PCP
PLANT & CELL PHYSIOLOGY

A Gatekeeper Residue of ClpS1 from *Arabidopsis thaliana* Chloroplasts Determines its Affinity Towards Substrates of the Bacterial N-End Rule

Clara V. Colombo[1], Germán L. Rosano[1,*], Axel Mogk[2] and Eduardo A. Ceccarelli[1]

[1]Instituto de Biología Molecular y Celular de Rosario (IBR), CONICET, Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Rosario 2000, Argentina
[2]Zentrum für Molekulare Biologie Heidelberg, Universität Heidelberg, INF 282, D-69120 Heidelberg, Germany
*Corresponding author: E-mail, rosano@ibr-conicet.gov.ar Fax, 54-341-4237070 ext. 607.
(Received March 30, 2017; Accepted January 23, 2018)

Regular Paper

Proteins that are to be eliminated must be proficiently recognized by proteolytic systems so that inadvertent elimination of useful proteins is avoided. One mechanism to ensure proper recognition is the presence of N-terminal degradation signals (N-degrons) that are targeted by adaptor proteins (N-recognins). The members of the caseinolytic protease S (ClpS) family of N-recognins identify targets bearing an N-terminal phenylalanine, tyrosine, tryptophan or leucine residue, and then present them to a protease system. This process is known as the 'bacterial N-end rule'. The presence of a ClpS protein in *Arabidopsis thaliana* chloroplasts (AtClpS1) prompted the hypothesis that the bacterial N-end rule exists in this organelle. However, the specificity of AtClpS1 is unknown. Here we show that AtClpS1 has the ability to recognize bacterial N-degrons, albeit with low affinity. Recognition was assessed by the effect of purified AtClpS1 on the degradation of fluorescent variants bearing bacterial N-degrons. In many bacterial ClpS proteins, a methionine residue acts as a 'gatekeeper' residue, fine-tuning the specificity of the N-recognin. In plants, the amino acid at that position is an arginine. Replacement of this arginine for methionine in recombinant AtClpS1 allows for high-affinity binding to classical N-degrons of the bacterial N-end rule, suggesting that the arginine residue in the substrate-binding site may also act as a gatekeeper for plant substrates.

Keywords: *Arabidopsis thaliana* • AtClpS1 • Chloroplasts • N-degron • N-end rule • Proteolysis.

Abbreviations: AAA+, ATPases associated with a variety of cellular activities; BCIP, 5-bromo-4-chloro-3-indolyl phosphate; Clp, caseinolytic protease; EGS, ethylene glycol bis(succinimidyl succinate); GFP, green fluorescent protein; GluTR, glutamyl-tRNA reductase; GST, glutathione *S*-transferase; ITC, isothermal titraton calorimetry; MALDI-TOF, matrix-assisted laser desorption/ionization time of flight; NBT, nitro blue tetrazolium chloride; PYROX, pyroxide oxidase domain; SEC, size exclusion chromatography; TBS, Tris-buffered saline.

Editor-in-Chief's Choice

Introduction

Proteolysis is one of the most important processes required to maintain cellular homeostasis. The regulated degradation of proteins is mainly performed by the AAA+ (ATPases associated with a variety of cellular activities) family of proteolytic machines. The ATP-dependent caseinolytic protease (Clp) system is present in bacteria and eukaryotic organelles of prokaryotic origin, such as mitochondria and chloroplasts (Ishikawa et al. 2009, Sauer and Baker 2011, Nishimura and van Wijk 2015). The bacterial ClpP proteolytic machine is composed of a proteolytic core formed by the protein ClpP, arranged into two opposed homoheptameric rings that can associate with an AAA+ chaperone (ClpA or ClpX in *Escherichia coli*). The chaperones can oligomerize into ring-shaped hexamers that dock onto the axial entrance gates of the ClpP core (Snider and Houry 2008). These gates are narrow enough to prevent inadvertent proteolysis of folded proteins. Consequently, substrates must be recognized, unfolded and presented to the protease by the chaperone moiety. For example, ClpA and ClpX recognize a C-terminal 11 amino acid extension (SsrA tag) that is added to polypeptides in the case of stalled translation (Gottesman et al. 1998, Lies and Maurizi 2008). In this way, by-products of aborted protein synthesis are recycled. In other cases, adaptor proteins deliver substrates to the chaperone–protease complex. The adaptors expand the substrate recognition capabilities of the complex, adding a new layer of regulation in the protein quality control task of the Clp system (Kirstein et al. 2009).

Some substrates are recognized when a destabilizing signal called "degron" is exposed or appended. Degrons can be located at the N- or C-terminus (like the SsrA tag) or in internal regions of the target protein. The detection of N-terminal degrons (N-degrons) by specific proteins and subsequent degradation of the target by proteases is called the N-end rule pathway. The N-end rule states that the identity of the N-terminal amino acid of a protein determines its stability (Bachmair et al. 1986, Dougan et al. 2012). Many variations of the N-end rule exist, which differ not only in the mechanistic aspects of

Plant Cell Physiol. 59(3): 624–636 (2018) doi:10.1093/pcp/pcy016, Advance Access publication on 1 February 2018, available FREE online at www.pcp.oxfordjournals.org

the process but also in the cellular type and compartment in which they operate. In the cytosol of eukaryotic cells, the N-end rule is divided into three branches: the Arg/N-end rule (targets unacetylated destabilizing amino acids located at the N-terminal end), the Ac/N-end rule (targets proteins for degradation by recognizing their N-terminal acetylated residues) and the newly discovered Pro/N-end rule (Bachmair et al. 1986, Hwang et al. 2010, Chen et al. 2017). The N-terminal residue is recognized by an N-recognin, a protein that either marks the protein for degradation or directs it to a protease system. In the eukaryotic N-end rule, the N-recognins and their cognate ubiquitin-conjugating enzyme polyubiquitylate N-end rule substrates, thereby targeting these proteins for degradation by the proteasome (Finley 2009). On the other hand, as prokaryotic cells do not possess an ubiquitylation/proteasome system, the N-end rule in Gram-negative bacteria (also called the Leu/N-end rule) involves a different set of N-recognins. In this case, the primary destabilizing N-terminal residues are unacetylated phenylalanine, tyrosine, tryptophan and leucine, which are recognized by N-recognins of the ClpS family. In *E. coli*, EcClpS is the discriminator of the N-end rule pathway as it recognizes, binds and delivers N-degrons to the ClpAP system for their processing (Erbse et al. 2006, Schmidt et al. 2009). The amino acids in the second position also play a prominent role in modulating recognition of the primary destabilizing residues. In particular, positively charged residues in the second position strongly enhanced EcClpS binding in peptide library screens, while the presence of negative charges weakened the interaction (Erbse et al. 2006). In line with these results, it was shown that in vivo, proteins with an N-terminal lysine or arginine can also be modified by an amino acyl transferase that adds a leucine or phenylalanine to the N-terminus, thereby yielding an 'optimal' bacterial N-end rule substrate (Varshavsky 2011). In contrast to the eukaryotic Ac/N-end-rule, N-terminal acetylation abolishes EcClpS recognition of its N-degrons, stressing the importance of a free N-terminus in the bacterial pathway (Erbse et al. 2006). Even though it is known as the 'bacterial N-end rule', this ClpS-mediated pathway has also been found in eukaryotes such as *Plasmodium falciparum* (AhYoung et al. 2016, Tan et al. 2016).

In chloroplasts, the N-terminal region is also a determinant of protein stability. Apel et al. (2010) used transplastomic tobacco plants expressing different green fluorescent protein (GFP) variants to test systematically the effect of all 20 amino acids in the position following the initiator methionine on the stability of the construct. They also investigated whether additional determinants reside in the N-terminal sequence by using GFP constructs starting with different N-terminal stretches from five plastid proteins and comparing their in vivo accumulation. Cysteine, histidine, aspartate, serine and proline in the second position were found to be highly destabilizing depending on the identity of the test construct, while glutamic acid, methionine and valine conferred stability (Apel et al. 2010). They concluded that the action of methionine aminopeptidase (which removes the initiator methionine and exposes the penultimate N-terminal amino acid residue), the existence of an N-end rule-like pathway and additional determinants in the N-terminal region are important factors that influence protein half-life in plastids. In addition, the presence of chloroplast ClpS (and other components of the Clp system) particularly suggests the presence of the bacterial Leu/N-end rule in these organelles, which probably inherited this mechanism from their bacterial ancestors (Olinares et al. 2011). *Arabidopsis thaliana* chloroplasts harbor one of the most intricate Clp systems, composed of hetero-oligomeric protease complexes containing a mix of ClpP1, 3, 4, 5 and 6 and ClpR1, 2, 3 and 4, three AAA+ chaperones (ClpC1/2 and ClpD), two accessory factors (ClpT1 and 2) and the adaptors ClpF and ClpS1 (AtClpS1) (Nishimura and van Wijk 2015, Nishimura et al. 2015). Given the sequence similarities between AtClpS1 and bacterial ClpS, it was expected that they would recognize similar degradation determinants (Nishimura et al. 2013, Varshavsky 2011). Eight candidate substrates of AtClpS1 were identified in stroma extracts from Δ*clps1* plants by affinity purification using immobilized AtClpS1–glutathione *S*-transferase (GST) (Nishimura et al. 2013). For two of them [glutamyl-tRNA reductase (GluTR) and pyroxide oxidase domain (PYROX)], semi-tryptic peptides close to the hypothetical transit peptide cleavage site predicted by the algorithm ChloroP (Emanuelsson et al. 1999) were found by tandem mass spectrometry. Thus, these peptides were proposed by the authors as the N-terminal sequences of the mature proteins. In contrast to the ClpS-dependent N-end rule in other organisms, aromatic residues were not found in the first position (glutamic acid in GluTR and serine in PYROX). Also, a recent survey of the plant N-terminome found that the most frequent amino acids at the N-termini of stromal proteins are serine, threonine, alanine, valine and glycine (Rowland et al. 2015). Aromatic amino acids such as phenylalanine and histidine are absent in N-terminal residues of proteins from the stroma, whereas tyrosine and tryptophan were observed only in one case each and in an acetylated state, suggesting that these N-termini may confer instability. Altogether, these studies suggest that a certain degree of overlap occurs in the identity of the N-terminal amino acid sequence that regulates protein half-life in plant chloroplasts and in other organisms where ClpS is present, but differences may exist. In light of this, biochemical analysis of AtClpS1 specificity was warranted. To meet this end, we purified and characterized mature AtClpS1 in a recombinant form. AtClpS1 specifically interacts with canonical bacterial N-end rule model substrates in vitro; however, binding affinity is low compared with EcClpS. Mutagenesis of a single amino acid at the substrate-binding cavity was sufficient to modify substrate binding characteristics of AtClpS1 significantly.

Results

AtClpS1 does not interact with the *E. coli* ClpAP system and weakly interacts with bacterial N-end rule substrates

Our first objective was to learn if AtClpS1 could replace EcClpS in the bacterial ClpAPS system. To this end, AtClpS1 was produced in a recombinant form. Mature AtClpS1 (from Ala45 in the full-length protein (Nishimura et al. 2013) was produced in

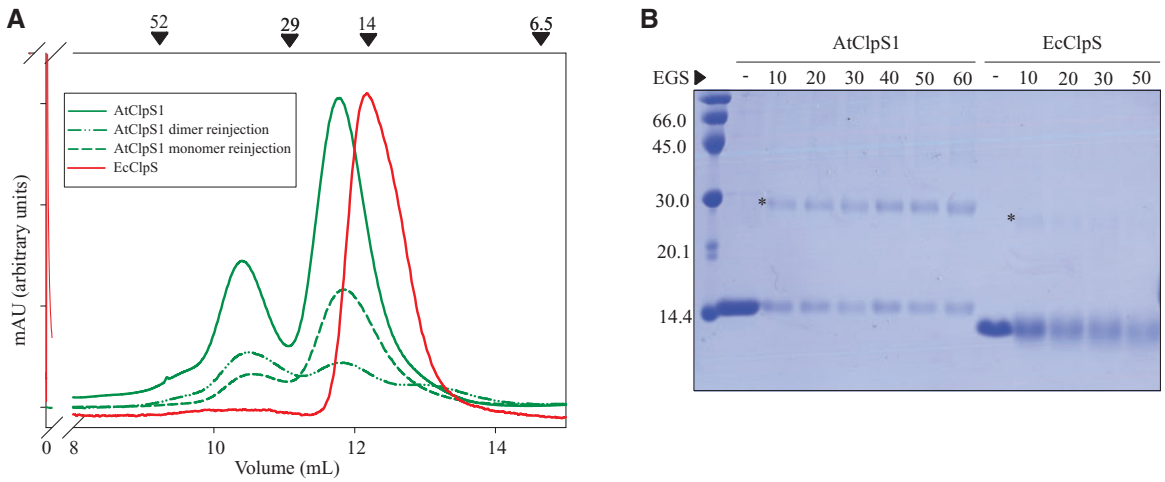

Fig. 1 Analysis of the oligomeric status of ClpS proteins. (A) SEC elution profile of purified proteins in a Superdex 75 column. Elution profiles of AtClpS1 are depicted in green and that of EcClpS in red. Chromatograms after reinjection of the collected AtClpS1 peaks are shown as broken lines as explained in the inset box. Arrows above the plot indicate elution of molecular weight standards. (B) SDS–PAGE of cross-linked ClpS proteins**.** Purified wild-type AtClpS1 and EcClpS were incubated with EGS for 30 min at 25°C. After treatment, samples were analyzed by 15% SDS–PAGE under denaturing conditions followed by Coomassie staining. All proteins were at 25 μM, and the molar fold excess of EGS used in each case is detailed on the top. Molecular weight standards are depicted on the left. Asterisks indicate the presence of oligomers.

E. coli and purified to near homogeneity (>97%) as determined by denaturing gel electrophoresis and mass spectrometry (Supplementary Fig. S1A, B). The band of pure AtClpS1 consistently ran above the 14.4 kDa band of the molecular weight marker in denaturing polyacrylamide gels; however, molecular weight determination by matrix-assisted laser desorption/ionization time of flight (MALDI-TOF) mass spectrometry showed that the experimental molecular weight coincided with the theoretical one. Moreover, the loss of the tag (1,209 Da) was clearly detected (Supplementary Fig. S1A, B).

During purification procedures (specifically, during size exclusion chromatography), two peaks were consistently detected: one major peak corresponding to a mol. wt. of 20 kDa and a minor second one centered at around 34 kDa (**Fig. 1A**). Collection of either peak and reinjection revealed that the species are in dynamic equilibrium, as the elution profiles again showed the presence of both peaks. Considering a theoretical mol. wt. of 13.2 kDa for recombinant AtClpS1, the peak at 20 kDa can be assigned to the presence of monomers or dimers, while dimers or trimers can explain the 34 kDa centered peak. To better elucidate the nature of the oligomers, cross-linking assays were carried out. AtClpS1 was incubated in the presence of different concentrations of the cross-linking agent ethylene glycol bis(succinimidyl succinate) (EGS) and the mixtures were subjected to SDS–PAGE. Only two well-defined bands were detected: one showing a molecular weight of approximately 15 kDa and another at 29 kDa (**Fig. 1B**), indicating that AtClpS1 exists as a mixture of monomers and dimers. Probably, the conformation of these species in solution is not completely globular, and deviations from the expected molecular weight arise in size exclusion chromatography (SEC) analysis. In fact, ClpS proteins are known to possess intrinsically disordered regions (Guo et al. 2002). Analogous assays were carried out with purified ClpS from *E. coli* (EcClpS) to compare its behavior with that of AtClpS1. As has already been shown, EcClpS displayed only one peak corresponding to a mol. wt. of 12 kDa (**Fig. 1A**) (Guo et al. 2002, Sauer and Baker 2011). However, when EcClpS was subjected to EGS treatment and analyzed by SDS–PAGE, a weak band was also detected at approximately 27 kDa, showing the presence of dimers (**Fig. 1B**). Apparently, EcClpS can dimerize, but the equilibrium is shifted substantially to the monomeric form of the protein.

Having successfully purified AtClpS1, we next tested if EcClpS could be replaced with AtClpS1 in the ClpAP system from *E. coli*. The entire Clp complex of *A. thaliana* is formed by almost a dozen different proteins that assemble in a determined stoichiometry to unfold and degrade substrates (Olinares et al. 2011). Reconstituting this complex in vitro with recombinant proteins would be a difficult task, whereas functional *E. coli* ClpAP complexes can be obtained from its purified constituents. This ClpAP system has been extensively characterized and used to unravel the N-end rule in bacteria (Dougan et al. 2002) by using a diverse set of N-end rule reporter proteins. These consist of variants of GFP with destabilizing residues fused at its N-terminus. A spacer sequence separating the N-terminal residues from the GFP moiety was also included, as a linker is needed for the transfer of EcClpS-bound N-end rule model substrates to ClpAP (Erbse et al. 2006). When these GFP variants were incubated in the presence of ClpAP, EcClpS and ATP, loss of fluorescence in those bearing an N-terminal phenylalanine, tyrosine, tryptophan or leucine was readily observed (Erbse et al. 2006, Schmidt et al. 2009).

To study the capacity of AtClpS1 in selecting and transferring said N-end rule model substrates in vitro (FR–GFP, YLFVQ–GFP, YKGEQ–GFP, WFCWS–GFP and LLWCR–GFP; **Fig. 2E**), we tested if AtClpS1 could replace EcClpS in substrate recognition and delivery to the bacterial ClpAP system. The fate of the GFP variants was followed by monitoring GFP

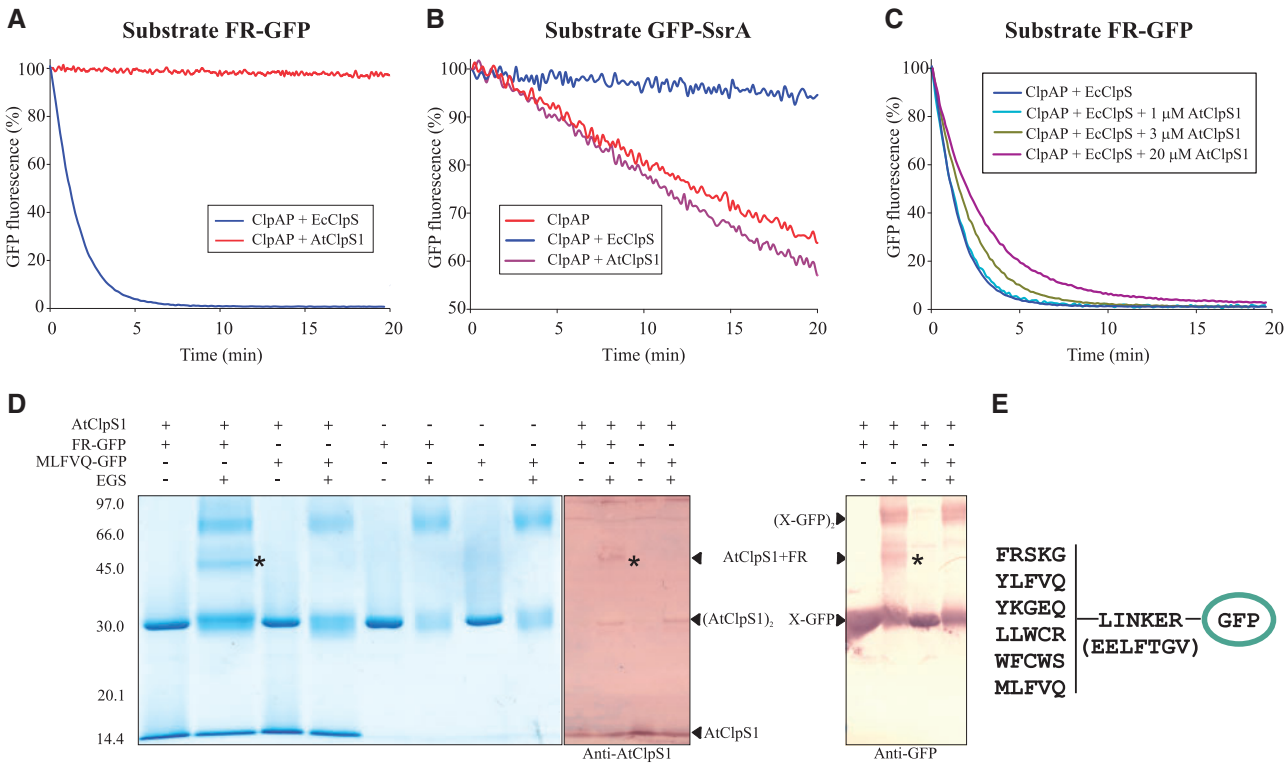

Fig. 2 Effect of AtClpS1 on the ClpAPS-assisted degradation of model substrates. (A) FR–GFP at 0.2 μM was incubated with 1 μM ClpAP + 1 μM EcClpS (blue) or ClpAP + 1 μM AtClpS1 (red). Degradation was determined by following GFP fluorescence. GFP fluorescence at $t = 0$ min was set as 100% and all remaining data points were calculated as a percentage of this initial value. (B) The inhibition of ClpAP-mediated degradation of GFP–SsrA by ClpS proteins was analyzed by incubating the ClpAP complex, the fluorescent substrate and EcClpS (blue) or AtClpS1 (purple). GFP–SsrA degradation without the addition of any ClpS protein (red) was used as a control. Experimental design and data processing were done as in (A). (C) Substrate binding by AtClpS1 in the presence of ClpAP + EcClpS was performed by adding a 20-fold molar excess of AtClpS1 over EcClpS to the mixture of ClpAP + EcClpS and FR–GFP (purple). The same mixture without the addition of AtClpS1 was used as a control (blue). Fluorescence measurements and data processing were done as in (A). (D) To stabilize the interaction of AtClpS1 with artificial substrates, mixtures of AtClpS1 and FR–GFP or MLFVQ–GFP (both referred as to X-GFP) were incubated in the absence (–) or presence (+) of the cross-linking reagent EGS at 1 mM for 30 min at room temperature. The mixtures were then analyzed using Coomassie-stained polyacrylamide gels (12%) under denaturing conditions. As controls, each GFP variant alone was incubated in the absence or presence of EGS. Replicates of cross-linking assays are shown in Supplementary Fig. S9. In another set of cross-linking experiments, two denaturing 12% polyacrylamide gels containing AtClpS1 + X-GFP (in the absence or presence of 1 mM EGS) were transferred to nitrocellulose membranes, one being revealed with anti-AtClpS1 antibodies and the other with anti-GFP antibodies. The band corresponding to the interaction between FR–GFP and AtClpS1 is marked with an asterisk. Molecular weights are depicted on the left. (E) Sequences of the GFP variants used in this work as artificial probes for the ClpS proteins. The N-terminal sequence relevant for their recognition is shown. The linker region is common to all and corresponds to amino acids 3–10 of GFP (Erbse et al. 2006, Schmidt et al. 2009).

fluorescence in the presence of ClpAP + AtClpS1, under the same experimental conditions as used in Schmidt et al. (2009). As previously reported, GFP variants incubated in the presence of ClpAP and EcClpS were rapidly and efficiently degraded. However, no degradation was detected within 30 min when EcClpS was replaced with AtClpS1 at an equal concentration (**Fig. 2A**, results shown for the model substrate FR–GFP). The lack of degradation of the fluorescent substrates by ClpAP in the presence of AtClpS1 can be explained by two non-mutually exclusive hypotheses: (i) AtClpS1 is not able to bind to ClpA and (ii) AtClpS1 cannot recognize the GFP variants. To test the former, a degradation assay was performed using the substrate GFP–SsrA, which is recognized and degraded solely by ClpAP. The addition of EcClpS prevents its degradation because binding of EcClpS to ClpAP prevents binding of GFP–SsrA to the proteolytic system and also triggers the release of any GFP–SsrA substrate bound to it (Dougan et al. 2002). Therefore, if AtClpS1 were able to bind to ClpA, it could also prevent GFP–SsrA degradation. The ClpAP-mediated proteolysis of GFP–SsrA was examined by following GFP fluorescence in the absence and presence of AtClpS1. A reaction including EcClpS served as a positive control. The addition of AtClpS1 did not significantly inhibit GFP–SsrA degradation, while EcClpS abolished its proteolysis, indicating that AtClpS1 does not bind to the protease in a productive manner (**Fig. 2B**). We next sought to analyze if AtClpS1 could recognize the N-end rule substrates (second hypothesis). As AtClpS1 does not bind to ClpA, substrate competition assays were performed with the aim of detecting a possible inhibitory effect on substrate degradation. Increasing amounts of AtClpS1 were used in the reaction mixture containing ClpAP, the GFP substrates and EcClpS at 1 μM.

While AtClpS1 at a 1- or 3-fold molar excess over EcClpS did not produce an apparent effect on the degradation kinetics of the GFP variants, a 20-fold molar excess of AtClpS1 over EcClpS caused a slight to moderate drop in substrate degradation in all cases, GFP variant starting with tryptophan being the the least affected (**Fig. 2C;** Supplementary Fig. S2). These results indicate that AtClpS1 can bind to the model substrates, thereby making them transiently inaccessible to recognition by EcClpS and subsequent degradation by ClpAP. This was also apparent when kinetic data were extracted from the degradation curves by curve fitting using integrated rate equations. K_m values for EcClpAPS and the different fluorescent substrates were in the low micromolar range (Supplementary Table S1), as previously reported (Schmidt et al. 2009). Addition of 20 μM AtClpS1 to the system did not significantly change K_m and V_{max} values, as would be expected if AtClpS1 does not interact in a productive manner with EcClpAPS. The amount of substrate available for EcClpAPS-mediated degradation would then be controlled by the dissociation constant (K_d) of the AtClpS1 + substrate equilibrium. Feeding this model to the fitting algorithm thus allows for a gross estimate of K_d values. For example, the K_d of AtClpS1 for FR–GFP should be >500 μM to explain its inhibitory effect on the degradation of FR–GFP by EcClpAPS, when present at a concentration 20× than that of EcClpS (Supplementary Table S1). As EcClpS presents a K_d for FR–GFP of 1.45 μM (measured by plasmon surface resonance; Erbse et al. 2006), AtClpS1 presents a much lower affinity for the model substrate than its bacterial counterpart. Despite AtClpS1 being present in a concentration 20 times than that of EcClpS, the latter could still bind to each substrate. This was evidenced by the fact that substrate degradation still ensued, as the final fluorescence reached control values. These observations indicate that the affinity of AtClpS1 for the substrates is much lower than that of its bacterial counterpart.

To confirm the binding of a model N-degron to AtClpS1, cross-linking assays were performed in order to stabilize AtClpS1 in complex with the FR–GFP model substrate. AtClpS1 and FR–GFP were incubated in the presence of EGS, and the mixture was analyzed by SDS–PAGE. In the cross-linking reaction mixture, a band of approximately 41 kDa was detected, which corresponds to FR–GFP (28 kDa) bound to AtClpS1 (13 kDa) (**Fig. 2D**). An extra band of approximately 56 kDa was also detected that could be attributed to a dimer of FR–GFP, as this band was also present in the control mixtures of this GFP variant + EGS in the absence of AtClpS1. The specificity of AtClpS1 was tested using the non-N-end rule MLFVQ–GFP variant (**Fig. 2E**) which is not recognized by bacterial ClpS proteins (from *E. coli* or *Caulobacter crescentus*) (Erbse et al. 2006, Roman-Hernandez et al. 2009). EGS cross-linking of mixtures of AtClpS1 and the protein MLFVQ–GFP followed by SDS–PAGE analysis showed that AtClpS1 could not bind this variant (**Fig. 2D**). Also, the mixtures were analyzed by Western blotting using anti-AtClpS1 or anti-GFP antibodies. The previously detected band at around 41 kDa in the SDS gels was the only one that appeared in both membranes revealed with either antibody (**Fig. 2D**). This shows that the 41 kDa band contained a complex of AtClpS1 with FR–GFP.

A gatekeeper residue controls substrate binding in AtClpS1

The difference in substrate affinity between AtClpS1 and EcClpS observed in the previous experiments may stem from variations in the substrate-binding site of each protein. Comparison of the amino acid sequences of EcClpS and AtClpS1 showed that some of the amino acid residues that are involved in substrate recognition are different (Nishimura et al. 2013). One of the most significant differences in the substrate-binding pocket of AtClpS1 is the presence of Arg50 instead of methionine (Met40 in EcClpS). The arginine residue at that position is largely conserved in algae, mosses and plants, and the same applies to methionine in bacterial ClpS (Lupas and Koretke 2003, Nishimura et al. 2013). The methionine residue in bacterial ClpS has been long recognized as a gatekeeper residue that is responsible for target selectivity (Wang et al. 2008). Considering the importance of these amino acids in ClpS target specificity, two mutant proteins were constructed: (i) EcClpSM40R, where the Met40 in EcClpS was changed to arginine; and (ii) AtClpS1R50M, where the Arg50 in AtClpS1 was changed to methionine. Both proteins were readily produced in recombinant form and displayed the same elution profiles in SEC in comparison with the wild types (Supplementary Fig. S3), implying that the overall structure remained unaltered.

EcClpSM40R was first used to screen a library of C-terminally coupled peptides as described (Erbse et al. 2006), in order to detect major changes in substrate recognition. Peptide binding of EcClpS wild type and EcClpSM40R was monitored by immunoblot analysis using EcClpS-specific antibodies after electrotransfer of bound proteins. Both EcClpS proteins were equally well recognized by the polyclonal antibodies (Supplementary Fig. S4A). Wild-type EcClpS served as the reference for pairwise comparison and bound strongly to peptides with destabilizing primary residues at the N-terminus (**Fig. 3**; Supplementary Fig. S4B, C). All N-end rule peptides (16 in total) were recognized by EcClpS. The binding pattern of EcClpSM40R was similar to that of EcClpS, but the amount of peptide-bound protein was reduced, indicating lower binding affinities. EcClpSM40R recognized 14 out of the 16 N-end rule peptides bound to EcClpS (**Fig. 3**). Both N-end rule peptides that were not recognized harbored a tyrosine residue in the first position and an acidic residue in the second position. We also observed binding of EcClpS to non-N-end rule peptides (17 peptides in total) as was previously reported (Erbse et al. 2006). The binding of EcClpS to these peptides was in general weaker, resulting in lower spot intensities of transferred EcClpS. All four non-N-end rule peptides that showed strong binding had a positive net charge, indicating that basic residues can promote EcClpS binding, in agreement with previous findings (**Fig. 3**) (Erbse et al. 2006). Binding of non-N-end rule substrates to EcClpSM40R was reduced (seven peptides). Taken together, EcClpSM40R still recognizes many N-end rule peptides specifically, yet with lower affinity as compared with wild-type EcClpS. Its lower

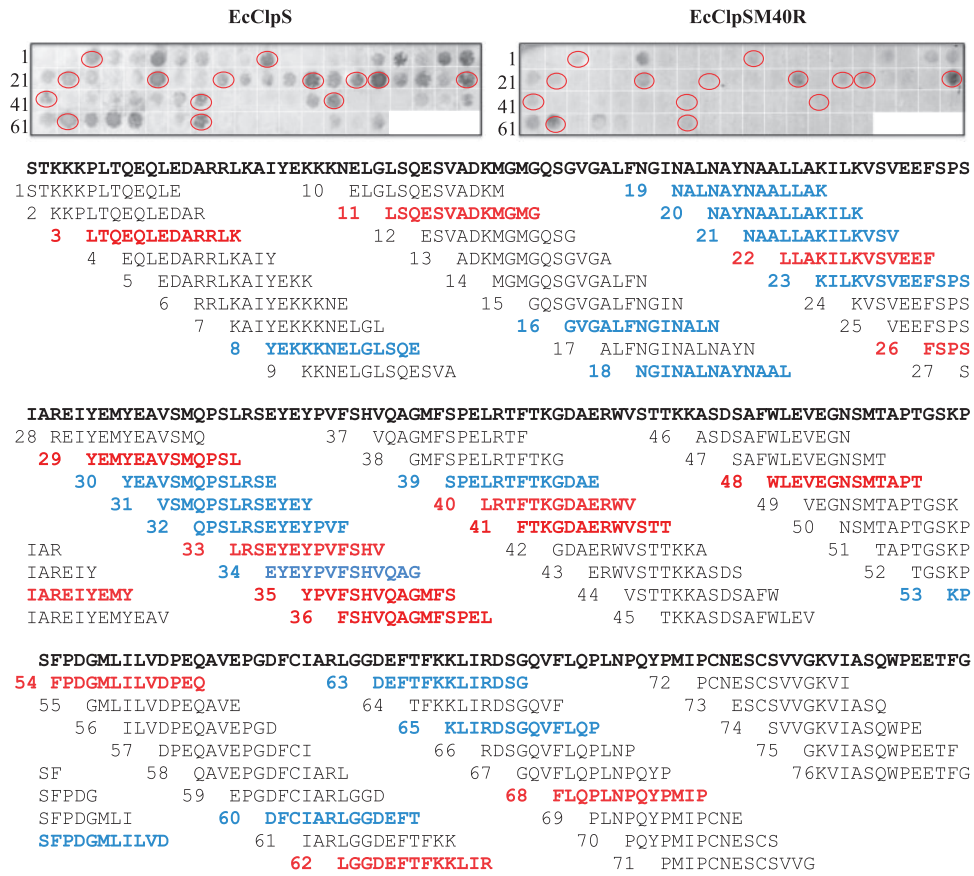

Fig. 3 Binding of EcClpS and EcClpSM40R to λcI peptide scanning libraries. Libraries consisted of 13-mer peptides overlapping by 10 residues covering the sequence of the λcI protein, as shown below the images. They were incubated with 500 nM EcClpSM40R, which was then electrotransferred to PVDF membranes. Binding of the protein was monitored by immunoblot analysis using EcClpS-specific antibodies. Libraries were regenerated and then incubated with 500 nM wild-type EcClpS, whose binding to the peptides was assessed as explained. Peptides starting with bacterial N-end rule amino acids that were recognized by both proteins are encircled in red and their sequences are colored in red, while the sequences of peptides recognized by EcClpS and no longer recognized by EcClpSM40R are colored in blue.

binding affinity makes EcClpSM40R more selective, reducing the number of non-N-end rule peptides recognized in a peptide library as compared with EcClpS.

As no mutations were introduced in the region that binds to ClpA, it was expected that the protein EcClpSM40R should have conserved its capacity to bind and deliver substrates to the chaperone–protease complex. Accordingly, binding of EcClpSM40R to ClpA was confirmed by its ability to inhibit the ClpAP-mediated degradation of GFP–SsrA (Supplementary Fig. S5). Hence, degradation assays of the N-end rule substrates were carried out with the ClpAP system and EcClpSM40R, using wild-type EcClpS as a control. A decrease in substrate degradation in the presence of EcClpSM40R was detected, but not all the substrates showed the same effect. FR–GFP, YKGEQ–GFP and WFCWS–GFP fluorescence was stable over time, indicating a complete loss of recognition by EcClpSM40R (**Fig. 4A;** Supplementary Fig. S6). However, YLFVQ–GFP and LLWCR–GFP (**Fig. 4B;** Supplementary Fig. S6, respectively) were degraded by ClpAP when EcClpSM40R was present, albeit more slowly and less efficiently compared with the control. For the YLFVQ–GFP substrate, the degradation curve could be fitted using the integrated rate equation. The global K_m for the EcClpAPSM40R system was 10 times higher than that of the wild-type system, suggesting a significant decrease of binding affinity (Supplementary Table S1). These results are in agreement with the peptide library assay and with the hypothesis that EcClpSM40R has diminished affinity towards N-degrons bearing bulky amino acids. For example, the rather small leucine in the first position of LLWCR–GFP is recognized by EcClpSM40R. Regarding the aromatic amino acids, the very voluminous tryptophan residue apparently cannot enter into the cavity, while phenylalanine and tyrosine can do so in some cases; whereas YLFVQ–GFP can be degraded, YKGEQ–GFP and FR–GFP cannot. Interestingly, these two substrates contain a positively charged residue in the second position, which may produce an electrostatic repulsion with arginine at position 40 in EcClpSM40R.

In AtClpS1, an arginine residue (Arg50) occupies the position of the gatekeeper methionine. We created a back-to-bacterial consensus variant (AtClpS1R50M) to test its importance in substrate discrimination and affinity. The interaction of AtClpS1R50M with bacterial N-end rule artificial substrates was analyzed by competition assays, as was done for the AtClpS1 wild-type protein. Likewise, a 20-fold molar excess of

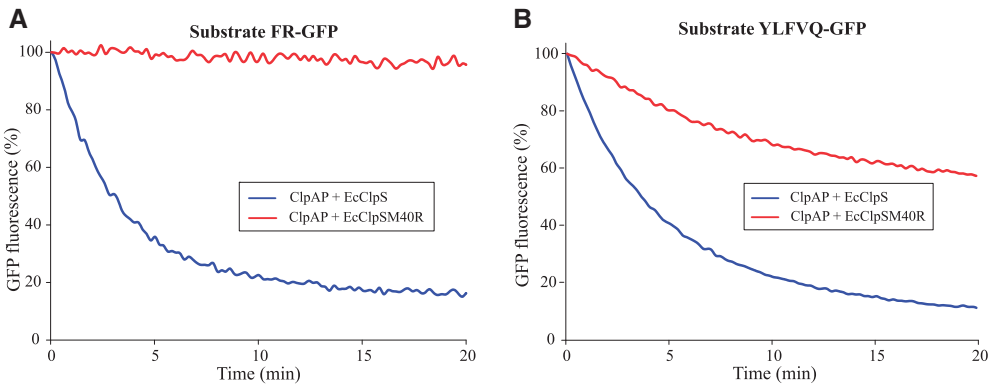

Fig. 4 N-degron recognition by EcClpS and EcClpSM40R. (A and B) Degradation assays of the model substrates FR–GFP (A) and YLFVQ–GFP (B) with ClpAP + EcClpS (control, blue) or ClpAP + EcClpSM40R (red). Experiments were undertaken as described in **Fig.2A**.

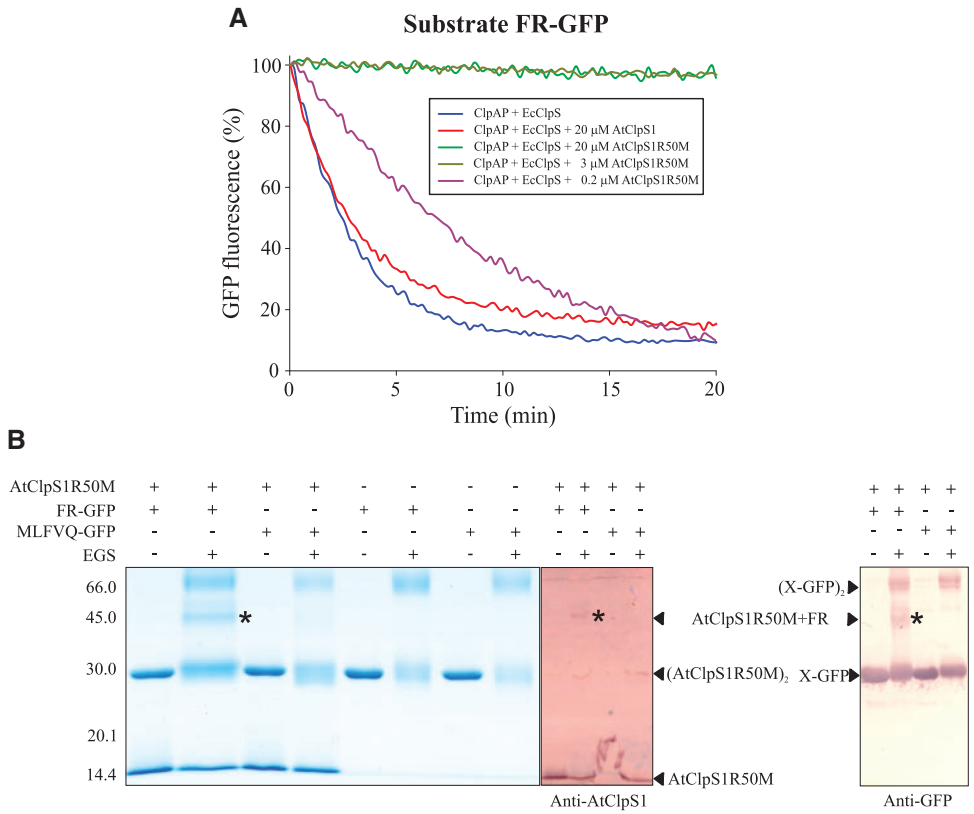

Fig. 5 Degradation of FR–GFP in the presence of AtClpS1R50M and complex formation by cross-linking treatment. (A) Effect of different concentrations of AtClpS1R50M in the ClpAPS-assisted degradation of FR–GFP. Controls included the 1 μM ClpAP + 1 μM EcClpS system without competitor (blue) and competition with wild-type AtClpS1 (red), as in **Fig.2C**, B. The interaction of AtClpS1R50M with GFP variants by cross-linking assays was performed and analyzed by denaturing 12% PAGE and Western blotting as described in **Fig.2D**. The band corresponding to the complex of FR–GFP and AtClpS1R50M is marked with an asterisk. Replicates of cross-linking assays are shown in Supplementary Fig. S9.

AtClpS1R50M over EcClpS was added to the degradation mixture (ClpAP, 1 μM EcClpS and 0.2 μM of each substrate) and GFP fluorescence was monitored for 20 min. In all cases, substrate degradation was almost completely abolished (**Fig. 5A** for substrate FR–GFP and Supplementary Fig. S7 for the other GFP variants), indicating a strong competition of AtClpS1R50M for the GFP variants, thereby reducing the fraction of substrate available for EcClpS binding. The same results were obtained when the concentration of AtClpS1R50M was lowered to a 3-fold molar excess. In the case of the variant FR–GFP, a competition assay was performed using 0.2 μM AtClpS1R50M. Even in these conditions where the level of AtClpS1R50M is much lower than that of EcClpS, competition was still observed (**Fig. 5A**). The half-life of the fluorescent substrate (time for 50% degradation) was approximately 7 min in the presence of 0.2 μM AtClpS1R50M. At this time point, degradation by the bacterial system in the absence of AtClpS1R50M was >80%, indicating that AtClpS1R50M interacts with high affinity with

Table 1 K_d and ΔH measurements of the binding of AtClpS1R50M and each GFP variant by ITC

GFP variant	K_d (μM)	ΔH ($\times 10^4$ cal mol^{-1})
FR–GFP	0.58 ± 0.13	−2.09 ± 0.09
YLFVQ–GFP	3.80 ± 0.35	−2.94 ± 0.86
YKGEQ–GFP	5.65 ± 1.63	−2.83 ± 0.37
LLWCR–GFP	0.17 ± 0.07	−1.04 ± 0.18
WFCWS–GFP	1.44 ± 0.07	−3.00 ± 0.05
FR–GFP (+EcClpS)[a]	1.45 ± 0.04	–

Values were obtained after curve fitting of the calorimetric data and represent the mean ± SD of two determinations.
[a]Determined by surface plasmon resonance (Erbse et al. 2006).

FR–GFP, in clear contrast to the wild-type plant protein. However, near-complete degradation of FR–GFP at the end of the experiment (20 min) suggests that the bound substrate eventually dissociates from AtClpS1R50M. As before, a model where substrate availability for the EcClpAPS system is influenced by the K_d of the AtClpSR50M + FR–GFP equilibrium can be used to estimate grossly the binding and kinetics constants. In this case, the K_d of AtClpS1R50M for FR–GFP was in the low micromolar range, almost 50 times lower than that of the wild-type protein (Supplementary Table S1).

To gain better insight into the thermodynamic parameters of the interaction between AtClpS1 and the R50M mutant, isothermal titration calorimetry (ITC) assays were undertaken. Each of the five substrates was injected multiple times into a solution of either AtClpS1 or AtClpS1R50M and the evolution of heat was recorded. The shape of an ITC curve is determined by the product of the association constant and the titrand concentration (c value). When c is <5, the curve becomes progressively more hyperbolic; values >5 result in sigmoidal curves (Turnbull and Daranas 2003). Under our conditions, hyperbolic curves were obtained and, in these cases, the stoichiometry of the reaction must be fixed to obtain K_d and ΔH values by curve fitting (Tellinghuisen 2008). In accordance with crystallographic data of bacterial ClpS proteins (Roman-Hernandez et al. 2011), the binding stoichiometry was fixed to 1. Evolution of heat was near zero when the substrates were added to AtClpS1, indicating that an interaction could not be detected under the experimental conditions tested. However, when the same experiments were performed using AtClpS1R50M, evolution of heat was clearly detected in all cases (Supplementary Fig. S8). The measured K_d values were similar to those reported by others for bacterial ClpS using surface plasmon resonance or ITC (**Table 1**) (Erbse et al. 2006, Schuenemann et al. 2009). The mutation in the gatekeeper residue thus lowered the K_d to values close to thode of bacterial ClpS proteins. Taken together, our data show that substitution of Arg50 for methionine gave AtClpS1 the ability to recognize bacterial N-end rule model substrates efficiently. The interaction and binding specificity were further assayed in the same way as was done for the wild-type protein by EGS cross-linking (see **Fig. 2D**). A 41 kDa cross-linked product was detected when a mixture of AtClpS1R50M and FR–GFP in the presence of EGS was analyzed by SDS–PAGE (**Fig. 5B**). Like wild-type AtClpS1, AtClpS1R50M did not cross-link to MLFVQ–GFP, indicating that the R50M mutation did not change the 'no binding' rule for substrates bearing an initial methionine residue.

Discussion

The study of adaptor proteins of the Clp system has escalated in the last few years. As ClpS proteins are the N-end rule selectors for this machinery, recent work has focused on elucidating the rules governing target selection, with the aim of establishing the N-end rule pathway in different organisms. In plant chloroplasts, the detailed knowledge of the molecular mechanism of this pathway is still under debate, largely because the sequence determinants that influence protein fate are still unknown (Gibbs et al. 2016). The aim of this work was to gain further insight into the selectivity of AtClpS1. For this, we employed a strategy based on the degradation of model substrates by bacterial ClpAP in the presence of different ClpS proteins obtained in a recombinant form. This approach has been proved to be successful in the investigation of the substrate specificity of ClpS in bacteria (Erbse et al. 2006, Tryggvesson et al. 2015, Stein et al. 2016) and eukaryotes (AhYoung et al. 2016, Tan et al. 2016). In all cases, recombinant ClpS was shown to behave as a monomer in solution. In contrast to this, AtClpS1 can clearly dimerize. EcClpS can also dimerize, as seen in cross-linking assays, but the equilibrium is displaced toward the monomer. The fold of EcClpS was shown to be very similar to that of the *E. coli* ribosomal protein L7/L12 (Guo et al. 2002, Zeth et al. 2002b, Lupas and Koretke 2003), a dimeric protein whose N-terminal domain is critical for oligomer formation. Given the structural similarity between L7/L12 and EcClpS (a root-mean-square deviation of 1.7 Å for 63 pairs of C-α atoms out of 68 residues in the structure) (Guo et al. 2002), it was speculated that EcClpS could also dimerize, yet only one report briefly commented on this possibility (Zeth et al. 2002a). In this work, the formation of AtClpS1 dimers was clearly detected. However, their physiological role, if any, remains to be elucidated.

The crystal structures of all ClpS proteins studied to date allow the distinction of two well-defined regions: an unstructured N-terminal extension implicated in substrate delivery to the cognate protease and a folded domain responsible for target recognition and binding (Guo et al. 2002, Zeth et al. 2002b, Wang et al. 2008, Schuenemann et al. 2009). This core harbors a hydrophobic cavity in which destabilizing residues dock through hydrogen bonds between ClpS and the α-NH_3^+ group of the target. The precise internal architecture of this binding pocket is what drives the identification of the amino acids that can be accommodated in it. An in silico analysis of the amino acids responsible for docking the substrate in the cavity showed that they are largely conserved between plants and bacteria (Nishimura et al. 2013). Of note, the gatekeeper methionine present in bacterial ClpS (Met40 in EcClpS) is replaced by an arginine residue (Arg50 in AtClpS1) in all ClpS1 proteins from angiosperms. Thus, the authors suggested that the substrate recognition mechanisms of plant ClpS1

proteins were partially retained throughout their evolutionary history in comparison with bacterial ClpS proteins. Our results provide the first experimental support for this hypothesis, as AtClpS1 can recognize in vitro the same set of bacterial N-end rule model substrates. However, the interaction is weaker in comparison with EcClpS, i.e. a 20-fold molar excess of AtClpS1 over EcClpS was needed to detect a decrease in the degradation rate of the fluorescent model substrates. Also, AtClpS1 did not interact with a GFP variant with a methionine residue in the first position. The fate of N-methionine-bearing proteins is an active area of research, and it includes proteins synthesized in the organelle. In chloroplasts, as in bacteria, nascent polypeptides start with N-formylmethionine, which can be eliminated by the sequential action of a peptide deformylase and methionine aminopeptidase (Giglione and Meinnel 2001). The latter exerts its action only if the previously deformylated methionine is followed by a residue no larger than valine (Frottin et al. 2006, Varshavsky 2011). As such, some plastidial proteins retain N-methionine (Giglione and Meinnel 2001). Our results suggest that they should not be targets of the Clp system as AtClpS1 (wild type and the R50M mutant) did not bind to MLFVQ–GFP in cross-linking assays. On the other hand, it was shown that a reporter protein is destabilized when its initial formylmethionine (followed by aspartic acid) cannot be effectively processed, which suggests that N-formylmethionine can act as a signal for degradation (Piatkov et al. 2015; for discussion, see Dohmen 2015. This was probably due to abolished peptide deformylase activity by the presence of a negatively charged residue in the second position (Hu et al. 1999). We did not test binding of N-formylmethionine bearing GFP variants to AtClpS1; however, this kind of modification is also not expected to be targeted by the Clp system, but probably by the FtsH complex (Giglione et al. 2003, Adam et al. 2011, Piatkov et al. 2015).

We sought to investigate the role of Met40 in EcClpS and Arg50 in AtClpS1, given the recognized importance of the methionine residue in bacterial ClpS and the strict conservation of both amino acids across the respective clades. Replacement of the gatekeeper methionine residue in bacterial ClpS by alanine enlarges the cavity opening, and so this protein was able to accommodate the bulky side chain of tryptophan better than wild-type EcClpS (Schuenemann et al. 2009). Based on experimental data and structural modeling of the M40A mutants, these authors concluded that Met40 is a key residue for ClpS substrate specificity. Also, its substitution by the smaller alanine residue also allows for the recognition of branched amino acids. This is the case for the mutant ClpSM53A from *C. crescentus* which can recognize isoleucine and valine, thus expanding the repertoire of targets (Wang et al. 2008). Our experiments using EcClpSM40R confirm the importance of Met40 as a discriminator in substrate selection, as this mutant could no longer bind to FR–GFP and promoted the proteolysis of the other substrates with low efficiency. It could be possible that the guanidinium group of the arginine in the cavity entrance may disfavor binding of bulky aromatic residues in the first position. Experimentally, the interaction with leucine-bearing substrates was less affected, as seen in the substrate degradation experiments and in the peptide library binding assays. Also, the presence of the positive charge of arginine may cause an electrostatic repulsion with substrates bearing positive charges in the second position which could impair efficient binding and subsequent degradation by the proteolytic system, as was the case for the substrates FR–GFP and YKGEQ–GFP in the degradation assays.

The prediction that the arginine residue in the cavity entrance may impose steric hindrance to bulky residues and repulsion to positively charged N-degrons can also be extended to AtClpS1 and may explain its lower affinity towards the bacterial model substrates. For example, the GFP variant starting with tryptophan was the least affected in the competition assays with EcClpS + ClpAP using AtClpS1 as competitor. To help understand better the role of Arg50 in AtClpS1, we modeled it in silico to compare and contrast its substrate-binding cavities with those of EcClpS, using the EcClpS structure PDB ID: 3O2B as a template (Roman-Hernandez et al. 2011). The global structures and the position of key residues involved in substrate recognition largely overlap (**Fig. 6A–C**). Comparison of the surface charge area of the binding pocket of both proteins revealed that in the case of EcClpS, the cavity where the N-degron binds is hydrophobic and close to a negative distribution of charges contributed by Asp36 and Glu41 (**Fig. 6D**). Instead, the pocket of AtClpS1, while also hydrophobic, is surrounded by positive charges due to the presence of Lys49 and Arg50 (**Fig. 6E**). Our study shows that the arginine to methionine substitution in the back-to-bacterial consensus mutant AtClpS1R50M allowed it to associate strongly with GFP variants in the substrate competition assays. This stronger interaction could also be detected in ITC determinations. K_d values for each GFP variant with AtClpS1R50M were in close agreement with those for bacterial ClpS, suggesting that the binding interaction is greatly influenced by the gatekeeper residue, and other differences in the substrate-binding site between bacterial and plastid ClpS proteins are not as important. Interestingly, a few α-proteobacteria (such as *Agrobacterium tumefaciens*, *Mesorhizobium loti* and *Rhodopseudomonas palustris*) also possess a second ClpS protein (called ClpS2) in which residues of the binding pocket are highly conserved, except for the gatekeeper methionine that is changed to arginine, as in plants (Lupas and Koretke 2003). A recent study comparing ClpS1 and ClpS2 from *A. tumefaciens* shows that while ClpS1 behaves exactly like other bacterial ClpS proteins (i.e. it recognizes aromatic residues and leucine), ClpS2 has a more stringent specificity, recognizing tyrosine and tryptophan with weaker affinity (Stein et al. 2016). According to these authors, the arginine residue clashes with substrates, starting with tyrosine, and charge repulsion also exists with the positively charged α-amino group of the N-degron. This observation gives support to the hypothesis that charge repulsion and steric clashes weaken bacterial N-degron binding in the presence of arginine in the binding pocket.

The N-terminal ends of two suggested substrates of AtClpS1 (PYROX and GluTR) were identified by mass spectrometry identification of proteins bound to a GST–AtClpS1-decorated glutathione agarose resin (Nishimura et al. 2013). The N-terminal end of GluTR apparently starts with glutamic acid and is followed by

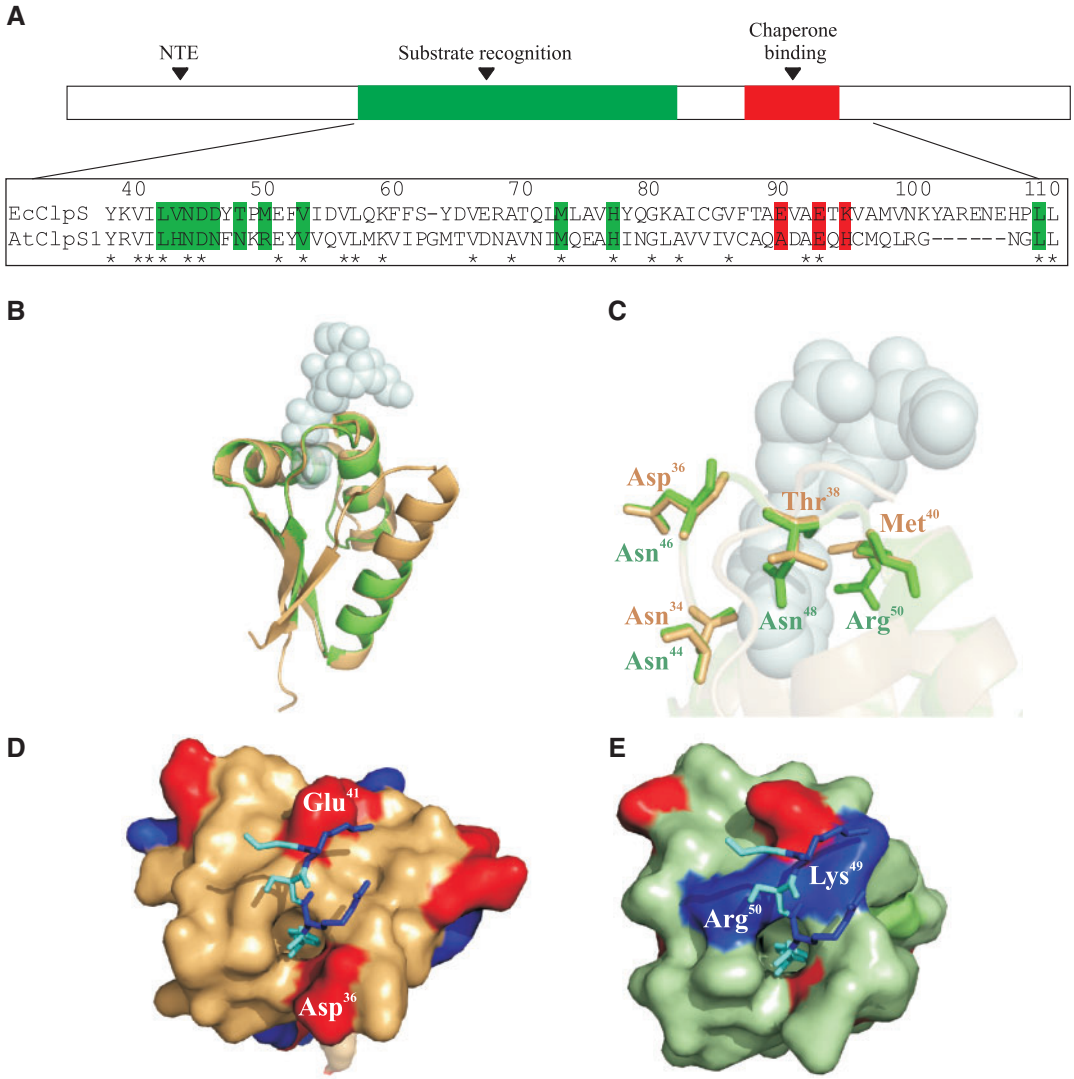

Fig. 6 Sequence conservation and comparative protein structure modeling of AtClpS1 and EcClpS. (A) Amino acid sequence alignment of AtClpS1 and EcClpS. Numbering is depicted according to the *Arabidopsis thaliana* protein. Amino acids important for substrate recognition and binding to chaperones are colored in green and red, respectively. Asterisks indicate amino acid conservation. (B) Overlapping structures of EcClpS (ribbon, beige, PDB ID: 3O2B) and AtClpS1 (ribbon, green, modeled in this work) with the sequence FR of the bound substrate (balls, light blue). (C) Close-up view of the binding pocket of ClpS proteins. Key residues involved in substrate recognition in EcClpS (beige) and the amino acids present in the same position in AtClpS1 (green) are indicated. (D and E) Molecular surface representations of EcClpS (D) and AtClpS1 (E) with the sequence FRSKG of the bound substrate (sticks) are colored to indicate charge distribution (blue, positive; red, negative).

a leucine, and PYROX starts with serine and is followed by alanine. These sequences differ from the sequence determinants of the bacterial N-end rule. However, this experiment was performed by co-incubation of stroma extracts with the AtClpS1-containing resin, so other adaptors (such as ClpF) were likely to be present. It has been recently found that AtClpS1 interacts with ClpF to deliver GluTR to the ClpCPR protease in vitro (Nishimura et al. 2015). So, this novel combined complex (which seems exclusive to photosynthetic eukaryotes) may have different substrate specificity from AtClpS1 acting alone. The binary complex of AtClpS1 and ClpF adds a new layer of complexity for N-degron recognition. How AtClpS1 and/or ClpF identify N-degrons warrants ultrastructural investigations on the 3-D conformation of AtClpS1 (and in concert with ClpF) with bound bona fide substrates, which will reveal the contributions of the different parts of the N-degron in the process. Actually, a recent large-scale survey of the degradation rates of over a thousand proteins from *A. thaliana* did not find any link between the identity of the amino acid in the first position of a protein and its half-life (Li et al. 2017). This highlights the fact that an N-degron is a complex structural feature in which the N-terminal amino acid plays only one part. Future studies will shed light on the molecular aspects of substrate selection of the Clp machinery, a key player in the process of plastid proteolysis.

Materials and Methods

Plasmid construction

AtClpS1 cDNA was obtained from the Riken cDNA bank (At1g68660, pda: 0148). The region coding for the mature protein was amplified and cloned

into the *Nco*I/*Eco*RI sites of a modified pET28a(+) plasmid which was described previously in Colombo et al. (2014), where the protein can be expressed as a fusion to a thrombin-cleavable C-terminal hexa-histidine tag. The primers used were: forward 5′-GT*CCATGG*CTGCTGCATTGGGTAAAGG-3′; and reverse 5′-TAA*GAATTC*TGCGCAGCCTCCACCATCAGG-3′ (restriction sites for *Nco*I and *Eco*RI, respectively, are in italics). After tag removal, recombinant mature AtClpS1 would bear eight extra amino acids at the C-terminus (AAEFLVPR). The construct was checked by DNA sequencing and the plasmid was submitted to the addgene repository (plasmid #98908). EcClpS and the GFP model substrates were obtained as previously described (Dougan et al. 2002, Schmidt et al. 2009). The mutant proteins AtClpS1R50M and EcClpSM40R were generated by site-directed mutagenesis of their corresponding expression plasmids.

Protein purification

The resulting plasmids containing AtClpS1 or AtClpS1R50M were transformed into the *E. coli* BL21(DE3) Codon Plus-RIL strain (Novagen). Cells were grown in 1 liter of Luria–Bertani medium at 37ºC until an optical density at 600 nm of 0.6–0.7 was reached. The temperature was lowered to 25ºC, and the inducer isopropyl-β-D-thiogalactopyranoside (IPTG) was added to a final concentration of 0.5 mM. After 6 h, cells were harvested by centrifugation and resuspended in cold lysis buffer [50 mM Tris–HCl pH 8.0, 400 mM NaCl, 1 mM benzamidine, 10% (v/v) glycerol] at a 25:1 ratio (ml of culture:ml of buffer). The cells were lysed by two passages through a French press (Aminco) and the soluble fraction was recovered by centrifugation (30,000×*g*, 1 h). The soluble fraction recovered was supplemented with 1 ml of equilibrated Ni^{2+}-NTA–agarose resin (Qiagen) and incubated for 1 h. The mixture was transferred to a column and washed with 30 column volumes of lysis buffer supplemented with 20 mM imidazole. Lysis buffer plus 250 mM imidazole was used to elute the recombinant protein in 100 µl fractions. These fractions were desalted by dialysis using a 3000 Da cut-off membrane against dialysis buffer [50 mM Tris–HCl pH 8.0, 100 mM NaCl, 10% (v/v) glycerol] for 8 h. To remove the hexa-histidine tag, 1 mg of recombinant protein was incubated in the presence of 3 U of thrombin at 10ºC for 16 h. The preparations were then loaded onto an Ni^{2+}-NTA–agarose column to remove free tags and undigested protein. Next, samples were subjected to SEC using a Superdex 75 column (see below). After this final step, a purity level of at least 97% was reached, as assayed by Coomassie-stained 15% gels under denaturing conditions. Molecular weight markers were from GE Healthcare (Low molecular weight calibration kit for SDS electrophoresis). Protein concentration was determined by the Bradford method using bovine serum albumin (BSA) as a standard protein.

The *E. coli* proteins ClpA, ClpP and EcClpS were purified as described (Erbse et al. 2006). EcClpSM40R was purified with the same protocol used for EcClpS. The GFP variant proteins were expressed and purified as described (Schmidt et al. 2009).

Gel filtration chromatography

Purified protein samples were loaded onto a Superdex 75 10/300 GL column (GE Healthcare) attached to an Äkta Prime chromatography system. Runs were performed at a flow rate of 0.5 ml min^{-1} using a degassed buffer made of 50 mM Tris–HCl pH 8.0 and 100 mM NaCl. Molecular weight standards were used to calibrate the column (MWGF1000 kit for molecular weights 29,000–700,000 and aprotinin, Sigma-Aldrich).

MALDI-TOF mass spectrometry

Determination of the exact molecular weight for proteins was carried out by MALDI-TOF mass spectrometry at the Analytical Biochemistry and Proteomics Unit, Institut Pasteur Montevideo, Uruguay. An aliquot of purified protein was desalted using C4-packed ZipTips (Millipore) according to the manufacturer's recommendations. Samples were mixed with an equal volume of a saturated solution of sinapinic acid (Sigma Aldrich) dissolved in 50% (v/v) acetonitrile in water with 1% (v/v) formic acid, and spotted onto a MALDI plate. The mass spectrometer (ABSciex 4800) was operated in linear mode. Results were analyzed with mMass 3 (Strohalm et al. 2010).

Cross-linking assays

The cross-linker EGS was dissolved in dimethyl sulfoxide at a concentration of 25 mM. Cross-linking experiments were carried out in a reaction mixture containing 50 mM HEPES pH 7.5, 100 mM NaCl and 25 µM of the corresponding protein or interacting mixture. EGS was added to the reaction mixture at different molar ratios, specified in each case. The reactions were incubated for 30 min at room temperature. Then, the cross-linking reaction was quenched by the addition of 50 mM Tris–HCl pH 7.5. Samples were analyzed by SDS–PAGE and Western blotting. For Western blots, nitrocellulose sheets were used. Membrane blocking was performed with 5% (w/v) non-fat dry milk in Tris-buffered saline (TBS), pH 7.4 for 1 h. Commercial antibodies produced in rabbits against the hexa-histidine tag (Santa Cruz Biotechnology, catalog No. sc-803, working dilution 1/1,000), GFP (Abcam, catalog No. ab290, working dilution 1/5,000) and alkaline phosphatase-conjugated mouse anti-rabbit IgG (Sigma-Aldrich, catalog No. A3687, working dilution 1/30,000) were used for protein detection. Anti-AtClpS1 antibodies were produced in house (see below). In all cases, the immune complexes were detected by 5-bromo-4-chloro-3-indolyl phosphate/nitro blue tetrazolium chloride (BCIP/NBT) substrate buffer [100 mM Tris–HCl pH 9.5, 100 mM NaCl, 5 mM $MgCl_2$, containing 0.03% (w/v) NBT and 0.015% (w/v) BCIP] and the reaction was stopped by washing with distilled water.

For the production of anti-AtClpS1 antibodies, rabbits were inoculated with purified recombinant AtClpS1. Animals were housed and maintained according to the Faculty of Biochemical and Pharmaceutical Sciences (UNR) experimental guidelines for animal studies; approval of the procedure was granted by the Faculty Ethics Committee. Specific antibodies against AtClpS1 were affinity purified from the serum by using the recombinant protein immobilized onto a nitrocellulose sheet. After binding, the loaded sheet was blocked as described above and then, extensively washed with TBS. The serum was then incubated with the membrane for 1 h at room temperature, which was then extensively washed to separate non-specific antibodies. Finally, anti-AtClpS1-specific antibodies were eluted with 0.1 M glycine-HCl, pH 2.8 for 2–3 min and neutralized to pH 8.0 with NaOH. Purified antibody specificity was tested by immunoblotting assays using purified AtClpS1 and protein extracts of *E. coli* cells producing recombinant AtClpS1. The antibodies were subsequently used at a working dilution of 1/1000.

In vitro degradation assays

Degradation of GFP variants (0.2 µM) was carried out by monitoring GFP fluorescence (excitation at 400 nm, emission at 510 nm) on an LS50B luminescence spectrometer (Perkin-Elmer). The buffer for the reactions contained 50 mM Tris–HCl pH 7.5, 300 mM NaCl, 20 mM $MgCl_2$, 2 mM dithiothreitol (DTT) and 2.5 mM ATP. ClpA, EcClpS, EcClpSM40R and ClpP were used at concentrations of 1 µM, while fluorescent substrates were at 0.2 µM. In substrate competition assays, competitors were added to the mixture at 20 µM (AtClpS1) or at 20, 3 and 0.2 µM (AtClpS1R50M). Each experiment was made at least in triplicate.

Isothermal titration calorimetry assays

ITC experiments were performed using a MicroCal VP-ITC-type microcalorimeter (MicroCal Inc.) at 25°C. All solutions were thoroughly degassed before use by stirring under vacuum. Protein samples (AtClpS1, the R50M mutant and the fluorescent probes) were prepared in the same buffer (50 mM sodium phosphate, pH 7.5, 50 mM NaCl). This buffer was used to load the reference cell in all cases. A typical titration experiment consisted of consecutive injections of 25 µl of 37 µM AtClpS1 or AtClpS1R50M in nine steps, at 6 min intervals into the titration cell, which was fully loaded with 1 µM GFP variant. The titration data were corrected for the small heat changes observed in the control titrations of each AtClpS1 protein (wild type or R50M mutant) into the buffer. Also, an initial 2 µl injection was discarded from each data set in order to remove the effect of titrant diffusion across the syringe tip during the equilibration process. Data analysis was performed with Origin 7.0 software, provided by MicroCal, using equations and curve-fitting analysis with the OneSites model to obtain least-square estimates of the binding enthalpy and binding constant. The weakness of the observed binding events required that the binding ratio be fixed at 1:1 before curve fitting of the thermogram. Experiments were performed in duplicate.

Screening of cellulose-bound peptide scans for EcClpS protein binding

λcI peptide libraries were prepared by automated spot synthesis by JPT Peptide Technologies GmbH (PepSpots™). The libraries were composed of 13-mer

peptides scanning the primary sequence of phage λ repressor protein cI (236 amino acids without the initial methionine, accession No. P03034) with an overlap of 10 residues between neighboring spots, giving rise to 76 distinct binding sites. Peptides were C-terminally attached to cellulose membranes via a $(\beta\text{-Ala})_2$ spacer. The overall variation of peptide loading between peptide spots is <25%, according to the manufacturer (personal communication). The quality control for peptide loading includes colorimetric reactions and measurements by MALDI-TOF mass spectrometry. Before screening, the dry membranes were washed in methanol for 2× 5 min, 3× 10 min in TBS and 1× 10 min in buffer A [10 mM Tris pH 7.5, 150 mM KCl, 20 mM $MgCl_2$, 5% (w/v) sucrose, 0.005% (v/v) Tween-20]. EcClpS or EcClpSM40R (500 nM) were incubated with the libraries for 30 min. Afterwards, buffer A was discarded and replaced by cold TBS. Fractionated Western blotting (4× 30 min using fresh blotting buffer after the second blot) allowed for transfer of EcClpS (wild type or R50M mutant) bound to peptide spots onto polyvinylidene difluoride (PVDF) membranes. Membranes were blocked with 5% (w/v) non-fat dry milk in TBS and then incubated with EcClpS-specific antibodies (1:5,000) (Dougan et al. 2002) and anti-rabbit IgG–alkaline phosphatase (1:20,000). Blots were scanned using ImageQuant LAS 4000 (GE Healthcare). Binding was considered positive if signal was detected above the background. The libraries were first tested with EcClpSM40R, regenerated after transfer as recommended by the vendor and then probed with EcClpS. Briefly, regeneration was performed by incubation with Tris buffer (62.5 mM, pH 6.7) containing 2% (w/v) SDS and 100 mM β-mercaptoethanol at 50°C four times, for 30 min each. The membrane was subsequently incubated in 10× phosphate-buffered saline (PBS; 3× 20 min), followed by a washing step in TBS supplemented with 0.05% (v/v) Tween-20 (20 min) and TBS (3× 10 min). Three λcI peptide libraries were independently incubated with both EcClpS wild type and EcClpSM40R to control for different peptide densities. EcClpS and EcClpM40R showed the same reactivity towards the polyclonal (rabbit) anti-EcClpS serum (Supplementary Fig. S4A).

Homology modeling of AtClpS1 structure

Pairwise sequence alignment was performed with ClustalX V2 (Larkin et al. 2007). Homology modeling of AtClpS1 was performed using the X-ray structure of EcClpS (PDB ID: 3O2B) as template. The mature sequence of AtClpS1 (NP_564937) was obtained from the NCBI protein database. The structure prediction was performed using the Swiss Model web server (Arnold et al. 2006). The quality of the obtained model was 'very good model' as indicated by the ProQ-Protein Quality Predictor (Wallner and Elofsson 2003). Structures were fit, superimposed and produced using PyMOL Version 1.8.0.0, Schrödinger, LLC.

Supplementary Data

Supplementary data are available at PCP online.

Funding

This work was supported by the Consejo Nacional de Investigaciones Científicas y Técnicas [grant No. PIP0345 to E.A.C.]; the Agencia Nacional de Promoción Científica y Tecnológica [grant Nos. PICT 2015-2955 to E.A.C. and PICT 2014-0825 to G.L.R]; the Deutsche Forschungsgemeinschaft [grant No. MO 970/4-2 to A.M.]; and the Deutscher Akademischer Austausch Dienst [to C.V.C.]

Acknowledgments

G.L.R. and E.A.C. are staff members of the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET, Argentina). C.V.C. is a fellow of that Institution. Also, G.L.R. is a Teaching Assistant and E.A.C. is a Professor of the Facultad de Ciencias Bioquímicas y Farmacéuticas, UNR, Argentina. We thank the staff of the Instituto de Investigaciones para el Descubrimiento de Fármacos de Rosario (Max Planck-IIDEFAR) for the use of the isothermal titration calorimeter, and especially Dr. María Eugenia Chesta for her technical help. We also thank Regina Zahn for technical assistance.

Disclosures

The authors have no conflicts of interest to declare.

References

Adam, Z., Frottin, F., Espagne, C., Meinnel, T. and Giglione, C. (2011) Interplay between N-terminal methionine excision and FtsH protease is essential for normal chloroplast development and function in *Arabidopsis*. *Plant Cell* 23: 3745–3760.

AhYoung, A.P., Koehl, A., Vizcarra, C.L., Cascio, D. and Egea, P.F. (2016) Structure of a putative ClpS N-end rule adaptor protein from the malaria pathogen *Plasmodium falciparum*. *Protein Sci.* 25: 689–701.

Apel, W., Schulze, W.X. and Bock, R. (2010) Identification of protein stability determinants in chloroplasts. *Plant J.* 63: 636–650.

Arnold, K., Bordoli, L., Kopp, J. and Schwede, T. (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* 22: 195–201.

Bachmair, A., Finley, D. and Varshavsky, A. (1986) In vivo half-life of a protein is a function of its amino-terminal residue. *Science* 234: 179–186.

Chen, S.J., Wu, X., Wadas, B., Oh, J.H. and Varshavsky, A. (2017) An N-end rule pathway that recognizes proline and destroys gluconeogenic enzymes. *Science* 355: eaal3655.

Colombo, C.V., Ceccarelli, E.A. and Rosano, G.L. (2014) Characterization of the accessory protein ClpT1 from *Arabidopsis thaliana*: oligomerization status and interaction with Hsp100 chaperones. *BMC Plant Biol.* 14: 228.

Dohmen, R.J. (2015) Starting with a degron: N-terminal formyl-methionine of nascent bacterial proteins contributes to their proteolytic control. *Microb. Cell* 2: 356–359.

Dougan, D.A., Micevski, D. and Truscott, K.N. (2012) The N-end rule pathway: from recognition by N-recognins, to destruction by AAA+ proteases. *Biochim. Biophys. Acta* 1823: 83–91.

Dougan, D.A., Reid, B.G., Horwich, A.L. and Bukau, B. (2002) ClpS, a substrate modulator of the ClpAP machine. *Mol. Cell* 9: 673–683.

Emanuelsson, O., Nielsen, H. and von Heijne, G. (1999) ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Sci.* 8: 978–984.

Erbse, A., Schmidt, R., Bornemann, T., Schneider-Mergener, J., Mogk, A. and Zahn, R. (2006) ClpS is an essential component of the N-end rule pathway in *Escherichia coli*. *Nature* 439: 753–756.

Finley, D. (2009) Recognition and processing of ubiquitin–protein conjugates by the proteasome. *Annu. Rev. Biochem.* 78: 477–513.

Frottin, F., Martinez, A., Peynot, P., Mitra, S., Holz, R.C., Giglione, C, et al. (2006) The proteomics of N-terminal methionine cleavage. *Mol. Cell. Proteomics* 5: 2336–2349.

Gibbs, D.J., Bailey, M., Tedds, H.M. and Holdsworth, M.J. (2016) From start to finish: amino-terminal protein modifications as degradation signals in plants. *New Phytol.* 211: 1188–1194.

Giglione, C. and Meinnel, T. (2001) Organellar peptide deformylases: universality of the N-terminal methionine cleavage mechanism. *Trends Plant Sci.* 6: 566–572.

Giglione, C., Vallon, O. and Meinnel, T. (2003) Control of protein life-span by N-terminal methionine excision. *EMBO J.* 22: 13–23.

Gottesman, S., Roche, E., Zhou, Y. and Sauer, R.T. (1998) The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. *Genes Dev.* 12: 1338–1347.

Guo, F., Esser, L., Singh, S.K., Maurizi, M.R. and Xia, D. (2002) Crystal structure of the heterodimeric complex of the adaptor, ClpS, with the N-domain of the AAA+ chaperone, ClpA. *J. Biol. Chem.* 277: 46753–46762.

Hu, Y.J., Wei, Y., Zhou, Y., Rajagopalan, P.T. and Pei, D. (1999) Determination of substrate specificity for peptide deformylase through the screening of a combinatorial peptide library. *Biochemistry* 38: 643–650.

Hwang, C.S., Shemorry, A. and Varshavsky, A. (2010) N-terminal acetylation of cellular proteins creates specific degradation signals. *Science* 327: 973–977.

Ishikawa, M., Fujiwara, M., Sonoike, K. and Sato, N. (2009) Orthogenomics of photosynthetic organisms: bioinformatic and experimental analysis of chloroplast proteins of endosymbiont origin in Arabidopsis and their counterparts in Synechocystis. *Plant Cell Physiol.* 50: 773–788.

Kirstein, J., Molière, N., Dougan, D.A. and Turgay, K. (2009) Adapting the machine: adaptor proteins for Hsp100/Clp and AAA+ proteases. *Nat. Rev. Microbiol.* 7: 589–599.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H, et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23: 2947–2948.

Li, L., Nelson, C.J., Trosch, J., Castleden, I., Huang, S. and Millar, A.H. (2017) Protein degradation rate in *Arabidopsis thaliana* leaf growth and development. *Plant Cell* 29: 207–228.

Lies, M. and Maurizi, M.R. (2008) Turnover of endogenous SsrA-tagged proteins mediated by ATP-dependent proteases in *Escherichia coli*. *J. Biol. Chem.* 283: 22918–22929.

Lupas, A.N. and Koretke, K.K. (2003) Bioinformatic analysis of ClpS, a protein module involved in prokaryotic and eukaryotic protein degradation. *J. Struct.Biol.* 141: 77–83.

Nishimura, K., Apitz, J., Friso, G., Kim, J., Ponnala, L., Grimm, B, et al. (2015) Discovery of a unique Clp component, ClpF, in chloroplasts: a proposed binary ClpF–ClpS1 adaptor complex functions in substrate recognition and delivery. *Plant Cell* 27: 2677–2691.

Nishimura, K., Asakura, Y., Friso, G., Kim, J., Oh, S.H., Rutschow, H, et al. (2013) ClpS1 is a conserved substrate selector for the chloroplast Clp protease system in *Arabidopsis*. *Plant Cell* 25: 2276–2301.

Nishimura, K. and van Wijk, K.J. (2015) Organization, function and substrates of the essential Clp protease system in plastids. *Biochim. Biophys. Acta* 1847: 915–930.

Olinares, P.D., Kim, J., Davis, J.I. and van Wijk, K.J. (2011) Subunit stoichiometry, evolution, and functional implications of an asymmetric plant plastid ClpP/R protease complex in *Arabidopsis*. *Plant Cell* 23: 2348–2361.

Piatkov, K.I., Vu, T.T., Hwang, C.S. and Varshavsky, A. (2015) Formyl-methionine as a degradation signal at the N-termini of bacterial proteins. *Microb. Cell* 2: 376–393.

Roman-Hernandez, G., Grant, R.A., Sauer, R.T. and Baker, T.A. (2009) Molecular basis of substrate selection by the N-end rule adaptor protein ClpS. *Proc. Natl. Acad. Sci. USA* 106: 8888–8893.

Roman-Hernandez, G., Hou, J.Y., Grant, R.A., Sauer, R.T. and Baker, T.A. (2011) The ClpS adaptor mediates staged delivery of N-end rule substrates to the AAA+ ClpAP protease. *Mol. Cell* 43: 217–228.

Rowland, E., Kim, J., Bhuiyan, N.H. and van Wijk, K.J. (2015) The *Arabidopsis* chloroplast stromal N-terminome: complexities of amino-terminal protein maturation and stability. *Plant Physiol.* 169: 1881–1896.

Sauer, R.T. and Baker, T.A. (2011) AAA+ proteases: ATP-fueled machines of protein destruction. *Annu. Rev. Biochem.* 80: 587–612.

Schmidt, R., Zahn, R., Bukau, B. and Mogk, A. (2009) ClpS is the recognition component for *Escherichia coli* substrates of the N-end rule degradation pathway. *Mol. Microbiol.* 72: 506–517.

Schuenemann, V.J., Kralik, S.M., Albrecht, R., Spall, S.K., Truscott, K.N., Dougan, D.A, et al. (2009) Structural basis of N-end rule substrate recognition in *Escherichia coli* by the ClpAP adaptor protein ClpS. *EMBO Rep*. 10: 508–514.

Snider, J. and Houry, W.A. (2008) AAA+ proteins: diversity in function, similarity in structure. *Biochem. Soc. Trans.* 36: 72–77.

Stein, B.J., Grant, R.A., Sauer, R.T. and Baker, T.A. (2016) Structural basis of an N-degron adaptor with more stringent specificity. *Structure* 24: 232–242.

Strohalm, M., Kavan, D., Novák, P., Volný, M. and Havlícek, V. (2010) mMass 3: a cross-platform software environment for precise analysis of mass spectrometric data. *Anal. Chem.* 82: 4648–4651.

Tan, J.L., Ward, L., Truscott, K.N. and Dougan, D.A. (2016) The N-end rule adaptor protein ClpS from *Plasmodium falciparum* exhibits broad substrate specificity. *FEBS Lett.* 590: 3397–3406.

Tellinghuisen, J. (2008) Isothermal titration calorimetry at very low c. *Anal. Biochem.* 373: 395–397.

Tryggvesson, A., Stahlberg, F.M., Topel, M., Tanabe, N., Mogk, A. and Clarke, A.K. (2015) Characterization of ClpS2, an essential adaptor protein for the cyanobacterium *Synechococcus elongatus*. *FEBS Lett*. 589: 4039–4046.

Turnbull, W.B. and Daranas, A.H. (2003) On the value of c: can low affinity systems be studied by isothermal titration calorimetry? *J. Amer. Chem. Soc.* 125: 14859–14866.

Varshavsky, A. (2011) The N-end rule pathway and regulation by proteolysis. *Protein Sci.* 20: 1298–1345.

Wallner, B. and Elofsson, A. (2003) Can correct protein models be identified? *Protein Sci.* 12: 1073–1086.

Wang, K.H., Roman-Hernandez, G., Grant, R.A., Sauer, R.T. and Baker, T.A. (2008) The molecular basis of N-end rule recognition. *Mol. Cell* 32: 406–414.

Zeth, K., Dougan, D.A., Cusack, S., Bukau, B. and Ravelli, R.B. (2002a) Crystallization and preliminary X-ray analysis of the *Escherichia coli* adaptor protein ClpS, free and in complex with the N-terminal domain of ClpA. *Acta Crystallogr. D Biol. Crystallogr.* 58: 1207–1210.

Zeth, K., Ravelli, R.B., Paal, K., Cusack, S., Bukau, B. and Dougan, D.A. (2002b) Structural analysis of the adaptor protein ClpS in complex with the N-terminal domain of ClpA. *Nat. Struct. Biol.* 9: 906–911.