

DISS. ETH NO. 29222

PUPYLATION SUBSTRATE DETERMINANTS

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

presented by

Matthias Fabian Block

M.Sc. in Biochemistry, EBERHARD KARLS University Tübingen

born on 29.04.1989

accepted on the recommendation of

Prof. Dr. Eilika Weber-Ban
Prof. Dr. Rudolf Glockshuber
Prof. Dr. Timm Maier

2023

Summary

All organisms rely on specialized systems to degrade proteins and maintain protein homeostasis. Several such systems have evolved throughout all domains of life, with compartmentalized proteases processing most of the proteins being degraded. Eukaryotes rely on a complex chain of different enzymes to target their proteasome through attachment of the small protein ubiquitin (Ub) to substrates. Bacteria, in contrast to eukaryotes, utilize different recognition methods to target their diverse arsenal of proteases. Actinobacteria, including *Mycobacterium tuberculosis* (Mtb), harbor an uncommon bacterial proteasome that was likely acquired from eukaryotes. A system has evolved around it to target substrates in a functionally similar but enzymatically different pathway than the eukaryotic Ub system. Actinobacteria utilize the small prokaryotic ubiquitin-like protein (Pup), which is covalently attached to target proteins by the proteasome accessory factor A (PafA). This ligase catalyzes the formation of an isopeptide bond between a surface-accessible lysine side chain of a target protein and the C-terminal glutamate of Pup. All mycobacteria, in contrast to many other actinobacterial species, encode Pup in a ligation-incompetent form with a C-terminal glutamine (PupQ). The C-terminal glutamine of PupQ needs to be deamidated in a first step by the deamidase of Pup (Dop), converting it to glutamate (PupE) and enabling its use for ligation. During the ligation reaction, PafA phosphorylates the glutamate side chain of PupE, generating a phosphorylated Pup intermediate at the expense of ATP. The isopeptide bond between substrate and Pup is formed in a nucleophilic substitution reaction between the γ -carbonyl carbon of Pup and the ϵ -amino group of a substrate lysine side chain. Pupylated proteins are recognized by the mycobacterial proteasomal ATPase (Mpa), which facilitates unfolding and threading of the unfolded substrate into the proteolytic core. Dop can alternatively depupylate substrates by hydrolyzing the isopeptide bond and hence rescuing substrates from their fate. Degradation of substrates allows recycling of amino acids as well as removal of aberrant proteins from the cytosol. It has been shown that the Pup-proteasome system (PPS) is not essential under standard culturing conditions but becomes essential under nutrient starvation. Mtb, for example, requires the pupylation pathway to overcome starvation in macrophages and to regulate the stress response induced by the host immune system. In contrast to hundreds of ligases in the ubiquitylation pathway, PafA is the only known ligase that carries out all the necessary steps of binding, recognition, and attachment of Pup to a given substrate. In mycobacteria, several pupylome studies identified hundreds of pupylation targets, raising the question of how a single ligase can recognize such plentitude of different substrates. Identified target proteins span a wide range of function, including metabolism, nutrient transport, replication and repair as well as targets involved in translation processes. Some substrates could be linked to observed phenotypes or were shown to accumulate in

absence of the PPS. Nevertheless, for most of them the exact implications of pupylation are unknown. Within these, some targets showed extensive pupylatability *in vitro*, while other could only be pupylated to a minimal degree or not at all.

Attempts identifying a common consensus or recognition motif around the targeted pupylation sites encoded in the primary structure have been unsuccessful. However, it has been shown that PafA discriminates between substrates and non-substrates, leading to the hypothesis of this thesis that PafA recognizes substrates via certain three-dimensional recognition sites on their surface. I identified an accumulation of negatively charged residues in the vicinity of pupylation sites and three conserved arginine patches at the rim of the active site of PafA. By performing site-directed mutagenesis of these potential interface residues, I was able to determine their effect on facilitating the interaction between the ligase and FabD. The residue R207 on the side of PafA was shown to be the crucial determinant for substrate recognition, with the other residues playing a supporting role. Several mutants of negatively charged residues close to the pupylation site of FabD showed a large reduction in pupylatability when mutated to alanine. To identify the direct interaction between the two, proposed participating charges were switched within the interface. By finding a combination of switched residues that retained pupylation of FabD, I was able to locate the interacting partners, including the required R207. Additionally, I was able to show the dependency on R207 as well as a few additional residues within the same patch for formation of the whole pupylome. By introducing alanine mutations of the identified arginine residues *in vivo*, I observed complete disappearance of the pupylome, indicating that PafA utilizes the same mechanism for substrate recognition with all its targets. Within my work, I was able to identify a minimal interface between the ligase PafA and the genuine substrate FabD which is established by charge-charge interactions between the two. Such a substrate determinant allows for broad but specific substrate recognition by PafA and subsequent targeting of these towards the bacterial proteasome, allowing Actinobacteria to adapt to their ever-changing environments.

Zusammenfassung

Alle Organismen benötigen spezialisierte Systeme, um Proteine abzubauen und um die Proteinhomeostase aufrecht zu erhalten. Verschiedenste Systeme haben sich hierzu in den unterschiedlichen Domänen der Lebewesen ausgebildet, wobei die kompartimentalisierenden Proteasen für den Abbau der meisten Proteine zuständig sind. Eukaryoten verwenden eine komplexe Kette unterschiedlicher Enzyme um Proteine mittels Anheftung des kleinen Proteins Ubiquitin (Ub) an ihr Proteasom zu leiten. Im Gegensatz dazu verwenden Bakterien verschiedenste Erkennungssignale, um ihre diversen Proteasen anzusteuern. Actinobakterien, einschließlich *Mycobacterium tuberculosis* (Mtb), enthalten ein ungewöhnliches bakterielles Proteasom welches wahrscheinlich aus Eukaryoten erworben wurde. Um dieses hat sich ein eigenständiges System entwickelt, das die gleiche Funktion wie das eukaryotische Ub System erfüllt, aber einer anderen enzymatischen Reaktion folgt. Actinobakterien verwenden das kleine prokaryotische Ubiquitin-ähnliche Protein (Pup) um damit Proteine für das bakterielle Proteasom zu markieren. Dazu wird Pup kovalent an Substrate durch den Proteasom-Zusatzfaktor A (PafA) gebunden. Diese Ligase katalysiert die Bildung einer Isopeptidbindung zwischen einer oberfläch zugänglichen Lysinseitenkette eines Substrats und dem C-terminalen Glutamat von Pup. Im Gegensatz zu vielen anderen actinobakteriellen Arten kodieren alle Mykobakterien Pup in einer ligationsinkompetenten Form mit einem C-terminalen Glutamin (PupQ). Das C-terminale Glutamin muss zuerst durch die Deamidase von Pup (Dop) deamidiert werden, wodurch das reaktive Glutamat (PupE) für die Ligation entsteht. Während der Ligation phosphoryliert PafA die Glutamatseitenkette von PupE unter Zuhilfenahme von ATP, wodurch phospho-Pup entsteht. Die Isopeptidbindung zwischen Substrat und Pup wird durch eine nukleophile Substitutionsreaktion zwischen dem γ -Carbonylkohlenstoff von Pup und der ϵ -Aminogruppe der Lysinseitenkette des Substrats geformt. Pupylierte Proteine werden von der mykobakteriellen proteasomalen ATPase (Mpa) erkannt, die das Entfalten und Einfädlung des ungefalteten Substrats in den proteasomalen Kern übernimmt. Alternativ kann Dop pupylierte Substrate vor ihrem Schicksal retten indem die Isopeptidbindung hydrolysiert wird. Abbau von Substraten ermöglicht das Recycling von Aminosäuren und die Entfernung von abnormalen zytosomalen Proteinen. Es konnte gezeigt werden, dass das Pup-Proteasom System (PPS) unter Standardkulturbedingungen nicht essenziell ist, aber unter Nährstoffmangel essenziell wird. Zum Beispiel verwendet Mtb das Pupylierungssystem um Nährstoffarmut in Makrophagen zu bewältigen und die vom Immunsystem des Wirts ausgelöste Stressreaktion zu regulieren. Im Gegensatz zu hunderten von Ligasen im Ubiquitylierungssystem ist PafA die einzige Ligase, die alle notwendigen Schritte der Bindung, Substraterkennung und Ligation von Pup an das Substrat durchführt. Nichtsdestotrotz konnten hunderte Pupylierungsziele bei

Pupylomstudien in Mykobakterien identifiziert werden. Dies wirft die Frage auf wie eine einzige Ligase alle diese Substrate erkennen kann. Die identifizierten Substrate decken ein breites Funktionsspektrum ab, darunter Proteine im Stoffwechsel, Nährstofftransport, Replikations-, Reparatur- sowie Proteine, die an Translationsprozessen beteiligt sind. Einige Substrate konnten mit beobachteten Phänotypen verknüpft werden oder akkumulieren in Abwesenheit des PPS. Für die meisten Substrate sind die genauen Auswirkungen der Pupylierung jedoch unbekannt. Einige dieser zeigen *in vitro* eine umfangreiche Pupylrierbarkeit, während andere nur minimal oder überhaupt nicht pupyliert werden können.

Versuche ein gemeinsames Erkennungsmotiv um die Pupylierungsstelle auf der Primärstruktur zu finden waren bislang nicht erfolgreich. Es konnte jedoch gezeigt werden, dass PafA zwischen Substraten und Nicht-Substraten unterscheiden kann. Basiert darauf wurde die Hypothese meiner Arbeit aufgestellt, dass PafA Substrate über dreidimensionale Motive auf der Oberfläche erkennt. Ich konnte eine Akkumulation von negativ geladenen Aminosäuren im Umkreis der Pupylierungsstelle auf der Oberfläche verschiedener Substrate, sowie drei konservierte argininreiche Felder am Rande von PafAs Aktivitätszentrum nachweisen. Diese potenziellen Interaktionsstellen wurden einer zielgerichteten Mutationsanalyse unterzogen, um den Effekt auf die Interaktion des Substrates FabD mit der Ligase zu untersuchen. Auf der Seite von PafA wurde der Rest R207 für die Substraterkennung als essenziell identifiziert, während anderen Resten innerhalb der verschiedenen Felder eine unterstützende Rolle zugeordnet werden konnte. Mehrere der negativ geladenen Reste um die Pupylierungsstelle von FabD zeigten ebenfalls eine Reduktion der Pupylrierbarkeit wenn sie zu Alanine mutiert wurden. Um die direkte Interaktion zwischen den beiden Partnern zu identifizieren, wurde die Ladung zwischen den beiden Seiten des Interfaces getauscht. So konnte eine Kombination ermittelt werden, die einen gewissen Grad von Pupylierung von FabD ermöglicht. Dadurch konnte ich zeigen, dass die zuvor identifizierten Reste, inklusive R207, innerhalb des Interfaces miteinander interagieren. Zusätzlich konnte ich demonstrieren, dass die Abhängigkeit von R207 und einiger weiteren Reste im gleichen Feld nicht limitiert ist auf FabD, sondern dass PafA den gleichen Mechanismus verwendet, um alle Substrate zu erkennen. Das Verschwinden des gesamten Pupyloms konnte beobachtet werden als ich die gleichen Mutationen von PafA *in vivo* untersuchte. Mit meiner Arbeit konnte ich ein minimales Interaktionsinterface zwischen der Ligase PafA und dem Substrat FabD basierend auf Ladungswchselwirkungen identifizieren. Eine solche Determinante erlaubt es PafA ein breites Spektrum von Substraten spezifisch zu erkennen und für das bakterielle Proteasom zu markieren. Somit erlaubt es Actinobakterien sich an ihre ständig verändernde Umwelt anzupassen.