

Diss. ETH No. 14175

Diversity in Combinatorial Libraries and Evolving Proteins

A Dissertation Submitted
to the

SWISS FEDERAL INSTITUTE OF TECHNOLOGY
ZURICH

for the Degree of
Doctor of Natural Sciences

presented by

Stephan Valentin Audétat

Dipl. Chem. ETH Zürich
born 13. September 1968
from Bern BE, Les Verrières and La Côte-aux-Fées NE

accepted on the recommendation of

Prof. Dr. G. Folkers, examiner
Prof. Dr. S. A. Benner, co-examiner and supervisor

Zürich, 2001

Summary

This dissertation focuses on the use of ideas drawn from biological evolution in the synthesis and analysis of biomolecules. The first part addresses structural diversity, a key to biological evolution, in small organic molecules. Methods for generating diversity in a solution-phase combinatorial library of small molecules are presented, where each molecule in the library can undergo rapid exchange reactions with others to form composite molecules. The goal was to generate a library of ligand fragments suitable for a "Receptor Assisted Combinatorial Synthesis" (RACS) experiment, a new method that combines the synthesis of potential ligands with their screening.

A combinatorial library of ketones was generated using a thiazolium-catalyzed coupling of two aldehydes to α -hydroxyketones. This condensation reaction was optimized for a variety of aliphatic and aromatic aldehydes as starting materials. Several libraries containing more than 25 individual members were synthesized and analyzed by GC-MS.

In a second step, methods to manipulate those α -hydroxyketone libraries were explored. Evidence arose that the carbonyl group and the hydroxyl group in a α -hydroxyketone do not react independently. This excluded a large number of transformations from being potentially useful in generating a RACS library. Two reactions were successfully developed, however, the oxidation to diketones and the reduction to 1,2-diols. These could be applied to generate libraries of ligand fragments. Borate esters of 1,2-diol libraries have been previously explored for RACS experiments, but difficulties with the analytical method (FTICR-MS) had arisen. A successful implementation of these envisaged RACS experiments awaits new tools for the analysis of complex mixtures.

In the second part of the dissertation, the protein family of phospholipase A₂ (PLA₂) was investigated from an evolutionary and computational / bioinformatics perspective. Phospholipase plays a crucial role in inflammatory processes and therefore comprises an important target for the development of novel drugs. In contrast to earlier work concerning the evolution of the PLA₂ family using protein sequence data, genetic sequence information of the phospholipases was used for various experiments. The finding of Davidson and Dennis, that the two major classes of venomous snakes, the

vipers and the cobras, generated venom independently through the independent recruitment of different homologues of PLA2 genes, was confirmed. Neutral evolutionary distances (NEDs) allowed the dating of events within the molecular record of the PLA2 family of proteins. The molecular dates were correlated with palaeontological records, suggesting that the NED method is a useful tool for geobiological dating where the fossil record is sparse.

The PLA2 family also displayed episodes of rapid sequence evolution suggestive of changing function. Analysis of the normalized ratio of non-synonymous to synonymous substitution in the PLA2 genes (the K_a/K_s ratio) indicated several episodes where function was changed. Notably, the K_a/K_s values were substantially higher in the branches leading from the common ancestor to the venom sequences than they are in the branches leading to mammalian sequences. This suggested that a venom protein be recruited from a pancreatic protein, not the other way around.

The identification of which amino acid positions are changing during a period of rapid sequence evolution suggested hypotheses about the importance of specific amino acid positions for the new function of a protein. Methods were developed to distinguish these sites from a background of "neutrally drifting" mutating sites that occur over time without changing the function of the protein. The analysis was performed on the PLA2 family and the results were mapped on an X-ray crystal structure of one member of the PLA2 family as a guide for the visualization in which region of the protein the changes occur.

Zusammenfassung

In der vorliegenden Dissertation wurden Methoden der biologischen Evolution in der Synthese und Analyse von Biomolekülen angewendet. Der erste Teil befasste sich mit struktureller Diversität (ein Schlüsselprinzip in biologischer Evolution) in kleinen organischen Molekülen. Methoden für die Einbeziehung von Diversität in Flüssigphasen-Bibliotheken kleiner Moleküle wurden präsentiert. Einzelne Komponenten der Molekülbibliotheken sollten Austauschreaktionen untereinander eingehen können, um zusammengesetzte Moleküle zu erhalten. Das Ziel war, eine Molekülbibliothek von Ligandbausteinen zu erhalten, die ein "Receptor Assisted Combinatorial Synthesis" (RACS) Experiment ermöglichen. Dies ist eine neue Methode, wo die Synthese von potentiellen Wirkstoffen mit ihrem Screening kombiniert wird.

Eine kombinatorische Bibliothek von α -Hydroxyketonen wurde mittels Kopplung von zwei Aldehyden mit einem Thiazoliumsals-Katalysator synthetisiert. Diese Kondensationsreaktion wurde für die Verwendung von verschiedenen aliphatischen und aromatischen Aldehyden als Edukte optimiert. Mehrere Bibliotheken mit mehr als 25 individuellen Komponenten wurden synthetisiert und im GC-MS analysiert.

In einem zweiten Schritt wurden Methoden für die Manipulation von diesen α -Hydroxyketon-Bibliotheken untersucht. Es stellte sich bald heraus, dass die Carbonylgruppe und die Hydroxylgruppe in einem α -Hydroxyketon nicht unabhängig voneinander reagieren. Dies führte zum Ausschluss von vielen Transformationen, die zu einer für ein RACS Experiment geeigneten Bibliothek geführt hätten. Zwei Reaktionen wurden hingegen erfolgreich angewendet: Die Oxidation zu Diketonen und die Reduktion zu 1,2-Diolen. Diese Reaktionen konnten zur Synthese von Ligandbaustein-Bibliotheken verwendet werden. Boratester von 1,2-Diolbibliotheken waren schon früher für die Verwendung in RACS Experimenten untersucht worden, wobei aber Schwierigkeiten in der Analytik (FTICR-MS) aufgetreten waren.

Im zweiten Teil der Dissertation wurde die Familie der Phospholipase A₂ (PLA₂) Proteine in einer numerisch-bioinformatischen Perspektive untersucht. Phospholipasen spielen eine grosse Rolle in Entzündungsprozessen und sind deshalb wichtige Zielgruppen für die Entwicklung von neuen Medikamenten. Im Gegensatz zu früheren Arbeiten über die Evolutionsgeschichte der PLA₂ Familie, wo Proteinsequenzen

verwendet wurden, wurden in dieser Arbeit Sequenzinformationen von PLA2 Genen zur Durchführung von verschiedenen Experimenten verwendet. Die Hypothese von Davidson und Dennis, dass die beiden Hauptklassen von Giftschlangen (Vipern und Kobras) ihr Gift durch unabhängige Rekrutierung von verschiedenen Homologen von PLA2 Genen generierten, wurde bekräftigt. Die "Neutral Evolutionary Distances" (NED) Methode erlaubte die Datierung von Ereignissen in der Molekulargeschichte der PLA2 Proteinfamilie. Die Datierungen stehen mit paläontologisch bekannten Zeitpunkten im Einklang, was die NED Methode zu einem nützlichen Werkzeug für geobiologische Datierungen macht, wenn nicht genug Fossile zur Datierung zur Verfügung stehen.

Die PLA2 Familie wies Episoden von beschleunigter Sequenzevolution vor, was einen Wechsel in der Funktion der Proteine suggeriert. Die Analyse des normalisierten Verhältnis von nicht-synonymen zu synonymen Substitutionen in den PLA2 Genen (das K_a/K_s Verhältnis) identifizierte mehrere Episoden, wo die Funktion wechselte. Beachtenswert war, dass die K_a/K_s Werte von Ästen, die vom gemeinsamen Vorfahren zu den Giftsequenzen führen, wesentlich höher waren als die von Ästen, die vom gemeinsamen Vorfahren zu den Säugetiersequenzen führen. Daher wurde vorgeschlagen, dass die Proteine für die Giftproduktion von den Pankreasproteinen rekrutiert wurden, und nicht umgekehrt.

Weiter wurde die Hypothese aufgestellt, dass der Einfluss spezifischer Aminosäurepositionen in einem Protein für dessen Funktion mittels einer Analyse der Sequenz herausgefunden werden kann. Die während der Evolution mutierenden Positionen wurden identifiziert. Dazu wurde eine Methode entwickelt, die es erlaubte, zwischen "neutral drifting" Positionen (mutierende Positionen ohne Einfluss auf die Funktion eines Proteins) und Positionen, die im Zusammenhang mit einem Wechsel der Funktion stehen, zu unterscheiden. Die Analyse wurde auf die PLA2-Daten angewendet. Die Resultate wurden auf eine Röntgenkristallstruktur von einer Phospholipase projiziert. Dies diente als Anschauungshilfe, in welcher Region im Protein die Mutationen auftreten.