

Abstract

Peptidases play a crucial role in all domains of life. Proteolytic processes build the counterpart to protein biosynthesis and thus are essential in posttranslational modification, protein degradation in digestion, as well as in regulation of activity of protein-protein interaction and localization of cleaved proteins. Most pathogens also employ proteolytic enzymes for various purposes, which therefore contribute to their pathophysiological effects. Consequently, those proteases are valuable pharmacological targets, e.g. the SARS-CoV2 main protease, which is targeted by Paxlovid. The understanding of their proteolytical mechanism plays a central role in the development of specific inhibitors and treatments of their caused diseases.

In pathogenic strains of *Vibrio cholerae* there is the putative metalloprotease *VcZmp1* encoded, which exhibits approximately 15 % sequence identity and a significant structural similarity to the well characterized zinc dependent metalloprotease (Zmp) proline-proline endopeptidase 1 (PPEP-1) from *Clostridioides difficile*. *VcZmp1* has been shown to modulate *V. cholerae* pathogenicity, but its function is unclear and proteolytic activity has remained elusive. In this thesis, a proteomic identification of the *VcZmp1* cleavage sites (PICS), a combinatorial peptide library as well as a small set of fluorogenic peptides resulted in two consensus motives and one additional putative substrate. Based on these results, fluorogenic peptides were designed and activity measurements were conducted. However, the proteolytic activity of *VcZmp1* could not be validated yet.

Therefore, the implementation of two combinatorial libraries is described in this thesis, employing a circularly permuted luciferase and split GFP as reporter proteins, thus allowing fast, cost effective and high throughput capable determination of protease activity *in vivo* in *E. coli*.

The circularly permuted luciferase was successfully cloned with the PPEP-1 cleavage site in a conformationally inactive restraining linker peptide. In presence of PPEP-1 wt, the luciferase indicated proteolytic activity by increased luminescence due to cleavage of the respective linker *in vitro* as well as in *in vivo* experiments. Consequently, a stepwise randomization of the linker peptide was introduced into the luciferase to build up a library step by step. Here, the two-fold randomized library revealed one specific amino acid sequence in the linker to be cleaved by endogenous *E. coli* proteases. The four-fold randomized luciferase library was even more prone

to hydrolysis by those proteases due to the diversity of presented sequences in the proteolysis site.

A second approach utilized split GFP, based on the restoration of fluorescence upon reconstitution of GFP. Here we used the altered elastase inhibitor eglin C which contained the β -sheet GFP11 followed by a proteolysis site in a surface exposed loop. By introduction of the known PPEP-1 recognition sequence in the proteolysis site, the split GFP was shown to regain its fluorogenic properties in presence of the protease PPEP-1 wt *in vitro* as well as *in vivo* in *E. coli*. In analogy to the previous attempt, a four-fold randomization was introduced in the proteolysis site resulting in an increase in cleavage by *E. coli* proteases.

Finally, a cell displayed split GFP library was established to minimize false positives using the AIDA-I autotransporter system. Fluorescence microscopy, as well as surface shedding by trypsin combined with fluorescence spectroscopy, confirmed the successful translocation of the eglinC-GFP11^(PPEP-1) reporter protein to the cell surface of *E. coli* BL21 C43. Furthermore, PPEP-1 activity could be determined in the split GFP assay using the cell display system, proving the combined methods to work as expected for the positive control.

To summarize, in this thesis new tools were developed with the potential to broaden the methods for protease substrate identification. However, further improvements and validations still need to be considered in future research.

Zusammenfassung

Peptidasen spielen eine entscheidende Rolle in allen Bereichen des Lebens. Proteolytische Prozesse bilden das Gegenstück zur Proteinbiosynthese und sind für die posttranslationale Modifikation, den Proteinabbau bei der Verdauung sowie für die Regulierung der Aktivität von Protein-Protein-Interaktion und Lokalisierung der gespaltenen Proteine von wesentlicher Bedeutung. Auch Krankheitserreger setzen proteolytische Enzyme für verschiedene Zwecke ein, die somit zu ihren pathophysiologischen Wirkungen beitragen. Folglich sind diese Proteasen wertvolle pharmakologische Ziele, wie zum Beispiel Paxlovid das gegen die Hauptprotease von SARS-CoV2 eingesetzt wird. Das Verständnis proteolytischer Mechanismen spielt eine zentrale Rolle bei der Entwicklung spezifischer Hemmstoffe und der Behandlung der entsprechend verursachten Krankheiten.

In pathogenen Stämmen von *Vibrio cholerae* wird die mutmaßliche Metalloprotease VcZmp1 kodiert, die etwa 15 % Sequenzidentität und eine erhebliche strukturelle Ähnlichkeit mit der gut charakterisierten zinkabhängigen Metalloprotease (Zmp) Prolin-Prolin-Endopeptidase 1 (PPEP-1) aus *Clostridioides difficile* aufweist. Es hat sich gezeigt, dass VcZmp1 die Pathogenität von *V. cholerae* moduliert, aber seine Funktion und seine proteolytische Aktivität blieb bisher unklar. In dieser Arbeit wurden durch eine Proteom-basierte Identifizierung Proteasespaltstellen (PICS), mittels einer kombinatorischen Peptidbibliothek sowie einer Sammlung fluorogener Substratpeptide zwei Konsensusmotive und ein zusätzliches mutmaßliches Substrat für VcZmp1 gefunden. Auf der Grundlage dieser Ergebnisse wurden wiederum fluorogene Peptide entworfen und Aktivitätsmessungen damit durchgeführt. Die proteolytische Aktivität von VcZmp1 konnte jedoch noch nicht validiert werden.

In dieser Arbeit wird die Implementierung von zwei kombinatorischen Bibliotheken beschrieben, die eine zirkulär permutierte Luciferase und gespaltenes GFP (split GFP) als Reporterproteine verwenden und so eine schnelle, kostengünstige und hochdurchsatzfähige Bestimmung von Proteaseaktivität *in vivo* in *E. coli* ermöglichen.

Die zirkulär permutierte Luciferase wurde erfolgreich mit der PPEP-1-Spaltstelle in ein Linker-Peptid kloniert, was die Luciferase in einer inaktiven Konformation hält. Bei Zugabe von PPEP-1 wt wurde der entsprechende Linker an der PPEP-1 Spaltstelle hydrolysiert und die Luciferase zeigte proteolytische Aktivität durch erhöhte Lumineszenz an. Dies konnte sowohl *in vitro* als

auch in *in vivo* Experimenten nachgewiesen werden. Folglich wurde eine schrittweise Randomisierung des Linker-Peptids durchgeführt, um nach und nach eine Bibliothek an Substraten zu generieren. Die zweifach randomisierte Bibliothek beinhaltete eine spezifische Aminosäuresequenz im Linker-Peptid, die von endogenen *E. coli* Proteasen gespalten wurde. Die vierfach randomisierte Luciferase-Bibliothek war aufgrund der Vielfalt der dargebotenen Sequenzen in der Proteolysestelle noch anfälliger für die Hydrolyse durch endogene Proteasen.

Ein zweiter Ansatz wurde mit split GFP gemacht, das auf der Wiederherstellung der Fluoreszenz nach der Rekonstitution von GFP beruht. Hier verwendeten wir den Elastase-Inhibitor Eglin C, der das β -Faltblatt GFP11 gefolgt von einer Proteolysestelle in einer oberflächenexponierten ungeordneten Tertiärstruktur enthielt. Nach Einführung der bekannten PPEP-1 Erkennungssequenz in die Proteolysestelle konnte nachgewiesen werden, dass das gespaltene GFP seine fluorogenen Eigenschaften in Gegenwart der Protease PPEP-1 wt sowohl *in vitro* als auch *in vivo* in *E. coli* wiederherstellte. In Analogie zum vorherigen Versuch wurde eine vierfache Randomisierung in die Proteolysestelle eingeführt, was wiederum zu häufigen Hydrolysen, also falsch Positiven durch *E. coli* Proteasen führte.

Um falsch positive Ergebnisse zu minimieren, wurde mit Hilfe des AIDA-I-Autotransportersystems eine oberflächenpräsentierte GFP-Bibliothek erstellt. Fluoreszenzmikroskopie sowie eine Kombination aus Trypsinierung der Oberflächenproteine und Fluoreszenzspektroskopie bestätigten die erfolgreiche Translokation des E glinC-GFP11^(PPEP-1) Reporterproteins auf die Zelloberfläche von *E. coli* BL21 C43. Darüber hinaus konnte im vorliegenden Experiment mittels PPEP-1 Aktivität eine funktionierende Positivkontrolle etabliert werden.

Zusammenfassend kann man sagen, dass in dieser Arbeit neue Methoden etabliert wurden, die das Potenzial haben, die Möglichkeiten zur Identifizierung von Proteasesubstraten zu erweitern, wenn auch noch Verbesserungen und Validierungen in Betracht gezogen werden müssen.