FEBS Letters

FEBS PRESS science publishing by scientists

Structural features of the plant N-recognin ClpS1 and sequence determinants in its targets that govern substrate selection

Dianela Aguilar Lucero[1], Alejo Cantoia[1], Carolina Sánchez-López[1], Andrés Binolfi[1,2], Axel Mogk[3], Eduardo A. Ceccarelli[1] and Germán L. Rosano[1]

1 CONICET, Facultad de Ciencias Bioquímicas y Farmacéuticas, Instituto de Biología Molecular y Celular de Rosario (IBR), Universidad Nacional de Rosario, Argentina
2 Plataforma Argentina de Biología Estructural y Metabolómica (PLABEM), Rosario, Argentina
3 DKFZ-ZMBH Alliance, Zentrum für Molekulare Biologie der Universität Heidelberg (ZMBH), Germany

Correspondence
G. L. Rosano, Instituto de Biología Molecular y Celular de Rosario (IBR), CONICET. Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario. Esmeralda y Ocampo S/N. Rosario (2000), Argentina
Tel: +54-341-4237070
E-mail: rosano@ibr-conicet.gov.ar

(Received 24 February 2021, revised 22 March 2021, accepted 23 March 2021, available online 16 April 2021)

doi:10.1002/1873-3468.14081

Edited by Ulf-Ingo Flügge

In the N-degron pathway of protein degradation of *Escherichia coli*, the N-recognin ClpS identifies substrates bearing N-terminal phenylalanine, tyrosine, tryptophan, or leucine and delivers them to the caseinolytic protease (Clp). Chloroplasts contain the Clp system, but whether chloroplastic ClpS1 adheres to the same constraints is unknown. Moreover, the structural underpinnings of substrate recognition are not completely defined. We show that ClpS1 recognizes canonical residues of the *E. coli* N-degron pathway. The residue in second position influences recognition (especially in N-terminal ends starting with leucine). N-terminal acetylation abrogates recognition. ClpF, a ClpS1-interacting partner, does not alter its specificity. Substrate binding provokes local remodeling of residues in the substrate-binding cavity of ClpS1. Our work strongly supports the existence of a chloroplastic N-degron pathway.

Keywords: adaptor; *Arabidopsis thaliana*; chloroplast; N-degron pathway; proteolysis

The elimination of proteins from the cellular milieu serves many purposes. One of them is removing abnormal, misfolded, or aggregated proteins, which constitutes an essential quality control mechanism that ensures proteostasis. Also, specific degradation of regulatory proteins is necessary to trigger or end genetic programs at specific times according to environmental cues or the state of the cell. The hydrolysis of the very stable peptide bond is catalyzed by proteases, a diverse family of proteins that can shred polypeptides to pieces. Proper recognition of the substrate is vital to avoid inadvertent elimination of useful proteins. For this reason, proteolytic complexes detect degradation signals in their targets. For example, sequence determinants called degrons mark proteins for their removal. In particular, specific residues at the N terminus (N-degrons) are a feature of short-lived proteins. The regulation of the half-life of a protein by the identity of its N-terminal region is known as the 'N-degron pathway' (formerly, the N-end rule) [1,2].

In many cases, it is not the protease itself that recognizes the substrate but other companion proteins,

Abbreviations
BCIP, 5-bromo-4-chloro-3-indolyl phosphate; Clp, caseinolytic protease; L-Phe-amide, L-phenylalaninamide; L-Trp-amide, L-tryptophanamide; NBT, nitroblue tetrazolium; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; NTE, N-terminal extension; PVDF, polyvinylidene fluoride; SOFAST-HMQC, band-selective optimized flip-angle short-transient heteronuclear multiple quantum coherence; TBS, Tris-buffered saline; TEV, tobacco etch virus.

such as molecular chaperones. In *E. coli*, ClpP is a major protease system that associates with the Hsp100 chaperones ClpA and ClpX [3–5]. Monomers of ClpP self-assemble into a barrel-shaped tetradecamer [6,7]. Inside the barrel's chamber, a catalytic triad of residues cleaves the peptide bonds of the incoming polypeptide. The axial pores of the ClpP barrel are too narrow to allow the entrance of folded proteins [7]. So, the Hsp100 chaperones are in charge of recognizing the target and then presenting it to ClpP. For example, ClpA oligomerizes into ring-shaped hexamers that anchor to the pores of ClpP [8,9]. Degrons bind to ClpA, thus initiating the first step in degradation [10]. Using the chemical energy from ATP hydrolysis, ClpA then mechanically unfolds and delivers the soon-to-be-degraded protein to ClpP.

Substrate specificity can be expanded by adaptor proteins that, for example, bind to targets *via* N-degrons and deliver them to ClpAP. In *E. coli* and other Gram-negative bacteria, ClpS is the cognate adaptor protein of ClpAP and recognizes four destabilizing N-terminal residues present in short-lived proteins: phenylalanine, tyrosine, tryptophan, and leucine [11,12]. ClpS is a monomeric 12 kDa protein that contains a disordered N-terminal extension (NTE) of 25 residues and a folded central core. The core has a typical globular appearance and presents a hydrophobic cavity in which destabilizing residues snugly fit [13]. The identity of the second residue also influences recognition, as positively charged side chains interact with the entrance of the cavity. Moreover, when an arginine or a lysine residue is present at the N terminus, Leu/Phe transferases conjugate a leucine or a phenylalanine residue, thus creating an optimal N-degron for ClpS [14,15]. Finally, successful recognition requires a free N terminus as N-terminal acetylation (a rare post-translational modification in bacteria) completely abrogates binding to ClpS [11].

Chloroplasts retain many of the molecular systems present in ancient bacteria that gave rise to plastids, being the caseinolytic protease (Clp) system one of them [16]. The three major constituents of a typical bacterial Clp complex (the protease, the Hsp100 unfoldase, and the adaptor) can be found in chloroplasts but with higher complexity in subunit composition. In *Arabidopsis thaliana* chloroplasts, the proteolytic chamber is made of nine different subunits (ClpP1, ClpP3-6, and ClpR1-4) [17]. Plants do not possess homologs to ClpA, but instead, Hsp100 chaperones of the ClpC (ClpC1 and ClpC2) and ClpD families fulfill its role [18–21]. The associated adaptor is ClpS1. Sequence alignments comparing ClpS1 to bacterial ClpS indicate that several of the residues involved in substrate recognition are conserved, suggesting that chloroplastic targets of ClpS1 may also follow the N-degron pathways of bacteria [22]. In that report, immobilized ClpS1-GST to glutathione agarose resin was used to identify possible natural substrates from chloroplast stroma isolated from *A. thaliana* seedlings. Two strong candidates emerged: glutamyl-tRNA reductase and pyridoxine oxidase domain protein. However, their N-terminal regions bear little resemblance to common bacterial N-degrons (ELSASS- and SAAQS-, respectively).

Adherence of ClpS1 to bacterial N-degron pathways was first tested *in vitro* using N-terminally engineered variants of the GFP bearing different N-terminal ends. ClpS1 recognized those starting with Phe, Tyr, Trp, and Leu but not Met, like other bacterial ClpS proteins [23]. The model substrates starting with Phe and Tyr were followed by Arg; so, they were optimal targets of *E. coli* ClpS. However, affinity toward *A. thaliana* ClpS1 was quite low. Later, ClpS1 was confirmed to bind to N-terminal variants of GFP starting with Phe, Trp, and Leu (with much less affinity) but not Tyr. So, it was suggested that *A. thaliana* ClpS1 has a more restricted specificity than its *E. coli* counterpart [24]. However, these experiments were carried out using a limited set of artificial substrates containing Arg in the second position. In stark contrast to *E. coli* ClpS, the X-ray crystallographic structure of ClpS1 showed that positive charges surround the cavity entrance [25]. So, electrostatic repulsions may prevent binding of substrates containing a basic residue in the second position as previously suggested by protein modeling experiments [23].

Finally, plant chloroplasts present another component of the ClpS1-Hsp100 complex named ClpF [26]. ClpF is not a substrate of ClpS1 because it interacts with a mutant version of ClpS1 defective in substrate binding. The function of ClpF is still ill-defined; it may regulate the specificity of ClpS1 or bind to N-degrons on its own. ClpF stimulates the binding of ClpS1 to ClpC chaperones; so, it may act as a facilitator of the ClpS1-Hsp100 interaction [26].

Overall, there are many open questions regarding the function, mode of action, and structural features of plant ClpS1. Its specificity is still under scrutiny, and the effects of target N-terminal acetylation and ClpF on substrate recognition are unknown. In this work, the sequence determinants that regulate the specificity of chloroplastic ClpS1 were analyzed using diverse approaches. A combination of peptide arrays, pull-down assays, fluorescence anisotropy, and nuclear magnetic resonance (NMR) was used to understand the process of substrate recognition by ClpS1. Using

these techniques, we report that ClpS1 does recognize the four cognate residues of the N-degron pathways present in many bacteria, but the second residue influences the interaction. As with bacterial ClpS, N-acetylation abolishes the interaction. Also, ClpF does not bind to N-terminal ends on its own, nor it alters the specificity of ClpS1. NMR analysis revealed the critical amino acids in ClpS1 responsible for substrate interaction and the rearrangement of the substrate cavity upon substrate binding. Our findings represent a new step in understanding target recognition by the Clp system in plant organelles.

Materials and methods

Protein production and purification

ClpS proteins

ClpS1 from *A. thaliana* and ClpS from *E. coli* were produced as described before [23,27]. For the recombinant production of ClpS1 C-terminally fused glutathione S-transferase (ClpS1-GST), the cDNA of ClpS1 was amplified and cloned into a pETGEXCT plasmid [28]. ClpS1-GST was purified using glutathione Sepharose 4B following the instructions of the manufacturer (GE Healthcare, Piscataway, NJ, USA). For the production of $^{15}N/^{13}C$ isotopically enriched ClpS1 for NMR experiments, the protocol described by Marley *et al.* was followed with minor modifications [29]. Briefly, a culture of *E. coli* cells containing the plasmid for the production of ClpS1 was grown in rich media until an optical density of 0.6–0.7 at 600 nm. The cells were harvested by centrifugation and resuspended in 1 L of M9 minimal media supplemented with 1 g ^{15}N-H_4Cl, 2 g ^{13}C-D-glucose, antibiotics, and 0.5 mM IPTG. The cultures were grown for 16 h at 18 °C. $^{15}N/^{13}C$ isotopically enriched ClpS1 was purified as described for ClpS1.

ClpF

The coding sequence of ClpF was amplified by PCR from a plasmid containing its cDNA (Riken cDNA bank pda:14908) using the following primers: forward 5′-CGGATCGGAAAACCTGTATTTTCAGGGAGTTGAAGCTAGATGGCCATTTC-3′; and reverse 5′-GGTGGCTCCAGCTGCCGGATCCTTAATCTTCGTCTTGTGAATCAAAA-3′. The PCR product codes for the predicted mature version of ClpF (i.e., without its transit peptide), from residue Val66 to Asp330 [22,26]. The coding sequence was cloned into a pET32a plasmid derivative from the vector suite previously described [30] by restriction-free cloning. The resulting plasmid was transformed into *E. coli* BL21-CodonPlus (DE3)-RIL strain (Stratagene, La Jolla, CA, USA). The culture was grown in LB medium supplemented with appropriate antibiotics at 37 °C until an optical density of 0.6–0.7 at 600 nm; then, 0.5 mM IPTG was added, and the culture was further incubated for 16 h at 18 °C. The cells were harvested by centrifugation, resuspended in cold lysis buffer (50 mM Tris/HCl pH 8.0, 400 mM NaCl, 1 mM benzamidine, 10% v/v glycerol), and disrupted by sonication. ClpF was obtained as a tobacco etch virus protease (TEV)-cleavable $(His)_6$-CelDnc fusion. The fusion protein was purified by affinity chromatography using an immobilized nickel resin according to the manufacturer's recommendations (Qiagen, Chatsworth, CA, USA), desalted by PD10 columns (GE Healthcare), and digested by the addition of $(His)_6$-TEV protease at a ratio of 1 : 300 (protease : recombinant protein) for 18 h at 4 °C. Undigested $(His)_6$-CelDnc-ClpF, $(His)_6$-CelDnc, and $(His)_6$-TEV were removed by affinity chromatography. For the production of anti-ClpF antibodies, rabbits were inoculated with purified recombinant ClpF. The animals were kept in the Faculty of Biochemical and Pharmaceutical Sciences (UNR) bioterium, which adheres to experimental and ethical guidelines for animal studies.

GFP derivatives

A pET32a derivative containing the coding sequence for $(His)_6$-[TEV]-GFP (being [TEV] the recognition sequence for the TEV protease) was used for producing wild-type GFP and as a starting template for constructing GFP derivatives [30]. By insertional mutagenesis using restriction-free cloning, the coding sequence for the 10-amino acid linker SKGEELFTGV was added between the TEV cleavage site and the first amino acid of GFP. This construct was used to generate two GFP derivatives by inserting the codons for Phe-Arg and Phe-Glu preceding the linker (resulting in sequences coding for $(His)_6$-[TEV]-XX-linker-GFP, where XX is either FR or FE). Constructions were confirmed by DNA sequencing. GFP variants were produced and purified as described for ClpF, with minor modifications.

In all cases, protein purity was determined by Coomassie-stained denaturing polyacrylamide gels. Protein concentrations were determined by the Bradford or bicinchoninic assays (Thermo Fisher, Waltham, MA, USA) using BSA as a standard.

Peptide libraries

The membranes consisted of peptides of 13 amino acids in length C-terminally attached to cellulose sheets. The overlapping λcI library was described in detail previously [23]. Two other libraries were prepared by automated spot synthesis by the manufacturer JPT Peptide Technologies GmbH (PepSpots™): another overlapping λcI library containing N-acetylated peptides, and a position scanning library of the peptide XSPSIAREIYEMY, where X is any of the 20 proteogenic amino acids. Before each use, the membranes were soaked 2× 5 min in methanol, 3× 10 min in Tris-buffered saline (TBS), 1 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), and incubated with *E. coli* ClpS, ClpS1, and/or ClpF

(2 μM) in buffer A for 30 min. Afterward, buffer A was discarded, and the membranes were thoroughly washed with TBS. Four consecutive electrotransfers to polyvinylidene fluoride (PVDF) membranes were made to transfer bound proteins (4× 30 min, using fresh blotting buffer after the second transfer). PVDF membranes were blocked with 3% w/v BSA in TBS and incubated with specific antibodies (1 : 1000). These were as follows: anti-ClpS [27], anti-ClpS1 [23], or anti-ClpF (this work). As secondary antibody, alkaline phosphatase-conjugated mouse anti-rabbit IgG (Sigma-Aldrich, St. Louis, MO, USA; Catalog No. A3687, working dilution 1 : 30 000) was used. The immunoblot was revealed using 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 $MgCl_2$, 0.015% w/v BCIP, and 0.03% w/v NBT). Experiments were performed at least in triplicates.

Pull-down assays

Pull-down assays were carried out in tubes previously coated with 5% w/v BSA [31]. Glutathione agarose resin was loaded with an excess of ClpS1-GST with gentle shaking at 4 °C for 1 h. The loaded resin was washed three times with five resin volumes of washing buffer (50 mM Tris/HCl pH 8.0, 200 mM NaCl, 10% v/v glycerol). Each GFP protein (1 μM final concentration) was added to 20 μL of ClpS1-GST-loaded glutathione agarose resin. Immediately, 10% (volume) of each mixture was retrieved and analyzed by SDS/PAGE to ensure equal loading ('input fraction'). The mixtures were incubated for 30 min with constant agitation at 4 °C. Next, the resin was washed four times with 10 volumes of washing buffer. Proteins were eluted with 20 μL 1× Laemmli sample buffer (50 mM Tris/HCl pH 6.8, 2% w/v SDS, 10% v/v glycerol, 1% v/v β-mercaptoethanol, 12.5 mM EDTA, and 0.02 % w/v bromophenol blue) followed by a heat treatment at 90 °C for 5 min. All samples were analyzed by SDS/PAGE [32] and western blots. Proteins were electrotransferred to nitrocellulose membranes that were then blocked with 5% w/v non-fat dry milk in TBS, pH 7.4, for 1 h. For protein detection, a GFP antibody was used at a 1 : 5000 working dilution (Abcam, Cambridge, MA, USA; Catalog No. ab290). Immunodetection was performed as described above. Experiments were performed at least in triplicates. Blots were digitalized, and band intensity was quantified using IMAGEJ (National Institute of Health, Bethesda, MD, USA). Statistical differences in band intensity (normalized to the intensity of GFP alone) were assessed by *t*-tests ($P < 0.05$).

Fluorescence anisotropy measurements

Fluorescent anisotropy measurements were made with a Cary Eclipse Fluorimeter at 25 °C. The instrument was set to 298 nm excitation, 353 mm emission, with slit widths of 5 nm for each monochromator, and automatic correction of G factor. ClpS1 (in 20 mM sodium phosphate, pH 7.0, 150 mM NaCl) was added in increasing concentrations to a 5 μM L-tryptophanamide (L-Trp-amide) solution. Samples were incubated for 10 min to allow the system to reach equilibrium, and then, fluorescent anisotropy was measured. After each determination, the emission spectrum of L-Trp-amide was recorded to check that fluorescent intensity was unaffected by binding. To measure the dissociation constant for L-phenylalaninamide (L-Phe-amide), a competition experiment was set up [33]. Five micromolar L-Trp-amide and 50 μM ClpS1 were kept constant, and increasing amounts of L-Phe-amide were added. Data were adjusted to the corresponding equations [33]. Experiments were performed at least in triplicates.

Nuclear magnetic resonance

All NMR spectra were acquired at 25 °C on a 700 MHz Bruker Avance III Spectrometer equipped with a triple-resonance inverse probe (5 mm $^{1}H/D$-$^{13}C/^{15}N$ TXI). For sequential assignment of ClpS1 backbone resonances, triple-resonance HNCO, HNCACO, CBCACONH, and HNCACB experiments were registered on 800 μM $^{15}N/^{13}C$ isotopically enriched ClpS1 dissolved in NMR buffer (20 mM phosphate buffer at pH 7.0, 150 mM NaCl, 1 mM DTT, and 10% v/v D_2O). Standard pulse sequences from the Bruker Topspin library with nonuniform sampling (NUS amount 25%) were used to decrease acquisition times and maximize signal-to-noise ratios [34]. Acquisition parameters for the experiments were as follows: HNCO, increments: 52 (^{15}N), 96 (^{13}C), and 2048 (^{1}H); spectral width (ppm): 30 (^{15}N), 10 (^{13}C), and 14 (^{1}H); number of scans 16; HNCACO: 52, 96, 2048; 30, 10, 14; 64; HNCACB: 52, 100, 2048; 30, 70, 14; 128; and CBCACONH: 44, 88, 2048; 30, 70, 14; 64. All spectra were zero-filled with twice the number of real points, multiplied with sine bell-shaped window functions, and baseline-corrected in all dimensions.

The secondary structure ClpS1 was delineated using the chemical shift method. Secondary chemical shifts were obtained by subtracting the measured CO and Cα chemical shifts to the empirical random coil values reported for each amino acid [35]. Referencing was done with 4,4-dimethyl-4-silapentane-1-sulfonic acid [36]. The formula $[3\Delta\delta(C\alpha) + 4\Delta\delta(C')]/7$ was used to calculate the average $^{13}C\alpha/CO$ chemical shift value for each residue. These values were plotted against the primary sequence of ClpS1 to obtain the secondary structure profiles [37]. Positive or negative values of $^{13}C\alpha/CO$ shifts indicate α-helix or β-strand conformations, respectively.

Titration experiments were performed to identify the residues in ClpS1 involved in the interaction with model amino acids. Sequential additions of 1.9-μL aliquots of L-Phe-amide or L-Trp-amide to 190 μM $^{15}N/^{13}C$ isotopically

enriched ClpS1 dissolved in NMR buffer were used to acquire ^{1}H-^{15}N SOFAST-HMQC spectra at 25 °C [38]. Spectra were registered with 96 scans and 1024 complex points for a sweep width of 16 p.p.m. for the ^{1}H dimension, and 128 complex points in the ^{15}N dimension for a sweep width of 26 p.p.m. Selective excitation of amide protons was done with 120° polychromatic PC9 pulses of 2200 ms and refocusing with 180° RSNOB pulses of 1000 ms. ^{13}C and ^{15}N decouplings were achieved with a smoothed chirp pulse of 0.5 ms (Crp60) and the garp4 sequence, respectively. Amide cross-peaks affected by L-Phe-amide or L-Trp-amide binding were identified by comparing their chemical shift values with those of the same signals in the spectrum of free ClpS1. Differences in the mean weighted chemical shift (MWΔCS) displacements for ^{1}H and ^{15}N were calculated as $[(\Delta\delta^{1}H)^2 + (\Delta\delta N/10)^2]^{1/2}$ and plotted against the protein's sequence [39].

Backbone dynamics of ClpS1 were determined at 25 °C on 190 μM samples dissolved in NMR buffer in the absence and presence of 3.0 equivalents of L-Phe-amide. A 3 : 1 of ligand:protein ratio was used to minimize ligand exchange effects on the relaxation rates. ^{15}N T_1 and T_2 relaxation and ^{1}H-^{15}N nuclear Overhauser effect (NOE) experiments were recorded using standard pulse sequences from the Bruker Topspin library [40]. 2D ^{1}H-^{15}N NMR spectra for ^{15}N T_1 relaxation analysis were obtained for the following relaxation delays (in ms): 60, 120, 240, 380, 760, 1200, and 1850. For ^{15}N T_2 experiments, a Carr–Purcell–Meiboom–Gill pulse train was used in the pulse sequence and the following relaxation delays (in ms): 40, 60, 80, 120, 240, 360, and 480. In both cases, relaxation delays were scrambled, and duplicate spectra were collected at several time points to estimate uncertainties. Experiments were recorded with 2048 complex points for a sweep width of 16 p.p.m. for the ^{1}H dimension and 128 complex points in the ^{15}N dimension for a sweep width of 26 p.p.m. Sixteen and 32 scans were used for T1 and T2 experiments, respectively. Relaxation rates were obtained by measuring the peak heights at each spectrum and fitting the values to a two-parameter exponential decay function where the independent variable was the relaxation delay. Steady-state ^{1}H-^{15}N NOE values were obtained by dividing the peak heights of paired spectra collected with and without an initial 4-s period of proton saturation. The same spectral resolution as for T1 and T2 experiments was used, and 32 scans were recorded.

For ^{1}H-^{15}N SOFAST-HMQC and ^{15}N relaxation experiments, processing was done by zero filling to 2K and 1K points in ^{1}H and ^{15}N, respectively, followed by sine-modulated window function multiplication and baseline correction in both dimensions. Acquisition, processing, and visualization of the spectra were made with TOPSPIN 3.5 (Bruker Biospin, Rheinstetten, Germany), Sparky [41], and CARA [42]. NMR assignments of ClpS1 resonances were deposited in the BioMagResBank (BMRB entry 50776).

Results

ClpS1 recognizes primary destabilizing residues of bacterial N-degron pathways

Thus far, the specificity of ClpS1 *in vitro* has been mainly assessed using GFP reporters bearing different N-terminal ends [23,24]. While this strategy yielded the first insights into the possible N-terminal composition of ClpS1 client proteins, surveying the universe of residue combinations requires a methodology more amenable to higher throughput. For many bacterial ClpS proteins, peptide arrays were successfully used to uncover their recognition patterns. In the arrays, peptides are attached C-terminally to a cellulose membrane by an Ala-Ala linker. An array of 76 13-mer peptides covering the sequence of the λcI protein was used to survey the specificity of ClpS1. The binding of ClpS1 was detected using ClpS1-specific antibodies after electrotransfer of ClpS1 from the library to PVDF sheets.

The λcI overlapping peptide library contains at least one peptide bearing an N-terminal residue of the *E. coli* N-degron pathway. ClpS1 recognized all peptides starting with Phe (five in total), Tyr (four), and Trp (one) (Fig. 1 and S1 for the complete list of sequences). The amount of ClpS1 detected in the immunoblots was highest in positions corresponding to Phe-starting peptides. Signals in positions corresponding to N-terminal Tyr tended to be of low intensity, except for position 35 (Tyr followed by Pro). On the contrary, ClpS1 recognized only three out of six peptides starting with Leu. The amount of ClpS1 bound to peptides in positions 33 and 40 (Leu in the first position and Arg in the second) presented similar intensity. In contrast, Thr, Ser, or Leu as second residues abolished ClpS1 recognition. Conversely, ClpS from *E. coli* readily recognized the Leu-Leu, Leu-Ser, and Leu-Thr bearing peptides in reports using the same library [11,23]. Differences in the binding of peptides starting with Leu most probably reflect the impact of the residue in the second position, as further downstream positions are not expected to influence binding greatly.

ClpS1 also recognized (with low affinity) a handful of peptides starting with residues not involved in bacterial N-degron pathways. This phenomenon was also described for bacterial ClpS proteins and may be attributed to nonspecific electrostatic interactions [11]. We next used a positional scanning library using a fixed sequence to avoid this background noise caused by irregular distribution of charged residues and differences in peptide sequences throughout the library. The chosen peptide sequence [position 26 (FSPSIAREIYEMY)] (a) displayed

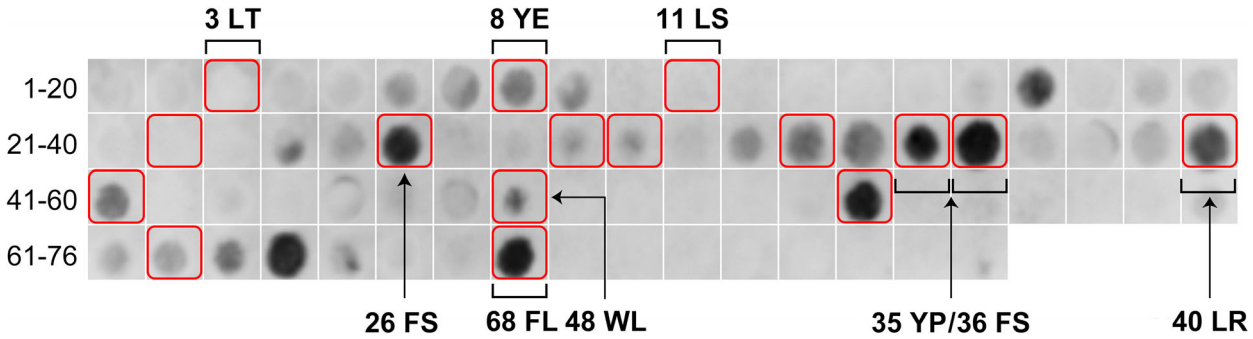

Fig. 1. Binding of ClpS1 to a λcI peptide scanning library. The library consisted of 13-mer peptides overlapping by 10 residues covering the sequence of the λcI protein. The array was incubated with 2 μM ClpS1, which was then electrotransferred to PVDF membranes. ClpS1 was detected using anti-ClpS1 antibodies. Peptides starting with destabilizing amino acids of the *E. coli* N-degron pathway (F, Y, W, and L) are marked in red boxes. The identity of the first two residues is shown for some spots, next to their position in the library. The full sequences of all peptides can be found in Fig. S1.

strong ClpS1 binding in the λcI array and (b) do not have charged residues near the N-terminal end. In this regard, this library allows for assessing binding to the very N-terminal residue specifically and exclusively. All other 19 amino acids substituted the first residue of the sequence. For comparison, this library was also interrogated with ClpS from *E. coli*. The binding patterns revealed differences in the specificity of both proteins (Fig. 2). ClpS1 readily bound peptides starting with the aromatic residues Phe, Tyr, and Trp. The Leu-starting peptide was not recognized when followed by Ser (in agreement with the λcI library, peptide in position 11). On the contrary, ClpS from *E. coli* recognized all of the classic destabilizing residues of its N-degron pathway when followed by Ser.

The entrance to the substrate-binding cavity of ClpS1 is surrounded by positive charges (Lys49 and Arg50, numbered according to the mature form) [25]. This disposition might deter the binding of substrates bearing basic residues in the second position. A mutant ClpS1 where the gatekeeper residue Arg50 was changed to Met binds the model substrate Phe-Arg-linker-GFP with higher affinity than ClpS1 [23]. As also explained above, the λcI peptide library does not allow judging the effect of charged residues in the second position as it does not contain peptides starting with the same destabilizing residue but with switching charges at the second. We resorted to pull-down assays with N-terminal variants of GFP to test whether ClpS1 binds to N termini enriched in acidic residues. The artificial substrate FR-linker-GFP (linker: SKGEELFTGV) is readily recognized by ClpS from *E. coli* due to the high-affinity interaction with the N-terminal Phe followed by Arg. The arginine residue in second position was substituted with a glutamate residue to create the variant FE-linker-GFP. Each GFP variant was incubated with a glutathione resin previously loaded with ClpS1-GST. After the washing steps, the amount of remaining GFP variant bound to the ClpS1-loaded resin was detected by western blot using anti-GFP antibodies. Wild-type GFP was used as a control for nonspecific interaction. As expected, FR-linker-GFP interacted with ClpS1-GST but rather weakly, as a minute amount remained bound to the resin. By contrast, ClpS1 interacted with FE-linker-GFP with higher affinity (Fig. 3). Twice as much FE-linker-GFP remained bound to the resin compared with FR-linker-GFP. This result indicates that client

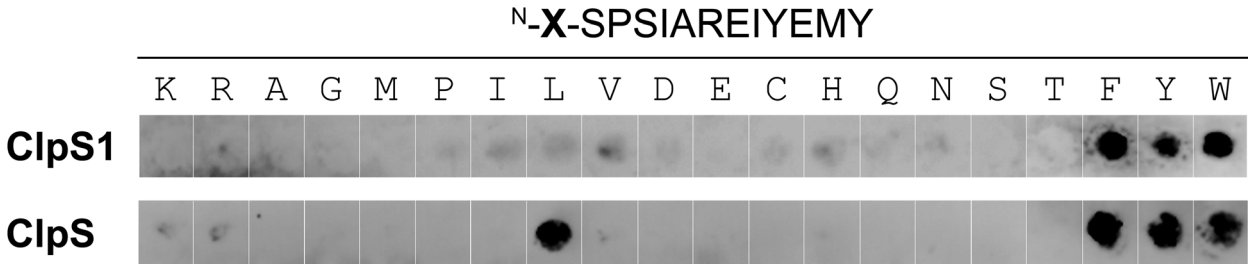

Fig. 2. Binding of ClpS1 from *A. thaliana* and ClpS from *E. coli* to a positional scanning library. The identity of the residue in the first position (X) in the peptide XSPSIAREIYEMY was surveyed. Treatment of the library with ClpS1 or ClpS was as explained in Fig. 1. Signal detection in the immunoblots was carried out with anti-ClpS1 or anti-ClpS antibodies.

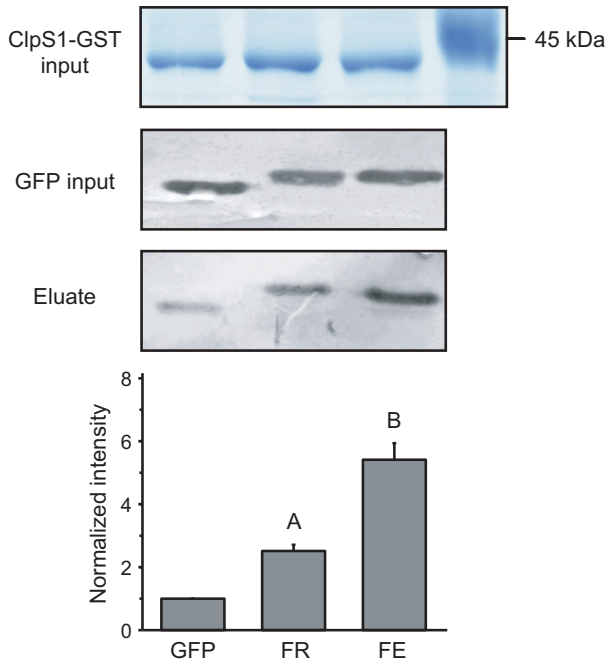

Fig. 3. Interaction of ClpS1 with N-terminal variants of GFP. Glutathione agarose beads were loaded with ClpS1-GST and evenly distributed in BSA-coated microcentrifuge tubes. 10% of resin from each tube was analyzed by SDS/PAGE to confirm equal ClpS1-GST loading (top image). The last lane shows the position of the molecular weight marker (ClpS1-GST, ca. 40 kDa). Each GFP variant, including wild-type GFP as a control, was added, and immediately, a 10% volume input was retrieved and later analyzed by western blotting using anti-GFP antibodies to verify equal GFP variant loading (middle image). Mixtures were incubated for 30 min under constant agitation. Next, the beads were thoroughly washed and finally boiled in 1× Laemmli sample buffer. Samples were analyzed by western blotting using anti-GFP antibodies (bottom image). The intensity of the bands from immunoblots (e.g., bottom image) was measured by densitometry. The intensity was normalized to the intensity of the (wild-type) 'GFP' lane. Bars represent the average of three independent replicates ± standard deviation. Letters represent statistically significant differences in normalized intensity when compared to the GFP condition. GFP variants naming in the bar plot: FR-linker-GFP (FR), FE-linker-GFP (FE).

N-terminal negative charges after Phe favor the interaction with ClpS1.

ClpF does not influence the specificity of ClpS1

The protein ClpF was shown to interact with ClpS1 and was proposed to recruit ClpS1 (with or without a bound substrate) to ClpC chaperones for substrate delivery. Alternatively, ClpF may alter the repertoire of N termini recognized by ClpS1, or may be specifically involved in the degradation of selected proteins by recognition of N-degrons or other regions of its target [26]. Using the λcI combinatorial library, we analyzed whether ClpF interacted with any of the 76 peptides (Fig. 4, top image). No binding signal was detected in any position of the library, not even nonspecific binding. Also, ClpF did not bind to any peptide in the positional scanning library shown in Fig. 2 (data not shown). Next, the λcI library was interrogated with an equimolar mixture of ClpS1 + ClpF. As shown in Fig. 4 (middle image), ClpF did not modify the repertoire of peptides recognized by ClpS1, as the same binding profile of ClpS1 alone was detected. Moreover, the signal intensity was quite similar, as in Fig. 1. Hence, ClpF did not influence the affinity of ClpS1 for this set of peptides.

N-terminal acetylation abrogates ClpS1 binding to N termini

In *A. thaliana* chloroplasts, N-terminal acetylation is a widespread protein modification. More than two hundred N-terminally acetylated chloroplastic proteins were so far identified in proteome-wide studies [43,44]. In *Chlamydomonas reinhardtii* plastids, around 30% of stromal proteins are N-acetylated [45]. This N-terminal modification could represent another layer of regulation in substrate recognition by the quality control system. An N-acetylated version of the λcI peptide array version was used to assess binding of ClpS1 to N-acetylated termini. The library was interrogated with ClpS1 (Fig. 4, bottom image), ClpF, or ClpS1 + ClpF (data not shown). No binding signal was detected in any case. Even the so-called nonspecific interactions were significantly reduced. It can be concluded that N-terminal acetylation completely prevents the recognition of N-terminal residues by ClpS1.

Degron binding induces local conformational rearrangements at the substrate cavity

ClpS1 presents a mainly disordered NTE and a well-folded core

The high-resolution structural features of bacterial ClpS proteins have been thoroughly described over the years. However, ClpS adaptors from higher organisms have not been studied in such detail. To gain better insight into the structure of ClpS1 and its ligand binding properties, we undertook a high-resolution solution-state NMR analysis. First, the ^{1}H-^{15}N SOFAST-HMQC spectrum of free ClpS1 was recorded; signals were sharp and well-dispersed over large ^{1}H and ^{15}N spectral widths, indicating that the protein is folded

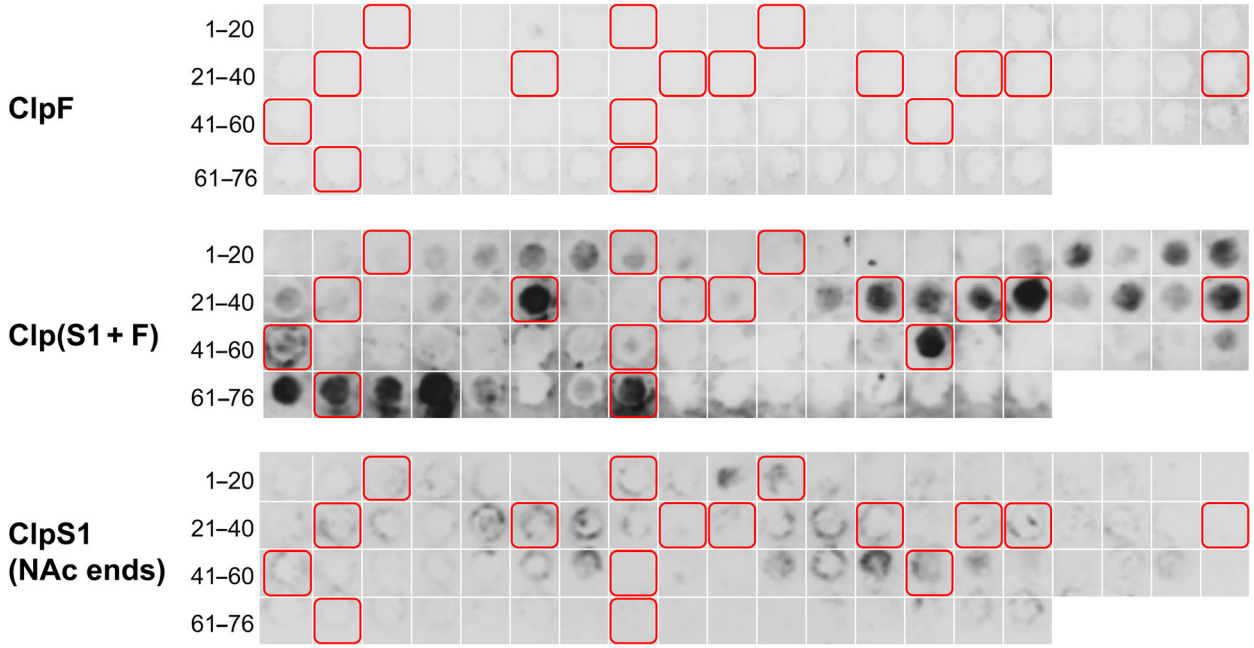

Fig. 4. Probing the role of ClpF and N-terminal acetylation in the recognition of peptides by ClpS1. The λcI library was interrogated with ClpF (top panel) or ClpS1 in the presence of equimolar quantities of ClpF (middle panel). In the bottom panel, peptides in the λcI combinatorial library were also synthesized in their N-terminal acetylated form (*NAc ends*) and interrogated with ClpS1. Array treatments with ClpF and ClpS1 and signal detection were performed using the corresponding antibodies as described in Fig. 1. Peptides starting with destabilizing amino acids of the *E. coli* N-degron pathway (F, Y, W, and L) are marked in red boxes.

(Fig. 5A). Backbone resonance assignment was possible for 105 out of 116 nonproline resonances while 11 remained unassigned, including the first two residues, which are typically not detected due to fast exchange with the solvent (Table S1). The chemical shifts of each amino acid were used to determine the secondary structure content of ClpS1. Positive or negative deviations in Cα and CO chemical shifts from their values in random coil conformations indicate α-helical or β-strand organization, respectively [35]. ClpS1 contains two distinctive regions, one intrinsically disordered that spans residues 1–35 and a folded region, rich in secondary structure elements that contain residues 38–109 (Fig. 5B and S2). Interestingly, the NTE of ClpS1 does not feature a pure random coil behavior but instead displays two subregions with propensities to populate β-sheet (residues 7–19) and α-helix (residues 20–29) secondary structures. The folded region of ClpS1 shows stable secondary structure elements in the order β-α-α-β-α-β, from N to C terminus.

Key residues at or near the central cavity interact with bound amino acids

The dynamics and the internal structural motions of ClpS1 upon substrate binding are not known. Each cross-peak in the NMR spectra represents one amino acid residue, and their positions and intensities depend on the chemical environment in their vicinity and the conformational properties of the protein. Thus, ligand binding affecting these features induces spectral changes that can be identified and mapped along the primary sequence [46]. The peptide library results clearly indicate that ClpS1 binds with high-affinity peptides starting with Phe and Trp. For this reason, we studied ligand binding using analogs of these two amino acids, L-Phe-amide and L-Trp-amide. Both have been previously used as model substrates for structural analysis of bacterial ClpS [47,48]. An amide group replaces the carboxyl group; so, these derivatives mimic peptide-bonded amino acids. Increasing amounts of L-Phe-amide were added to isotopically enriched ClpS1 to follow changes in the ^{1}H-^{15}N correlation NMR spectra. Substoichiometric additions of L-Phe-amide caused the site-selective broadening of a group of ClpS1 resonances, mainly around residues 45–54 and 74–80. Also, a decrease in intensity was detected for resonances 67–69 and 105 (Fig. 6A,B). At higher L-Phe-amide:ClpS1 ratios, a new set of resonances appeared, whose intensities increased with further additions of the substrate (Fig. 6C,D). These results indicate that ClpS1 binds to L-Phe-amide and

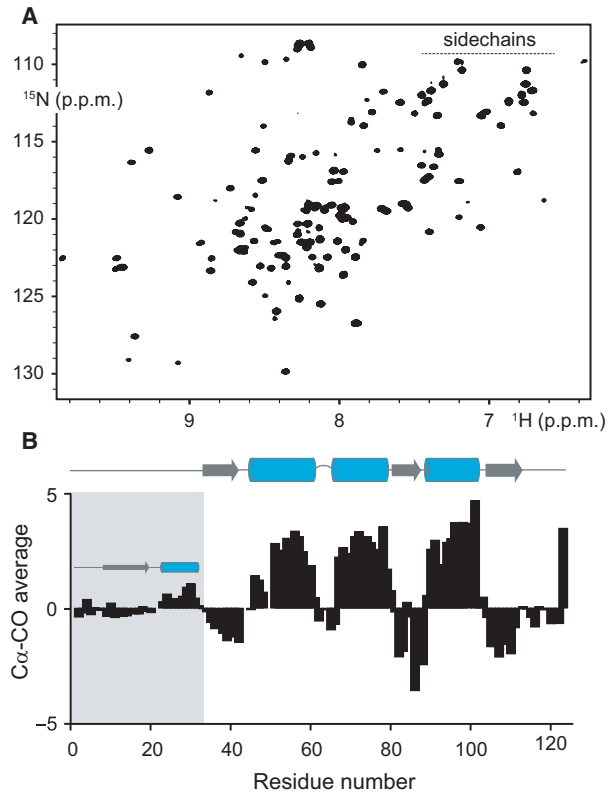

Fig. 5. NMR characterization of ClpS1 conformational features. (A) ^{1}H-^{15}N SOFAST-HMQC spectrum of ^{15}N/^{13}C isotopically enriched ClpS1. (B) Average ^{13}Cα-^{13}CO secondary chemical shift plot of ClpS1. Stable secondary structure elements identified along ClpS1 sequence are indicated on the top. The area shaded in gray delimits the intrinsically disordered domain at the N terminus of ClpS1. Regions with propensities to populate β-sheet and α-helical structures within the NTE are also indicated.

that the interaction is in slow exchange compared with the chemical shifts timescale, suggesting that the affinity of the protein for this substrate is high (K_d < 10 μM) [39] The chemical shift perturbation analysis upon L-Phe-amide addition along the primary sequence of ClpS1 revealed that the binding site is located between the first and second α-helices and the N-terminal portion of the first β-sheet (Fig. 6E). The most affected regions were 45–53, 59–61, and 72–80, in good agreement with the intensity profile shown in Fig. 6B. Small perturbations of resonances in the loop preceding the last β-sheet (102–106) were also detected. NMR titration analysis using L-Trp-amide as a substrate gave similar profiles, although the effects induced on ClpS1 were more pronounced (Fig. S3). As the Trp side chain is bulkier than Phe, it has a more significant influence ratio. As an aid to better visualize the changes in the structure of ClpS1 detected by NMR, residues showing chemical shift perturbations upon model substrate binding are depicted in Fig. 7, over the crystallographic structure obtained by Kim *et al.* [25]. The images nicely show that the region most affected by L-Phe-amide binding is the hydrophobic pocket present in ClpS1.

Binding of destabilizing residues of bacterial N-degron pathways does not elicit significant conformational changes to ClpS1

NMR residue-resolved ^{15}N R_1 ($1/T_1$, spin–lattice relaxation), R_2 ($1/T_2$, spin–spin relaxation), and ^{1}H-^{15}N NOEs are useful probes to monitor protein backbone fluctuations in the ps-ns (T_1, T_2, and NOE) and ms time scales (T_2) [49]. We applied this strategy to characterize the backbone mobility of ClpS1 and determine whether ligand binding modifies these motions. Initially, T_1, T_2, and ^{1}H-^{15}N NOE experiments were recorded for free ClpS1, where two well-defined protein domains with significant differences in backbone mobility were detected (Fig. 8). The first 36 residues displayed R_1, R_2, and NOE values between 2–3.5 Hz, 2–10 Hz, and −1.7 and 0.2, respectively, indicating a highly flexible, disordered conformation. On the other hand, residues 38–109 showed ^{15}N relaxation values fully compatible with a folded, globular protein domain with restricted internal motions. This is especially evident in the ^{1}H-^{15}N NOE profile. A closer inspection of the R_2 values in this domain showed more significant fluctuations than R_1 and NOE. These fluctuations correlated well with the secondary structure content determined by ^{13}C chemical shifts analysis, indicating that the regions with consolidated secondary structure, particularly the helical regions, presented restricted motions as opposed to the more flexible interconnecting loops.

In the presence of 3.0 molar equivalents of L-Phe-amide, the R_1, R_2, and ^{1}H-^{15}N NOE profiles of ClpS1 were similar to those of the free protein, indicating that substrate binding does not elicit significant perturbations in ClpS1 dynamics. However, residues Asp45, Gln74, and Glu75 presented higher R_2 values compared with the uncomplexed protein. These residues are located in the two main interaction regions. So, the increase in R_2 may be the result of decreased mobility and/or exchange processes due to ligand binding.

Bound amino acids interact with ClpS1 with high affinity

NMR titration experiments with L-Trp-amide and L-Phe-amide revealed that the K_d for each interactor lies in the low μM range. Fluorescence anisotropy was used

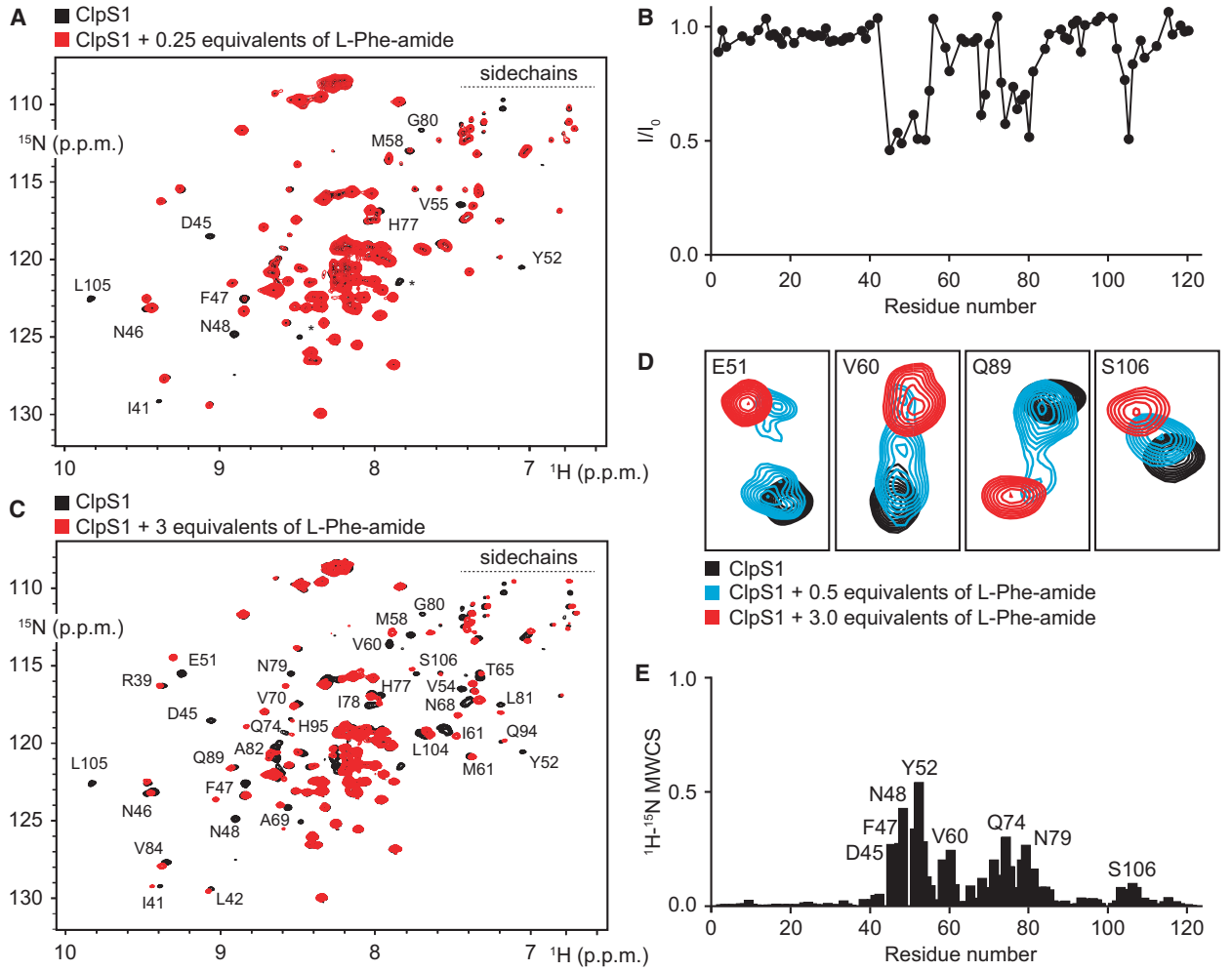

Fig. 6. Interaction between ClpS1 and L-Phe-amide. (A) ^{1}H-^{15}N SOFAST-HMQC spectrum of ClpS1 in the absence (black) and presence (red) of 0.25 equivalents of L-Phe-amide. Shifted peaks are labeled. (B) I/I_0 profile of cross-peak broadening upon substrate addition quantified along the primary sequence of ClpS1. (C) ^{1}H-^{15}N SOFAST-HMQC spectrum of ClpS1 in the absence (black) and presence (red) of 3 equivalents of L-Phe-amide. Shifted peaks are labeled. (D) Enlargement of cross-peaks corresponding to different residues of ClpS1 in the absence (black) and in the presence of 0.5 (cyan) and 3.0 (red) equivalents of L-Phe-amide. Cross-peaks for the free and ligand-bound protein states at the different ligand:protein ratios are visible within the selected spectral region. (E) ^{1}H-^{15}N MWΔCS s (^{1}H-^{15}N MWCS) of ClpS1 residues in the presence of 3.0 equivalents of L-Phe-amide.

to more closely determine binding affinities to these interactors. For ClpS1, we benefited from the fact that it lacks tryptophan in its primary sequence. So, the fluorescence anisotropy of L-Trp-amide was recorded while varying the concentration of ClpS1. The experimental points were fitted to a hyperbolic curve in which the factor K_d could be calculated (Fig. S4A). ClpS1 bound to L-Trp-amide with a K_d of 2.20 ± 0.33 μM, well within the expected value as evidenced from NMR titration experiments. An anisotropy competition assay was set up to calculate the K_d for the binding of ClpS1 to L-Phe-amide. In this case, increasing concentrations of the competitor L-Phe-amide were added to ClpS1 + L-Trp-amide and L-Trp-amide fluorescence anisotropy was measured. Of note, L-Phe-amide fluorescence does not bleed into the emission wavelength of L-Trp-amide. A decrease in L-Trp-amide fluorescence anisotropy was detected when increasing amounts of L-Phe-amide were added to the mixture, indicating that L-Phe-amide effectively displaced L-Trp-amide from its binding site (Fig. S4B). Curve fitting allowed us to extract the necessary parameters to calculate the K_d of ClpS1 for L-Phe-amide, according to Ref. [33]. ClpS1 bound to L-Phe-amide with similar affinity (K_d = 1.83 ± 0.21 μM) as L-Trp-amide.

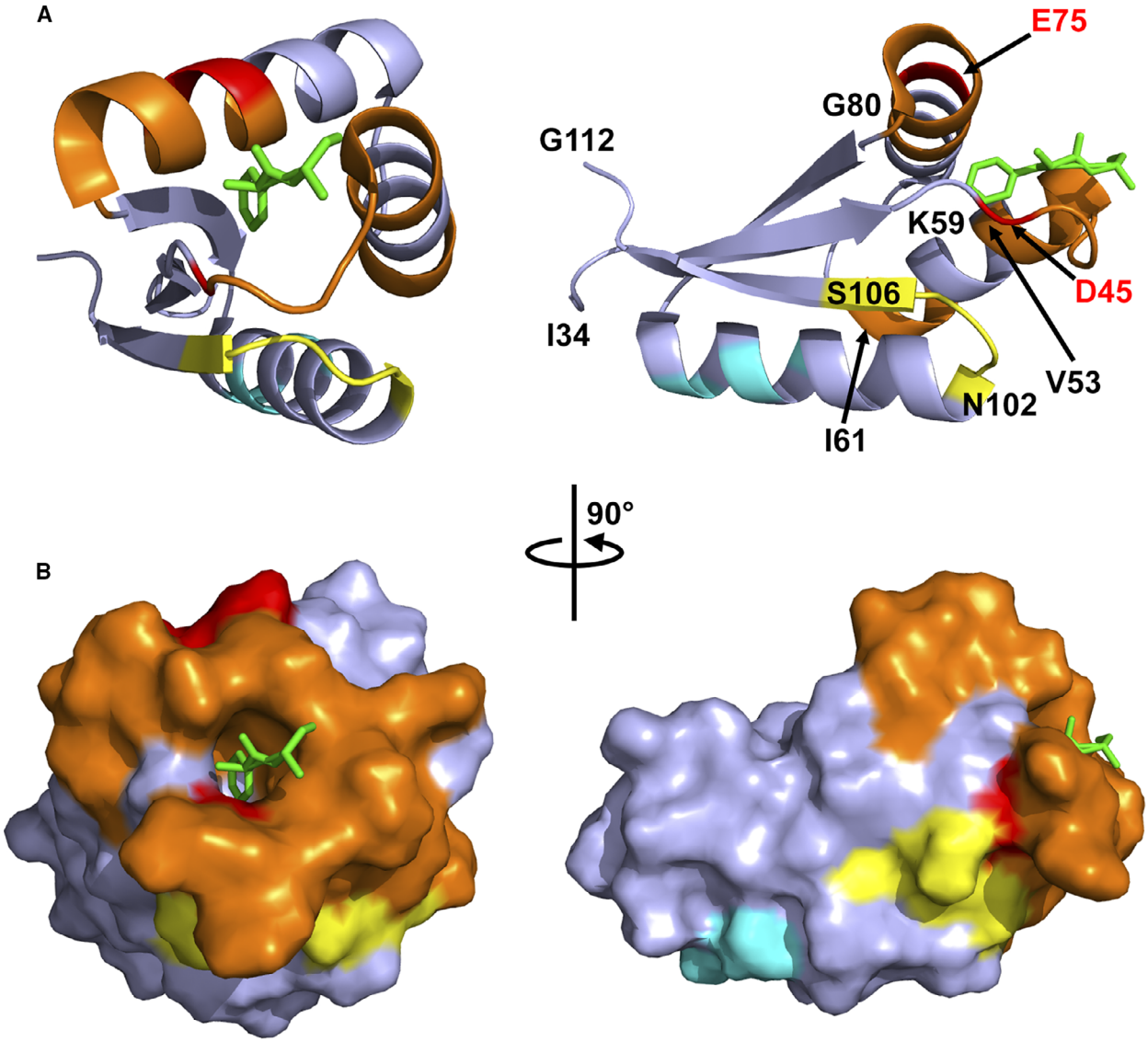

Fig. 7. Backbone perturbations upon substrate binding. The chemical shift perturbations detected by NMR analysis are represented as a color scale on the crystallographic structure of ClpS1 recently obtained by Kim *et al.* (PDB: 7d34, chain A) [25]. Residues suggested to be involved in binding to Hsp100 chaperones are colored in cyan. Residues showing pronounced, medium, and small chemical shift perturbations are colored red, orange, and yellow, respectively. The bound dipeptide Phe-Ala is depicted in green. (A) Ribbon representation and (B) surface representation.

Discussion

The elimination of most proteins by proteolytic machines is not a random process. Potential substrates for degradation possess structural and sequence determinants recognized by adaptors proteins, molecular chaperones, and proteases. Processing by the proteolytic machine then results in the selective elimination of damaged, truncated, misfolded, or unneeded proteins and polypeptides. Knowing the rules governing substrate selection for degradation is important to understand key cellular mechanisms and has biotechnological applications [50]. The half-life of a given protein could be fiddled at will, allowing for controlled manipulation of metabolic pathways or genetic programs. However, after almost 20 years of the discovery of ClpS, the substrate selector for the *E. coli* ClpAP system, its specificity is still not completely defined. That is also true for eukaryotic ClpS proteins, where research into their selectivity and structural features are beginning to surface. In this work, we report new insights into the recognition patterns of chloroplastic

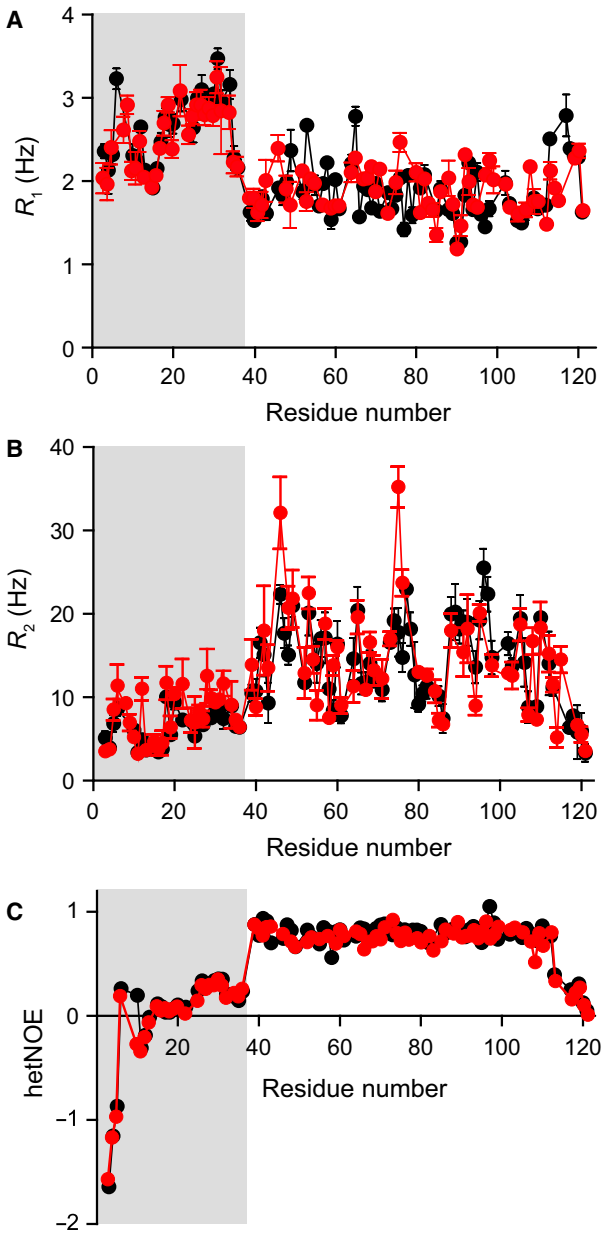

Fig. 8. Backbone dynamics of free and ligand-bound ClpS1. R_1 (A), R_2 (B), and ^1H-^{15}N NOE (C) experiments of ClpS1 in the absence (black) and in the presence of 3.0 equivalents of L-Phe-amide (red). Areas shaded in gray denote the NTE of ClpS1.

ClpS1 from *A. thaliana*, showing that it closely follows the constraints of the N-degron pathway from Gram-negative bacteria but with some differences that may alter the repertoire of recognized substrates compared with its bacterial counterparts.

In vitro, the specificity of ClpS1 has been hitherto analyzed by interaction assays using recombinant ClpS1 and GFP variants as possible interactors. We have previously shown that ClpS1 could bind to all GFP variants starting with primary destabilizing residues of the *E. coli* N-degron pathway (Phe, Tyr, Trp, and Leu) but with low affinity [23]. The N-terminal ends of the GFP variants used in that study were known to bind to bacterial ClpS proteins with high affinity [10,12]. While pull-down assays are useful to uncover potential ClpS1 binders, the experimental setup is not amenable to high throughput. In the past, peptide arrays were successfully used to survey a more extensive set of interacting sequences to bacterial ClpS proteins [11,47]. Using this strategy, we were able to detect ClpS1 binding to primary destabilizing N-terminal residues, analyze the effect of N-terminal acetylation, study the influence of ClpF on substrate recognition by ClpS1, and uncover subtle differences in specificity of ClpS1 and ClpS from *E. coli*.

All N-terminal primary destabilizing residues of the N-degron pathway of many bacteria are represented in the λcI permutation library. ClpS1 recognized all peptides beginning with Phe (five peptides), Tyr (four peptides), and Trp (one peptide). In contrast, ClpS1 bound to three out of six peptides starting with Leu. Montandon *et al.* [24] used pull-down assays with N-terminal GFP variants to probe binding to N-terminal Phe, Trp, Tyr, and Leu, with Arg being the residue in second position. No binding to the GFP variant starting with Tyr-Arg was detected. Our data showed ClpS1 binding to N-terminal Tyr-bearing peptides in the λcI library, although the signal was rather weak (except in spot 35, Tyr followed by Pro). The pattern of recognition of Leu-bearing peptides suggests that some residues immediately after the N-terminal residue can abolish the interaction [e.g., Thr (spot 3), Ser (spot 11), and Leu (spot 22)]. Interestingly, those three peptides were readily recognized by *E. coli* ClpS when the same peptide library was used [12]. Moreover, a GFP derivative starting with Leu-Leu interacted with ClpS1 with extremely low affinity [23].

When surveying the specificity of ClpS adaptors using peptide libraries, binding to noncanonical N-terminal ends is usually detected, but with much less signal than in positions reflecting true interactions. ClpS1 could bind to peptides starting with noncanonical N-terminal ends (positions 6, 7, 9, 17, 32, 34, 61, 63, and 64 in the λcI library). This is probably the result of electrostatic interactions and a high local concentration of peptides. However, we did note that in all cases where ClpS1 bound unspecifically, at least one of the common destabilizing residues of bacterial N-degron pathways lied among the first five amino acids in the sequence. It was recently demonstrated that ClpS from *Salmonella enterica* could bind to destabilizing residues

in the fourth position [51]. However, an N-terminal methionine was also a requirement, which was not the case in the λcI library. More research is needed to establish whether ClpS proteins can recognize destabilizing residues in internal positions within the N-terminal end of their targets.

Consequently, a scanning library was used to conclusively establish the binding pattern of ClpS1 using a peptide that does not contain any charges or cognate destabilizing residues of bacterial N-degron pathways between positions 2 and 5. Using this library, ClpS1 only interacted with peptides starting with Phe, Tyr, and Trp. No nonspecific interactions were noted. ClpS1 did not bind to the peptide starting with Leu-Ser (thus confirming the result observed with the λcI library), whereas ClpS from *E. coli* did. ClpS1 did not bind to other hydrophobic N-terminal amino acids in this library and the λcI array, such as Val, Ile, and Met.

Another requirement for ClpS recognition is the requirement for a free (e.g., unacetylated) N terminus. In bacterial ClpS, N-terminal acetylation abolishes the interaction of the adaptor to its substrates [11]. This post-translational modification is rare in bacteria, as < 2% of all proteins are N-terminally acetylated [52,53]. However, in chloroplasts, N-terminal acetylation is more prominent [44,45]. ClpS1 did not bind to any position of a λcI library containing N-acetylated peptides. It is tempting to speculate that N-terminal acetylation protects proteins bearing primary destabilizing residues for degradation by the Clp system. This is consistent with the observation that chloroplastic mature proteins starting with Tyr or Trp are N-terminally acetylated [44].

ClpF is a newly discovered component of the Clp system that is unique to plants. ClpF interacts with ClpS1, apparently forming a binary complex that potentiates the interaction of ClpS1 to Hsp100 chaperones [26]. ClpF may also alter the repertoire of ClpS1-recognized substrates. Using the peptide libraries, we observed that ClpF did not alter the recognition pattern of ClpS1 in any way. Also, ClpF did not bind to N termini on its own. However, this does not preclude that ClpF could bind proteins by recognizing other regions. Thus, accumulated data and results from this work indicate that ClpF acts in substrate delivery to the Clp system, in conjunction with ClpS1 and not in substrate recognition.

Undoubtedly, the molecular basis for these substrate recognition trends must stem from the structural features of the substrate cavity and other regions. The crystallographic structure of ClpS1 bound to a Phe-Ala dipeptide was recently elucidated. It showed that the hydrogen bonds and hydrophobic interactions of the cavity with its bound degron were similar to other bacterial ClpS [25]. Moreover, the entrance to the substrate cavity is positively charged due to the presence of lysine and arginine residues (Lys49-Arg50), which may favor the interaction with negative charges in the N-terminal end of the substrate [23]. This was actually the case for the artificial substrate FE-linker-GFP, as it binds to ClpS1 more strongly than the model bacterial substrate FR-linker-GFP.

The experimental evidence on the specificity of ClpS1 is presented in Table 1, which compiles results from this work and others. Many trends clearly arise, which can be summarized as follows: (a) ClpS1 can bind to the four cognate N-terminal residues of the *E. coli* N-degron pathway, (b) ClpS1 can bind to N-terminal Phe and Trp, regardless of the residue in the second position, (c) binding to N-terminal Phe is enhanced by a negative charge in the second position, (d) binding to N-terminal Tyr is restricted and depends on the residue in the second position, (e) binding to N-terminal Leu is even more restricted but not prohibited, (f) ClpS1 does not seem to bind to N-terminal Ile or Met, and (g) N-terminal acetylation inhibits the binding of ClpS1 to its targets.

The recombinant ClpS1 that was used to solve its crystallographic structure did not contain the NTE, which may play a role in substrate binding or change conformation upon target recognition. Also, the apo form structure was not obtained; so, details of cavity rearrangement upon substrate binding cannot be drawn. Here, we used NMR spectroscopy to further the knowledge of the structure of ClpS1. NMR analysis demonstrated that substrate binding does not elicit significant conformational changes in the globular protein core and the disordered NTE. Indeed, both the NTE and the core domain remained disordered and folded, respectively, upon ligand binding and did not experience major perturbations in their backbone dynamics. This suggests that before binding, the cavity is already formed. We did observe an effect in the last β-sheet (102–106), indicating that substrate binding to the main pocket elicits small conformational perturbations at this region. Residues involved in the interaction with substrates included all those predicted by sequence alignments to bacterial ClpS [22] and those interacting with the Phe-Ala dipeptide in the X-ray structure [25]. However, shifted peaks of nearby residues were also detected, suggesting backbone rearrangements beyond the cavity upon substrate binding. Moreover, no structural perturbations were noticed in residues involved in binding to Hsp100 chaperones (Glu90, Glu93, and Lys95), which implies that

Table 1. Summary of *A. thaliana* ClpS1 specificity. Sequences only account for the first two positions of the targets. ITC, isothermal titration calorimetry.

	Binding	Method	Reference		Binding	Method	Reference
Phe-Ala	++	ITC/X-ray	[25]	Tyr-Ala	++	ITC	[25]
Phe-Arg	+	Pull-down	This work, [23,24]	Tyr-Arg	–	Pull-down	[24]
Phe-Glu	++	Pull-down	This work	Tyr-Glu	+/–	Array	This work
Phe-Leu	++	Array	This work	Tyr-Leu	+	Pull-down	[23]
Phe-Pro	++	Array	This work	Tyr-Lys	+/–	Pull-down	[23]
Phe-Ser	++	Array	This work	Tyr-Pro	++	Array	This work
Phe-Thr	+	Array	This work	Tyr-Ser	++	Array	This work
Phe-amide	++	Anisotropy	This work				
Leu-Ala	–	ITC	[25]	Trp-Ala	++	ITC	[25]
Leu-Arg	++	Array/pull-down	This work, [24]	Trp-Arg	++	Pull-down	[24]
Leu-Gly	+	Array	This work	Trp-Leu	++	Array/pull-down	This work, [24]
Leu-Leu	–	Array / pull-down	This work, [23]	Trp-Phe	+/–	Pull-down	[23]
Leu-Ser	–	Array	This work	Trp-Ser	++	Array	This work
Leu-Thr	–	Array	This work	Trp-amide	++	Anisotropy	This work
Ile-Ala	–	Array/ITC	This work, [25]				
Ile-Arg	–	Pull-down	[24]				
Ile-Leu	–	Array/pull-down	This work, [24]				
Met-Leu	–	Pull-down	[23]				

substrate binding may not be a signal for association with the Hsp100 chaperones.

Solution state NMR spectroscopy complements other ultrastructural techniques such as X-ray crystallography. For example, the subtle conformational changes detected in this study may not be visible in rigid crystals. Also, due to its high mobility, the full NTE does not tend to be visible in electron density maps [54–56]. Using NMR, we characterized the conformational properties and backbone dynamics of this disordered domain and demonstrated that it is not involved in ligand binding, suggesting that it may play other roles in the Clp machinery. According to current models, the NTE of bacterial ClpS is pulled into the axial pore of ClpA, causing a rearrangement of ClpS architecture resulting in the handoff of the substrate. Also, deletions at the NTE domain of bacterial ClpS prevent efficient substrate delivery to the ClpA pore and decrease the affinity of the ClpS-ClpA complex in *Caulabacter crescentus* [48]. In this way, the disordered NTE plays a crucial role in initiating the formation of the ClpS-ClpA complex that allows for substrate delivery to the protease. Our results showed that the NTE does not adopt a full random coil conformation but has transient β-sheet and α-helical secondary structure elements at defined positions. These regions may fold upon binding to the Hsp100 chaperone to stabilize a productive complex, as observed for other intrinsically disordered protein regions experiencing disordered-to-ordered transitions upon complex formation [57]. We are now using NMR combined with secondary structure destabilizing point mutants at the NTE to test this hypothesis in the plant system.

In conclusion, ClpS1 from *A. thaliana* chloroplasts recognizes the four cognate primary destabilizing residues of the *E. coli* N-degron pathway, but the chemical identity of the residue in the second position can significantly influence the interaction, and N-terminal acetylation completely annuls it. This set of constraints is compatible with N-terminome data of stromal proteins [44]. This study found that the most frequent N-terminal residues were serine, threonine, alanine, valine, and glycine. Aromatic amino acids were 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. Finally, the structural characteristics of ClpS1 show great juxtaposition to features present in bacterial homologs. All of this evidence indicates that an N-degron pathway, akin to bacterial ones, operates in chloroplasts, but with some variations that reflect the evolutionary path of organelles and bacteria.

Acknowledgements

We thank Dr. Hyun Kyu Song for providing us the coordinates of ClpS1 (PDB 7d34) prior to its release. DAL gratefully acknowledges Deutscher Akademischer Austauschdienst (DAAD) for financial support to visit A.M. laboratory. CSL acknowledges SECTEI CDMX, Mexico, for a postdoctoral fellowship. We thank Alejandro Gago and Andrea Coscia for the

maintenance of the NMR facility. This work was supported by Agencia Nacional de Promoción de la Investigación, el Desarrollo Tecnológico y la Innovación (Grant numbers PICT 2018-4494 to EAC and PICT 2014-0825 to GLR), and Universidad Nacional de Rosario (Grant Number UNR BIO497 to EAC).

Data Accessibility

The data that support the findings of this study are available in the Supporting information and from the corresponding author (rosano@ibr-conicet.gov.ar) upon reasonable request.

Author contributions

DAL performed most of the experimental work, under the supervision of AM, EAC, and GLR. AC contributed to the body of experimental work. CSL and AB performed the NMR analysis. EAC and GLR conceptualized the project. All authors were involved in the writing of the manuscript and approved its final version.

References

1 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.
2 Varshavsky A (2011) The N-end rule pathway and regulation by proteolysis. *Protein Sci* **20**, 1298–1345.
3 Wickner S, Gottesman S, Skowyra D, Hoskins J, McKenney K and Maurizi MR (1994) A molecular chaperone, ClpA, functions like DnaK and DnaJ. *Proc Natl Acad Sci USA* **91**, 12218–12222.
4 Seol JH, Woo KM, Kang MS, Ha DB and Chung CH (1995) Requirement of ATP hydrolysis for assembly of ClpA/ClpP complex, the ATP-dependent protease Ti in *Escherichia coli*. *Biochem Biophys Res Commun* **217**, 41–51.
5 Gottesman S, Clark WP, de Crecy-Lagard V and Maurizi MR (1993) ClpX, an alternative subunit for the ATP-dependent Clp protease of *Escherichia coli*. Sequence and *in vivo* activities. *J Biol Chem* **268**, 22618–22626.
6 Shin DH, Lee CS, Chung CH and Suh SW (1996) Molecular symmetry of the ClpP component of the ATP-dependent Clp protease, an *Escherichia coli* homolog of 20 S proteasome. *J Mol Biol* **262**, 71–76.
7 Wang J, Hartling JA and Flanagan JM (1997) The structure of ClpP at 2.3 A resolution suggests a model for ATP-dependent proteolysis. *Cell* **91**, 447–456.
8 Kessel M, Maurizi MR, Kim B, Kocsis E, Trus BL, Singh SK and Steven AC (1995) Homology in structural organization between *E. coli* ClpAP protease and the eukaryotic 26 S proteasome. *J Mol Biol* **250**, 587–594.
9 Maurizi MR, Singh SK, Thompson MW, Kessel M and Ginsburg A (1998) Molecular properties of ClpAP protease of *Escherichia coli*: ATP-dependent association of ClpA and clpP. *Biochemistry* **37**, 7778–7786.
10 Wang KH, Sauer RT and Baker TA (2007) ClpS modulates but is not essential for bacterial N-end rule degradation. *Genes Dev* **21**, 403–408.
11 Erbse A, Schmidt R, Bornemann T, Schneider-Mergener J, Mogk A, Zahn R, Dougan DA and Bukau B (2006) ClpS is an essential component of the N-end rule pathway in *Escherichia coli*. *Nature* **439**, 753–756.
12 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.
13 Roman-Hernandez G, Grant RA, Sauer RT and Baker TA (2009) Molecular basis of substrate selection by the N-end rule adaptor protein ClpS. *Proc Natl Acad Sci USA* **106**, 8888–8893.
14 Mogk A, Schmidt R and Bukau B (2007) The N-end rule pathway for regulated proteolysis: prokaryotic and eukaryotic strategies. *Trends Cell Biol* **17**, 165–172.
15 Ninnis RL, Spall SK, Talbo GH, Truscott KN and Dougan DA (2009) Modification of PATase by L/F-transferase generates a ClpS-dependent N-end rule substrate in *Escherichia coli*. *EMBO J* **28**, 1732–1744.
16 Bouchnak I and van Wijk KJ (2021) Structure, function and substrates of Clp AAA+ protease systems in cyanobacteria, plastids and apicoplasts; a comparative analysis. *J Biol Chem* **296**, 100338.
17 Olinares PD, Kim J, Davis JI and van Wijk KJ (2011) Subunit stoichiometry, evolution, and functional implications of an asymmetric plant plastid ClpP/R protease complex in Arabidopsis. *Plant Cell* **23**, 2348–2361.
18 Peltier JB, Ripoll DR, Friso G, Rudella A, Cai Y, Ytterberg J, Giacomelli L, Pillardy J and van Wijk KJ (2004) Clp protease complexes from photosynthetic and non-photosynthetic plastids and mitochondria of plants, their predicted three-dimensional structures, and functional implications. *J Biol Chem* **279**, 4768–4781.
19 Rosano GL, Bruch EM and Ceccarelli EA (2011) Insights into the CLP/HSP100 chaperone system from chloroplasts of *Arabidopsis thaliana*. *J Biol Chem* **286**, 29671–29680.
20 Rosano GL, Bruch EM, Colombo CV and Ceccarelli EA (2012) Toward a unified model of the action of CLP/HSP100 chaperones in chloroplasts. *Plant Signal Behav* **7**, 672–674.
21 Bruch EM, Rosano GL and Ceccarelli EA (2012) Chloroplastic Hsp100 chaperones ClpC2 and ClpD interact *in vitro* with a transit peptide only when it is

located at the N-terminus of a protein. *BMC Plant Biol* **12**, 57.
22 Nishimura K, Asakura Y, Friso G, Kim J, Oh SH, Rutschow H, Ponnala L and van Wijk KJ (2013) ClpS1 is a conserved substrate selector for the chloroplast Clp protease system in Arabidopsis. *Plant Cell*. **25**, 2276–2301.
23 Colombo CV, Rosano GL, Mogk A and Ceccarelli EA (2018) A gatekeeper residue of ClpS1 from *Arabidopsis thaliana* chloroplasts determines its affinity towards substrates of the bacterial N-End Rule. *Plant Cell Physiol* **59**, 624–636.
24 Montandon C, Dougan DA and van Wijk KJ (2019) N-degron specificity of chloroplast ClpS1 in plants. *FEBS Lett* **593**, 962–970.
25 Kim L, Heo J, Kwon DH, Shin JS, Jang SH, Park ZY and Song HK (2020) Structural basis for the N-degron specificity of ClpS1 from *Arabidopsis thaliana*. Protein science: a publication of the Protein Society.
26 Nishimura K, Apitz J, Friso G, Kim J, Ponnala L, Grimm B and van Wijk KJ (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.
27 Dougan DA, Reid BG, Horwich AL and Bukau B (2002) ClpS, a substrate modulator of the ClpAP machine. *Mol Cell* **9**, 673–683.
28 Sharrocks AD (1994) A T7 expression vector for producing N- and C-terminal fusion proteins with glutathione S-transferase. *Gene* **138**, 105–108.
29 Marley J, Lu M and Bracken C (2001) A method for efficient isotopic labeling of recombinant proteins. *J Biomol NMR* **20**, 71–75.
30 Correa A, Ortega C, Obal G, Alzari P, Vincentelli R and Oppezzo P (2014) Generation of a vector suite for protein solubility screening. *Front Microbiol* **5**, 67.
31 Felgner PL and Wilson JE (1976) Hexokinase binding to polypropylene test tubes. Artifactual activity losses from protein binding to disposable plastics. *Anal Biochem* **74**, 631–635.
32 Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685.
33 Rossi AM and Taylor CW (2011) Analysis of protein-ligand interactions by fluorescence polarization. *Nat Protoc* **6**, 365–387.
34 Kazimierczuk K and Orekhov V (2015) Non-uniform sampling: post-Fourier era of NMR data collection and processing. *Magn Reson Chem* **53**, 921–926.
35 Wishart DS and Sykes BD (1994) The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data. *J Biomol NMR* **4**, 171–180.
36 Wishart DS, Bigam CG, Yao J, Abildgaard F, Dyson HJ, Oldfield E, Markley JL and Sykes BD (1995) 1H, 13C and 15N chemical shift referencing in biomolecular NMR. *J Biomol NMR* **6**, 135–140.
37 Bertoncini CW, Rasia RM, Lamberto GR, Binolfi A, Zweckstetter M, Griesinger C and Fernandez CO (2007) Structural characterization of the intrinsically unfolded protein beta-synuclein, a natural negative regulator of alpha-synuclein aggregation. *J Mol Biol* **372**, 708–722.
38 Schanda P, Kupce E and Brutscher B (2005) SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds. *J Biomol NMR* **33**, 199–211.
39 Cavanagh J, Fairbrother WJ, Palmer AGI, Rance M and Skelton NJ (2007) Protein NMR Spectroscopy: Principles and Practice, 2nd edn. Elsevier Academic Press, San Diego, CA.
40 Farrow NA, Zhang O, Forman-Kay JD and Kay LE (1997) Characterization of the backbone dynamics of folded and denatured states of an SH3 domain. *Biochemistry* **36**, 2390–2402.
41 Goddard TD and Kneller DG (2002) Sparky 3 (San Francisco: University of California).
42 Keller RLJ (2004) The Computer Aided Resonance Assignment Tutorial. (Cantina Verlag and Rochus Keller, Goldau, Switzerland).
43 Bienvenut WV, Sumpton D, Martinez A, Lilla S, Espagne C, Meinnel T and Giglione C (2012) Comparative large scale characterization of plant versus mammal proteins reveals similar and idiosyncratic N-alpha-acetylation features. *Mol Cell Proteomics* **11** (M111), 015131.
44 Rowland E, Kim J, Bhuiyan NH and van Wijk KJ (2015) The arabidopsis chloroplast stromal N-terminome: complexities of amino-terminal protein maturation and stability. *Plant Physiol* **169**, 1881–1896.
45 Bienvenut WV, Espagne C, Martinez A, Majeran W, Valot B, Zivy M, Vallon O, Adam Z, Meinnel T and Giglione C (2011) Dynamics of post-translational modifications and protein stability in the stroma of *Chlamydomonas reinhardtii* chloroplasts. *Proteomics* **11**, 1734–1750.
46 Becker W, Bhattiprolu KC, Gubensak N and Zangger K (2018) Investigating protein-ligand interactions by solution nuclear magnetic resonance spectroscopy. *Chem Phys Chem* **19**, 895–906.
47 Stein BJ, Grant RA, Sauer RT and Baker TA (2016) Structural basis of an N-degron adaptor with more stringent specificity. *Structure* **24**, 232–242.
48 Roman-Hernandez G, Hou JY, Grant RA, Sauer RT and Baker TA (2011) The ClpS adaptor mediates staged delivery of N-end rule substrates to the AAA+ ClpAP protease. *Mol Cell* **43**, 217–228.
49 Boehr DD, Dyson HJ and Wright PE (2006) An NMR perspective on enzyme dynamics. *Chem Rev* **106**, 3055–3079.

50 Dissmeyer N (2019) Conditional protein function via N-degron pathway-mediated proteostasis in stress physiology. *Annu Rev Plant Biol* **70**, 83–117.
51 Gao X, Yeom J and Groisman EA (2019) The expanded specificity and physiological role of a widespread N-degron recognin. *Proc Natl Acad Sci USA* **116**, 18629–18637.
52 Christensen DG, Baumgartner JT, Xie X, Jew KM, Basisty N, Schilling B, Kuhn ML and Wolfe AJ (2019) Mechanisms detection, and relevance of protein acetylation in prokaryotes. *MBio* **10**, e02708-18.
53 Schmidt A, Kochanowski K, Vedelaar S, Ahrne E, Volkmer B, Callipo L, Knoops K, Bauer M, Aebersold R and Heinemann M (2016) The quantitative and condition-dependent *Escherichia coli* proteome. *Nat Biotechnol* **34**, 104–110.
54 Zeth K, Ravelli RB, Paal K, Cusack S, Bukau B and Dougan DA (2002) Structural analysis of the adaptor protein ClpS in complex with the N-terminal domain of ClpA. *Nat Struct Biol* **9**, 906–911.
55 Guo F, Esser L, Singh SK, Maurizi MR 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.
56 Xia D, Esser L, Singh SK, Guo F and Maurizi MR (2004) Crystallographic investigation of peptide binding sites in the N-domain of the ClpA chaperone. *J Struct Biol* **146**, 166–179.
57 Theillet FX, Binolfi A, Frembgen-Kesner T, Hingorani K, Sarkar M, Kyne C, Li C, Crowley PB, Gierasch L, Pielak GJ *et al.* (2014) Physicochemical properties of cells and their effects on intrinsically disordered proteins (IDPs). *Chem Rev* **114**, 6661–6714.

Supporting information

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

Fig. S1. Sequences of peptides in the λcI library.

Fig. S2. ClpS1 secondary structure.

Fig. S3. Interaction between ClpS1 and L-Trp-amide.

Fig. S4. Fluorescence anisotropy measurements to determine dissociation constants of ClpS1 to model interactors.

Table S1. NMR chemical shifts table of ClpS1 backbone resonances.