

DISS. ETH No. 28801

Stereoselective Tripeptide Organocatalysis – Highly Reactive Electrophiles and Machine Learning

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES

(Dr. sc. ETH Zurich)

presented by

Martin Schnurr

Master of Science in Chemistry, Ruprecht-Karls-Universität Heidelberg

Born on 16.12.1992

Citizen of Germany

accepted on the recommendation of

Prof. Dr. Helma Wennemers, examiner

Prof. Dr. Bill Morandi, co-examiner

2022

Abstract

Tripeptides of the Pro-Pro-Xaa type are highly reactive and stereoselective organocatalysts for conjugate addition reactions. This thesis investigated the use of tripeptides to overcome catalyst deactivation, caused by highly reactive nitroolefins (*Chapters 3 and 4*), and challenged chemoinformatic workflows, to predict and optimize stereoselective tripeptide catalysts (*Chapters 5 and 6*).

In the first part of the thesis, we investigated H-DPro-Pro-Glu-NH₂ and related peptides in the conjugate addition reactions of aldehydes with fluorinated nitroolefins and nitroacrylates. While these electrophiles typically deactivate secondary amine catalysts by *N*-alkylation, mechanistic studies revealed that this deactivation is reversible for Pro-Pro-Glu type peptides, due to the presence of a well-positioned proton donor in the catalyst. Detailed kinetic analysis showed that the *N*-alkylation proceeds slow compared to the enamine-formation, and that the formation of the enamine-intermediate is the rate-limiting step. We identified H-DPro- α MePro-Glu-NH₂ as highly reactive and stereoselective catalyst for the conjugate addition reaction of aldehydes with fluorinated nitroolefins and obtained a broad substrate scope and chiral downstream derivatives. These reactions were carried out with only 0.5 mol% catalyst loading and provided the products in generally high yields and stereoselectivities. Furthermore, we identified H-(4*S*)Mep-DPro-DGlu-NH₂ as even more stereoselective catalyst than H-DPro-Pro-Glu-NH₂ for the reaction of aldehydes with nitroacrylates and provided the first comprehensive substrate scope for this reaction, with catalyst loadings of as little as 0.05 mol%. These low catalyst loadings for reactions with highly electron-deficient nitroolefins are remarkable and highlight the excellent reactivity of Pro-Pro-Glu type peptides, and the value of thorough mechanistic and kinetic investigations.

In the second part of this thesis, we challenged state-of-the-art machine learning methods to predict and optimize conformationally flexible tripeptide organocatalysts. First, we showed that machine learning algorithms can be trained to predict the same highly selective tripeptides of the Pro-Pro-Glu type that were previously developed by rational design. Next, we constructed a *in silico* library with >30000 tripeptide catalysts and synthesized a representative subset of 161 peptides. We tested these peptides on a dienamine-mediated model reactions and obtained a broad experimental data set. This first round of experiments also revealed a stereoselective Pro-Pro-Glu type catalyst that, however, could only be marginally optimized over four consecutive rounds of supervised machine learning. Thus, this work highlights the current

limitations of machine learning methods and emphasizes the need for more sophisticated chemoinformatic workflows.

We also addressed the question whether and how seven electronegative C γ -substituents at the *N*-terminal proline of H-Pro-DPro-DGlu-NH₂ influence the catalyst reactivity and stereoselectivity. First, proline methyl ester model compounds were computationally modelled by the two independent programs MacroModel and CREST. Comparison of the predicted conformational and electronic properties showed a high linear correlation between both programs and experimental validation with dihedral angles, determined by NMR spectroscopy for hydroxyproline methyl esters in solution, were in good agreement with the computational data. In a second step, we synthesized the C γ -substituted tripeptides and tested them as catalysts in two model reactions. While we observed a strong influence of the C γ -substituents on the catalyst performance, we could not see a clear relationship between the computationally predicted properties and the reaction outcome. Thus, this research emphasizes that isolated properties predicted by molecular modeling do not suffice to explain the complex influence of C γ -substituents on tripeptide catalysis.

Kurzzusammenfassung

Tripeptide vom Typ Pro-Pro-Xaa sind hochreaktive und stereoselektive Organokatalysatoren für konjugierte Additionsreaktionen. In dieser Arbeit wurde der Einsatz von Tripeptiden zur Überwindung der durch hochreaktive Nitroolefine verursachten Katalysatordeaktivierung untersucht (*Kapitel 3 und 4*) und chemoinformatische Arbeitsabläufe zur Vorhersage und Optimierung stereoselektiver Tripeptidkatalysatoren (*Kapitel 5 und 6*) getestet.

Im ersten Teil der Arbeit untersuchten wir H-DPro-Pro-Glu-NH₂ und verwandte Peptide in den konjugierten Additionsreaktionen von Aldehyden mit fluorierten Nitroolefinen und Nitroacrylaten. Während diese Elektrophile normalerweise sekundäre Aminkatalysatoren durch *N*-Alkylierung deaktivieren, zeigten mechanistische Studien, dass diese Deaktivierung für Peptide des Typs Pro-Pro-Glu reversibel ist, was auf das Vorhandensein eines gut positionierten Protonendonors im Katalysator zurückzuführen ist. Detaillierte kinetische Analysen zeigten, dass die *N*-Alkylierung im Vergleich zur Enaminbildung langsam verläuft und dass die Bildung des Enamin-Zwischenprodukts der geschwindigkeitsbestimmende Schritt ist. Wir identifizierten H-DPro- α MePro-Glu-NH₂ als hochreaktiven und stereoselektiven Katalysator für die konjugierte Additionsreaktion von Aldehyden mit fluorierten Nitroolefinen und synthetisierten ein breites Substratspektrum und davon abgeleitete chirale Verbindungen. Diese Reaktionen wurden mit nur 0.5 mol% Katalysatorladung durchgeführt und wir erhielten die Produkte in allgemein hohen Ausbeuten und Stereoselektivitäten. Darüber hinaus identifizierten wir H-(4*S*)Mep-DPro-DGlu-NH₂ als noch stereoselektiveren Katalysator als H-DPro-Pro-Glu-NH₂ für die Reaktion von Aldehyden mit Nitroacrylaten und synthetisierten das erste umfassende Substratspektrum für diese Reaktion, mit Katalysatorbeladungen von nur 0.05 mol%. Diese niedrigen Katalysatorbeladungen für Reaktionen mit hoch elektronenarmen Nitroolefinen sind bemerkenswert und unterstreichen die ausgezeichnete Reaktivität von Peptiden des Typs Pro-Pro-Glu, sowie den Wert gründlicher mechanistischer und kinetischer Untersuchungen.

Im zweiten Teil dieser Arbeit haben wir aktuelle Methoden des maschinellen Lernens eingesetzt um Tripeptid-Organokatalysatoren, welche eine hohe Flexibilität der Konformation aufweisen, vorherzusagen und zu optimieren. Zunächst haben wir gezeigt, dass Algorithmen des maschinellen Lernens so trainiert werden können, dass sie dieselben hochselektiven Tripeptide vom Typ Pro-Pro-Glu vorhersagen, die zuvor durch rationales Katalysatordesign entwickelt wurden. Anschließend erstellten wir eine *in-silico*-Bibliothek mit mehr als 30000

Tripeptid-Katalysatoren und synthetisierten eine repräsentative Teilmenge von 161 Peptiden. Wir testeten diese Peptide an einer Dienamin-vermittelten Modellreaktion und erhielten einen breiten experimentellen Datensatz. In dieser ersten Versuchsrunde wurde ein stereoselektiver Katalysator vom Typ Pro-Pro-Glu entdeckt, der in vier aufeinanderfolgenden Runden des überwachten maschinellen Lernens nur geringfügig optimiert werden konnte. Diese Arbeit zeigt die derzeitigen Grenzen maschineller Lernmethoden auf und unterstreicht den Bedarf an weiterentwickelten chemoinformatischen Arbeitsabläufen.

Wir gingen auch der Frage nach, ob und wie sieben elektronegative $C\gamma$ -Substituenten am *N*-terminalen Prolin von H-Pro-DPro-DGlu-NH₂ die Katalysatorreaktivität und Stereoselektivität beeinflussen. Zunächst wurden Prolin-Methylester-Modellverbindungen mit den beiden unabhängigen Programmen MacroModel und CREST rechnerisch modelliert. Der Vergleich der vorhergesagten konformationellen und elektronischen Eigenschaften zeigte eine hohe lineare Korrelation zwischen beiden Programmen, und die experimentelle Validierung mit den durch NMR-Spektroskopie für Hydroxyprolinmethylester in Lösung bestimmten Flächenwinkeln stimmte gut mit den berechneten Daten überein. In einem zweiten Schritt haben wir die $C\gamma$ -substituierten Tripeptide synthetisiert und sie als Katalysatoren in zwei Modellreaktionen getestet. Während wir einen starken Einfluss der $C\gamma$ -Substituenten auf die Katalysatorleistung feststellten, konnten wir keine klare Beziehung zwischen den rechnerisch vorhergesagten Eigenschaften und dem Reaktionsergebnis erkennen. Diese Forschungsarbeit unterstreicht somit, dass isolierte, durch molekulare Modellierung vorhergesagte Eigenschaften nicht ausreichen, um den komplexen Einfluss von $C\gamma$ -Substituenten auf die Tripeptidkatalyse zu erklären.