

DISS. ETH NO. 19896

**Directed evolution as a versatile tool to investigate natural and
artificial aldolases**

A dissertation submitted to

ETH ZURICH

for the degree of

DOCTOR OF SCIENCES

presented by

LARS GIGER

Master of Science in Chemistry, ETH Zurich

born July 9, 1981

citizen of

Busserach SO, Switzerland

Accepted on the recommendation of:

Prof. Dr. D. Hilvert, examiner

Prof. Dr. J. Bode, co-examiner

2011

Abstract

Enzymes are the most proficient catalysts we know and the number of processes that take advantage of their enormous rate accelerations is constantly growing. The precision achieved by the tightly controlled conversion of a substrate into a desired product has fundamental importance in the synthesis of important medical and pharmaceutical compounds. For such applications, the wrong isomer can have lethal physiological consequences. However, the high selectivity of natural enzymes comes at the price of a narrow substrate range and thereby limits their broad utility as biocatalysts. Furthermore, the complexity of their mode of action hampers rational reengineering. As described in Chapter 1, designing artificial biocatalyst based on our current understanding of structure-function relationships is a long-standing scientific dream, which is slowly being realized through the development of sophisticated computer algorithms. In this thesis, focusing on natural and artificial aldolases as model systems, biochemical tools were developed to gain further insights into enzyme catalysis. Directed evolution, in particular, was used to study mechanism, tailor enzyme function, and amplify modest initial activities.

In Chapter 2 we took advantage of the key role played by pyridoxal-5' phosphate (PLP)-dependent enzymes in amino acid metabolism to develop a novel genetic selection system for PLP-dependent aldolases. The simultaneous inactivation of three genes involved in the biosynthesis of glycine enabled growth-based selection experiments with β -hydroxy amino acids as sole carbon source. We showed that expression of a gene encoding a previously uncharacterized L-threonine-aldolase from *Caulobacter crescentus* CB15 (Cc-LTA) using the tunable tetracycline promoter restores glycine prototrophy by converting β -hydroxy- α -amino acids provided in the minimal medium to glycine and an aldehyde. This system was used to test the hypothesis that a conserved Glu-His dyad, identified in a multiple sequence alignment of more than 100 known and putative threonine aldolases, might facilitate proton abstraction from the substrate β -hydroxyl group. Subsequent selection experiments on L-threonine and L-*allo*-threonine as sole carbon source demonstrated that histidine is invariable at its position, but the glutamate is not an essential interaction partner for catalysis. The glutamate may orient the histidine for catalysis but it is not required for activation.

Such PLP-dependent aldolases have attracted growing interest in recent years since they provide a direct synthetic route to β -hydroxy- α -amino acids, which can serve as precursors for pharmaceutical and agrochemical fine chemicals. This structural motif is broadly found in natural antibiotics such as thiamphenicol or chloramphenicol, as well as immunosuppressants, HIV-inhibitors and other compounds. We therefore evaluated the synthetic potential of a reengineered PLP-dependent alanine racemase from *Bacillus stearothermophilus*, which contains a single point mutation (Bs-AR-Y265A) that results in the loss of most starting isomerase activity but in the gain of novel aldolase activity (Chapter 3). A broad range of aromatic acceptor aldehydes was tested in combination with glycine as the donor substrate for the stereoselective synthesis of β -hydroxy- α -amino acids. The desired products were obtained with complete stereochemical control at C_α (ee > 99%, D) and moderate to high selectivity at C_β (up to 97% de). Moreover, the additional room around the α -carbon allows Bs-AR-Y265A to accept alternative donor substrates. The activation of D-alanine as substrate enables this artificial threonine aldolase to produce an expanded scope of biologically active building blocks.

Chapters 4 to 6 explore the potential of three different retro-aldolases that were designed by the Rosetta algorithm. The possibility of designing tailor made enzymes *in silico* for a desired function or substrate is of fundamental academic and industrial interest since such catalysts can serve as synthetic tools for preparing densely functionalized building blocks. However, our incomplete understanding of enzyme catalysis makes the design of minimalistic active sites extremely challenging. Small variations in the position of catalytic and binding residues relative to the transition state lead to substantially lowered catalytic efficiencies. These low activities have indeed been observed for computationally designed enzymes. However, these catalysts can be improved by site-directed mutagenesis at the active site and, more generally, through directed evolution. We chose three retro-aldolase designs for detailed studies. RA45 and RA95 were designed in the same $(\beta\alpha)_8$ -barrel scaffold but have different active site compositions. RA110, on the other hand, was designed into the structurally distinct NTF2-fold of ketosteroid isomerase (KSI). We were interested in how these three designs might evolve under identical laboratory conditions and what impact the introduced mutations would have on activity.

Chapter 4 describes the optimization of RA45, one of the first Rosetta designs. It catalyzes the retro-aldol cleavage of (\pm)-methanol to 6-methoxy-2-naphthaldehyde and

acetone using the mechanism of a Class I aldolase with a rate acceleration of more than 6000-fold over background. The minimalistic active site comprises a catalytic lysine embedded in a hydrophobic environment and a glutamate to stabilize a catalytic water molecule. Using site directed mutagenesis of active site residues, the initial RA45.0 design was improved 15-fold to give a $k_{\text{cat}}/K_m = 32 \text{ M}^{-1} \text{ min}^{-1}$. We subsequently optimized RA45.2 by error-prone PCR and DNA shuffling using a robust fluorescence-based assay to identify improved variants in crude cell lysates. The best clone obtained after ten rounds of Darwinian evolution (RA45.2-10) had accumulated 17 mutations and showed a 100-fold higher catalytic efficiency than its progenitor. The final k_{cat} and k_{cat}/K_m values of 0.22 min^{-1} and $2,800 \text{ M}^{-1} \text{ min}^{-1}$, respectively, make the evolved catalyst only ten-fold less active than the commercially available catalytic antibody 38C2. Moreover, over the course of evolution, the initial poor solubility of RA45.0 ($1\text{-}2 \text{ mg L}^{-1}$) improved by more than a factor of 20, highlighting the strength of directed evolution as a tool for simultaneously improving multiple parameters.

The activity of a second-generation retro-aldolase RA95 was improved more than 60-fold by site-directed mutagenesis of active site residues. During this process, a second lysine was introduced to further shift the pK_a of the catalytic lysine, a mutation that boosted activity roughly five-fold. The resulting clone, RA95.5, was subjected to eight rounds of directed evolution, which increased catalytic efficiency a further 100-fold. In contrast to RA45, the best variant of RA95, RA95.5-8, has a k_{cat} value of 21.5 min^{-1} and a k_{cat}/K_m of $97,700 \text{ M}^{-1} \text{ min}^{-1}$. It achieves a rate acceleration of 5×10^7 over background, which is one order of magnitude higher than the commercially available catalytic antibody 38C2. In addition, laboratory evolution turned the slight (S)-enantiomer selectivity of the starting catalyst into a greater than 14:1 preference for the (R)-methodol.

To monitor the structural impact of these improvements, the crystal structures of the original design, the optimized RA95.5 intermediate, and a variant from the fifth round of evolution (RA95.5-5) were solved in collaboration with Prof. Nenan Ban (Institute of Molecular Biology, ETHZ) as complexes with a mechanism-based inhibitor. Analysis of the crystal structure of the starting RA95.0 design demonstrated that the inhibitor binds to the catalytic lysine in the predicted active site pocket. Moreover, the experimental structure agrees for most parts of the protein with the model structure predicted by Rosetta. However, significant conformational changes were observed in three loops that make

contact with the ligand, resulting in rotation of the ligand around its long axis and orientation of the polar carbonyl group away from the “catalytic” glutamate that was supposed to mediate proton transfers during catalysis via an ordered water molecule. Surprisingly, the crystal structure of the optimized RA95.5 revealed a second, alternative reaction center involving the lysine that had been introduced during the optimization procedure to shift the pK_a of the catalytic lysine. Subsequent rounds of directed evolution favored this novel reaction site at the expense of the original lysine, which lost its catalytic function. The large deviations found in loop regions close to the active site suggest that these dynamic elements limit the accuracy of the computation design process. However, the 7,000-fold enhanced catalytic efficiency obtained after the experimental refinement stresses the complementary strength of Darwinian evolution to locate beneficial mutations remote from the active site and thereby inaccessible to the design algorithms. The combination of both tools is a promising way to gain additional insights into enzyme evolution. Feeding this information back into the design process may eventually allow a continual refinement of the Rosetta algorithms.

The last design studied, RA110, is housed in the much smaller KSI scaffold. In contrast to the TIM-barrel designs, it showed the largest improvements (80-fold) during the binding pocket optimization process (RA110.4). In the subsequent six rounds of laboratory evolution, 8 mutations accumulated, predominantly on the protein surface, which improved catalytic efficiency only 10-fold. The best variant obtained in these experiments, RA110.4-6, is similar to RA45.2-10 with respect to activity. Its structure in complex with the mechanism-based inhibitor reveals that the ligand binds completely differently than predicted; the inhibitor is rotated roughly 90° around its vertical axis placing the naphthyl group in an alternative hydrophobic pocket. This new binding mode prompted investigation of alternative, sterically more demanding substrate analogs to exploit interactions with the original KSI binding pocket. We detected burst phase kinetics for a broad range of substrate analogs and their fitted structures were also in agreement with the experimentally derived active site of RA110.4-6. However, we did not observe a simple correlation between electron donating or withdrawing substituents and k_{cat} , K_m or k_{cat}/K_m . Nevertheless, these analogs may serve as interesting alternative substrates for further rounds of evolution in the future.

Our current understanding of how enzymes achieve enormous rate accelerations remains incomplete. The methods developed in this thesis may help to investigate the mechanisms of existing enzymes or install novel activities in otherwise inert scaffolds. Computation-based approaches, like the presented Rosetta design protocol, can successfully predict novel active sites, but resulting activities are typically many orders of magnitude below those of natural enzymes. Using three *de novo* designed retro-aldolases we showed that these low initial activities can be readily improved by directed evolution and yield, in the best cases, catalysts that are superior to state-of-the-art catalytic antibodies. To finally reach the catalytic efficiency of natural aldolases, however, it will likely be necessary to screen larger numbers of clones to identify additional function groups and to refine the designed active sites for optimal catalysis. Promising new approaches that could fulfill these requirements are becoming increasingly available and will simultaneously enable the parallel evolution of many more designs. Detailed analysis of these improved clones will enhance our understanding of structure-function relationships in enzymes and also further mature the design process.

Zusammenfassung

Enzyme sind die effizientesten Katalysatoren die wir kennen. Sie finden wachsende Anerkennung in industriellen Prozessen, die sich die enorme Beschleunigung der katalysierten Reaktion zunutze machen. Die Präzision, mit der die einzelnen Bausteine umgewandelt werden, hat eine große Bedeutung bei der Herstellung medizinischer und pharmazeutischer Produkte, bei welchen das falsche Isomer fatale physiologische Wirkungen haben kann. Die hohe Selektivität der Enzyme kommt jedoch mit dem Preis der Substrat-Exklusivität, womit eine breite Anwendung dieser Biokatalysatoren stark eingeschränkt wird. Die Komplexität der Wirkungsweise von Enzymen macht das gezielte Designen bzw. Umdesignen dieser schwierig. Wie im Kapitel 1 beschrieben, ist es ein bereits altes Ziel der Forschung, ausgehend von bereits gewonnenen Erkenntnissen und Bauplänen natürlicher Enzyme, ein künstliches bzw. ein in der Natur nicht vorhandenes Enzym zu kreieren. In dieser Arbeit wurden einerseits Tools entwickelt (Screening und Selektion) welche dazu beitragen sollen, die Wirkungsweise von Enzymen, im speziellen die der Aldolasen, besser verstehen zu lernen und andererseits mittels gerichteter Evolution nach Wunsch zu verbessern.

In Kapitel 2 machen wir uns die Schlüsselposition von Pyridoxal-5'-phosphat (PLP)-abhängigen Enzymen im Aminosäuren-Metabolismus zunutze, um ein *in vivo* Selektionssystem für PLP-abhängige Aldolasen zu entwickeln. Selektives Inaktivieren von drei Enzymen, welche die Glycin-Versorgung der Zelle regeln, macht den Selektionsstamm direkt von einer bioorthogonalen Kohlenstoffquelle abhängig. Wir konnten am Beispiel der kontrollierten Expression einer noch nicht charakterisierten L-Threonin-Aldolase von *Caulobacter crescentus* CB15 (Cc-LTA) mittels regulierbarem Tetracyclin-Promoter zeigen, dass die Versorgungslücke von Glycin durch eine funktionelle Aldolase kompensiert werden kann, sofern diese die β -Hydroxy-Aminosäuren im Medium via Retro-Aldol Reaktion zu Glycin und einem Aldehyd umwandeln kann. Eine umfangreiche Sequenzanalyse von über 100 bekannten bzw. vermuteten Threonin-Aldolasen haben zu der Hypothese geführt, dass PLP-abhängige Aldolasen möglicherweise eine katalytische Dyade (Glu-His Dyade) zur Abstraktion des Protons verwenden. Selektionsexperimente auf L-Threonin und L-*allo*-Threonin als einzige Kohlenstoff-Quelle haben gezeigt, dass Histidin zwar in seiner Position unverzichtbar ist, aber nicht zwingend ein Glutamat als Wechselwirkungspartner benötigt,

um eventuell die Funktion einer Base auszuüben. Vermutlich spielt diese zweite Position mehr eine unterstützende Rolle zur Orientierung des Histidin-Rests als zu dessen Aktivierung.

PLP-abhängige Aldolasen haben in den vergangenen Jahren ein deutliches Interesse als Biokatalysatoren geweckt, da sie einen direkten Zugang zu pharmazeutisch und agrochemisch bedeutenden Bausteinen bieten. Vorstufen zu Antibiotika wie Thiamphenicol oder Chloramphenicol aber auch Immunsuppressiva, HIV-Inhibitoren usw. sind mit dieser Klasse von Enzymen stereoselektiv zugänglich. Wir haben daher die synthetischen Möglichkeiten einer *de novo* designten Aldolase untersucht, welche durch eine einzige Punktmutation in der aktiven Tasche von einer Isomerase zu einer Aldolase (*Bacillus stearothermophilus* Alanin-Racemase zu Bs-AR-Y265A) umfunktioniert werden konnte (Kapitel 3). Die Eigenschaften dieser Designer-Aldolase zur stereoselektiven Synthese von β -Hydroxy-Aminosäuren wurden durch eine breite Palette von Substraten mit aromatischen Aldehyden als Akzeptor und Glycin bzw. Alanin als Donor untersucht. Die detektierten Produkte zeigen hohe Kontrolle der Stereochemie an C_α (ee > 99%, D) als auch an C_β (bis zu 97% de) und untermauert das gemeinsame Erbe von Alanin Racemasen und natürlichen D-Threonin Aldolasen mit vergleichbaren Stereoselektivitäten und Substratspezifitäten. Interessanterweise erlaubt der zusätzliche Raum um C_α , welcher die Selektivität nicht beeinträchtigt, die Verwendung alternativen Donor-Substrate. Bs-AR-Y265A ermöglicht durch die Aktivierung von Alanin als Substrat eine Erweiterung der Substratpalette zu biologisch aktiven Bausteinen.

Die Kapitel 4 bis 6 befassen sich mit drei unterschiedlichen *in silico* designten Retro-Aldolasen. Das Potential, Enzyme für eine gewünschte Reaktion bzw. Substrate am Computer zu designen, ist von fundamentalem akademischen aber auch praktischen Interesse. Die Möglichkeiten, *de novo* Enzyme als synthetische Werkzeuge für hochfunktionalisierte pharmazeutische Wirkstoffe einzusetzen, unterstreicht die Wichtigkeit dieser zukunftsweisenden Methode. Jedoch ist unser Verständnis von Enzymkatalyse ausschlaggebend für den Erfolg des Computer basierten Designs von Enzymen. Unser Wissen findet eine direkte Umsetzung in der Definition der minimalistischen aktiven Taschen zur Stabilisierung von hochliegenden Energiezuständen, welche unausweichlich während einer Reaktion auftreten. Die moderaten Aktivitäten der jetzigen Designs reflektieren jedoch einerseits das enorme Ausmass der Komplexität des Vorhabens. Andererseits bieten

sich die Möglichkeit mit Hilfe von experimentellen Mitteln, wie die positionsbezogene Optimierung der aktiven Tasche, die primäre Aktivität zu amplifizieren. Wir haben drei unterschiedliche Designs zur detaillierten biochemischen Untersuchung ausgewählt und anschliessend auf deren Potential sich durch gerichtete Evolution verbessern zu lassen untersucht. RA45 und RA95 sind Retro-Aldolase-Designs im identischen TIM-Barrel-Protein, jedoch mit einer unterschiedlichen Konstitution der aktiven Tasche. Auf der anderen Seite wurde das aktive Zentrum von RA110 in das halb so grosse Gerüst der Ketosteroid-Isomerase (KSI) modelliert. Wir wollten untersuchen, ob und wie gut sich alle drei Enzyme unter denselben Bedingungen im Labormassstab evolvieren lassen und wie sie sich dabei verändert haben.

Kapitel 4 beschreibt eines der ersten erfolgreichen Rosetta-Designs vom Jahre 2008, RA45 im TIM-Barrel-Gerüst, welches das zur Berechnung verwendete (\pm)-Methodol mit einer mehr als 6000-fachen Beschleunigung zu 6-Methoxy-2-Naphthaldehyd und Aceton umsetzt. Die minimalistische katalytische Einheit, bestehend aus einem Iminium-Ion formenden Lysin vergraben in einer hydrophoben Tasche und einem Glutamat zur Stabilisierung eines katalytischen Wassers, erreichte nach der gezielten experimentellen Optimierung der bereits eingeführten Resten eine katalytische Effizienz von $32 \text{ M}^{-1} \text{ min}^{-1}$. Das Potential von RA45.2 für weitere Verbesserungen wurde anschliessend durch zufällig eingeführte Mutationen und die Rekombination von RA45-Varianten mit verbessertem Phänotyp in einem robusten Zelllysate-Screening-Assay getestet. Die biochemische Analyse von RA45.2 und der besten Variante nach zehn Runden Darwin'scher Evolution (RA45.2-10) konnte zeigen, dass sich die primäre Aktivität von RA45.2 durch 17 Mutationen um das 100-fache verbessern lies. Mit einer derartigen Steigerung der katalytischen Effizienz von RA45.2-10 zu einem k_{cat} von 0.22 min^{-1} und $k_{\text{cat}}/K_m = 2'800 \text{ M}^{-1} \text{ min}^{-1}$ ergibt nur etwa einen Faktor drei hinter dem k_{cat} des kommerziell erhältlichen katalytischen Antikörpers (38C2). Des Weiteren konnten wir zeigen, dass sich die anfänglich schlechte Löslichkeit des ursprünglichen designten Proteins ($1\text{-}2 \text{ mg L}^{-1}$) um mehr als einen Faktor 20 verbessern lies und damit die Möglichkeiten der gerichteten Evolution unterstreicht.

Die zweite Generation von Retro-Aldolase-Designs, im speziellen RA95, konnte die anfängliche Aktivität in der Routine-Optimierung (11 Mutationen) der aktiven Tasche um das 60-fache gesteigert werden. Das gezielte Einführen eines zweiten Lysins, mit der Absicht den $\text{p}K_a$ -Wert des katalytischen Lysins noch weiter zu senken, steigerte die Aktivität um

einen zusätzlichen Faktor fünf. Acht zusätzliche Mutationen, die in acht Runden von gerichteter Evolution erzielt wurden, bewirkten eine weitere 100-fache Verbesserung der katalytischen Effizienz. Die beste Variante, RA95.5-8, zeigt einen k_{cat} -Wert von 21.5 min^{-1} und einen k_{cat}/K_m -Wert von $97'700 \text{ M}^{-1} \text{ min}^{-1}$. Die Geschwindigkeitsbeschleunigung gegenüber der Hintergrundreaktion beträgt 5×10^7 . Dieser Katalysator übersteigt die Aktivität des katalytischen Antikörpers 38C2 um das 2-fache mit einem 20-fach grösseren k_{cat} und repräsentiert somit momentan die aktivste, nicht-natürliche Aldolase. Während der Evolution veränderte sich auch die anfängliche geringe (S)-Enantiomer Selektivität zu einer 14:1 Bevorzugung des (R)-Enantiomers in RA95.5-8. Unter Zuhilfenahme einer hochauflösenden Kristallstruktur des ursprünglichen Designs hat sich bestätigt, dass ein auf dem Mechanismus basierender Inhibitor, wie vorhergesagt in der simulierten Tasche am Lysin bindet. Überdies konnte eine generell gute Übereinstimmung mit dem Modell für den grössten Teil des Enzymes festgestellt werden. Die grössten Abweichungen sind in Loop-Regionen, die in unmittelbarer Nähe der aktiven Tasche liegen. Wahrscheinlich induzierten diese konformationellen Änderungen die beobachtete Rotation des Liganden um die eigene Achse und verhindern somit eine vorteilhafte Wechselwirkung zwischen dem Glutamat und dem Substrat. Eine Überraschung ergab die Struktur von RA95.5 mit dem zweiten Lysin aus der Optimierung. Im Gegensatz zur beabsichtigten sekundären Rolle etablierte dieses Lysin ein zweites, alternatives Reaktionszentrum neben dem ursprünglich designten. Die nachträglich durch die gerichtete Evolution eingebrachten Mutationen haben die Funktion dieser neuen Tasche weiter verfeinert, während die ursprüngliche ihre Funktion nahezu einbüsste. Die starken Abweichungen der Loop-Regionen vom berechneten Modell betonen die Bedeutsamkeit solcher dynamischer Elemente für die atomare Präzision des Designs in der direkten Umgebung des Liganden. Die über 7'000-fache Aktivitätssteigerung in diesem Beispiel vom ursprünglichen Design zeigen das Potential dieser komplementären Methoden zur Erforschung der Enzymkatalyse als auch der fortlaufende Verbesserung des Design-Prozesses auf.

Im letzten von uns untersuchten Design handelt es sich um RA110, das im Vergleich zu RA45 und RA95 in der nur halb so grossen Ketosteroid-Isomerase (KSI) eingebettet ist. Obwohl die Optimierung der aktiven Tasche in einer mehr als 80-fachen Steigerung der Aktivität resultierte, konnten wir mittel sechs Runden Darwin'scher Evolution und 8 hauptsächlich auf der Oberfläche gefundenen Mutationen nur noch einen Faktor 10

erreichen. Schlussendlich erreichte RA110.4-6 dieselbe katalytische Effizienz wie RA45. Zudem zeigt die Kristallstruktur der evolvierten Variante den Reaktionsmechanismus basierenden Inhibitor nicht in der ursprünglich berechneten Orientierung sondern über 90° vom Modell weggedreht. Die neue, offene Positionierung des Liganden und die noch vorhandene Substratbindungstasche von KSI erlaubte uns die Untersuchung alternativer, räumlich anspruchsvollere Substrate. In der Tat konnte katalytische Aktivität für eine breite Palette von Substraten beobachtet werden, die sich mit der Struktur und Position des Liganden vereinbaren lassen. Obwohl sich keine simple Korrelation zwischen den Substituenten im Substrat (Elektronen spendende und entziehenden Gruppen) und den k_{cat} , K_m - und k_{cat}/K_m -Werten ergaben, könnten sich diese Substrate als wertvolle Alternativen für weitere Runden der Evolution erweisen.

Unser Verständnis, wie Enzyme die gewaltigen Geschwindigkeitsbeschleunigungen der von ihnen katalysierten Reaktion erreichen, ist noch sehr lückenhaft. Die unterschiedlichen Methoden die hier vorgestellt wurden, können dazu beitragen sowohl die mechanistischen Möglichkeiten existierender Enzyme zu erforschen als auch neue Aktivitäten zu kreieren. Computer gestützte Berechnungen, wie der vorgestellte Rosetta Algorithmus, können dabei helfen neue aktive Taschen in einem natürlichen Enzyme vorhersagen, jedoch zur Zeit mit nur eingeschränkter Aktivität. Wir konnten am Beispiel dreier Retro-Aldolasen-Designs zeigen, dass sich diese anfänglichen moderaten Aktivitäten um mehrere Größenordnungen durch gerichtete Evolution steigern lassen und die Aktivitäten von entsprechenden Antikörpern übertreffen. Um diese Aktivität jedoch noch auf das Level von natürlichen Aldolasen zu evolvieren muss höchstwahrscheinlich die Anzahl der gescreenten Varianten vergrößert werden. Solche Methoden sind in der Entwicklung und könnte sowohl bei der Weiterentwicklung des Designprozesses helfen, als auch unserem Verständnis der enzymatischen Katalyse vertiefen.