–273.
- Hurley, T.D. and Bosron, W.F. 1992. Human alcohol dehydrogenase: Dependence of secondary alcohol oxidation on the amino acids at positions 93 and 94. *Biochem. Biophys. Res. Commun.* **183**: 93–99.
- Jörnvall, H. 1977. Differences between alcohol dehydrogenases—structural properties and evolutionary aspects. *Eur. J. Biochem.* **72**: 443–452.
- Jörnvall, H., Eklund, H., and Branden, C.I. 1978. Subunit conformation of yeast alcohol dehydrogenase. *J. Biol. Chem.* **253**: 8414–8419.
- Jörnvall, H., Hoog, J.O., and Persson, B. 1999. SDR and MDR: Completed genome sequences show these protein families to be large, of old origin, and of complex nature. *FEBS Lett.* **445**: 261–264.
- Jörnvall, H., Nordling, E., and Persson, B. 2003. Multiplicity of eukaryotic ADH and other MDR forms. *Chem. Biol. Interact.* **143–144**: 255–261.
- Kabsch, W. and Sander, C. 1983. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**: 2577–2637.
- MacKintosh, R.W. and Fewson, C.A. 1988. Benzyl alcohol dehydrogenase and benzaldehyde dehydrogenase II from *Acinetobacter calcoaceticus*. Purification and preliminary characterization. *Biochem. J.* **250**: 743–751.
- Meijers, R., Morris, R.J., Adolph, H.W., Merli, A., Lamzin, V.S., and Cedergren-Zeppezauer, E.S. 2001. On the enzymatic activation of NADH. *J. Biol. Chem.* **276**: 9316–9321.

- Nordling, E., Jörnvall, H., and Persson, B. 2002. Medium-chain dehydrogenases/reductases (MDR). Family characterizations including genome comparisons and active site modeling. *Eur. J. Biochem.* **269**: 4267–4276.
- Norin, A., Van Ophem, P.W., Piersma, S.R., Persson, B., Duine, J.A., and Jörnvall, H. 1997. Mycothiol-dependent formaldehyde dehydrogenase, a prokaryotic medium-chain dehydrogenase/reductase, phylogenetically links different eukaryotic alcohol dehydrogenases—primary structure, conformational modelling and functional correlations. *Eur. J. Biochem.* **248**: 282–289.
- Pei, J. and Grishin, N.V. 2001. AL2CO: Calculation of positional conservation in a protein sequence alignment. *Bioinformatics* **17**: 700–712.
- Persson, B., Zigler Jr., J.S., and Jörnvall, H. 1994. A super-family of medium-chain dehydrogenases/reductases (MDR). Sub-lines including zeta-crystallin, alcohol and polyol dehydrogenases, quinone oxidoreductase enoyl reductases, VAT-1 and other proteins. *Eur. J. Biochem.* **226**: 15–22.
- Pleiss, J., Fischer, M., Peiker, M., Thiele, C., and Schmid, R.D. 2000. Lipase engineering database—understanding and exploiting sequence–structure–function relationships. *J. Mol. Catal., B Enzym.* **10**: 491–508.
- Rice, P., Longden, I., and Bleasby, A. 2000. EMBOSS: The European molecular biology open software suite. *Trends Genet.* **16**: 276–277.
- Riveros-Rosas, H., Julian-Sanchez, A., Villalobos-Molina, R., Pardo, J.P., and Pina, E. 2003. Diversity, taxonomy and evolution of medium-chain dehydrogenase/reductase superfamily. *Eur. J. Biochem.* **270**: 3309–3334.
- Shafiqat, J., Hoog, J.O., Hjelmqvist, L., Oppermann, U.C., Ibanez, C., and Jörnvall, H. 1999. An ethanol-inducible MDR ethanol dehydrogenase/ acetaldehyde reductase in *Escherichia coli*: Structural and enzymatic relationships to the eukaryotic protein forms. *Eur. J. Biochem.* **263**: 305–311.
- Thompson, J.D., Higgins, D.G., and Gibson, T.J. 1994. 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**: 4673–4680.
- Xie, P.T. and Hurley, T.D. 1999. Methionine-141 directly influences the binding of 4-methylpyrazole in human σ alcohol dehydrogenase. *Protein Sci.* **8**: 2639–2644.
- Ziegelmann-Fjeld, K.I., Musa, M.M., Phillips, R.S., Zeikus, J.G., and Vieille, C. 2007. A *Thermoanaerobacter ethanolicus* secondary alcohol dehydrogenase mutant derivative highly active and stereoselective on phenylacetone and benzylacetone. *Protein Eng. Des. Sel.* **20**: 47–55.

4.5

Publikation erschienen in *Bioinformatics* **23**: 2015-2017

Fischer, M. Knoll, M., Sirim, D., Wagner, F., Funke, S., Pleiss, J., 2007. The Cytochrome P450 Engineering Database: a navigation and prediction tool for the cytochrome P450 protein family.

Sequence analysis

The Cytochrome P450 Engineering Database: a navigation and prediction tool for the cytochrome P450 protein family

Markus Fischer¹, Michael Knoll², Demet Sirim², Florian Wagner², Sonja Funke² and Juergen Pleiss^{2,*}

¹Department of Biochemistry & Molecular Biophysics, Columbia University, 1130 St. Nicholas Ave, New York, NY 10032, USA and ²Institute of Technical Biochemistry, University of Stuttgart, Allmandring 31, 70569 Stuttgart, Germany

Received on March 20, 2007; revised on May 7, 2007; accepted on May 10, 2007

Advance Access publication May 17, 2007

Associate Editor: Alex Bateman

ABSTRACT

Summary: The Cytochrome P450 Engineering Database (CYPED) has been designed to serve as a tool for a comprehensive and systematic comparison of protein sequences and structures within the vast and diverse family of cytochrome P450 monooxygenases (CYPs). The CYPED currently integrates sequence and structure data of 3911 and 25 proteins, respectively. Proteins are grouped into homologous families and superfamilies according to Nelson's classification. Nonclassified CYP sequences are assigned by similarity. Functionally relevant residues are annotated. The web accessible version contains multisequence alignments, phylogenetic trees and HMM profiles. The CYPED is regularly updated and supplies all data for download. Thus, it provides a valuable data source for phylogenetic analysis, investigation of sequence–function relationships and the design of CYPs with improved biochemical properties.

Abbreviations: Cytochrome P450 Engineering Database, CYPED; cytochrome P450 monooxygenase, CYP; Hidden Markov Model, HMM.

Availability: www.cyped.uni-stuttgart.de

Contact: Juergen.Pleiss@itb.uni-stuttgart.de

(<http://www.imm.ki.se/CYPalleles/>), the Directory of P450-containing Systems (<http://www.icgeb.trieste.it/>), the P450 Knowledgebase (<http://cpd.ibmh.msk.su/>) (Lisitsa *et al.*, 2001), the Arabidopsis P450 database (<http://www.p450.kvl.dk/>), the Insect P450 Site (<http://p450.antibes.inra.fr/>) or the P450s in PROMISE (<http://metallo.scripps.edu/PROMISE/P450.html>), a common data structure enabling the integration of available information on protein sequence and structure is still lacking. Therefore the Cytochrome P450 Engineering Database (CYPED) was implemented using the data warehouse system DWARF (Fischer *et al.*, 2006). The underlying data model assists the systematic analysis of the relationship of sequence, structure, and function of this vast and highly diverse protein family. The CYPED is the first cytochrome P450 data resource that combines information on sequences, sequence alignments, annotation and structures of CYPs. For data retrieval sequence, structure and annotation information is extracted from GenBank (Benson *et al.*, 2003) and PDB (<http://www.pdb.org/>). Functional annotation information is extended by an automated annotation transfer and was manually validated and enriched. Besides, the online accessible version, which is publicly available, supports the classification of unknown sequences by performing a BLAST search against the CYPED or by alignment to family-specific HMM profiles.

1 INTRODUCTION

Cytochrome P450 monooxygenases (CYPs) are heme containing enzymes that metabolize physiologically important compounds in many species of microorganisms, plants, animals and humans. CYPs catalyse the oxidation of a wide range of endogenous compounds in biosynthetic and biodegradation pathways, as well as xenobiotics such as drugs and environmental contaminants (Montellano, 1995). Therefore, understanding the substrate specificities of human CYPs is crucial for successful drug development (Raucy *et al.*, 2001). Although there are already numerous resources dedicated to CYPs like the Cytochrome P450 Homepage (<http://drnelson.utmem.edu/CytochromeP450.html>), the Homepage of the Human Cytochrome P450 (CYP) Allele Nomenclature Committee

2 DEVELOPMENT AND CONSTRUCTION

Seed sequences of CYPs were extracted from the Cytochrome P450 Homepage (Nelson *et al.*, 2002) and assigned to homologous families and superfamilies according to the Nelson classification scheme (Nelson, 2006). For each seed sequence, a BLAST search (Altschul *et al.*, 1997) was performed in the non-redundant sequence database at GenBank (Benson *et al.*, 2003) with a low E-value ($E = 10^{-100}$) to prevent overlapping hits among different superfamilies. For each hit, information on sequence, position-specific annotations, functional descriptions and the source organism was extracted and loaded by an automated retrieval system into an in-house developed relational database system (Fischer and Pleiss, 2003). New protein entries are assigned to homologous families and superfamilies according to their

*To whom correspondence should be addressed.

sequence similarity. The parameters are chosen as specified by Nelson. Proteins sharing a sequence identity of ≥ 40 or $\geq 55\%$ are members of the same superfamily or homologous family, respectively. About 2% of sequences were individually assigned to a family according to the recommendations of the P450 Nomenclature Committee. This procedure was applied to sequences which do not share a identity of $\geq 40\%$ with members of a homologous family but still belong to this family by definition. About 30% of the proteins within the database have not been classified by the nomenclature committee yet, and thus are assigned to the corresponding family by sequence similarity and named as 'homologous protein of family X (by similarity)'. They will be reassigned automatically during a database update in case of the classification information changes. Therefore, in contrast to existing P450 resources, the CYPED includes additional information which is expected to deepen the understanding in sequence-structure-function relationship of the CYP protein family and to apply the new gained knowledge to the design of improved CYPs.

CYP sequences that originate from the same organism and share a sequence identity of at least 98% are assigned to a single protein entry. For each protein entry the longest sequence was defined as reference sequence of the respective protein. For GenBank entries representing protein structures, monomers were extracted from the ExPDB database (Schwede *et al.*, 2000) and deposited as structure entries. Secondary structure information was calculated using DSSP (Kabsch and Sander, 1983), stored and annotated. To improve consistency and quality of the data, the classification into families and superfamilies for those protein entries that have not been classified was validated by performing multisequence alignments and a phylogenetic analysis. Multisequence alignment was also used to enrich annotation information and to control annotation quality. It was assumed that conserved sequence motifs should align for each superfamily. Therefore annotation information was transferred from one sequence to all other sequences in an alignment if the respective residues were conserved.

The CYPED is updated regularly. By an automated Perl script (Fischer *et al.*, 2006) new sequences are retrieved. New annotation information at GenBank for existing sequence entries and structure information is updated as well. Additionally, the update script takes care of changes of classification information.

3 DATA CONTENT AND ANNOTATED MULTISEQUENCE ALIGNMENTS

The CYPED contains sequence data on 3911 proteins. For 25 proteins of 20 different homologous families crystal structures are deposited. Since the CYPED provides a protein analysis tool to investigate the relationship between protein sequences, structures and their function, the data content is limited to sequences, structures and annotation information on amino acid level. The protein entries are assigned to 1111 homologous families, which are grouped into 531 superfamilies. Superfamilies and homologous families are represented as

multisequence alignments generated by CLUSTALW (Thompson *et al.*, 1994).

Common CYP motifs not present in GenBank entries were extracted from literature and annotated within CYPED entries. These motifs are the proline-rich region at the N terminus of CYPs (Kemper, 2004), the motif at the C-terminal end of helix K, the AGXXT motif present in helix I, and the functionally essential cysteine (Mestres, 2005). Amino acids linked to functional annotations are coloured within the alignments, and the residue number and further information is displayed upon moving the cursor over the respective amino acid. For each alignment, each column is coloured by the amino acid conservation score as calculated by PLOTCON (Rice *et al.*, 2000). For each homologous family and superfamily family-specific HMM profiles (<http://hmmer.janelia.org/>) are supplied.

4 WEB ACCESSIBILITY

The CYPED is accessible at <http://www.cyped.uni-stuttgart.de> by any JavaScript capable WWW browser. It can be browsed by superfamilies and homologous families, organisms, protein structures and the systematic CYP nomenclature. Protein names, source organisms, identifier codes and links to the corresponding GenBank entries are presented as tables. Protein sequences and trees can either be displayed with their accession codes as identifiers or their systematic names and source organisms. The hits are linked to the respective sequence entry, superfamily and homologous family. This functionality can also be applied to classify unknown sequences by performing a BLAST search on the web interface against the CYPED or by alignment to family-specific HMM profiles, which can be downloaded. All multisequence alignments, phylogenetic trees and structural monomers have been pre-calculated, and can be visualized or downloaded from this web site, too. Additionally, an archive can be downloaded comprising sequences, structures, alignments and phylogenetic trees grouped by families, and a formatted text file listing all protein information.

ACKNOWLEDGEMENTS

We acknowledge valuable contributions by Michael Krahn. This work was supported by the German Federal Ministry of Education and Research (project PTJ 0313080) and the German Research Foundation (SFB 706). We further are grateful for the unknown reviewer's comments which contributed greatly to the improvement of data content and the user interface of the CYPED. Funding to pay the Open Access publication charges was provided by the German Research Foundation.

Conflict of Interest: none declared.

REFERENCES

- Altschul, S.F. *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389-3402.

- Benson,D.A. *et al.* (2003) GenBank. *Nucleic Acids Res.*, **31**, 23–27.
- Fischer,M. and Pleiss,J. (2003) The Lipase Engineering Database: a navigation and analysis tool for protein families. *Nucleic Acids Res.*, **31**, 319–321.
- Fischer,M. *et al.* (2006) DWARF – a data warehouse system for analyzing protein families. *BMC Bioinformatics*, **7**, 495.
- Kabsch,W. and Sander,C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.
- Kemper,B. (2004) Structural basis for the role in protein folding of conserved proline-rich regions in cytochromes P450. *Toxicol Appl. Pharmacol.*, **199**, 305–315.
- Lisitsa,A.V. *et al.* (2001) Cytochrome P450 database. *SAR QSAR Environ Res.*, **12**, 359–366.
- Mestres,J. (2005) Structure conservation in cytochromes P450. *Proteins*, **58**, 596–609.
- Montellano,O.d. (1995) Cytochrome P450: structure, mechanism and biochemistry. New York, Plenum Press.
- Nelson,D.R. (2006) Cytochrome P450 nomenclature, 2004. *Methods Mol. Biol.*, **320**, 1–10.
- Nelson,D.R. *et al.* (2002) Mining databases for cytochrome P450 genes. *Methods Enzymol.*, **357**, 3–15.
- Raucy,J.L. and Allen,S.W. (2001) Recent advances in P450 research. *Pharmacogenomics J.*, **1**, 178–186.
- Rice,P. *et al.* (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.*, **16**, 133–154.
- Schwede,T. *et al.* (2000) Protein structure computing in the genomic era. *Res Microbiol.*, **151**, 107–112.
- Thompson,J.D. *et al.* (1994) 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**, 4673–4680.

5 Abschließende Betrachtung

Anhand systematischer Strukturvergleiche verschiedener Vertreter der Familie der Thiamindiphosphat-abhängigen Enzyme konnten einfache Modelle abgeleitet werden, die es ermöglichten, die im Experiment beobachteten Substratspezifitäten und Selektivitäten hinreichend zu erklären. Darüber hinaus konnten diese einfachen Regeln prädiktiv zur Vorhersage von Mutanten mit erweitertem Substratspektrum angewendet werden. Besondere Bedeutung kommt der vorliegenden Arbeit zu, da mithilfe der erstellten einfachen Regeln (*S*)-Enantiomere der Produkte der sonst (*R*)-spezifischen Enzyme zugänglich werden. Die im Rahmen dieser Arbeit abgeleiteten Regeln trugen daher wesentlich zum Verständnis der Sequenz-Struktur-Funktionsbeziehung dieser Familie bei. Um diese Regeln zu verallgemeinern und auf weitere Proteine der Familie anzuwenden, befindet sich eine Proteinfamiliendatenbank, die eine umfassende systematische Analyse dieser Enzymfamilie ermöglicht, im Aufbau.

Eine solche Proteinfamiliendatenbank konnte im Rahmen dieser Arbeit für die diverse Familie der mittelkettigen Alkoholdehydrogenasen erstellt werden (MDRED). Diese erlaubte eine systematische Untersuchung der Proteinfamilie. Eine einfache Klassifizierung anhand eines in der Länge variablen Strukturelements (QSDL) wurde etabliert. Diese ermöglicht es, gezielt für Proteine der einzelnen Klassen wichtige Aminosäuren für die Form der Bindetasche zu identifizieren und Mutanten zur Veränderung der Substratspezifität vorherzusagen. Dies ist von großer Bedeutung, da Alkoholdehydrogenasen in der industriellen Anwendung weit verbreitet sind. Durch die Klassifizierung der Proteine in wohl definierte Familien und der familienspezifischen Analyse, konnte darüber hinaus ein Zusammenhang der Länge des QSDL der einzelnen Familienmitglieder und der Quartärstruktur beobachtet werden. Dies ermöglicht es für alle Familienmitglieder, deren Quartärstruktur bislang nicht untersucht ist, diese vorauszusagen.

Für die Familie der Cytochrom P450 Monooxygenasen wurde ebenfalls eine Proteinfamiliendatenbank etabliert (CYPED), bei welcher vor allem die Klassifizierung und Integration der vorhandenen Daten im Vordergrund stand. Die CYPED ermöglicht es zum Beispiel dem Forschungsfeld der humanen Cytochrom P450 Monooxygenasen systematische Familienanalysen durchzuführen, was hinsichtlich der bekannten Polymorphismen, die zu

unterschiedlichen Verträglichkeiten von Medikamenten führen eine wichtige Rolle spielt. Die CYPED wird am Institut für Technische Biochemie der Universität Stuttgart verwendet, um die Sequenz-Struktur-Funktionsbeziehung dieser Familie zu verstehen. Somit können auftretende Polymorphismen erkannt, und deren Auswirkung verstanden werden.

Im Rahmen dieser Arbeit konnten die Methoden der vergleichenden Sequenz- und Strukturanalyse zur Untersuchung familienspezifischer Eigenschaften erfolgreich auf drei unterschiedliche Enzymfamilien angewendet werden. Am Beispiel der Thiamindiphosphat-abhängigen Enzyme konnte gezeigt werden, dass abgeleitete einfache Regeln direkt zur erfolgreichen Vorhersage von Enzymvarianten herangezogen werden können. Durch Weiterentwicklung und Anwendung des am Institut für Technische Biochemie der Universität Stuttgart entwickelten Datenbanksystems DWARF (Fischer et al. 2006) konnten Proteinfamiliendatenbanken für die zwei großen Proteinfamilien der Cytochrom P450 Monooxygenasen (CYPED) und der mittelkettigen Alkoholdehydrogenasen (MDRED) erstellt werden. Die Datenbanken erlauben eine Integration unterschiedlicher biologischer Daten und erweitern die Möglichkeiten der systematischen Sequenzanalyse. Sie bieten daher eine ideale Grundlage zur Vorhersage von Enzymvarianten mit verbesserten Eigenschaften.

6 Gesamtliteraturverzeichnis

- Collaborative Computational Project Number 4, The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. Sec. D: Biol. Crystallogr.* **50**: 760-763.
- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389-3402.
- Arjunan, P., Umland, T., Dyda, F., Swaminathan, S., Furey, W., Sax, M., Farrenkopf, B., Gao, Y., Zhang, D., and Jordan, F. 1996. Crystal structure of the thiamin diphosphate-dependent enzyme pyruvate decarboxylase from the yeast *Saccharomyces cerevisiae* at 2.3 Å resolution. *J Mol Biol* **256**: 590-600.
- Arnold, F.H. 2001. Combinatorial and computational challenges for biocatalyst design. *Nature* **409**: 253-257.
- Attwood, T.K., Bradley, P., Flower, D.R., Gaulton, A., Maudling, N., Mitchell, A.L., Moulton, G., Nordle, A., Paine, K., Taylor, P., et al. 2003. PRINTS and its automatic supplement, prePRINTS. *Nucleic Acids Res* **31**: 400-402.
- Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Wheeler, D.L. 2003. GenBank. *Nucleic Acids Research* **31**: 23-27.
- Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Wheeler, D.L. 2007. GenBank. *Nucleic Acids Res* **35**: D21-25.
- Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. 2000. The Protein Data Bank. *Nucleic Acids Res* **28**: 235-242.
- Berthold, C.L., Gocke, D., Wood, D., Leeper, F.J., Pohl, M., and Schneider, G. 2007. Structure of the branched-chain keto acid decarboxylase (KdcA) from *Lactococcus lactis* provides insights into the structural basis for the chemoselective and enantioselective carbonylation reaction. *Acta Crystallogr D* **63**: 1217-1224.
- Bornscheuer, U.T., and Pohl, M. 2001. Improved biocatalysts by directed evolution and rational protein design. *Curr Opin Chem Biol* **5**: 137-143.
- Bradford, M.M. 1976. 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.
- Bradshaw, C.W., Lalonde, J.J., and Wong, C.H. 1992. Enzymatic-Synthesis of (R) and (S) 1-Deuteriohexanol. *Appl Biochem Biotech* **33**: 15-24.
- Bringer-Meyer, S., Schimz, K.L., and Sahn, H. 1986. Pyruvate Decarboxylase from *Zymomonas-Mobilis* - Isolation and Partial Characterization. *Archives of Microbiology* **146**: 105-110.

- Bürgi, H.B., and Dunitz, J.D. 1983. From Crystal Statics to Chemical-Dynamics. *Accounts Chem Res* **16**: 153-161.
- Chang, A.K., Nixon, P.F., and Duggleby, R.G. 2000. Effects of deletions at the carboxyl terminus of *Zymomonas mobilis* pyruvate decarboxylase on the kinetic properties and substrate specificity. *Biochemistry-Us* **39**: 9430-9437.
- Chase, T. Jr. 1999. Alcohol Dehydrogenases: Identification and Names for Gene Families. *Plant Molecular Biology Reporter* **17**: 333-350.
- Chothia, C., and Lesk, A.M. 1986. The relation between the divergence of sequence and structure in proteins. *Embo J* **5**: 823-826.
- Creaser, E.H., Murali, C., and Britt, K.A. 1990. Protein engineering of alcohol dehydrogenases: effects of amino acid changes at positions 93 and 48 of yeast ADH1. *Protein Eng* **3**: 523-526
- Danielsson, O., Atrian, S., Luque, T., Hjelmqvist, L., Gonzalezduarte, R., and Jörnvall, H. 1994. Fundamental Molecular Differences between Alcohol-Dehydrogenase Classes. *P Natl Acad Sci USA* **91**: 4980-4984.
- Dawes, E.A., Ribbons, D.W., and Large, P.J. 1966. The route of ethanol formation in *Zymomonas mobilis*. *Biochem J* **98**: 795-803.
- DeLano, W.L. 2002. The PyMOL Molecular Graphics System. <http://www.pymol.org>, San Carlo, CA, USA.
- Demir, A.S., Dünwald, T., Iding, H., Pohl, M., and Müller, M. 1999. Asymmetric benzoin reaction catalyzed by benzoylformate decarboxylase. *Tetrahedron-Asymmetry* **10**: 4769-4774.
- Demir, A.S., Pohl, M., Janzen, E., and Muller, M. 2001. Enantioselective synthesis of hydroxy ketones through cleavage and formation of acyloin linkage. Enzymatic kinetic resolution via C-C bond cleavage. *J Chem Soc Perk T 1*: 633-635.
- Demir, A.S., Sesenoglu, O., Dünkemann, P., and Muller, M. 2003. Benzaldehyde lyase-catalyzed enantioselective carbonylation of aromatic aldehydes with mono- and dimethoxy acetaldehyde. *Org Lett* **5**: 2047-2050.
- Demir, A.S., Sesenoglu, O., Eren, E., Hosrik, B., Pohl, M., Janzen, E., Kolter, D., Feldmann, R., Dunkelmann, P., and Muller, M. 2002. Enantioselective synthesis of alpha-hydroxy ketones via benzaldehyde lyase-catalyzed C-C bond formation reaction. *Adv Synth Catal* **344**: 96-103.
- Dietrich, A., and König, S. 1997. Substrate activation behaviour of pyruvate decarboxylase from *Pisum sativum* cv Miko. *Febs Lett* **400**: 42-44.
- Dobritzsch, D., König, S., Schneider, G., and Lu, G. 1998. High resolution crystal structure of pyruvate decarboxylase from *Zymomonas mobilis*. Implications for substrate activation in pyruvate decarboxylases. *J Biol Chem* **273**: 20196-20204.

- Dominguez de Maria, P., Pohl, M., Gocke, D., Groger, H., Trauthwein, H., Stillger, T., Walter, L., and Muller, M. 2007. Asymmetric synthesis of aliphatic 2-hydroxy ketones by enzymatic carboligation of aldehydes. *Eur J Org Chem*: 2940-2944.
- Dominguez de Maria, P., Stillger, T., Pohl, M., Wallert, S., Drauz, K., Gröger, H., Trauthwein, H., and Liese, A. 2006. Preparative enantioselective synthesis of benzoin and (R)-2-hydroxy-1-phenylpropane using benzaldehyde lyase. *J. Molec. Cat. B: Enzymatic* **38**: 43-47.
- Dünkemann, P., Kolter-Jung, D., Nitsche, A., Demir, A.S., Siegert, P., Lingen, B., Baumann, M., Pohl, M., and Müller, M. 2002. Development of a donor-acceptor concept for enzymatic cross-coupling reactions of aldehydes: The first asymmetric cross-benzoin condensation. *J Am Chem Soc* **124**: 12084-12085.
- Dünkemann, P., Pohl, M., and Müller, M. 2004. Enantiomerically pure 2-hydroxy carbonyl compounds through enzymatic C-C bond formation. *Chim. Oggi/Chemistry Today supplement Chiral Catalysis* **22**: 24-28.
- Dünnwald, T., Demir, A.S., Siegert, P., Pohl, M., and Müller, M. 2000. Enantioselective synthesis of (S)-2-hydroxypropanone derivatives by benzoylformate decarboxylase catalyzed C-C bond formation. *Eur J Org Chem*: 2161-2170.
- Dünnwald, T., and Müller, M. 2000. Stereoselective Formation of Bis(α -hydroxy ketones) via Enzymatic Carboligation. *J. Org. Chem.* **65**: 8608 -8612.
- Eklund, H., Nordström, B., Zeppezauer, E., Söderlund, G., Ohlsson, I., Boiwe, T., Söderberg, B.-O., Tapia, O., Brändén, C.-I., and Åkeson, Å. 1976. Three-dimensional structure of horse liver alcohol dehydrogenase at 2.4 Å resolution. *J. Mol. Biol.* 102: 27-59.
- Emsley, P., and Cowtan, K. 2004. Coot: model-building tools for molecular graphics. *Acta Crystallogr D* **60**: 2126-2132.
- Enright, A.J., Van Dongen, S., and Ouzounis, C.A. 2002. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* **30**: 1575-1584.
- Faber, K., and Kroutil, W. 2005. New enzymes for biotransformations. *Curr Opin Chem Biol* **9**: 181-187.
- Finn, R.D., Tate, J., Mistry, J., Coghill, P.C., Sammut, S.J., Hotz, H.R., Ceric, G., Forslund, K., Eddy, S.R., Sonnhammer, E.L., et al. 2008. The Pfam protein families database. *Nucleic Acids Res* **36**: D281-288.
- Fischer, M., Knoll, M., Sirim, D., Wagner, F., Funke, S., and Pleiss, J. 2007. The Cytochrome P450 Engineering Database: a navigation and prediction tool for the cytochrome P450 protein family. *Bioinformatics* **23**: 2015-2017.
- Fischer, M., and Pleiss, J. 2003. The Lipase Engineering Database: a navigation and analysis tool for protein families. *Nucleic Acids Research* **31**: 319-321.

- Fischer, M., Thai, Q.K., Grieb, M., and Pleiss, J. 2006. DWARF--a data warehouse system for analyzing protein families. *BMC Bioinformatics* **7**: 495.
- Frank, R.A.W., Leeper, F.J., and Luisi, B.F. 2007. Structure, mechanism and catalytic duality of thiamine-dependent enzymes. *Cell Mol Life Sci* **64**: 892-905.
- Gaasterland, T. 1998. Structural genomics taking shape. *Trends Genet* **14**: 135.
- Gerstein, M. 2000. Integrative database analysis in structural genomics. *Nat Struct Biol* **7 Suppl**: 960-963.
- Gocke, D., Berthold, C., Graf, T., Brosi, H., Frindi-Wosch, I., Knoll, M., Stillger, T., Walter, L., Müller, M., Pleiss, J., et al. unpublished results.
- Gocke, D., Nguyen, C.L., Pohl, M., Stillger, T., Walter, L., and Mueller, M. 2007. Branched-chain keto acid decarboxylase from *Lactococcus lactis* (KdcA), a valuable thiamine diphosphate-dependent enzyme for asymmetric C-C bond formation. *Adv Synth Catal* **349**: 1425-1435.
- Gocke, D., Walter, L., Gauchenova, E., Kolter, G., Knoll, M., Berthold, C.L., Schneider, G., Pleiss, J., Müller, M., and Pohl, M. 2008. Rational Protein Design of ThDP-dependent Enzymes: Engineering Stereoselectivity. *Chembiochem*: in press.
- Gonzalez, B., and Vicuna, R. 1989. Benzaldehyde lyase, a novel thiamine PPi-requiring enzyme, from *Pseudomonas fluorescens* biovar I. *J Bacteriol* **171**: 2401-2405.
- Gonzalez, E., Fernandez, M.R., Larroy, C., Sola, L., Pericas, M.A., Pares, X., and Biosca, J.A. 2000. Characterization of a (2R,3R)-2,3-butanediol dehydrogenase as the *Saccharomyces cerevisiae* YAL060W gene product. Disruption and induction of the gene. *J Biol Chem* **275**: 35876-35885.
- Guex, N., and Peitsch, M.C. 1997. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18**: 2714-2723.
- Guy, J.E., Isupov, M.N., and Littlechild, J.A. 2003. The structure of an alcohol dehydrogenase from the hyperthermophilic archaeon *Aeropyrum pernix*. *J Mol Biol* **331**: 1041-1051.
- Hansen, J. 2006. Wissenschaftliche Datenbanken in der Genomforschung. *BIOspektrum* **04.06, 12. Jahrgang**: 387-389.
- Hasemann, C.A., Kurumbail, R.G., Boddupalli, S.S., Peterson, J.A., and Deisenhofer, J. 1995. Structure and function of cytochromes P450: a comparative analysis of three crystal structures. *Structure* **3**: 41-62.
- Hasson, M.S., Muscate, A., McLeish, M.J., Polovnikova, L.S., Gerlt, J.A., Kenyon, G.L., Petsko, G.A., and Ringe, D. 1998. The crystal structure of benzoylformate decarboxylase at 1.6 angstrom resolution: Diversity of catalytic residues in thiamin diphosphate-dependent enzymes. *Biochemistry-US* **37**: 9918-9930.

- Hegeman, G.D. 1970. [89] Benzoylformate decarboxylase (*Pseudomonas putida*)
Methods in Enzymology. In *Metabolism of Amino Acids and Amines Part A*, Volume 17, Part 1 ed. (ed. H.T.a.C.W. Tabor), pp. 674-678. Academic Press.
- Heidlas, J., and Tressl, R. 1990. Purification and characterization of a (R)-2,3-butanediol dehydrogenase from *Saccharomyces cerevisiae*. *Arch Microbiol* **154**: 267-273.
- Henke, E., Pleiss, J., and Bornscheuer, U.T. 2002. Activity of lipases and esterases towards tertiary alcohols: insights into structure-function relationships. *Angew Chem Int Ed Engl* **41**: 3211-3213.
- Hildebrand, F., Kuhl, S., Pohl, M., Vasic-Racki, D., Muller, M., Wandrey, C., and Lutz, S. 2007. The production of (R)-2-hydroxy-1-phenyl-propan-1-one derivatives by benzaldehyde lyase from *Pseudomonas fluorescens* in a continuously operated membrane reactor. *Biotechnol Bioeng* **96**: 835-843.
- Hildebrandt, G., and Klavehn, W. 1930. Verfahren zur Herstellung von 1-L-Phenyl-2-methylaminopropan-1-ol, Knoll AG Chemische Fabriken in Ludwigshafen, Ger. Patent DE 548459.
- Hinrichsen, P., Gomez, I., and Vicuna, R. 1994. Cloning and sequencing of the gene encoding benzaldehyde lyase from *Pseudomonas fluorescens* biovar I. *Gene* **144**: 137-138.
- Hischer, T., Gocke, D., Fernandez, M., Hoyos, P., Alcantara, A.R., Sinisterra, J.V., Hartmeier, W., and Ansorge-Schumacher, M.B. 2005. Stereoselective synthesis of novel benzoinis catalysed by benzaldehyde lyase in a gel-stabilised two-phase system. *Tetrahedron* **61**: 7378-7383.
- Hübner, G., Weidhase, R., and Schellenberger, A. 1978. Mechanism of Substrate Activation of Pyruvate Decarboxylase - 1st Approach. *European Journal of Biochemistry* **92**: 175-181.
- Hulo, N., Bairoch, A., Bulliard, V., Cerutti, L., De Castro, E., Langendijk-Genevaux, P.S., Pagni, M., and Sigrist, C.J. 2006. The PROSITE database. *Nucleic Acids Res* **34**: D227-230.
- Hurley, T.D., and Bosron, W.F. 1992. Human alcohol dehydrogenase: dependence of secondary alcohol oxidation on the amino acids at positions 93 and 94. *Biochem Biophys Res Commun* **183**: 93-99.
- Iding, H., Dünwald, T., Greiner, L., Liese, A., Müller, M., Siegert, P., Grotzinger, J., Demir, A.S., and Pohl, M. 2000. Benzoylformate decarboxylase from *Pseudomonas putida* as stable catalyst for the synthesis of chiral 2-hydroxy ketones. *Chem-Eur J* **6**: 1483-1495.
- Iding, H., Siegert, P., Mesch, K., and Pohl, M. 1998. Application of alpha-keto acid decarboxylases in biotransformations. *Bba-Protein Struct M* **1385**: 307-322.
- Ingelman-Sundberg, M. 2004. Human drug metabolising cytochrome P450 enzymes: properties and polymorphisms. *N-S Arch Pharmacol* **369**: 89-104.

- Janzen, E., Muller, M., Kolter-Jung, D., Kneen, M.M., McLeish, M.J., and Pohl, M. 2006. Characterization of benzaldehyde lyase from *Pseudomonas fluorescens*: A versatile enzyme for asymmetric C-C bond formation. *Bioorg Chem* **34**: 345-361.
- Jordan, F. 2003. Current mechanistic understanding of thiamin diphosphate-dependent enzymatic reactions. *Nat Prod Rep* **20**: 184-201.
- Jordan, F., Liu, M., Sergienko, E., Zhang, Z., Brunskill, A., Arjunan, P., and Furey, W. 2004. Yeast pyruvate decarboxylase: New features of the structure and mechanism. In *Thiamine-Catalytic Mechanisms in Normal and Disease States*. (eds. F. Jordan, and R.N. Patel). Marcel Dekker Inc., New York / Basel.
- Jordan, F., and Nemeria, N.S. 2005. Experimental observation of thiamin diphosphate-bound intermediates on enzymes and mechanistic information derived from these observations. *Bioorg Chem* **33**: 190-215.
- Jörnvall, H. 1977. Differences between Alcohol Dehydrogenases - Structural-Properties and Evolutionary Aspects. *European Journal of Biochemistry* **72**: 443-452.
- Jörnvall, H., Eklund, H., and Branden, C.I. 1978. Subunit conformation of yeast alcohol dehydrogenase. *J Biol Chem* **253**: 8414-8419.
- Jörnvall, H., Hoog, J.O., and Persson, B. 1999. SDR and MDR: completed genome sequences show these protein families to be large, of old origin, and of complex nature. *Febs Lett* **445**: 261-264.
- Jörnvall, H., Nordling, E., and Persson, B. 2003. Multiplicity of eukaryotic ADH and other MDR forms. *Chem Biol Interact* **143-144**: 255-261.
- Kabsch, W., and Sander, C. 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**: 2577-2637.
- Kalscheuer, R., Stolting, T., and Steinbuchel, A. 2006. Microdiesel: *Escherichia coli* engineered for fuel production. *Microbiol-Sgm* **152**: 2529-2536.
- Kemper, B. 2004. Structural basis for the role in protein folding of conserved proline-rich regions in cytochromes P450. *Toxicol Appl Pharm* **199**: 305-315.
- Kern, D., Kern, G., Neef, H., Tittmann, K., KillenbergJabs, M., Wikner, C., Schneider, G., and Hubner, G. 1997. How thiamine diphosphate is activated in enzymes. *Science* **275**: 67-70.
- Kim, S.H., Shin, D.H., Choi, I.G., Schulze-Gahmen, U., Chen, S., and Kim, R. 2003. Structure-based functional inference in structural genomics. *J Struct Funct Genomics* **4**: 129-135.
- Kneen, M.M., Pogozheva, I.D., Kenyon, G.L., and McLeish, M.J. 2005. Exploring the active site of benzaldehyde lyase by modeling and mutagenesis. *Bba-Proteins Proteom* **1753**: 263-271.

- Knoll, M., Muller, M., Pleiss, J., and Pohl, M. 2006. Factors mediating activity, selectivity, and substrate specificity for the thiamin diphosphate-dependent enzymes benzaldehyde lyase and benzoylformate decarboxylase. *ChemBiochem* **7**: 1928-1934.
- Konig, S. 1998. Subunit structure, function and organisation of pyruvate decarboxylases from various organisms. *Bba-Protein Struct M* **1385**: 271-286.
- Korz, D.J., Rinas, U., Hellmuth, K., Sanders, E.A., and Deckwer, W.D. 1995. Simple fed-batch technique for high cell density cultivation of Escherichia coli. *J Biotechnol* **39**: 59-65.
- Krause, A., Stoye, J., and Vingron, M. 2005. Large scale hierarchical clustering of protein sequences. *BMC Bioinformatics* **6**: 15.
- Krogh, A., Brown, M., Mian, I.S., Sjolander, K., and Haussler, D. 1994. Hidden Markov models in computational biology. Applications to protein modeling. *J Mol Biol* **235**: 1501-1531.
- Laskowsky, R.A., McArthur, M.W., Moss, D.S., and Thornton, J.M. 1993. PROCHECK. *J. Appl. Crystallogr.* **26**: 282-291.
- Lee, D., Redfern, O., and Orengo, C. 2007. Predicting protein function from sequence and structure. *Nat Rev Mol Cell Biol* **8**: 995-1005.
- Leisola, M., and Turunen, O. 2007. Protein engineering: opportunities and challenges. *Appl Microbiol Biot* **75**: 1225-1232.
- Leslie, A.G.W. 1992. Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 + ESF-EAMCB Newsl. Protein Crystallogr* **26**.
- Liebeton, K., Zonta, A., Schimossek, K., Nardini, M., Lang, D., Dijkstra, B.W., Reetz, M.T., and Jaeger, K.E. 2000. Directed evolution of an enantioselective lipase. *Chem Biol* **7**: 709-718.
- Liese, A., Seelbach, K., and Wandrey, C. 2006. *Industrial Biotransformations*. Wiley-VCH.
- Lingen, B., Grötzinger, J., Kolter, D., Kula, M.R., and Pohl, M. 2002. Improving the carboligase activity of benzoylformate decarboxylase from *Pseudomonas putida* by a combination of directed evolution and site-directed mutagenesis. *Protein Eng* **15**: 585-593.
- Lingen, B., Kolter-Jung, D., Dunkelmann, P., Feldmann, R., Grotzinger, J., Pohl, M., and Muller, M. 2003. Alteration of the substrate specificity of benzoylformate decarboxylase from *Pseudomonas putida* by directed evolution. *ChemBiochem* **4**: 721-726.
- Lisitsa, A.V., Gusev, S.A., Karuzina, I.I., Archakov, A.I., and Koymans, L. 2001. Cytochrome P450 database. *Sar Qsar Environ Res* **12**: 359-366.

- Lowe, S.E., and Zeikus, J.G. 1992. Purification and Characterization of Pyruvate Decarboxylase from *Sarcina-Ventriculi*. *J Gen Microbiol* **138**: 803-807.
- Lu, G.G., Dobritzsch, D., Baumann, S., Schneider, G., and Konig, S. 2000. The structural basis of substrate activation in yeast pyruvate decarboxylase - A crystallographic and kinetic study. *European Journal of Biochemistry* **267**: 861-868.
- Lu, G.G., Dobritzsch, D., Konig, S., and Schneider, G. 1997. Novel tetramer assembly of pyruvate decarboxylase from brewer's yeast observed in a new crystal form. *Febs Lett* **403**: 249-253.
- MacKintosh, R.W., and Fewson, C.A. 1988. Benzyl alcohol dehydrogenase and benzaldehyde dehydrogenase II from *Acinetobacter calcoaceticus*. Purification and preliminary characterization. *Biochem J* **250**: 743-751.
- May, O., Nguyen, P.T., and Arnold, F.H. 2000. Inverting enantioselectivity by directed evolution of hydantoinase for improved production of L-methionine. *Nat Biotechnol* **18**: 317-320.
- Meijers, R., Morris, R.J., Adolph, H.W., Merli, A., Lamzin, V.S., and Cedergren-Zeppezauer, E.S. 2001. On the enzymatic activation of NADH. *J Biol Chem* **276**: 9316-9321.
- Mestres, J. 2005. Structure conservation in cytochromes P450. *Proteins* **58**: 596-609.
- Montellano, O.d. 1995. *Cytochrome P450: structure, mechanism and biochemistry*. Plenum Press, New York.
- Mosbacher, T.G., Mueller, M., and Schulz, G.E. 2005. Structure and mechanism of the ThDP-dependent benzaldehyde lyase from *Pseudomonas fluorescens*. *Febs J* **272**: 6067-6076.
- Mulder, N.J., Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Binns, D., Bork, P., Buillard, V., Cerutti, L., Copley, R., et al. 2007. New developments in the InterPro database. *Nucleic Acids Res* **35**: D224-228.
- Murshudov, G.N., Vagin, A.A., and Dodson, E.J. 1997. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D* **53**: 240-255.
- Murzin, A.G., Brenner, S.E., Hubbard, T., and Chothia, C. 1995. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol* **247**: 536-540.
- Nakamura, K., and Matsuda, T. 1998. Asymmetric reduction of ketones by the acetone powder of *Geotrichum candidum*. *J Org Chem* **63**: 8957-8964.
- Nakamura, K., Yamanaka, R., Matsuda, T., and Harada, T. 2003. Recent developments in asymmetric reduction of ketones with biocatalysts. *Tetrahedron-Asymmetry* **14**: 2659-2681.

- Neale, A.D., Scopes, R.K., Wettenhall, R.E.H., and Hoogenraad, N.J. 1987. Pyruvate Decarboxylase of *Zymomonas-Mobilis* - Isolation, Properties, and Genetic Expression in *Escherichia-Coli*. *Journal of Bacteriology* **169**: 1024-1028.
- Nelson, D.R. 2002. Mining databases for cytochrome P450 genes. *Methods Enzymol* **357**: 3-15.
- Nelson, D.R. 2006. Cytochrome P450 nomenclature, 2004. *Methods Mol Biol* **320**: 1-10.
- Neuberg, C., and Hirsch, J. 1921. Über ein Kohlenstoffketten knüpfendes Ferment (Carboligase). *Biochem. Zeitschrift* **115**: 282.
- Nicolaou, K.C., Yang, Z., Liu, J.J., Ueno, H., Nantermet, P.G., Guy, R.K., Claiborne, C.F., Renaud, J., Couladouros, E.A., Paulvannan, K., et al. 1994. Total Synthesis of Taxol. *Nature* **367**: 630-634.
- Nordling, E., Jörnvall, H., and Persson, B. 2002. Medium-chain dehydrogenases/reductases (MDR). Family characterizations including genome comparisons and active site modeling. *Eur J Biochem* **269**: 4267-4276.
- Norin, A., Van Ophem, P.W., Piersma, S.R., Persson, B., Duine, J.A., and Jörnvall, H. 1997. Mycothiol-dependent formaldehyde dehydrogenase, a prokaryotic medium-chain dehydrogenase/reductase, phylogenetically links different eukaryotic alcohol dehydrogenases--primary structure, conformational modelling and functional correlations. *Eur J Biochem* **248**: 282-289.
- Okamoto, T., Taguchi, H., Nakamura, K., Ikenaga, H., Kuraishi, H., and Yamasato, K. 1993. *Zymobacter palmae* gen. nov., sp. nov., a new ethanol-fermenting peritrichous bacterium isolated from palm sap. *Arch Microbiol* **160**: 333-337.
- Orengo, C.A., Michie, A.D., Jones, S., Jones, D.T., Swindells, M.B., and Thornton, J.M. 1997. CATH--a hierarchic classification of protein domain structures. *Structure* **5**: 1093-1108.
- Pei, J., and Grishin, N.V. 2001. AL2CO: calculation of positional conservation in a protein sequence alignment. *Bioinformatics* **17**: 700-712.
- Persson, B., Zigler, J.S., Jr., and Jörnvall, H. 1994. A super-family of medium-chain dehydrogenases/reductases (MDR). Sub-lines including zeta-crystallin, alcohol and polyol dehydrogenases, quinone oxidoreductase enoyl reductases, VAT-1 and other proteins. *Eur J Biochem* **226**: 15-22.
- Pleiss, J., Fischer, M., Peiker, M., Thiele, C., and Schmid, R.D. 2000. Lipase engineering database - Understanding and exploiting sequence-structure-function relationships. *J Mol Catal B-Enzym* **10**: 491-508.
- Pohl, M. 1997. Protein design on pyruvate decarboxylase (PDC) by site-directed mutagenesis. Application to mechanistical investigations, and tailoring PDC for the use in organic synthesis. *Adv Biochem Eng Biotechnol* **58**: 15-43.

- Pohl, M., Grotzinger, J., Wollmer, A., and Kula, M.R. 1994. Reversible Dissociation and Unfolding of Pyruvate Decarboxylase from *Zymomonas-Mobilis*. *European Journal of Biochemistry* **224**: 651-661.
- Pohl, M., Lingen, B., and Muller, M. 2002. Thiamin-diphosphate-dependent enzymes: new aspects of asymmetric C-C bond formation. *Chemistry* **8**: 5288-5295.
- Pohl, M., Mesch, K., Rodenbrock, A., and Kula, M.R. 1995. Stability Investigations on the Pyruvate Decarboxylase from *Zymomonas-Mobilis*. *Biotechnol Appl Bioc* **22**: 95-105.
- Pohl, M., Siegert, P., Mesch, K., Bruhn, H., and Grotzinger, J. 1998. Active site mutants of pyruvate decarboxylase from *Zymomonas mobilis*--a site-directed mutagenesis study of L112, I472, I476, E473, and N482. *Eur J Biochem* **257**: 538-546.
- Pohl, M., Sprenger, G.A., and Muller, M. 2004. A new perspective on thiamine catalysis. *Curr Opin Biotechnol* **15**: 335-342.
- Polovnikova, E.S., McLeish, M.J., Sergienko, E.A., Burgner, J.T., Anderson, N.L., Bera, A.K., Jordan, F., Kenyon, G.L., and Hasson, M.S. 2003. Structural and kinetic analysis of catalysis by a thiamin diphosphate-dependent enzyme, benzoylformate decarboxylase. *Biochemistry-Us* **42**: 1820-1830.
- Raj, K.C., Ingram, L.O., and Maupin-Furlow, J.A. 2001. Pyruvate decarboxylase: a key enzyme for the oxidative metabolism of lactic acid by *Acetobacter pasteurianus*. *Archives of Microbiology* **176**: 443-451.
- Raj, K.C., Talarico, L.A., Ingram, L.O., and Maupin-Furlow, J.A. 2002. Cloning and characterization of the *Zymobacter palmae* pyruvate decarboxylase gene (*pdc*) and comparison to bacterial homologues. *Appl Environ Microb* **68**: 2869-2876.
- Ramaswamy, S., Eklund, H., and Plapp, B.V. 1994. Structures of Horse Liver Alcohol-Dehydrogenase Complexed with Nad(+) and Substituted Benzyl Alcohols. *Biochemistry-Us* **33**: 5230-5237.
- Raucy, J.L., and Allen, S.W. 2001. Recent advances in P450 research. *Pharmacogenomics J* **1**: 178-186.
- Reetz, M.T., and Jaeger, K.E. 2000. Enantioselective enzymes for organic synthesis created by directed evolution. *Chem-Eur J* **6**: 407-412.
- Reid, M.F., and Fewson, C.A. 1994. Molecular Characterization of Microbial Alcohol Dehydrogenases. *Crit Rev Microbiol* **20**: 13-56.
- Rice, P., Longden, I., and Bleasby, A. 2000. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* **16**: 276-277.
- Riveros-Rosas, H., Julian-Sanchez, A., Villalobos-Molina, R., Pardo, J.P., and Pina, E. 2003. Diversity, taxonomy and evolution of medium-chain dehydrogenase/reductase superfamily. *Eur J Biochem* **270**: 3309-3334.

- Sanchez-Gonzalez, M., and Rosazza, J.P.N. 2003. Mixed aromatic acyloin condensations with recombinant benzaldehyde lyase: Synthesis of alpha-hydroxydihydrochalcones and related alpha-hydroxy ketones. *Adv Synth Catal* **345**: 819-824.
- Scheib, H., Pleiss, J., Stadler, P., Kovac, A., Potthoff, A.P., Haalck, L., Spener, F., Paltauf, F., and Schmid, R.D. 1998. Rational design of *Rhizopus oryzae* lipase with modified stereoselectivity toward triacylglycerols. *Protein Eng* **11**: 675-682.
- Schellenberger, A. 1998. Sixty years of thiamin diphosphate biochemistry. *Bba-Protein Struct M* **1385**: 177-186.
- Schmid, R.D. 2002. *Taschenatlas der Biotechnologie und Gentechnik*. Wiley-VCH.
- Schomburg, I., Chang, A., Ebeling, C., Gremse, M., Heldt, C., Huhn, G., and Schomburg, D. 2004. BRENDA, the enzyme database: updates and major new developments. *Nucleic Acids Res* **32**: D431-433.
- Schwede, T., Diemand, A., Guex, N., and Peitsch, M.C. 2000. Protein structure computing in the genomic era. *Res Microbiol* **151**: 107-112.
- Servant, F., Bru, C., Carrere, S., Courcelle, E., Gouzy, J., Peyruc, D., and Kahn, D. 2002. ProDom: automated clustering of homologous domains. *Brief Bioinform* **3**: 246-251.
- Shafqat, J., Hoog, J.O., Hjelmqvist, L., Oppermann, U.C., Ibanez, C., and Jörnvall, H. 1999. An ethanol-inducible MDR ethanol dehydrogenase/acetaldehyde reductase in *Escherichia coli*: structural and enzymatic relationships to the eukaryotic protein forms. *Eur J Biochem* **263**: 305-311.
- Shukla, V.B., and Kulkarni, P.R. 2000. L-Phenylacetylcarbinol (L-PAC): biosynthesis and industrial applications. *World J Microb Biot* **16**: 499-506.
- Siegert, P. 2000. Vergleichende Charakterisierung der Decarboxylase- und Carboligasereaktion der Benzoylformiatdecarboxylase aus *Pseudomonas putida* und der Pyruvatdecarboxylase aus *Zymomonas mobilis* mittels gerichteter Mutagenese - doctoral thesis. Heinrich-Heine University (Düsseldorf).
- Siegert, P., McLeish, M.J., Baumann, M., Iding, H., Kneen, M.M., Kenyon, G.L., and Pohl, M. 2005. Exchanging the substrate specificities of pyruvate decarboxylase from *Zymomonas mobilis* and benzoylformate decarboxylase from *Pseudomonas putida*. *Protein Engineering Design & Selection* **18**: 345-357.
- Stermitz, F.R., Lorenz, P., Tawara, J.N., Zenewicz, L.A., and Lewis, K. 2000. Synergy in a medicinal plant: Antimicrobial action of berberine potentiated by 5'-methoxyhydrnocarpin, a multidrug pump inhibitor. *P Natl Acad Sci USA* **97**: 1433-1437.
- Stillger, T., Pohl, M., Wandrey, C., and Liese, A. 2006. Reaction engineering of benzaldehyde lyase from *Pseudomonas fluorescens* catalyzing enantioselective C-C bond formation. *Org Process Res Dev* **10**: 1172-1177.

- Straathof, A.J., Panke, S., and Schmid, A. 2002. The production of fine chemicals by biotransformations. *Curr Opin Biotechnol* **13**: 548-556.
- Talarico, L.A., Ingram, L.O., and Maupin-Furlow, J.A. 2001. Production of the Gram-positive *Sarcina ventriculi* pyruvate decarboxylase in *Escherichia coli*. *Microbiol-Sgm* **147**: 2425-2435.
- Thompson, J.D., Higgins, D.G., and Gibson, T.J. 1994. 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**: 4673-4680.
- Thornton, J.M., Todd, A.E., Milburn, D., Borkakoti, N., and Orengo, C.A. 2000. From structure to function: approaches and limitations. *Nat Struct Biol* **7 Suppl**: 991-994.
- Tittmann, K., Golbik, R., Uhlemann, K., Khailova, L., Schneider, G., Patel, M., Jordan, F., Chipman, D.M., Duggleby, R.G., and Hubner, G. 2003. NMR analysis of covalent intermediates in thiamin diphosphate enzymes. *Biochemistry-Us* **42**: 7885-7891.
- Tittmann, K., Vyazmensky, M., Hubner, G., Barak, Z., and Chipman, D.M. 2005. The carboligation reaction of acetohydroxyacid synthase II: Steady-state intermediate distributions in wild type and mutants by NMR. *P Natl Acad Sci USA* **102**: 553-558.
- Tsou, A.Y., Ransom, S.C., Gerlt, J.A., Buechter, D.D., Babbitt, P.C., and Kenyon, G.L. 1990. Mandelate Pathway of *Pseudomonas-Putida* - Sequence Relationships Involving Mandelate Racemase, (S)-Mandelate Dehydrogenase, and Benzoylformate Decarboxylase and Expression of Benzoylformate Decarboxylase in *Escherichia-Coli*. *Biochemistry-Us* **29**: 9856-9862.
- Vagin, A., and Teplyakov, A. 1997. MOLREP: an automated program for molecular replacement. *J. Appl. Crystallogr.* **30**: 1022-1025.
- Weiss, P.M., Garcia, G.A., Kenyon, G.L., Cleland, W.W., and Cook, P.F. 1988. Kinetics and Mechanism of Benzoylformate Decarboxylase Using C-13 and Solvent Deuterium-Isotope Effects on Benzoylformate and Benzoylformate Analogs. *Biochemistry-Us* **27**: 2197-2205.
- Wilcocks, R., Ward, O.P., Collins, S., Dewdney, N.J., Hong, Y., and Prosen, E. 1992. Acyloin formation by benzoylformate decarboxylase from *Pseudomonas putida*. *Appl Environ Microbiol* **58**: 1699-1704.
- Wolfson, H.J., Shatsky, M., Schneidman-Duhovny, D., Dror, O., Shulman-Peleg, A., Ma, B., and Nussinov, R. 2005. From structure to function: methods and applications. *Curr Protein Pept Sci* **6**: 171-183.
- Wu, C.H., Apweiler, R., Bairoch, A., Natale, D.A., Barker, W.C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., et al. 2006. The Universal Protein Resource (UniProt): an expanding universe of protein information. *Nucleic Acids Res* **34**: D187-191.

- Xie, P.T., and Hurley, T.D. 1999. Methionine-141 directly influences the binding of 4-methylpyrazole in human sigma sigma alcohol dehydrogenase. *Protein Sci* **8**: 2639-2644.
- Yano, T., Oue, S., and Kagamiyama, H. 1998. Directed evolution of an aspartate aminotransferase with new substrate specificities. *Proc Natl Acad Sci U S A* **95**: 5511-5515.
- Yep, A., Kenyon, G.L., and McLeish, M.J. 2006. Determinants of substrate specificity in KdcA, a thiamin diphosphate-dependent decarboxylase. *Bioorg Chem* **34**: 325-336.
- Yona, G., Linial, N., and Linial, M. 2000. ProtoMap: automatic classification of protein sequences and hierarchy of protein families. *Nucleic Acids Res* **28**: 49-55.
- Ziegelmann-Fjeld, K.I., Musa, M.M., Phillips, R.S., Zeikus, J.G., and Vieille, C. 2007. A *Thermoanaerobacter ethanolicus* secondary alcohol dehydrogenase mutant derivative highly active and stereoselective on phenylacetone and benzylacetone. *Protein Eng Des Sel* **20**: 47-55.

Danksagung

Bei Herrn Prof. Dr. Rolf D. Schmid möchte ich mich für die Möglichkeit, die vorliegende Arbeit an seinem Institut unter hervorragenden Arbeitsbedingungen anfertigen zu können, bedanken. Die Tätigkeiten am Institut für Technische Biochemie ermöglichten es mir, über die wissenschaftliche Arbeit hinaus, sehr viel für meinen weiteren Lebensweg zu lernen.

Besonderer Dank gilt Herrn Prof. Dr. Jürgen Pleiss für die Überlassung des Themas, die wissenschaftliche Betreuung und die stetige Diskussionsbereitschaft sowie für den Gestaltungsfreiraum, den er mir bei der Bearbeitung des Themas überlassen hat. Bedanken möchte ich mich für die Möglichkeit der Mitwirkung an zahlreichen spannenden Projekten und die damit verbundenen internationalen Kooperationen sowie für die mehrmalige Entsendung nach Vietnam, wodurch ich dieses schöne Land kennen lernen durfte.

Herrn Prof. Dr. Georg Sprenger danke ich für die Bereitschaft den Prüfungsvorsitz zu übernehmen.

Für die erfolgreiche und freundliche Zusammenarbeit auf dem Gebiet der Thiamindiphosphat-abhängigen Enzyme bin ich Apl. Prof. Dr. Martina Pohl und Dr. Dörte Gocke vom Institut für Molekulare Enzymtechnologie der Heinrich-Heine Universität Düsseldorf sehr dankbar.

Den Mitgliedern der Bioinformatikgruppe gilt mein Dank für ein mehr als angenehmes Arbeitsklima. Besonders zu erwähnen sind dabei die stets geselligen und unvergessenen Unternehmungen des Ältestenrats mit Dr. Fabian Bös, Dr. Stephan Tatzel, Dr. Peter Trodler und Dipl. Biol. (t.o.) Alexander Steudle.

Allen Mitarbeitern des Instituts für Technische Biochemie sei für die offene und freundliche Zusammenarbeit gedankt. Bedanken möchte ich mich besonders bei Dr. Markus Fischer für die geduldige Einführung in die Welt der biologischen Datenbanken und seine grenzenlose Hilfsbereitschaft bei Problemen aller Art sowie bei Sven Richter für so manches gesellige Gespräch fern ab jeglicher wissenschaftlicher Themen im *“front office“*.

An dieser Stelle will ich mich auch bei den zahlreichen wissenschaftlichen Hilfskräften und Praktikanten, die mich während meiner Promotion bei so mancher Auswertung tatkräftig

unterstützt haben, bedanken. Besonderer Dank geht dabei an Florian Wagner für die hervorragende Zusammenarbeit und so manche durchprogrammierte Nacht im Dienste der biologischen Datenbanken.

Schließlich bin ich meinen Eltern für ihre unermüdliche Unterstützung, nicht nur finanzieller Art, während meines Studiums und meiner Promotion besonders dankbar, ohne diese die vorliegende Arbeit in dieser Form nicht möglich gewesen wäre. Ihnen sowie meiner „Schwiegerfamilie“ und allen Freunden danke ich für ihren Rückhalt in den letzten Jahren und dafür, dass alle stets an mich geglaubt haben.

Meiner Freundin und Lebensgefährtin Stephanie danke ich herzlich für ihr grenzenloses Verständnis für die zahlreichen vor dem Computer verbrachten Nächte, jedoch besonders für die liebevolle Unterstützung in den letzten Jahren und die gemeinsame schöne Zeit, nun zu dritt mit Ajoscha Simeon.

Erklärung

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

Stuttgart, Januar 2008