

# Structural basis for the N-degron specificity of ClpS1 from *Arabidopsis thaliana*

Leehyeon Kim<sup>1</sup>  | Jiwon Heo<sup>1</sup> | Do Hoon Kwon<sup>1</sup>  | Jin Seok Shin<sup>1</sup> |  
Se Hwan Jang<sup>2</sup> | Zee-Yong Park<sup>2</sup> | Hyun Kyu Song<sup>1</sup> 

<sup>1</sup>Department of Life Sciences, Korea University, Seoul, South Korea

<sup>2</sup>School of Life Sciences, Gwangju Institute of Science and Technology, Gwangju, South Korea

## Correspondence

Hyun Kyu Song, Department of Life Sciences, Korea University, 145 Anam-ro, Seongbuk-gu, Seoul 02841, South Korea.  
Email: hksong@korea.ac.kr

## Present address

Do Hoon Kwon, Department of Biochemistry, Duke University School of Medicine, Durham, North Carolina, USA.

## Funding information

National Research Foundation of Korea, Grant/Award Numbers: 2020R1A2C3008285, 2020R1A5A1019023

## Abstract

The N-degron pathway determines the half-life of proteins in both prokaryotes and eukaryotes by precisely recognizing the N-terminal residue (N-degron) of substrates. ClpS proteins from bacteria bind to substrates containing hydrophobic N-degrons (Leu, Phe, Tyr, and Trp) and deliver them to the caseinolytic protease system ClpAP. This mechanism is preserved in organelles such as mitochondria and chloroplasts. Bacterial ClpS adaptors bind preferentially to Leu and Phe N-degrons; however, ClpS1 from *Arabidopsis thaliana* (AtClpS1) shows a difference in that it binds strongly to Phe and Trp N-degrons and only weakly to Leu. This difference in behavior cannot be explained without structural information due to the high sequence homology between bacterial and plant ClpS proteins. Here, we report the structure of AtClpS1 at 2.0 Å resolution in the presence of a bound N-degron. The key determinants for  $\alpha$ -amino group recognition are conserved among all ClpS proteins, but the  $\alpha$ 3-helix of eukaryotic AtClpS1 is significantly shortened, and consequently, a loop forming a pocket for the N-degron is moved slightly outward to enlarge the pocket. In addition, amino acid replacement from Val to Ala causes a reduction in hydrophobic interactions with Leu N-degron. A combination of the fine-tuned hydrophobic residues in the pocket and the basic gatekeeper at the entrance of the pocket controls the N-degron selectivity of the plant ClpS protein.

## KEYWORDS

ClpS, complex structure, N-degron pathway, N-end rule, plant chloroplast, type-2 substrate, X-ray crystallography

## 1 | INTRODUCTION

The amount of proteins in the cells is elaborately regulated to maintain cellular homeostasis. A mechanism for this precise regulation in eukaryotes is the ubiquitin/proteasome system (UPS),<sup>1,2</sup> and bacteria utilize different degradation systems.<sup>3–5</sup> One of the well-characterized degradation mechanisms is the N-degron

pathway (formerly N-end rule), which is a process that determines the half-life of proteins based on the N-terminal residue, called the “N-degron.”<sup>6–8</sup> The N-degrons are classified into two different types of residues and recognized by adaptor proteins (N-recognin), which deliver them to proteases for degradation.<sup>9–12</sup> This N-degron pathway is common in both eukaryotes and prokaryotes.<sup>13–17</sup> In eukaryotes, primary destabilizing

residues, type-1N-degrons (positively charged residues: Arg, Lys, and His), are recognized by the UBR box motif, and type-2N-degrons (bulky hydrophobic residues: Leu, Phe, Tyr, Trp, and Ile) are recognized by the ClpS-homology domain.<sup>7,8,18–21</sup> The bound substrate is then ubiquitylated and ultimately degraded by the 26S proteasome.<sup>2,22,23</sup> Furthermore, recent studies have reported that an autophagy adapter, SQSTM1/p62, recognizes both type-1 and type-2 substrates via its ZZ domain, and ultimately, cargo molecules are targeted to the lysosome for degradation.<sup>24–26</sup>

Unlike eukaryotes, bacteria do not possess a UPS; instead, the caseinolytic protease (Clp) system exists in bacteria as well as in organelles such as mitochondria and chloroplasts in eukaryotes.<sup>27–31</sup> The Clp protease is an energy-dependent protease system that consists of two distinct functional components, a barrel-shaped tetradecameric serine protease, ClpP, and hexameric AAA+ chaperones, ClpA, ClpC, ClpE, and ClpX.<sup>5,27,31–34</sup> This two-component system shares some characteristics with the 26S proteasome.<sup>35–37</sup> The ClpS adaptor binds and delivers N-degron substrates for degradation upon association with the ClpAP proteolytic machine.<sup>4,38–41</sup> This ClpS adaptor has been studied extensively in *Escherichia coli* and some other bacterial systems.<sup>4,21,38,42–44</sup> Structural and biochemical studies of bacterial ClpS revealed how it recognizes type-2N-degrons with a preference for Leu and Phe.<sup>38–40,42,43,45–47</sup> In prokaryotes, these primary destabilizing residues are attached by leucyl/phenylalanyl-tRNA-protein transferase (L/F-transferase).<sup>48</sup> However, the structure of the ClpS-homology domain in eukaryotic N-recognins, such as Ubr1 E3 ubiquitin ligase, remains elusive. Moreover, characterization of the N-degron pathways in mitochondria and chloroplasts is also quite limited.<sup>49–51</sup>

In contrast to the wealth of structural information on ClpS from bacteria, only one structure of eukaryotic ClpS from the protozoan parasite *Plasmodium falciparum* (PfClpS) is available.<sup>52</sup> This PfClpS is present in the apicoplast of the parasite, an organelle that is evolutionarily related to plastids/chloroplasts of algae. Intriguingly, in chloroplasts, the plastid Clp system contributes to embryogenesis, plastid biogenesis, and plant development.<sup>53–58</sup> Furthermore, ClpS1 in *Arabidopsis thaliana* (AtClpS1) showed a distinct N-degron specificity that is linked to the translational capacity of plants.<sup>59</sup> It has been reported that AtClpS1 prefers the bulky aromatic residues Phe and Trp at the N-terminus and shows weak binding to Leu and very weak or no binding to Ile and Tyr N-termini.<sup>60</sup> This is not easy to explain because of the high-sequence homology between bacterial and plant ClpS.<sup>47,60</sup> Therefore, we determined the structure of AtClpS1 in complex with N-degron and provided a

framework for understanding the substrate specificity of the N-degron pathway in higher eukaryote plant chloroplasts.

## 2 | RESULTS AND DISCUSSION

### 2.1 | Overall structure of AtClpS1

We determined the structure of AtClpS1 (residues 79–159) in complex with the Phe-Ala peptide at 2.0 Å resolution (Table 1). There are two molecules in the asymmetric unit; however, only one subunit (chain A) showed a clear electron density, while the other subunit (chain B) did not. Therefore, all structural comparisons and displays are performed with only the chain A model. It consists of three  $\alpha$ -helices and three  $\beta$ -strands in a cone shape (Figure 1a), and there is an obvious cavity on the surface that is able to accommodate the N-degron substrate (Figures 1b and S1a). Compared with other ClpS proteins, plant ClpS1 shares a very similar folding pattern, which might be expected based on the sequence identity of approximately 30% among them (Figure 1c). Among *E. coli* (Ec), *Caulobacter crescentus* (Cc), *Plasmodium falciparum* ClpS (PfClpS), and *Agrobacterium tumefaciens* ClpS2 (AtuClpS2), the structure of AtClpS1 is the most similar to that of PfClpS. The root-mean-square deviation (RMSD) for 73 matching C $\alpha$  atoms between AtClpS1 and PfClpS is only 0.961 Å (Figure 1d). Comparing the matching C $\alpha$  atoms of AtClpS1 with those of EcClpS (79 atoms), CcClpS (75 atoms), and AtuClpS2 (77 atoms) shows RMSDs below 1.5 Å, confirming that the structures of ClpS from all different species, including prokaryotes and eukaryotes, are very similar. However, there are unique structural features as well. In particular, the third  $\alpha$ -helix ( $\alpha$ 3) is shortened (Figure 1c,d), which can be inferred to have occurred during the evolutionary process. Another salient difference is that the entrance of the cavity has a positive charge, in contrast to other ClpS proteins, since the beginning of the first helix consists of Lys93 and Arg94 (Figures 1b and S1)

### 2.2 | The N-degron binding pocket of AtClpS1

To understand the substrate specificity of AtClpS1,<sup>60,61</sup> a high-resolution structure of AtClpS in the presence of bound peptide was needed. Initially, we tried to determine AtClpS1 in complex with an N-degron using the LC3B fusion technique by attaching the Leu residue at the N-terminal region of AtClpS1 for easier crystallization.<sup>62</sup> Unfortunately, the expected N-degron segment

**TABLE 1** Data collection and refinement statistics of the AtClpS1-peptide complex

AtClpS1-peptide complex	
<b>Data collection</b>	
X-ray source	Home (CuK $\alpha$ , $\lambda = 1.54178 \text{ \AA}$ )
Space group	$P 2_1 2_1 2$
Cell parameters [ $a, b, c$ ( $\text{\AA}$ )/ $\alpha, \beta, \gamma$ ( $^\circ$ )]	81.1, 38.0, 61.4/90, 90, 90
Resolution ( $\text{\AA}$ )	50.00–2.00 (2.04–2.00) <sup>a</sup>
$R_{\text{sym}}$	0.066 (0.387) <sup>a</sup>
$I/\sigma I$	47.8 (6.06) <sup>a</sup>
$CC_{1/2}$	0.995 (0.922) <sup>a</sup>
Completeness (%)	99.8 (100.0) <sup>a</sup>
Redundancy	6.3 (6.2) <sup>a</sup>
<b>Refinement</b>	
Resolution ( $\text{\AA}$ )	32.32–2.00 (2.08–2.00) <sup>a</sup>
No. reflections	12,932 (1,238) <sup>a</sup>
$R_{\text{work}}/R_{\text{free}}$	0.275/0.299
No. atoms	1,240
Protein	1,208 (chain A: 604; chain B: 604)
Ligands	20 (FA: 16; ACY: 4)
Water	12
B-factors ( $\text{\AA}^2$ )	41.28
Protein	41.48 (Chain A: 31.17; Chain B: 51.79)
Ligands	35.82 (FA: 34.35; ACY: 41.72)
Water	29.88
Wilson B-factor	31.45
<b>R.m.s deviations</b>	
Bond lengths ( $\text{\AA}$ )	0.010
Bond angles ( $^\circ$ )	1.30
<b>Ramachandran (%)</b>	
Favored	99.35
Allowed	0.65
Outliers	0.00
<b>PDB ID</b>	7d34

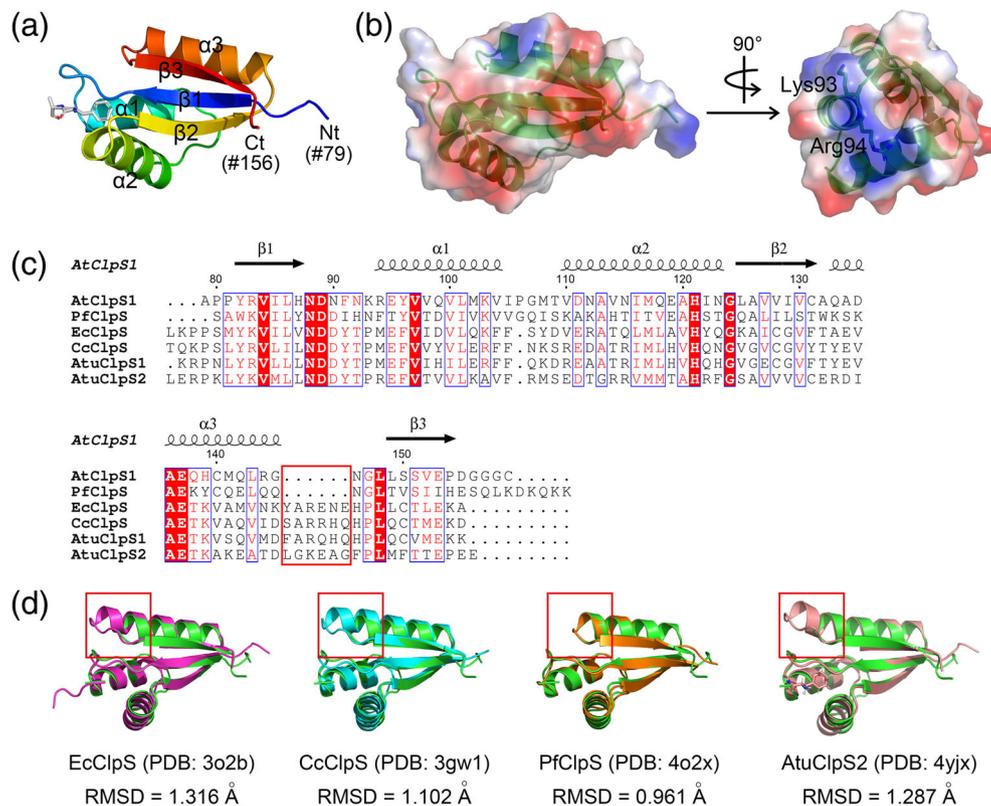
<sup>a</sup>Highest-resolution shell.

was not found in the binding pocket of AtClpS1 of the neighboring molecule in the crystalline lattice. To our surprise, our AtClpS1 structure has a clear electron density map at this binding site without exogenous addition (Figure S2), which must originate from the bacterial culture. For accurate identification of the unknown dipeptide-like ligand by qualitative analysis, we employed mass spectrometry for analysis (Figure S3). Interestingly,

the molecule was a mixture of several Phe-containing dipeptides. By careful investigation, we built the Phe-Ala peptide into the electron density map because we could not observe the electron density beyond the C $\beta$  atom, and it fits perfectly (Figure S2b). It already occupied the binding pocket, which partially explains why we could not obtain the Leu-bound AtClpS1 complex structure using the LC3B fusion technique. This is additional evidence that the binding affinity of AtClpS1 to Phe N-degron is stronger than that to Leu N-degron.<sup>60</sup> This is consistent with the known results that among type-2 substrates, phenylalanine fits the pocket best, binds the most strongly, and degrades fastest.<sup>42</sup> Strictly conserved polar side chains (Asn88 and His121) and main chains of Asn92 and Arg94 residues form hydrogen bonds with backbone atoms of the Phe-Ala peptide (Figure 2a,c), which is a conserved structural feature among the ClpS proteins.<sup>42,45,47,63</sup> The hydrophobic pocket (Leu86, Val97, Met117, Ala120, and Leu148) for recognizing the type-2N-degron is also well conserved (Figure 1c). The three-dimensional arrangement of the hydrophobic residues provides a suitable volume for accommodating the hydrophobic N-degrons (Figure S1). When the solvent-accessible area and volume were calculated with a probe radius of 1.4  $\text{\AA}$ ,<sup>64</sup> the cavity area and volume of AtClpS1 were larger than those of bacterial ClpS proteins but smaller than those of AtClpS2, which possesses stringent specificity (Figure S1).

### 2.3 | The binding specificity of AtClpS1

It is known that AtClpS1 is able to recognize type-2N-degrons; however, it has a different specificity for N-degrons, which must originate from its unique structural features. Competition in degradation assays reported that the binding efficiency of AtClpS1 to type-2N-degrons was very low and clearly demonstrated the role of the positively charged residue on the surface as a gatekeeper (Figure 1b).<sup>61</sup> Another study conducted binding assays with N-terminal-modified green fluorescence protein reporters, and the results showed that AtClpS1 binds strongly to N-degrons starting with Phe or Trp; however, very intriguingly, the favorable Leu N-terminus in bacterial ClpS is recognized only weakly by AtClpS1.<sup>60,63</sup> In an effort to quantitatively determine the binding affinity, we measured the  $K_D$  values between AtClpS1 and several type-2N-degron tripeptides using the ITC method (Figure S4). Peptides bearing phenylalanine or tryptophan at the N-terminus showed tight binding to AtClpS1, with dissociation constants  $K_D$  ranging from 1  $\mu\text{M}$  to 2  $\mu\text{M}$ , although purified AtClpS1 possessed a phenyl peptide at the binding site (Figures 2, S2, and S3). In contrast



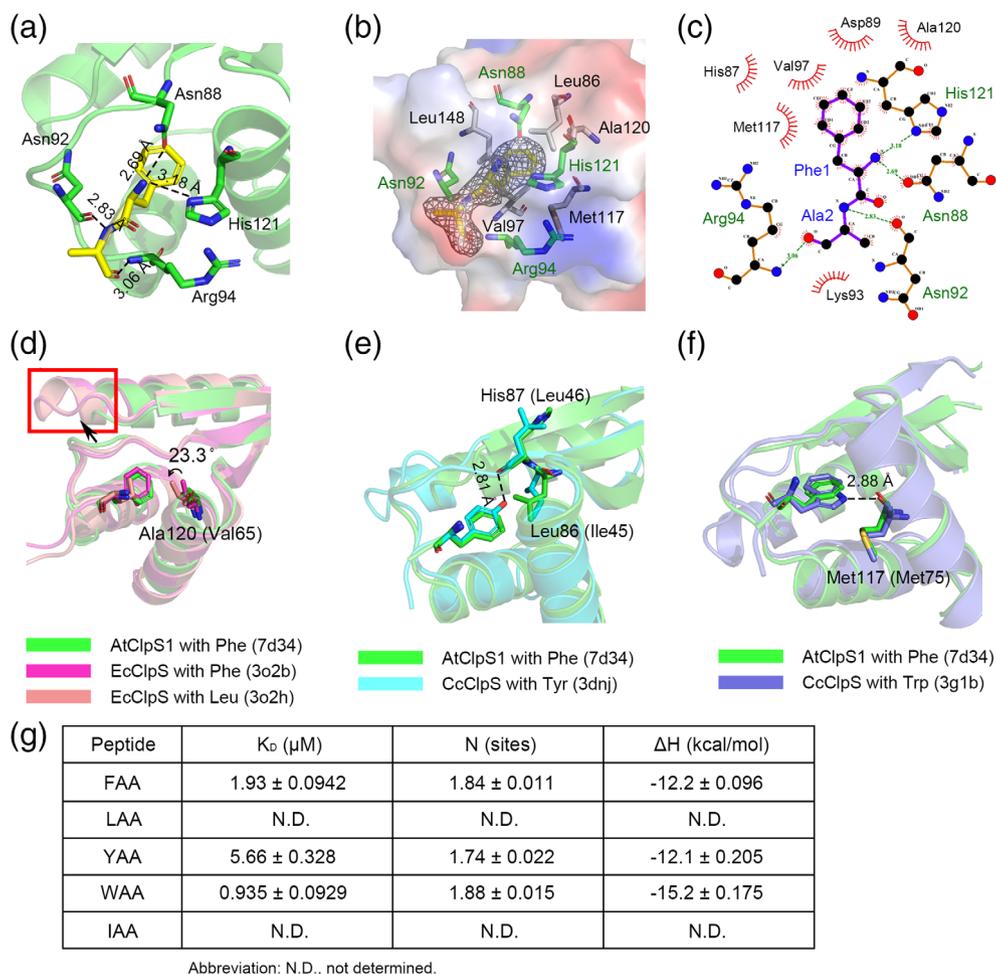
**FIGURE 1** Structure of AtClpS1. (a) Structure of the AtClpS1-peptide complex. Secondary structures are labeled in numerical order. N- and C-termini are indicated as Nt and Ct, respectively, with the residue number in parentheses, and the ribbon color changes from blue (Nt) to red (Ct) gradually. The bound N-degron (Phe-Ala) is shown in the stick model. (b) Transparent molecular surface with electrostatic potentials shows the distribution of positively and negatively charged surfaces, colored blue and red, respectively. The entrance of the N-degron binding pocket possesses positively charged residues, which are known as the gatekeeper. (c) Sequence alignment of ClpS structures in the Protein Data Bank (3o2b: *Escherichia coli* ClpS; 3gw1: *Caulobacter crescentus* ClpS; 4o2x: *Plasmodium falciparum* ClpS; and 4yjm: *Agrobacterium tumefaciens* ClpS2). The characteristic shortened helix in the eukaryotic ClpS proteins is boxed in red. (d) Comparison of the overall structure of AtClpS1 with structures of other ClpS proteins. The calculated root-mean-square deviation (RMSD) values between AtClpS1 and other ClpS proteins were 1.316 Å (EcClpS; 79 matching C $\alpha$  atoms), 1.102 Å (CcClpS; 75 matching C $\alpha$  atoms), 0.961 Å (PfClpS; 73 matching C $\alpha$  atoms), and 1.413 Å (AtuClpS2; 77 matching C $\alpha$  atoms). The red boxes represent the shortened  $\alpha$ 3-helix, which is the most structurally divergent region

to a previous report,<sup>60</sup> we also found relatively tight binding ( $\sim 5.7$   $\mu$ M) for the YAA peptide. However, LAA and IAA peptides showed there is no binding detectable at all, which is consistent with a previous report.

Hydrophobic residues located at the entrance of the pocket in bacterial ClpS were replaced with positively charged residues (Pro39 and Met40 in EcClpS to Lys93 and Arg94 in AtClpS1), which reduced the accessibility to the N-degrons containing a positively charged residue at the second position, which is known to be abundant in bacterial N-degrons because of the hierarchical enzymatic reaction of L/F-transferase.<sup>13,65</sup> However, the binding pocket of AtClpS1 is larger than that of bacterial ClpSs (Figure S1). The loop (residues Asn88-Asn92) forming a wall of the pocket is shifted outward (Figure 2d), presumably because of the shortened  $\alpha$ 3 helix (Figures 1c and 2d). Actually, this loop region might be flexible in

eukaryotic ClpS, and it occludes the cavity entrance in the closed conformation of PfClpS (Figure S1d).<sup>52</sup> In bacterial ClpSs, the side chain of Val65 showed different orientations depending on various N-degron complexes (Phe in PDB ID: 3o2b, Leu in PDB ID: 3o2h).<sup>63</sup> The valine is replaced with a smaller alanine in AtClpS1 (Figures 1c and 2d). The hydrophobic interaction between the valine and the N-terminal leucine of the N-degron might be weakened by the change to Ala120 in AtClpS1, which is a plausible explanation of the preference for Phe over Leu residue at the N-terminus.<sup>60</sup> Intriguingly, a similar preference was also found in AtuClpS2, which also possesses alanine at the equivalent position (Figure 1c).<sup>47</sup>

It is known that the  $\beta$ -branched hydrophobic residue isoleucine cannot be accommodated in the pocket of any ClpS protein due to steric hindrance.<sup>43,46</sup> In the bacterial



**FIGURE 2** N-degron recognition of AtClpS1. (a) Close-up view of AtClpS1 in complex with Phe-Ala. The conserved Asn88 and His121 form hydrogen bonds with the  $\alpha$ -amino group of the N-terminal phenylalanine residue. The Asn92 and Arg94 residues are not conserved in bacterial ClpS proteins, but the interactions are sequence-independent backbone-backbone hydrogen bonds. (b) Residues for accommodating N-degron and electron-density maps for the bound Phe-Ala peptide. Polar interacting residues (Asn88, Asn92, Arg94, and His121) of AtClpS1 are shown in green. The hydrophobic pocket for N-degron binding consists of several hydrophobic amino acid residues (Leu86, Val97, Met117, Ala120, and Leu148) shown in gray. (c) A schematic diagram showing the interactions between AtClpS1 and the Phe-Ala peptide. Hydrogen bond interactions are shown as green dashed lines with hydrogen bonding distances, and hydrophobic interactions are denoted by red starbursts. (d) The aliphatic side chain of Val65 of EcClpS rotates approximately 23° toward the Leu N-degron substrate compared with Phe N-degron. The equivalent residue in AtClpS1 is shorter Ala and thus may cause much reduced hydrophobic interactions. Furthermore, the shortened helix (red box) allows movement of the loop (residue Asn88-Asn92; movement direction marked with an arrow) and consequently creates a larger space for the N-degron substrate. (e) The hydroxyl oxygen of Tyr N-degron forms a hydrogen bond with the carbonyl oxygen of Leu46 in the CcClpS structure (cyan), and this interaction might be preserved. However, there may be steric hindrance between the side chain of Leu86 in AtClpS1 and that of the tyrosine residue in the N-degron. (f) The nitrogen atom in the indole ring of the Trp N-degron forms a hydrogen bond with the carbonyl oxygen of Met75 of EcClpS. The interaction must be conserved in the recognition of the Trp N-degron by AtClpS1. The residues in parentheses are the equivalents of comparing the ClpS protein sequences to that of AtClpS1. (g) Summary of the binding affinities of AtClpS1 to N-degron peptides using ITC (see Figure S4 for details)

ClpS proteins, the hydroxyl oxygen of the tyrosine N-degron forms a hydrogen bond with the carbonyl oxygen of Leu46 (Figure 2e),<sup>43,45</sup> and this interaction might be preserved. However, the side chain atoms of Leu86 in AtClpS1 seem to be rigid because neighboring hydrophobic residues, including Leu143, prevent the movement of

Leu86. Interestingly, the equivalent Leu or Ile residues in bacterial ClpS proteins showed sufficient plasticity to accommodate tyrosine residues, but the same was not observed in AtClpS1. Therefore, the Tyr N-degron showed slightly weaker binding affinity to AtClpS1 than the Phe and Trp N-degrons. The tryptophan N-degron

can be well accommodated because the bulkiest indole ring of Trp might be well fitted. Most likely, the hydrogen bond between the nitrogen atom of Trp and the carbonyl oxygen of Met75 in CcClpS (Met117 in AtClpS1) is preserved (Figure 2f). In conclusion, a combination of the spacious cavity allowed by the shortened  $\alpha$ 3-helix, the fine-tuned hydrophobic residues in the pocket, and the basic gatekeeper at the entrance of the pocket regulates the unique N-degron selectivity of plant ClpS.

### 3 | MATERIALS AND METHODS

#### 3.1 | Cloning

A construct of AtClpS1 (residues Ala79–Cys159) was designed based on sequence alignment using the ESPript server<sup>66</sup> and secondary structure prediction using the PSIPRED server.<sup>67</sup> The AtClpS1 construct was amplified using PCR from the codon-optimized gBlock gene fragments (IDT, Integrated DNA Technologies) and then cloned into KpnI and XhoI restriction sites of a modified-pET-His-LC3B vector.<sup>62</sup> The resulting plasmid was transformed into *E. coli* BL21(DE3) cells.

#### 3.2 | Protein expression and purification

AtClpS1-transformed *E. coli* BL21(DE3) cells were incubated in LB media at 37°C until  $OD_{600nm} = 0.5$  and then induced with a final concentration of 0.5 mM isopropyl- $\beta$ -D-thiogalactoside at 18°C for 20 h. The cells were harvested by centrifugation and resuspended in lysis buffer containing 50 mM Tris–HCl pH 8.0, 1 mM Tris (2-carboxyethyl)phosphine (TCEP), and 200 mM NaCl. Cell lysis was performed using ultrasonication, and then the insoluble fraction was removed by centrifugation at 17,000 rpm for 2 h. The supernatant was loaded onto a pre-equilibrated HisTrap<sup>TM</sup> column (GE Healthcare, 17-5255-01) and eluted by gradually increasing the concentration of imidazole to 500 mM. Ion exchange chromatography using a HiTrap<sup>TM</sup> Q HP column (GE Healthcare, 17-1154-01) was used for further purification. The N-terminal His-LC3B tag was cleaved by human ATG4B protease at 25°C overnight, resulting in Leu at the N-terminus, which was followed by an authentic AtClpS1 sequence (<sub>81</sub>APPY---: Figure 1c). The ATG4B protease was prepared as previously reported.<sup>26,62,68</sup> The eluted proteins were concentrated by ultrafiltration (Amicon Ultra 3K NMWL, Millipore) and loaded onto a HiLoad<sup>TM</sup> Superdex<sup>TM</sup> 75 16/600 pg (GE Healthcare, 28-9893-33) column equilibrated with

buffer, 20 mM Ultra-Tris–HCl pH 7.5, 100 mM NaCl, and 1 mM TCEP.

#### 3.3 | Crystallization and data collection

Purified AtClpS1 protein was concentrated to 12 mg/mL and crystallized at 22°C using the sitting drop vapor diffusion method. A Gryphon machine (Art Robbins Instrument) was used for initial crystallization screening. The protein sample was mixed with an equal volume of reservoir solution. The reservoir containing 0.1 M sodium acetate trihydrate pH 4.6 (Hampton Research, HR2-731) and 2.0 M NaCl (Hampton Research, HR2-637) was optimized for the best crystal. The crystals were flash-frozen in liquid nitrogen with 35% (w/v) glycerol as a cryoprotectant in the original mother liquor. AtClpS1 data were collected using a laboratory X-ray diffractometer (Rigaku RU300) with an R-axis IV detector. Diffraction data were indexed, integrated, and scaled using HKL2000 software.<sup>69</sup>

#### 3.4 | Structure determination and refinement

The crystal structure of AtClpS1 was solved by molecular replacement (MR) using the program PHASER-MR in the Phenix software package.<sup>70</sup> The coordinates of PfClpS (PDB ID: 4o2x) were used as a search model.<sup>52</sup> The bound N-degron was quite evident in the initial electron density map calculated by the MR phases (Figure S2). The initial model was rebuilt manually using COOT<sup>71</sup> and refined in iterative cycles using PHENIX.<sup>70</sup> Statistics for the collected data and refinement are summarized in Table 1. All structure figures were drawn using PyMOL (<http://www.pymol.org/>).

#### 3.5 | LC–MS/MS analysis

LC–MS/MS analysis was performed using an Agilent 1100 high-performance liquid chromatography (HPLC) system (Agilent Technologies) coupled to a linear ion trap mass spectrometer (LTQ, Thermo Fischer Scientific). Chromatographic separation was performed with an in-house packed C<sub>18</sub> column (100  $\mu$ m  $\times$  70 mm) using gradient elution (solvent A: 0.1% formic acid; solvent B: 80% acetonitrile, 0.1% formic acid; 0 min 5% B, 10 min 25% B, 11 min 100% B, 11–14 min 100% B, 15 min 5% B). All MS/MS spectra assigned to peptides were manually curated and referenced to the METLIN mass spectral database.<sup>72</sup>

### 3.6 | Isothermal titration calorimetry (ITC)

For the ITC binding experiments, ITC buffer (50 mM Tris-HCl pH 7.5, 50 mM NaCl, and 1 mM TCEP) was used. AtClpS1 protein was diluted to a concentration of 21  $\mu$ M in ITC buffer, while type-2N-degron peptides (FAA, YAA, WAA, LAA, and IAA) were dissolved in the same buffer at a concentration of 600  $\mu$ M. The experiment was conducted at 25°C using a MicroCal PEAQ-ITC (Malvern). Each peptide was injected 19 times (2  $\mu$ L each) into 280  $\mu$ L of AtClpS1. The experimental data were calculated using MicroCal PEAQ-ITC analysis software. At least three experiments were performed.

#### ACKNOWLEDGMENTS

This work was also supported by National Research Foundation of Korea grants (NRF-2020R1A2C3008285 and NRF-2020R1A5A1019023) from the Korean government. L.K. is a recipient of a POSCO Science Fellowship of POSCO TJ Park Foundation and the Korea University Graduate School Junior Fellow Research Grant.

#### AUTHOR CONTRIBUTIONS

**Leehyeon Kim:** Conceptualization; data curation; investigation; methodology; validation; visualization; writing-original draft. **Jiwon Heo:** Data curation; investigation; methodology; validation; visualization; writing-original draft. **Do Hoon Kwon:** Data curation; investigation; methodology; resources; validation; visualization. **Jin Seok Shin:** Data curation; validation. **Se Hwan Jang:** Data curation; investigation; validation; visualization. **Zee-Yong Park:** Methodology; supervision; writing-review and editing. **Hyun Kyu Song:** Conceptualization; funding acquisition; investigation; project administration; resources; supervision; validation; visualization; writing-original draft; writing-review and editing.

#### CONFLICT OF INTEREST

The authors declare no potential conflicts of interest.

#### ORCID

Leehyeon Kim  <https://orcid.org/0000-0002-1266-7828>

Do Hoon Kwon  <https://orcid.org/0000-0002-3214-3394>

Hyun Kyu Song  <https://orcid.org/0000-0001-5684-4059>

#### REFERENCES

- Ciechanover A. Proteolysis: From the lysosome to ubiquitin and the proteasome. *Nat Rev Mol Cell Biol.* 2005;6:79–87.
- Schrader EK, Harstad KG, Matouschek A. Targeting proteins for degradation. *Nat Chem Biol.* 2009;5:815–822.
- Jastrab JB, Darwin KH. Bacterial proteasomes. *Annu Rev Microbiol.* 2015;69:109–127.
- Dougan DA, Reid BG, Horwich AL, Bukau B. ClpS, a substrate modulator of the ClpAP machine. *Mol Cell.* 2002;9:673–683.
- Baker TA, Sauer RT. ATP-dependent proteases of bacteria: Recognition logic and operating principles. *Trends Biochem Sci.* 2006;31:647–653.
- Bachmair A, Finley D, Varshavsky A. In vivo half-life of a protein is a function of its amino-terminal residue. *Science.* 1986;234:179–186.
- Varshavsky A. The N-end rule: Functions, mysteries, uses. *Proc Natl Acad Sci U S A.* 1996;93:12142–12149.
- Varshavsky A. N-degron and C-degron pathways of protein degradation. *Proc Natl Acad Sci U S A.* 2019;116:358–366.
- Bachmair A, Varshavsky A. The degradation signal in a short-lived protein. *Cell.* 1989;56:1019–1032.
- Bartel B, Wunning I, Varshavsky A. The recognition component of the N-end rule pathway. *EMBO J.* 1990;9:3179–3189.
- Xie Y, Varshavsky A. Physical association of ubiquitin ligases and the 26S proteasome. *Proc Natl Acad Sci U S A.* 2000;97:2497–2502.
- Varshavsky A. Discovery of cellular regulation by protein degradation. *J Biol Chem.* 2008;283:34469–34489.
- Tobias JW, Shrader TE, Rocap G, Varshavsky A. The N-end rule in bacteria. *Science.* 1991;254:1374–1377.
- Shrader TE, Tobias JW, Varshavsky A. The N-end rule in *Escherichia coli*: Cloning and analysis of the leucyl, phenylalanyl-tRNA-protein transferase gene. *J Bacteriol.* 1993;175:4364–4374.
- Potuschak T, Stary S, Schlogelhofer P, Becker F, Nejnskaia V, Bachmair A. Prt1 of *Arabidopsis thaliana* encodes a component of the plant N-end rule pathway. *Proc Natl Acad Sci U S A.* 1998;95:7904–7908.
- Mogk A, Schmidt R, Bukau B. The N-end rule pathway for regulated proteolysis: Prokaryotic and eukaryotic strategies. *Trends Cell Biol.* 2007;17:165–172.
- Tasaki T, Kwon YT. The mammalian N-end rule pathway: New insights into its components and physiological roles. *Trends Biochem Sci.* 2007;32:520–528.
- Varshavsky A. The N-end rule pathway and regulation by proteolysis. *Protein Sci.* 2011;20:1298–1345.
- Choi WS, Jeong BC, Joo YJ, et al. Structural basis for the recognition of N-end rule substrates by the UBR box of ubiquitin ligases. *Nat Struct Mol Biol.* 2010;17:1175–1181.
- Tasaki T, Mulder LC, Iwamatsu A, et al. A family of mammalian e3 ubiquitin ligases that contain the UBR box motif and recognize N-degrons. *Mol Cell Biol.* 2005;25:7120–7136.
- Lupas AN, Koretke KK. Bioinformatic analysis of clps, a protein module involved in prokaryotic and eukaryotic protein degradation. *J Struct Biol.* 2003;141:77–83.
- Finley D. Recognition and processing of ubiquitin-protein conjugates by the proteasome. *Annu Rev Biochem.* 2009;78:477–513.
- Dougan DA, Truscott KN, Zeth K. The bacterial N-end rule pathway: Expect the unexpected. *Mol Microbiol.* 2010;76:545–558.
- Cha-Molstad H, Yu JE, Feng Z, et al. P62/SQSTM1/sequestosome-1 is an N-recognin of the N-end rule pathway which modulates autophagosome biogenesis. *Nat Commun.* 2017;8:102.
- Kwon DH, Kim L, Song HK. pH-dependent regulation of SQSTM1/p62 during autophagy. *Autophagy.* 2019;15:180–181.

26. Kwon DH, Park OH, Kim L, et al. Insights into degradation mechanism of N-end rule substrates by p62/SQSTM1 autophagy adapter. *Nat Commun.* 2018;9:3291.
27. Gottesman S, Maurizi MR, Wickner S. Regulatory subunits of energy-dependent proteases. *Cell.* 1997;91:435–438.
28. Gottesman S, Roche E, Zhou Y, Sauer RT. The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. *Genes Dev.* 1998;12:1338–1347.
29. Adam Z, Adamska I, Nakabayashi K, et al. Chloroplast and mitochondrial proteases in *Arabidopsis*. A proposed nomenclature. *Plant Physiol.* 2001;125:1912–1918.
30. van Wijk KJ. Protein maturation and proteolysis in plant plastids, mitochondria, and peroxisomes. *Annu Rev Plant Biol.* 2015;66:75–111.
31. Liu K, Ologbenla A, Houry WA. Dynamics of the ClpP serine protease: A model for self-compartmentalized proteases. *Crit Rev Biochem Mol Biol.* 2014;49:400–412.
32. Wang J, Hartling JA, Flanagan JM. The structure of ClpP at 2.3 Å: A resolution suggests a model for ATP-dependent proteolysis. *Cell.* 1997;91:447–456.
33. Lee BG, Kim MK, Song HK. Structural insights into the conformational diversity of ClpP from *Bacillus subtilis*. *Mol Cells.* 2011;32:589–595.
34. Lee BG, Park EY, Lee KE, et al. Structures of ClpP in complex with acyldepsipeptide antibiotics reveal its activation mechanism. *Nat Struct Mol Biol.* 2010;17:471–478.
35. Kessel M, Maurizi MR, Kim B, et al. Homology in structural organization between *E. coli* ClpAP protease and the eukaryotic 26S proteasome. *J Mol Biol.* 1995;250:587–594.
36. Voges D, Zwickl P, Baumeister W. The 26S proteasome: A molecular machine designed for controlled proteolysis. *Annu Rev Biochem.* 1999;68:1015–1068.
37. Striebel F, Kress W, Weber-Ban E. Controlled destruction: Aaa + ATPases in protein degradation from bacteria to eukaryotes. *Curr Opin Struct Biol.* 2009;19:209–217.
38. Erbse A, Schmidt R, Bornemann T, et al. ClpS is an essential component of the N-end rule pathway in *Escherichia coli*. *Nature.* 2006;439:753–756.
39. Zeth K, Ravelli RB, Paal K, Cusack S, Bukau B, Dougan DA. Structural analysis of the adaptor protein ClpS in complex with the N-terminal domain of ClpA. *Nat Struct Biol.* 2002;9:906–911.
40. Guo F, Esser L, Singh SK, Maurizi MR, Xia D. Crystal structure of the heterodimeric complex of the adaptor, ClpS, with the N-domain of the AAA+ Chaperone, ClpA. *J Biol Chem.* 2002;277:46753–46762.
41. Hou JY, Sauer RT, Baker TA. Distinct structural elements of the adaptor ClpS are required for regulating degradation by ClpAP. *Nat Struct Mol Biol.* 2008;15:288–294.
42. Schuenemann VJ, Kralik SM, Albrecht R, et al. Structural basis of N-end rule substrate recognition in *Escherichia coli* by the ClpAP adaptor protein ClpS. *EMBO Rep.* 2009;10:508–514.
43. Wang KH, Roman-Hernandez G, Grant RA, Sauer RT, Baker TA. The molecular basis of N-end rule recognition. *Mol Cell.* 2008;32:406–414.
44. Torres-Delgado A, Kotamarthi HC, Sauer RT, Baker TA. The intrinsically disordered N-terminal extension of the ClpS adaptor reprograms its partner AAA+ ClpAP protease. *J Mol Biol.* 2020;432:4908–4921.
45. Roman-Hernandez G, Grant RA, Sauer RT, Baker TA. Molecular basis of substrate selection by the N-end rule adaptor protein ClpS. *Proc Natl Acad Sci U S A.* 2009;106:8888–8893.
46. Wang KH, Oakes ES, Sauer RT, Baker TA. Tuning the strength of a bacterial N-end rule degradation signal. *J Biol Chem.* 2008;283:24600–24607.
47. Stein BJ, Grant RA, Sauer RT, Baker TA. Structural basis of an N-degron adaptor with more stringent specificity. *Structure.* 2016;24:232–242.
48. Dong X, Kato-Murayama M, Muramatsu T, et al. The crystal structure of leucyl/phenylalanyl-tRNA-protein transferase from *Escherichia coli*. *Protein Sci.* 2007;16:528–534.
49. Vogtle FN, Wortelkamp S, Zahedi RP, et al. Global analysis of the mitochondrial N-proteome identifies a processing peptidase critical for protein stability. *Cell.* 2009;139:428–439.
50. Calvo SE, Julien O, Clauser KR, et al. Comparative analysis of mitochondrial N-termini from mouse, human, and yeast. *Mol Cell Proteomics.* 2017;16:512–523.
51. Bouchnak I, van Wijk KJ. N-degron pathways in plastids. *Trends Plant Sci.* 2019;24:917–926.
52. AhYoung AP, Koehl A, Vizcarra CL, Cascio D, Egea PF. Structure of a putative ClpS N-end rule adaptor protein from the malaria pathogen *Plasmodium falciparum*. *Protein Sci.* 2016;25:689–701.
53. Nishimura K, van Wijk KJ. Organization, function and substrates of the essential Clp protease system in plastids. *Biochim Biophys Acta.* 1847;2015:915–930.
54. Kim J, Rudella A, Ramirez Rodriguez V, Zybailov B, Olinares PD, van Wijk KJ. Subunits of the plastid clppr protease complex have differential contributions to embryogenesis, plastid biogenesis, and plant development in *Arabidopsis*. *Plant Cell.* 2009;21:1669–1692.
55. Rudella A, Friso G, Alonso JM, Ecker JR, van Wijk KJ. Down-regulation of ClpR2 leads to reduced accumulation of the ClpPRS protease complex and defects in chloroplast biogenesis in *Arabidopsis*. *Plant Cell.* 2006;18:1704–1721.
56. Sjogren LL, Stanne TM, Zheng B, Sutinen S, Clarke AK. Structural and functional insights into the chloroplast ATP-dependent Clp protease in *Arabidopsis*. *Plant Cell.* 2006;18:2635–2649.
57. Koussevitzky S, Stanne TM, Peto CA, et al. An *Arabidopsis thaliana* virescent mutant reveals a role for ClpR1 in plastid development. *Plant Mol Biol.* 2007;63:85–96.
58. Olinares PD, Kim J, van Wijk KJ. The Clp protease system: A central component of the chloroplast protease network. *Biochim Biophys Acta.* 1807;2011:999–1011.
59. Nishimura K, Asakura Y, Friso G, et al. ClpS1 is a conserved substrate selector for the chloroplast Clp protease system in *Arabidopsis*. *Plant Cell.* 2013;25:2276–2301.
60. Montandon C, Dougan DA, van Wijk KJ. N-degron specificity of chloroplast ClpS1 in plants. *FEBS Lett.* 2019;593:962–970.
61. Colombo CV, Rosano GL, Mogk A, Ceccarelli EA. A gatekeeper residue of ClpS1 from *Arabidopsis thaliana* chloroplasts determines its affinity towards substrates of the bacterial N-end rule. *Plant Cell Physiol.* 2018;59:624–636.
62. Kim L, Kwon DH, Heo J, Park MR, Song HK. Use of the LC3B-fusion technique for biochemical and structural studies of proteins involved in the N-degron pathway. *J Biol Chem.* 2020;295:2590–2600.

63. Roman-Hernandez G, Hou JY, Grant RA, Sauer RT, Baker TA. The ClpS adaptor mediates staged delivery of N-end rule substrates to the AAA+ ClpAP protease. *Mol Cell*. 2011;43:217–228.
64. Tian W, Chen C, Lei X, Zhao J, Liang J. Castp 3.0: Computed atlas of surface topography of proteins. *Nucleic Acids Res*. 2018;46:W363–W367.
65. Ninnis RL, Spall SK, Talbo GH, Truscott KN, Dougan DA. Modification of patase by 1/f-transferase generates a ClpS-dependent N-end rule substrate in *Escherichia coli*. *EMBO J*. 2009;28:1732–1744.
66. Robert X, Gouet P. Deciphering key features in protein structures with the new endscript server. *Nucleic Acids Res*. 2014;42:W320–W324.
67. Buchan DWA, Jones DT. The psipred protein analysis workbench: 20 years on. *Nucleic Acids Res*. 2019;47:W402–W407.
68. Kwon DH, Kim S, Jung YO, et al. The 1:2 complex between RavZ and LC3 reveals a mechanism for deconjugation of LC3 on the phagophore membrane. *Autophagy*. 2017;13:70–81.
69. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol*. 1997;276:307–326.
70. Adams PD, Afonine PV, Bunkoczi G, et al. Phenix: A comprehensive python-based system for macromolecular structure solution. *Acta Cryst D*. 2010;66:213–221.
71. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of coot. *Acta Cryst D*. 2010;66:486–501.
72. Guijas C, Montenegro-Burke JR, Domingo-Almenara X, et al. Metlin: A technology platform for identifying knowns and unknowns. *Anal Chem*. 2018;90:3156–3164.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Kim L, Heo J, Kwon DH, et al. Structural basis for the N-degron specificity of ClpS1 from *Arabidopsis thaliana*. *Protein Science*. 2021;30:700–708. <https://doi.org/10.1002/pro.4018>