

DISS. ETH NO. 25733

**Directed Evolution of Tobacco Etch Virus Protease  
Towards Higher In Vitro Activity**

A thesis submitted to attain the degree of  
DOCTOR OF SCIENCES of ETH ZURICH  
(Dr. sc. ETH Zurich)

presented by

LUZIUS MICHAEL PESTALOZZI

MSc ETH Biotech., ETH Zurich

born on 05.03.1986

citizen of Zurich (ZH)

accepted on the recommendation of

Prof. Dr. Sven Panke  
Prof. Dr. Victor de Lorenzo  
Prof. Dr. Florian Hollfelder

2019

## SUMMARY

Selective proteases enable the targeted hydrolysis of polypeptides and are widely used to cleave off peptide tags after protein purification. A commonly used protease for this purpose is the tobacco etch virus protease (TEVp) which selectively cleaves within the amino acid sequence ENLYFQ|S (single letter amino acid code, the scissile bond is marked). This amino acid motif is specific enough such that accidental off-target cleavage at similar sequences is unlikely.

A drawback of using TEVp for targeted hydrolysis are unfavorable kinetic parameters  $k_{\text{cat}}$  and  $K_M$ , which are in the order of  $0.1 \text{ s}^{-1}$  and  $50 \mu\text{M}$ , respectively. Therefore, TEVp mediated cleavage is slow and additionally slowed down when substrate concentrations fall below the  $K_M$ . Often, TEVp-mediated hydrolysis stops before all substrate is cleaved. In this work, we addressed this limitation by engineering TEVp towards higher in vitro activity.

To this end, we generated naïve and focused libraries of TEVp genes and screened their gene products for improved activity. High throughput and precision for TEVp activity measurements was achieved by applying two activity assays in sequence. First, TEVp activity was connected in vivo to the half-life of a fluorescent protein, allowing direct measurement in *E. coli* and thus high-throughput screening of large libraries using fluorescence-assisted cell sorting. Second, the activity of promising TEVp variants was analyzed in cell-free extracts in a microtiter plate format allowing for precise measurements. After six rounds of directed evolution we obtained TEVp VI containing nine amino acid substitutions and exhibiting an activity about a magnitude higher than the parental TEVp.

However, subsequent characterization of the initial reaction rates of this variant with a set of all possible recognition site with single amino acid substitutions showed, that TEVp VI had lost selectivity at one of the substrate peptide position. This was attributed to mutation N176I leading to fewer hydrogen bonds with the substrate. The reversal of this mutation restored selectivity, however, also led to a decreased activity. To compensate for the loss in activity, we engineered TEVp further by screening a focused library in which mutations were limited to positions lining the substrate tunnel and active site. This led to variant TEVp VII with thirteen mutations which was as selective as the parental TEVp variant and exhibited again an increased activity about a magnitude higher than the parental TEVp.

After successfully addressing the activity of TEVp, we focused on relaxing the selectivity specifically for the C-terminal amino acid of the substrate peptide which in the canonical recognition sequence is occupied with serine. A protease that would still show selectivity for a multi-amino acid residue motif but have this motif located completely on one side of the scissile bond would remove a wide range of limitations that the use of selective proteases still have. For example, it could be used to leave any N-terminus desired with the C-terminal hydrolysis product.

To this end, we screened for TEVp variants in a combinatorial library in which positions of TEVp were mutated that are in proximity of the C-terminal amino acid residue of the substrate peptide. First, we identified functional TEVp variants using the in vivo assay and the canonical

substrate. Second, we tested promising TEVp variants *in vitro* on a mixture of substrate peptides that differed at the C-terminal position of the canonical recognition sequence. For the latter step, we improved the protocol for preparation of cell-free extracts containing TEVp variants for *in vitro* testing, which was dependent on the production of a cell envelope destroying protein (coliphage  $\Phi$ X174 protein E) in the late exponential to stationary growth phase.

With the improved protocol for the activity measurements at hand we identified TEVp variants with relaxed selectivity. One of them cleaved the substrate peptide containing isoleucine or tryptophan instead of the original serine with similar activity. This variant could be used to induce protein degradation inside *E. coli*, since N-terminal tryptophan residues are known to mark proteins for degradation.

In conclusion, we established protocols for the directed evolution of TEVp that allowed for both high-throughput screening and precise activity measurements. This approach was successfully applied for engineering TEVp variants with higher activity and reduced selectivity at the C-terminal position of the substrate peptide.

## ZUSAMMENFASSUNG

Selektive Proteasen ermöglichen die gezielte Hydrolyse von Polypeptiden und werden routinemässig zur Abspaltung von Peptid-Tags nach der Proteinaufreinigung verwendet. Eine hierzu oft eingesetzte Protease ist die Tabak-Ätz-Mosaik-Virus Protease (TEVp), die selektiv innerhalb der Aminosäuresequenz ENLYFQ|S spaltet (Einbuchstaben-Aminosäurecode, die spaltbare Bindung ist markiert). Dieses Aminosäuremotiv ist spezifisch genug, dass eine zufällige Off-Target-Spaltung bei ähnlichen Sequenzen unwahrscheinlich ist. Nach der TEVp-vermittelten Spaltung verbleibt das C-terminale Serin der Erkennungssequenz auf einem der beiden Proteinfragmente und bildet einen neuen N-Terminus. Diese Position ist von besonderer Bedeutung, da die Art des N-Terminus *in vivo* die Proteinhalfwertszeit bestimmt und in einigen Enzymen eine katalytisch aktive Entität enthält.

Ein Nachteil der TEVp für die gezielte Hydrolyse sind die ungünstigen kinetischen Parameter  $k_{cat}$  und  $K_M$ , die in der Größenordnung von  $0,1 \text{ s}^{-1}$  bzw.  $50 \text{ }\mu\text{M}$  liegen. Die TEVp-vermittelte Spaltung ist daher langsam und wird, wenn die Substratkonzentration unter den  $K_M$  fällt, zusätzlich verlangsamt. Zudem stoppt die TEVp-vermittelte Hydrolyse oft bevor das gesamte Substrat gespalten ist. In dieser Arbeit haben wir die TEVp in Richtung höherer *In-vitro*-Aktivität evolviert und weiterentwickelt, um diese Einschränkungen zu adressieren.

Zu diesem Zweck haben wir naive und fokussierte Genbanken von TEVp generiert und deren Produkte auf gesteigerte Aktivität überprüft. Durch die Anwendung von zwei sequentiellen Aktivitätsprüfungen haben wir einen hohen Durchsatz, sowie eine hohe Präzision für TEVp Aktivitätsmessungen erreicht.

Zunächst haben wir die TEVp-Aktivität an die *in vivo* Halbwertszeit eines Fluoreszenzproteins gekoppelt. Diese direkte Aktivitätsmessung innerhalb von *E. coli* hat uns ermöglicht Hochdurchsatz-Screenings von Genbanken von TEVp-Varianten mittels Durchflusszytometrie (FACS) durchzuführen. Die Aktivität vielversprechender TEVp-Varianten in zellfreien Extrakten wurde anschliessend in genaueren Messungen in Mikrotiterplatten analysiert.

Nach sechs Runden gerichteter Evolution haben wir eine TEVp-Variante (TEVp VI) erhalten, die neun Aminosäuren-Substitutionen enthalten hat und deren Aktivität rund eine Grössenordnung höher war als die der Ursprungsvariante. Die nachfolgende Charakterisierung der anfänglichen Reaktionsgeschwindigkeit mit allen möglichen Erkennungssequenzen, die jeweils eine einzelne Aminosäure-Substitution enthalten haben, hat jedoch gezeigt, dass TEVp VI die Selektivität an einer Substratposition verloren hat. Dieser Verlust der Selektivität wurde der Mutation N176I zugeschrieben, da diese mit dem Substrat weniger Wasserstoffbrücken bildet. Die Umkehrung dieser Mutation stellte die Selektivität wieder her, führte jedoch auch zu einer verringerten Aktivität. Um für diesen Aktivitätsverlust zu kompensieren, haben wir daher eine fokussierte Genbank von TEVp analysiert, in der Positionen mutiert wurden, die den Substrattunnel und das aktive Zentrum säumen. Dies hat zur Identifizierung von Variante TEVp VII, die dreizehn Mutationen enthält, geführt. Diese war ebenso selektiv wie die ursprüngliche TEVp-Variante und zeigte dieser gegenüber eine etwa achtfach höhere Aktivität.

Nachdem wir die Aktivität von TEVp erfolgreich erhöht hatten, konzentrierten wir uns darauf, die Selektivität speziell für die C-terminale Aminosäure des Substratpeptids zu verringern. In der kanonischen Erkennungssequenz ist an dieser Position ein Serin. Eine Protease, die die Selektivität für ein Motiv mit mehreren Aminosäureresten beibehält, wenn dieses Motiv vollständig auf einer Seite der spaltbaren Bindung angeordnet ist, würde eine breite Palette von Limitierungen, die selektive Proteasen immer noch mit sich bringen, beseitigen. Zum Beispiel könnte eine solche benutzt werden, um irgendeinen erwünschten N-Terminus an dem C-terminalen Hydrolyse-Produkt zu erhalten.

Daher haben wir eine systematische Analyse einer kombinatorischen Genbank für TEVp-Varianten durchgeführt, in welcher die TEVp-Aminosäurereste, die sich in der Nähe der C-terminalen Aminosäure des Erkennungspeptids befinden, mutiert wurden. Zunächst haben wir funktionelle TEVp-Varianten mittels dem oben beschriebenen *in vivo* Assay und der kanonischen Erkennungssequenz identifiziert. Anschliessend haben wir vielversprechende TEVp-Varianten an einer Reihe von Substratpeptiden, die sich in der C-terminalen Position der kanonischen Erkennungssequenz unterscheidet haben, *in vitro* getestet. Für diesen Schritt haben wir das Protokoll zur Präparation von zellfreien Extrakten verbessert, welche zum Testen der TEVp-Varianten *in vitro* verwendet werden. Dieses basiert auf der Produktion eines Zellhüllen-zerstörenden Proteins (coliphage  $\Phi$ X174 protein E) in der späten exponentiellen bis stationären Wachstumsphase.

Mit dem vorliegenden verbesserten Protokoll für die Aktivitätsmessungen haben wir TEVp-Varianten mit relaxierter Selektivität identifiziert. Eine dieser Varianten schneidet das Substratpeptid mit Isoleucin oder Tryptophan anstelle von Serin mit ähnlicher Aktivität. Diese TEVp-Variante könnte zur Regulieren von Proteindegradierung in *E. coli* benutzt werden, da bekannt ist, dass N-terminale Tryptophanreste Proteine für den Abbau markieren. Zusammenfassend haben wir Protokolle zur gerichteten Evolution von TEVp etabliert, welche uns erlaubt haben sowohl Hochdurchsatz-Screenings *in vivo* als auch präzise Aktivitätsmessungen *in vitro* durchzuführen. Anschliessend wurden diese erfolgreich zur Weiterentwicklung von TEVp-Varianten mit höherer Aktivität und reduzierter Selektivität an der C-terminalen Position der Erkennungssequenz verwendet.