ETH zürich

ETH Library

Novel Methods to Engineer the Substrate Specificity of Proteases

Doctoral Thesis

Author(s):
Huber, Lukas

Publication date:
2024

Permanent link:
https://doi.org/https://doi.org/10.3929/ethz-b-000674445

Rights / license:
In Copyright - Non-Commercial Use Permitted

This page was generated automatically upon download from the ETH Zurich Research Collection.
For more information, please consult the Terms of use.

DISS. ETH NO. 29968

NOVEL METHODS TO ENGINEER THE SUBSTRATE SPECIFICITY OF PROTEASES

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES

(Dr. sc. ETH Zurich)

presented by

LUKAS HUBER

M.Sc. Molecular Biology, University of Vienna

born on 30.04.1992

accepted on the recommendation of

Prof. Dr. Sven Panke

Prof. Dr. Markus Jeschek

Prof. Dr. Bernhard Hauer

Prof. Dr. Sai Reddy

2024

Abstract

Highly specific proteases enable the targeted cleavage of polypeptides and have found various applications in fields such as biotechnology, biological research, and medicine. However, commercially used proteases still mainly rely on natively available substrate specificities. The ability to engineer proteases to selectively cleave new targets would substantially broaden their utility. For example, in the field of proteomics, engineered proteases cleaving sites that contain post-translational modifications have already led to new discoveries. A particularly exciting prospect is the application of designer proteases for therapeutic interventions, where they could once be used to cleave disease-related targets.

In recent years, directed evolution has led to remarkable achievements in equipping proteins with improved or novel functions. Key to such engineering efforts is testing a diverse pool of variants for the desired activity and isolating improved candidates. In **Chapter 1**, we review currently available high-throughput platforms particularly suitable for the development of proteases with reprogrammed specificity. We describe two major limitations of those methods and address them by developing novel high-throughput platforms, each addressing one of the shortcomings.

Firstly, standard screening assays do not allow integration of the testing for cleavage of multiple undesired substrates during the screening process. This often leads to variants with expanded, rather than truly shifted specificity. To integrate such controls, we aimed to reinvent how protease activity is assessed, and developed a novel *in vivo* assay in **Chapter 2**. We present a DNA recorder for protease activity that enables testing of thousands of protease variants on a plethora of substrate variants in a one vessel-pool of *Escherichia coli* cells. Relying on next generation sequencing as a read out, we identify each individual protease-substrate combination along with the resulting cleavage efficiency in parallel. We demonstrate the suitability of the method by assessing the activity of around 30’000 variants of Tobacco Etch Virus protease (TEVp) as a model on up to 134 substrates concomitantly, representing ~600’000 protease-substrate combinations. First, we performed a deep mutational scanning of TEVp with all single mutants of its wildtype substrate peptide. This allowed us to identify TEVp positions critical for the acceptance of certain variations in the substrate. Furthermore, we combinatorically mutated three selected positions to change the specificity at the substrate position succeeding the scissile bond (P1’ position). We could identify several TEVp variants with distinct amino acid preferences at P1’, including a variant exhibiting a specificity shift away from the amino acid residues preferred by the wildtype TEVp as well as promiscuous variants that may serve as starting points for further re-engineering efforts. Importantly, the amount and quality of data that can be acquired by the presented DNA recorder has so far been inaccessible, opening up exciting possibilities for data-driven protease engineering. We explore this potential by leveraging the unique sequence-activity data through machine learning and provide a demonstration of the *in-silico* prediction and design of proteolytic properties.

Another limitation of current methods for protease engineering concerns the development of variants that cleave substrates containing non-canonical amino acids, for instance those with post-translational modifications. The difficulty of expressing such substrates *in vivo* or importing them into cells renders methods relying on intracellular protease testing impractical. Cell surface display (CSD) of proteases

bypasses this problem, but existing methods primarily rely on substrate retention on the cell surface to keep the genotype-phenotype linkage and allow for screening via fluorescence activated cell sorting. Alternatively, modern *in vitro* approaches enable enzyme-substrate testing within microcompartments but involve complex and laborious workflows due to necessity of cell-free transcription and translation. In **Chapter 3**, we present a platform that combines cell surface display and microfluidic droplet technologies to circumvent these individual limitations and unify the advantages of both methodologies. First, we successfully displayed functional TEVp on the surface of *E. coli*. We further encapsulated single cells together with a fluorogenic substrate, specifically choosing microfluidic double emulsions for compartmentalization to allow for screening with standard flow-cytometric devices. We validated the feasibility of our method by enriching droplets containing cells that display active TEVp from a mixture with inactive variants. Finally, we screened a library of 250'000 TEVp variants for activity on a substrate containing D-aspartate in its recognition sequence. We conducted four consecutive rounds of encapsulation and sorting and employed next generation sequencing to accurately assess the enrichment in every round and to select promising variants. While we were unsuccessful in identifying a TEVp candidate that could cleave a D-aspartate-containing substrate, this work provides valuable methodological insights into the challenges of developing a high-throughput, microfluidic droplet-based method for protease engineering. We detail the complex interplay of the identified factors critical to the success of such screenings, and thus provide a valuable resource for future efforts to engineer proteases or other enzymes with substrates that cannot be readily supplied within the cell.

Zusammenfassung

Hochspezifische Proteasen ermöglichen die gezielte Spaltung von Polypeptiden und haben zahlreiche Anwendungen in Bereichen wie der Biotechnologie, der biologischen Forschung und der Medizin gefunden. Kommerziell eingesetzte Proteasen stützen sich jedoch immer noch hauptsächlich auf die von Natur aus vorhandenen Substratspezifitäten. Die Fähigkeit, Proteasen so zu verändern, dass sie selektiv neue Zielsubstrate spalten können, würde ihren Einsatzbereich erheblich erweitern. So haben beispielsweise im Bereich der Proteomik modifizierte Proteasen, die Substrate mit posttranslationalen Modifikationen spalten, bereits zu neuen Entdeckungen geführt. Eine besonders vielversprechende Aussicht ist die Anwendung von Designer-Proteasen für therapeutische Eingriffe, bei denen sie zur Spaltung krankheitsrelevanter Ziele eingesetzt werden könnten.

In den letzten Jahren hat gerichtete Evolution zu bemerkenswerten Erfolgen bei der Ausstattung von Proteinen mit verbesserten oder neuen Funktionen geführt. Der Schlüssel zu solchen Entwicklungen ist das Testen eines diversen Reservoirs von Varianten auf die gewünschte Aktivität und die Isolierung verbesserter Kandidaten. In **Kapitel 1** geben wir einen Überblick über die derzeit verfügbaren Hochdurchsatzmethoden, die sich besonders für die Entwicklung von Proteasen mit umprogrammierter Spezifität eignen. Wir kommen zu dem Schluss, dass diese Methoden zwei wesentliche Einschränkungen aufweisen, die wir durch die Entwicklung zweier neuartiger Hochdurchsatzplattformen angehen, wobei jede jeweils eines der Defizite adressiert.

Zunächst führte bisher die Unfähigkeit, auf die Spaltung mehrerer unerwünschter Substrate zu testen, häufig zu Varianten mit erweiterter anstatt tatsächlich verlagerter Spezifität. Um unerwünschte Spezifitätsänderungen während der Proteaseentwicklung zu überwachen, beabsichtigten wir die Art, wie Proteaseaktivität in Zellen gemessen wird, neu zu erfinden. In **Kapitel 2** präsentieren wir einen DNA-Rekorder für Proteaseaktivität, der das Testen von Tausenden von Proteasevarianten auf eine Vielzahl von Substratvarianten in einem einzigen Reservoir in *Escherichia coli* ermöglicht. Auf der Grundlage von Next Generation Sequencing können wir jede einzelne Protease-Substrat-Kombination zusammen mit der resultierenden Spaltungseffizienz parallel ermitteln. Wir demonstrieren die Leistungsfähigkeit der Methode, indem wir die Aktivität von rund 30'000 Varianten der Tobacco-Etch-Virus-Protease (TEVp) als Modell auf bis zu 134 Substraten (~600'000 Protease-Substrat-Kombinationen) untersuchen. Als erstes führten wir ein tiefgreifendes Mutationsscanning von TEVp mit allen Einzelmutanten ihres Wildtyp-Substratpeptids durch. Dadurch konnten wir TEVp-Positionen identifizieren, die für die Akzeptanz bestimmter Variationen im Substrat überaus wichtig sind. Darüber hinaus haben wir drei ausgewählte Positionen kombinatorisch mutiert, um die Spezifität an der auf die Spaltbindung folgenden Substratposition (P1'-Position) zu verändern. Wir konnten mehrere TEVp-Varianten mit unterschiedlichen Aminosäurepräferenzen an P1' identifizieren, darunter eine Variante, die eine Spezifitätsverlagerung weg von den Aminosäureresten aufweist, die die Wildtyp-TEVp bevorzugt, sowie promiskuitive Varianten, die als Ausgangspunkte für weitere Entwicklungsbemühungen dienen können. Hervorzuheben ist, dass die Menge und Qualität der Daten, die mit unserem DNA-Rekorder erfasst werden können, bisher unzugänglich waren, und sich somit nun spannende Möglichkeiten für datengesteuerte Proteasenentwicklung ergeben. Wir erkunden dieses Potenzial, indem wir die einzigartigen Sequenz-Aktivität-Daten mit Hilfe von maschinellem Lernen verwerten und eine

Demonstration über die *In-silico*-Vorhersage und Neugestaltung von proteolytischen Eigenschaften vorlegen.

Eine weitere Einschränkung der derzeitigen Methoden für Proteasenentwicklung betrifft die Schaffung von Varianten, welche Substrate spalten, die nicht-kanonische Aminosäuren enthalten, zum Beispiel solche mit posttranslationalen Modifikationen. Die Schwierigkeit, solche Substrate *in vivo* zu exprimieren oder in Zellen zu importieren, macht Methoden, die sich auf intrazelluläre Proteasetests stützen, unbrauchbar. Mit der Zelloberflächenpräsentation von Proteasen lässt sich dieses Problem umgehen, jedoch beruhen die bestehenden Methoden in erster Linie auf der Substratretention an der Zelloberfläche, um ein Durchmusterungsverfahren mittels fluoreszenzaktivierter Zellsortierung zu ermöglichen. Alternativ dazu ermöglichen moderne *In-vitro*-Ansätze Enzym-Substrat-Tests innerhalb von Mikrokompartimenten, erfordern jedoch aufgrund der Notwendigkeit einer zellfreien Transkription und Translation komplexe und aufwändige Arbeitsabläufe. **In Kapitel 3** stellen wir eine Plattform vor, die Zelloberflächenpräsentation und tropfenbasierte Mikrofluidik kombiniert, um diese individuellen Limitationen zu umgehen und die Vorteile beider Methoden zu vereinen. Zunächst gelang es uns, funktionell intakte TEVp auf der Oberfläche von *E. coli* zu präsentieren. Anschließend verkapselten wir einzelne Zellen zusammen mit einem fluorogenen Substrat, wobei wir uns für mikrofluidische Doppelemulsionen zur Kompartimentierung entschieden, um ein Durchmusterungsverfahren mit Standard-Durchflusszytometern zu ermöglichen. Wir validierten die Praxistauglichkeit unserer Methode durch Anreicherung von Tröpfchen, welche Zellen mit aktiver TEVp enthalten, aus einer Mischung mit Tröpfchen mit inaktiven Varianten. Als Letztes untersuchten wir eine Bibliothek von 250'000 TEVp-Varianten hinsichtlich ihrer Aktivität auf einem Substrat, das D-Aspartat in seiner Erkennungssequenz enthält. Wir führten vier aufeinanderfolgende Runden der Verkapselung und Sortierung durch und setzten Next Generation Sequencing ein, um die Anreicherung in jeder Runde genau zu bewerten und vielversprechende Varianten auszuwählen. Obwohl es uns nicht gelungen ist, eine Kandidatin zu identifizieren, welche D-Aspartat akzeptiert, bietet dieses Kapitel wertvolle methodische Einblicke in die Herausforderungen bei der Entwicklung einer tropfenbasierten mikrofluidischen Hochdurchsatz-Methode. Wir erläutern das komplexe Zusammenspiel der identifizierten Faktoren, die für den Erfolg solcher Durchmusterungsverfahren entscheidend sind, und stellen damit eine wertvolle Ressource für künftige Bemühungen um die Entwicklung von Proteasen oder anderen Enzymen mit Substraten dar, die innerhalb der Zelle nicht ohne weiteres bereitgestellt werden können.