

# SCIENTIFIC REPORTS



OPEN

## Crystal structures reveal N-terminal Domain of *Arabidopsis thaliana* ClpD to be highly divergent from that of ClpC1

Received: 02 September 2016

Accepted: 07 February 2017

Published: 13 March 2017

Chinmayee Mohapatra\*, Manas Kumar Jagdev\* &amp; Dileep Vasudevan

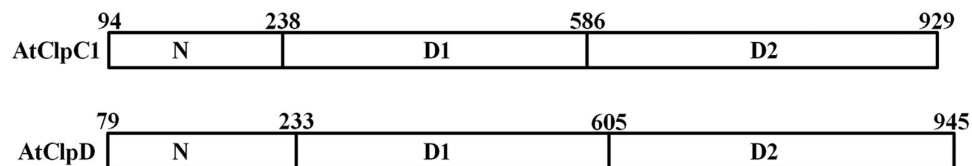
The caseinolytic protease machinery associated chaperone protein ClpC is known to be present in bacteria, plants and other eukaryotes, whereas ClpD is unique to plants. Plant ClpC and ClpD proteins get localized into chloroplast stroma. Herein, we report high resolution crystal structures of the N-terminal domain of *Arabidopsis thaliana* ClpC1 and ClpD. Surprisingly, AtClpD, but not AtClpC1, deviates from the typical N-terminal repeat domain organization of known Clp chaperones and have only seven  $\alpha$ -helices, instead of eight. In addition, the loop connecting the two halves of AtClpD NTD is longer and covers the region which in case of AtClpC1 is thought to contribute to adaptor protein interaction. Taken together, the N-terminal domain of AtClpD has a divergent structural organization compared to any known Clp chaperones which hints towards its specific role during plant stress conditions, as opposed to that in the maintenance of chloroplastic homeostasis by AtClpC1. Conservation of residues in the NTD that are responsible for the binding of the cyclic peptide activator - Cyclomarin A, as reported for mycobacterial ClpC1 suggests that the peptide could be used as an activator to both AtClpC1 and AtClpD, which could be useful in their detailed *in vitro* functional characterization.

Plants being sessile organisms are at the risk of getting exposed to a multitude of environmental stress conditions which can cause accumulation and aggregation of misfolded protein species. In this context, the plant caseinolytic protease (Clp) complex plays an important role in degradation of accumulated and misfolded proteins in chloroplasts wherein the chaperones aid in ATP-dependent unfolding of protein substrates to be degraded by the associated protease machinery<sup>1</sup>. Like its counterpart in eubacteria, Clp machinery of higher plants also possesses a barrel-shaped protease core, capped by a homo-hexamers of Clp chaperones belonging to the AAA<sup>+</sup> superfamily of HSP100 proteins<sup>1-2</sup>. In higher plants four different Clp chaperones namely ClpB3, C1, C2 and D are known to get localized into chloroplastic stroma and one each in cytoplasm (ClpB1) and mitochondria (ClpB4) of which ClpC1, C2 and D are known to associate with Clp protease barrel<sup>1,3-5</sup>.

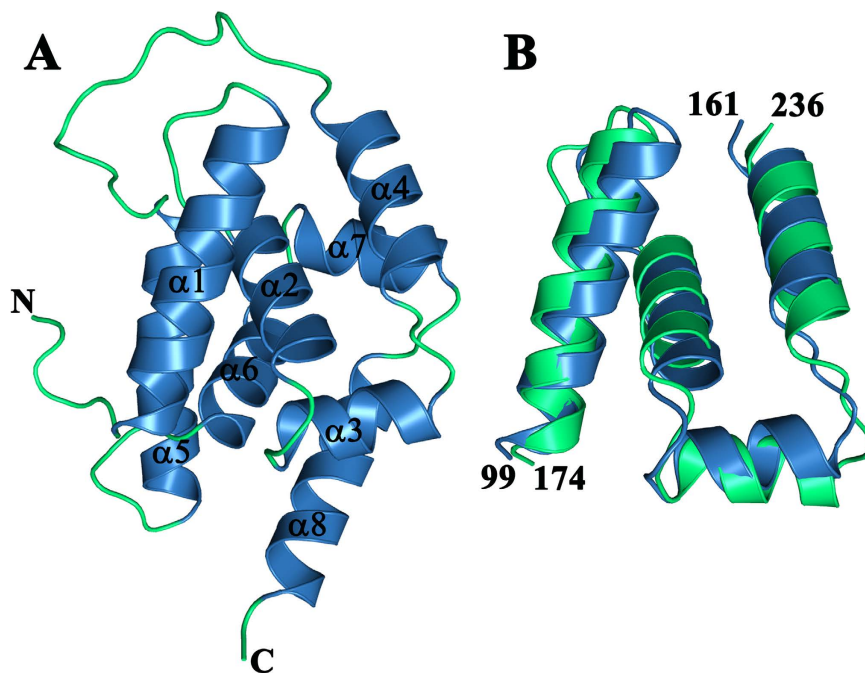
The function of chloroplastic Clp chaperones in plants does not restrict to housekeeping duties that are essential for chloroplast development and maintenance as performed by ClpC1<sup>6</sup> and ClpC2, but also include special duties performed during stress-conditions as performed by ClpB3 and ClpD<sup>4,7</sup>. AtClpC1 & AtClpC2 are expressed constitutively, whereas constitutive level of ClpD is relatively low and its expression gets upregulated only during stress conditions such as high salt, dehydration, cold, darkness-induced etiolation and senescence<sup>7-10</sup>, possibly hinting towards their different substrate preferences. In fact, ClpD was originally named as Early Responsive to Dehydration 1 (ERD1), suggesting its increased expression in response to dehydration. Over-expression of rice ClpD1 protein was found to enhance tolerance to salt and desiccation stresses in transgenic *Arabidopsis* plants<sup>11</sup>.

AtClpC1 and AtClpC2 share 92% identity at the amino acid sequence level, indicating functional overlap. However, the two proteins share only about 48% sequence identity with AtClpD<sup>12</sup>. ClpC1 is highly conserved among the various plant species with more than 90% sequence identity, suggesting high functional importance in plants, whereas, ClpD which has been reported only in plants<sup>13,14</sup> is less conserved across plant species (up to as low as 70% sequence identity). The structured, mature forms of AtClpC1/C2/D consist of three distinct domains such as a short N-terminal domain (NTD) and two ATPase domains<sup>10</sup> (Fig. 1), a feature very similar to bacterial

Institute of Life Sciences, Bhubaneswar - 751023, Odisha State, India. \*These authors contributed equally to this work. Correspondence and requests for materials should be addressed to D.V. (email: dileep@ils.res.in)



**Figure 1. Domain organization of AtClpC1 and AtClpD.** The organization of functional domains of AtClpC1 and AtClpD are shown with their respective amino acid residue boundaries. N corresponds to the N-terminal domain, D1 the first ATPase domain and D2 the second ATPase domain.



**Figure 2. Structure of AtClpC1 N-terminal domain.** (A) Cartoon representation of the structure of *A. thaliana* ClpC1 NTD with the helices (blue) and loops (green). The N and C termini as well as the eight  $\alpha$ -helices are labeled. (B) Structural alignment of AtClpC1 residues 99–161 (blue) with residues 174–236 (green).

ClpC proteins<sup>15</sup>. The ATPase domains are essential to bring about ATP-dependent unfolding of proteins to be degraded by the protease core.

Interestingly, more sequence variation between AtClpC1/C2 and AtClpD occurs at the N-terminal domain (34% identity) as compared to the remaining stretch (53% identity). It is the N-terminal domain of Clp chaperones that comes into direct contact with adaptor proteins (if any). The N-terminal domain of ClpC1 is known to perform the regulatory function of substrate recognition, either directly or indirectly through the adaptor proteins ClpS1/ClpF for which it becomes a binding platform<sup>16–18</sup>, whereas no adaptor proteins have so far been reported for ClpD. Even though *in vitro* ATPase activity level of AtClpD has been reported to be lower compared to AtClpC2<sup>14</sup>, it appears that it is the regulatory N-terminal domain and its structure that primarily differentiates the two chaperones, as a primary requirement to recognize and recruit different protein substrates. So we wanted to know whether the N-terminal domain structures of AtClpC1 and AtClpD are considerably different. Here we report and analyse the crystal structures of AtClpC1 residues 94 to 238 and AtClpD residues 79 to 233 (hereafter mentioned as AtClpC1 NTD and AtClpD NTD respectively) that correspond to the ClpC NTD of eubacteria.

## Results and Discussion

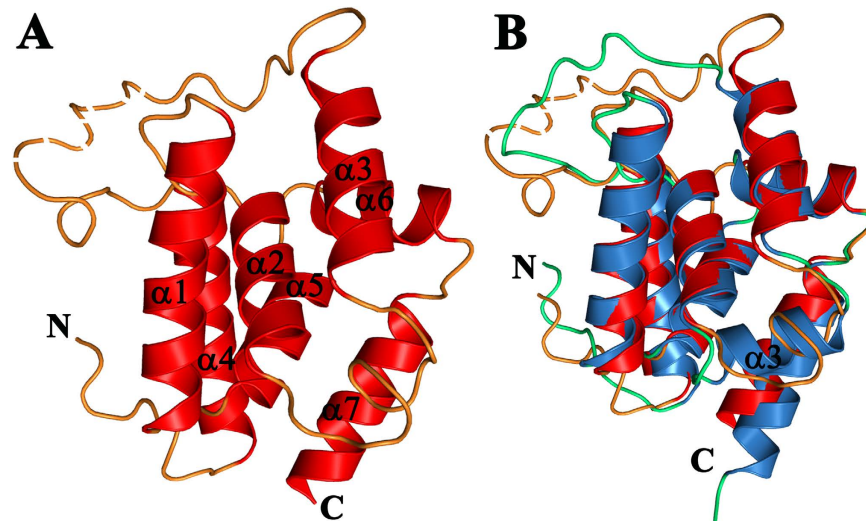
**AtClpC1 NTD structure.** The 1.20 Å structure showed AtClpC1 N-terminal domain (residues 94–238; Fig. 2A) to be a  $\alpha$ -helical domain, having 66% helical content and made up of eight tightly packed  $\alpha$ -helices:  $\alpha 1$  (E100–L116),  $\alpha 2$  (T123–E133),  $\alpha 3$  (I137–M145),  $\alpha 4$  (L149–I160),  $\alpha 5$  (P175–L191),  $\alpha 6$  (S198–E208),  $\alpha 7$  (V212–L220) and  $\alpha 8$  (P224–N238). The two central helices:  $\alpha 2$  and  $\alpha 6$  which are rich in the hydrophobic residues Leu and Ile contribute significantly to the packing. All residues of the NTD, as well as two residues from the linker of C-terminal His tag could also be seen in the electron density, allowing a complete model to be built. In addition, five phosphate ions were also fit into the density. The final model was refined to an  $R$  factor of 15.6% and an  $R_{\text{free}}$  of 17.5% for all data between 17.54 Å and 1.20 Å resolution. Ramachandran plot values revealed no outliers for the structure. The data collection and refinement statistics can be found in Table 1.

Parameter	AtClpC1 NTD	AtClpD NTD
<b>Data Collection</b>		
Beamline	RRCAT-BL21	ESRF-BM14
Detector type	Marmosaic 225 mm CCD	Marmosaic 225 mm CCD
Wavelength (Å)	0.9794	0.9537
Data collection temperature (K)	103	100
Space group	$P2_12_12_1$	$P2_1$
a, b, c (Å)	40.10, 44.29, 97.56	38.08, 37.31, 99.98
$\alpha, \beta, \gamma$ (°)	90, 90, 90	90, 96.08, 90
Resolution (Å)	17.54–1.20 (1.26–1.20)	37.87–1.60 (1.63–1.60)
$R_{\text{meas}}$ (%)	7.5 (63.4)	8.8 (78.1)
$I/\sigma I$	9.4 (2.7)	9.8 (2.2)
CC(1/2) (%)	99.8 (80.5)	99.5 (71.7)
Total number of reflections	262145	351287
Mosaicity (°)	0.43	0.41
Completeness (%)	98.4 (100)	100 (100)
Multiplicity	4.8 (4.7)	4.2 (4.1)
Wilson B-factor (Å <sup>2</sup> )	11.9	18.2
Matthews coefficient (Å <sup>3</sup> /Da)	2.66	2.00
Solvent content (%)	53.75	38.63
No. of molecules in ASU	1	2
<b>Refinement</b>		
No. of unique reflections	54126	37033
$R_{\text{work}}/R_{\text{free}}$ (%)	15.6/17.5	20.5/22.3
Total no. of non-H atoms	1459	2450
No. of water molecules	130	104
No. of ligands	5 phosphate ions	0
Mean B-factor (Å <sup>2</sup> )	18.0	23.0
<b>R.m.s. deviations</b>		
Bond lengths (Å)	0.016	0.012
Bond angles (°)	1.931	1.370
<b>Ramachandran plot values (%)</b>		
Favoured/Outliers	100/0	100/0

**Table 1. Data collection and refinement statistics for AtClpC1 NTD and AtClpD NTD.** Numbers in parentheses correspond to the last resolution shell.  $R_{\text{meas}} = \sum_h (n/n - 1)^{1/2} \sum_i |I_i(h) - \langle I(h) \rangle| / \sum_h \sum_i I_i(h)$ , where  $I_i(h)$  and  $\langle I(h) \rangle$  are the  $i$ th and mean measurement of the intensity of reflection  $h$ .  $R_{\text{work}} = \sum_h |F_{\text{obs}}(h) - |F_{\text{calc}}(h)|| / \sum_h |F_{\text{obs}}(h)|$ , where  $F_{\text{obs}}(h)$  and  $F_{\text{calc}}(h)$  are the observed and calculated structure factors, respectively.  $R_{\text{free}}$  is the R value obtained for a test set of reflections consisting of a randomly selected 5% subset of the data set excluded from refinement.

The fold of AtClpC1 NTD contains two repeats of a 4-helix motif that share 46% sequence identity and 84.1% similarity in a region of 63-residue overlap. A 14-residue long loop between  $\alpha 4$  and  $\alpha 5$  connects the two repeats. The two halves that stay in close proximity are related by a pseudo two-fold axis of symmetry. Residues 99–161 and 174–236 superpose with a r.m.s. deviation of about 1.77 Å for 46 C $\alpha$  atoms (Fig. 2B). Strong sequence identity and conservation of secondary structure between the two halves suggest the possibility of an early evolutionary gene duplication event that could have resulted in this repeated motif, as has been predicted for the N-terminal domain of bacterial ClpA and other ClpC proteins<sup>19</sup>. A query to the DALI database<sup>20</sup> and PDBeFold server<sup>21</sup> with AtClpC1 NTD structure coordinates yielded NTD of mycobacterial ClpC1 and other Gram-positive bacterial ClpC as the most similar structures. It also showed structural similarity to AtClpT1/T2 proteins, as well as the NTD of malarial and bacterial ClpB proteins and bacterial ClpA proteins.

**AtClpD NTD structure.** The 1.6 Å structure of AtClpD N-terminal domain (residues 79–233; Fig. 3A) revealed a  $\alpha$ -helical structure, having 61% helical content. The domain is made up of seven helices:  $\alpha 1$  (E85-L101),  $\alpha 2$  (T106-E118),  $\alpha 3$  (I132-D144),  $\alpha 4$  (I170-M186),  $\alpha 5$  (P191-V203),  $\alpha 6$  (S207-L215) and  $\alpha 7$  (M219-K231). The crystal structure included two molecules in an asymmetric unit. The two molecules had almost identical structures; aligning with an r.m.s. deviation of 0.19 Å for 112 C $\alpha$  atoms. One of the molecules had better electron density and hence could fit more residues into the structure. For this molecule, all residues of the NTD, except four in the long middle loop and one at C-terminus could be fit into the density. In addition, four residues from the N-terminal linker stretch after GST-tag cleavage could also be fit into the structure. The molecule had an average B value of 21.91 Å<sup>2</sup>. The second molecule did not have any visible electron density for eleven residues of the



**Figure 3. Structure of AtClpD N-terminal domain.** (A) Cartoon representation of the structure of *A. thaliana* ClpD NTD with the helices (red) and loops (orange). The N and C termini as well as the seven  $\alpha$ -helices are labeled. The four-residue long stretch in the long middle loop (that could not be seen in electron density) was fitted into the structure and is shown with intermittent gaps. (B) Structural alignment of AtClpD NTD (red helices and orange loops) with AtClpC1 NTD (blue helices and green loops). The N and C termini as well as  $\alpha 3$  of AtClpC1 NTD are labeled.

middle loop and two residues of the C-terminus and had a slightly higher average B value of  $26.75 \text{ \AA}^2$ . The final model was refined to an  $R$  factor of 20.5% and an  $R_{\text{free}}$  of 22.3% for all data between  $37.87 \text{ \AA}$  and  $1.6 \text{ \AA}$  resolution. There were no Ramachandran plot value outliers in the structure. The data collection and refinement statistics can be found in Table 1.

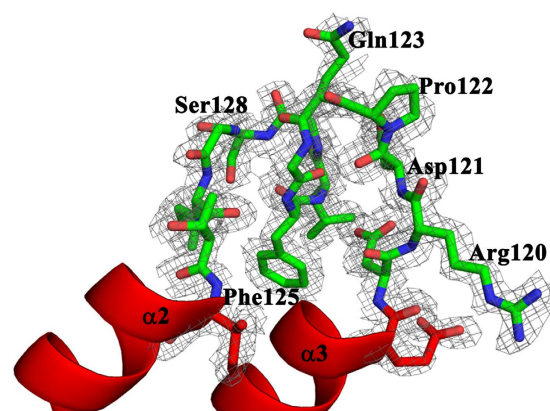
The structure reveals a longer middle loop for AtClpD NTD when compared to AtClpC1 NTD. This feature was obvious when the NTD sequences were aligned to each other. The crystal structure additionally shows the 26-residue long middle loop to take a different conformation compared to the 14-residue long loop of AtClpC1 NTD (Fig. 3B). Interestingly, the sequence of this long middle loop is the least conserved across the ClpD isoforms from different plant species, with its length varying between 19 and 29 residues in the sequences we analyzed (Fig. 4). Also, unlike in AtClpC1, the two halves of AtClpD NTD do not form the typical four-helix motif repeats. In AtClpD, residues 122–128, which makes the short stretch corresponding to  $\alpha 3$  of AtClpC1 NTD forms a loop instead of helix (Fig. 3B). The loop region spanning residues Asp119 to Thr131 had unambiguous electron density in the difference Fourier omit map, supporting the absence of a helix (Fig. 5). To rule out the possibility of a deformity caused due to packing in  $P2_1$  space group, we have looked at the packing of molecules within the crystal (Fig. S1). The loop in question of AtClpD structure that corresponds to  $\alpha 3$  of ClpA/B/C family proteins is not in close contact with the adjacent molecules in the crystal, thereby ruling out the possibility of a crystallographic artefact. In addition, the structural alignment of AtClpD NTD with AtClpC1 NTD (Fig. 3B) shows the loop to be projecting outwards as compared to the helix of AtClpC1 NTD, which may not have happened if indeed adjacent molecules of AtClpD NTD come in close contact due to crystal packing in this region and deforms the secondary structure. Also, the two molecules of ClpD NTD in the asymmetric unit align with a minimal r.m.s. deviation of about  $0.19 \text{ \AA}$  and the loop in question aligns very well for the two molecules (Fig. S2). There is very little chance that a crystal packing deformity will affect the two molecules the same way when they are not related by symmetry (Fig. S2). The fact that there are helix breaker residues such as a Pro and two Gly within this stretch supports the experimental evidence for the presence of loop instead of a helix. Though only semi-conserved, the corresponding stretch in ClpD isoforms from other plants also does not seem to support the formation of a helix (Fig. 4). The presence of longer middle loop and the absence of a helix in the stretch corresponding to  $\alpha 3$  of AtClpC1 NTD make AtClpD NTD very different from any known Clp chaperone NTD structures and perhaps might attribute its functional significance of having to recognize specific substrates during stress conditions, without the involvement of any adaptor proteins. This becomes the first structure of a Clp chaperone wherein the NTD does not have a 4-helix motif repeat organization.

A query to the DALI database and PDBFold server with AtClpD NTD structure coordinates yielded NTD of Gram-positive bacterial ClpC and mycobacterial ClpC1 as the most similar structures. AtClpD NTD showed some degree of structural similarity to AtClpT1/T2 proteins, as well as the NTD of malarial and bacterial ClpB proteins and bacterial ClpA proteins.

**Structural comparison with ClpC NTD.** The N terminal domain of structurally characterized prokaryotic ClpA/B/C proteins all have the four-helical structural repeats conserved. The N terminal domain of bacterial Clp chaperones are involved in substrate binding and interaction with adaptor proteins (e.g., MecA, YpbH, and McsB for ClpC and ClpS for ClpA)<sup>22–25</sup>. Crystal structures of ClpC N-terminal domain from five different bacteria

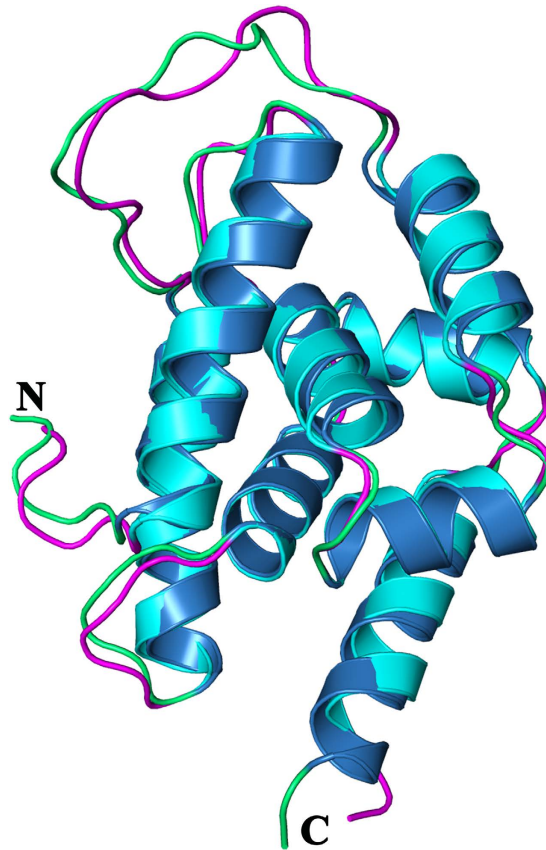
		$\alpha 1$	$\alpha 2$	$\alpha 3$	
AtClpD	079	VFERFTERAIRAIIIFSQKEAKSLGKDMVYTPQHLLGLIAEDRD	QGFGLGSGITIDKAREA		138
BoClpD	076	VFERFTERAIRAIIIFSQKEAKSLGKDMVYTPQHLLGLIAEDRDERGFLGSGVTVDKAREA			135
CsClpD	081	VFERFTERAVKAVIFSQREAKSLGKDMVFTQHLLGLIAEDRHENGFLESGITIDKAREA			140
GhClpD	080	VFERFTERAIKAVILSQREAKSLGKDMVFTQHLLGLIGEDRDPNGFLGSGLKTIENARDA			139
NtClpD	096	VFERFTERSIAKAVMFSQKEAKALGKDMVYTPQHLLGLIAEDRSPGGFLGSRITIDKAREA			155
RcClpD	079	VFERFTERAIKVVIFSQREARALGKDMVFTQHLLGLIGEDRDPDGFGLSGIKIDKAREI			138
VvClpD	077	VFERFTERAIKAVIFSQREAKALGRNMVFTQHLLGLVAEDRSLDGFGLGSGITIDARDA			136
OsClpD2	070	VFERFTERAVKAVVFSQREARGMGDETVAAPHLLGLVAEDRSELGFLASGVRVERAREA			129
OsClpD1	073	VFERFTERAVKAVVLSQREAKLGEVAVAPRHLLGLIAEDRSAGGFLSSGINTERAREE			132
ZmClpD	073	VFERFTERAVKAVVLSQREARGLGEPVAVAPRHLLGLVAEDRSSGGFLSSGINTERAREE			132
			$\alpha 4$		
AtClpD	139	VWSIWDEANSDSKQEEAS---	STSYSKSTDMPFSSISTKRVFEAAVEYSRTMDCQYIAP	EH	195
BoClpD	136	VSSIWDEASSDDSGLES---	SSYSRSTDMPFSSISTKRVFEAAVEYSRTLECCQYIAP	EH	193
CsClpD	141	VVSIWHSTNNQDTDDAAQ--	GKPFSSAAKMPFSSISTKRVFEAAVEYSRSRGYNFIAP	EH	198
GhClpD	140	VRSIWQSSNHGEDLDNKQ---	GSIVSSTDPVPSISTKRVFEAAVEYSRTMGYNFIAP	EH	196
NtClpD	156	VRSIWHDDVEDDKAKLASQ-	DSGSATSDVAFSSSTKRVFEAAVEYSRTMGYNFIAP	EH	214
RcClpD	139	VQNIWSSDGDGTNASGSS	TGKSGGGGSATDVPEAISTKRVFEAAVEYSRTMGYNFIAP	EH	198
VvClpD	137	VRSIWHDYNDSSIIISGIPS-	SQTSVASSTDPVPSISTKRVFEAAIEYSRTMGYNFIAP	EH	195
OsClpD2	130	CRAAVGKEGLAQAPVGL-----	ATDPVPSGASKRVFEAAVEFSRNMGCNFI	SP	179
OsClpD1	133	CRGIGARDLTPGAPSPSGS-----	GLEMIPFSGSCKRVFEVAVEFSRNMGC	SFISPEH	186
ZmClpD	133	CRGIAASARDADSATAASR--	PGSSGLDTPVPEAAPTQVFDVAVVLSKNMGSS	FVSP	190
		$\alpha 5$	$\alpha 6$	$\alpha 7$	
AtClpD	196	IAVGLFTVDDGSAGRVLKRLGANMNL	TAAALTRLKGE		233
BoClpD	194	IAVGLFTVDDGSAGKVLKRLGANMNL	TAAALTRLKGE		231
CsClpD	199	IALGLFTVDDGSAGRVLKRLGVDVNH	LAAVAVSRLQGE		236
GhClpD	197	IAIGLFTVDDGSASRVLKRLGANIN	HLAAEAVIRLQGE		234
NtClpD	215	IAFGLFTVDDGNATRVLKRLGVNV	NRLAAEAVSRLQGE		252
RcClpD	199	IAIGLLTVDDGSASRVLKRLGANL	DDLATAAVARLQGE		236
VvClpD	196	IAIGLFTVDDGSAGRVLKRLGANV	NHLAAVAVSRLQGE		233
OsClpD2	180	IALGLFNLDPTNNVLKSLGVDSS	QLAKQALTRVQGE		217
OsClpD1	187	LALALFTLDDPTNSLRLSLGADP	SQLASVALTRLQAE		224
ZmClpD	191	LALALFTLDDPTNNLRLSLGADP	SHLASVAVDRLQAE		228

**Figure 4.** Sequence alignment of AtClpD NTD with the NTD of putative ClpD from different plants. The sequences have been aligned using ClustalW. For AtClpD, the residues forming helices are underlined and the helices are labelled  $\alpha 1$  to  $\alpha 7$ . The residues conserved among all ten sequences are highlighted in red, those conserved among eight or nine sequences are highlighted in green and those conserved in five, six or seven sequences are highlighted in grey. The residues of regions that are substantially different in AtClpD structure as compared to ClpC have been given in blue. Bo, Cs, Gh, Nt, Rc, Vv, Os and Zm corresponds to *Brassica oleracea*, *Citrus sinensis*, *Gossypium hirsutum*, *Nicotiana tabacum*, *Ricinus communis*, *Vitis vinifera*, *Oryza sativa* and *Zea mays* respectively.



**Figure 5.** Electron density map for the loop region between helices  $\alpha 2$  and  $\alpha 3$  in the crystal structure of AtClpD. The  $F_o - F_c$  omit map for the residues in the loop region between helices  $\alpha 2$  and  $\alpha 3$ , contoured to 1.0 sigma is represented in grey colour. Helices and loop residues are labeled.

are available in the protein data bank: *Mycobacterium tuberculosis* (PDB id: 3WDB), *Bacillus subtilis* (PDB id: 2Y1Q), *Corynebacterium glutamicum* (PDB id: 2FH2), *Bacillus lehensis* (PDB id: 4P15) and *Clostridium difficile* (PDB id: 3FES).



**Figure 6. Structural alignment of AtClpC1 NTD with MtClpC1 NTD.** AtClpC1 NTD is shown in blue helices and green loops, MtClpC1 NTD is shown in cyan helices and magenta loops. The N and C termini are labeled.

The *A. thaliana* ClpC1 N-terminal domain structure is very similar to the structure of NTD from bacterial ClpC and aligns with r.m.s. deviations of 0.66 Å for 102 C $\alpha$  atoms with *M. tuberculosis* ClpC1 NTD (PDB id: 3WDB), 0.72 Å for 105 C $\alpha$  atoms with *B. subtilis* ClpC NTD (PDB id: 2Y1Q), 0.76 Å for 103 C $\alpha$  atoms with *B. lehensis* ClpC NTD (PDB id: 4P15), 1.57 Å for 104 C $\alpha$  atoms with *C. glutamicum* ClpC NTD (PDB id: 3FH2) and 0.93 Å for 81 C $\alpha$  atoms with *C. difficile* ClpC NTD (PDB id: 3FES). The structural alignment of AtClpC1 NTD with *M. tuberculosis* ClpC1 NTD has been shown as a representative in Fig. 6.

AtClpD NTD structure does not have the four-helical structural repeat organization and it has a longer middle loop. As such, it aligns poorly with the other known ClpC NTD structures which have two halves of four-helix motif. Alignment of AtClpD NTD gives r.m.s. deviations of 0.72 Å for 75 C $\alpha$  atoms with *C. difficile* ClpC NTD (PDB id: 3FES), 1.05 Å for 99 C $\alpha$  atoms with *M. tuberculosis* ClpC1 NTD (PDB id: 3WDB), 1.09 Å for 85 C $\alpha$  atoms with AtClpC1 NTD (PDB id: 3GUI; reported in this work), 1.21 Å for 84 C $\alpha$  atoms with *B. subtilis* ClpC NTD (PDB id: 2Y1Q), 1.27 Å for 83 C $\alpha$  atoms with *B. lehensis* ClpC NTD (PDB id: 4P15) and 2.15 Å for 116 C $\alpha$  atoms with *C. glutamicum* ClpC NTD (PDB id: 3FH2). Figure 7 shows the structure-based sequence alignment of AtClpC1 NTD and AtClpD NTD with bacterial ClpC NTD.

**Structural comparison with ClpA and ClpB NTD.** The overall structure of AtClpC1 NTD is similar to that of NTD from bacterial ClpA and ClpB proteins as well. AtClpC1 NTD has its secondary structure organization quite similar to that of *E. coli* ClpA and ClpB -NTD, however the helices and loops of AtClpC1 NTD and those of ClpA/B proteins are oriented slightly different and hence the structure superpose with a relatively poor r.m.s. deviation of 3.82 Å (for 99 C $\alpha$  atoms) and 7.31 Å (for 81 C $\alpha$  atoms) to *E. coli* ClpB NTD (PDB id: 1KHY) and *E. coli* ClpA NTD (PDB id: 1R6C), respectively. The NTD of malarial ClpB2 protein (PDB id: 4IOD) also has similar secondary structural features, but superposes with AtClpC1 NTD with a high r.m.s. deviation of 7.14 Å (for 74 C $\alpha$  atoms). AtClpD aligns with ClpB and ClpA proteins with even higher r.m.s. deviations.

**Structural comparison with ClpT1/T2.** The chloroplast ClpT1 and ClpT2 proteins show high sequence similarity with each other and are likely derived from ClpC chaperones given their significant sequence similarity (31% sequence identity across 93 to 98 residues) to the NTD of chloroplast ClpC1/C2 chaperones<sup>26</sup>. However, unlike ClpC1, ClpT1 and ClpT2 are unique proteins seen only in higher plants and their interaction with the ClpPR core is predicted to be an adaptation to the plastid proteome and/or Clp protease system of higher plants<sup>26</sup>. Structural characterization has shown ClpT1 and ClpT2 to possess the four-helical structural repeats similar to ClpC NTD<sup>26</sup>. The structure of AtClpT1 (PDB id: 4Y0B) and AtClpT2 (PDB id: 4Y0C) aligns with AtClpC1 NTD

		$\alpha 1$		$\alpha 2$		$\alpha 3$	
AtClpC1	094	<u>M</u> <u>F</u> <u>E</u> <u>R</u> <u>F</u> <u>T</u> <u>E</u> <u>K</u> <u>A</u> <u>I</u> <u>K</u> <u>V</u> <u>I</u> <u>M</u> <u>L</u> <u>A</u> <u>Q</u> <u>E</u> <u>E</u> <u>A</u> <u>R</u> <u>R</u> <u>L</u> <u>G</u> <u>H</u> <u>N</u> <u>F</u> <u>V</u> <u>G</u> <u>T</u> <u>E</u> <u>Q</u> <u>I</u> <u>L</u> <u>L</u> <u>G</u> <u>L</u> <u>I</u> <u>G</u> <u>E</u> <u>G</u> <u>T</u> <u>G</u> <u>I</u> <u>A</u> <u>A</u> <u>K</u> <u>V</u> <u>L</u> <u>K</u> <u>S</u> <u>M</u> <u>G</u> <u>I</u> <u>N</u> <u>L</u> <u>K</u> <u>D</u> <u>A</u> <u>R</u>	153				
MtClpC1	001	<u>M</u> <u>F</u> <u>E</u> <u>R</u> <u>F</u> <u>T</u> <u>D</u> <u>R</u> <u>A</u> <u>R</u> <u>R</u> <u>V</u> <u>V</u> <u>L</u> <u>A</u> <u>Q</u> <u>E</u> <u>E</u> <u>A</u> <u>R</u> <u>M</u> <u>L</u> <u>N</u> <u>H</u> <u>N</u> <u>Y</u> <u>I</u> <u>G</u> <u>T</u> <u>E</u> <u>H</u> <u>I</u> <u>L</u> <u>L</u> <u>G</u> <u>L</u> <u>I</u> <u>H</u> <u>E</u> <u>G</u> <u>E</u> <u>G</u> <u>V</u> <u>A</u> <u>A</u> <u>K</u> <u>S</u> <u>L</u> <u>E</u> <u>S</u> <u>L</u> <u>G</u> <u>I</u> <u>S</u> <u>L</u> <u>E</u> <u>G</u> <u>V</u> <u>R</u>	060				
CgClpC	001	<u>M</u> <u>F</u> <u>E</u> <u>R</u> <u>F</u> <u>T</u> <u>D</u> <u>R</u> <u>A</u> <u>R</u> <u>R</u> <u>V</u> <u>I</u> <u>V</u> <u>L</u> <u>A</u> <u>Q</u> <u>E</u> <u>E</u> <u>A</u> <u>R</u> <u>M</u> <u>L</u> <u>N</u> <u>H</u> <u>N</u> <u>Y</u> <u>I</u> <u>G</u> <u>T</u> <u>E</u> <u>H</u> <u>I</u> <u>L</u> <u>L</u> <u>G</u> <u>L</u> <u>I</u> <u>H</u> <u>E</u> <u>G</u> <u>E</u> <u>G</u> <u>V</u> <u>A</u> <u>A</u> <u>K</u> <u>A</u> <u>L</u> <u>E</u> <u>S</u> <u>M</u> <u>G</u> <u>I</u> <u>S</u> <u>L</u> <u>D</u> <u>A</u> <u>V</u> <u>R</u>	060				
CdClpC	001	<u>N</u> <u>F</u> <u>N</u> <u>R</u> <u>F</u> <u>T</u> <u>Q</u> <u>R</u> <u>A</u> <u>K</u> <u>K</u> <u>A</u> <u>I</u> <u>D</u> <u>L</u> <u>A</u> <u>F</u> <u>S</u> <u>A</u> <u>K</u> <u>S</u> <u>L</u> <u>G</u> <u>H</u> <u>N</u> <u>I</u> <u>V</u> <u>G</u> <u>S</u> <u>E</u> <u>H</u> <u>I</u> <u>L</u> <u>L</u> <u>G</u> <u>L</u> <u>L</u> <u>R</u> <u>E</u> <u>E</u> <u>G</u> <u>I</u> <u>A</u> <u>A</u> <u>K</u> <u>V</u> <u>L</u> <u>S</u> <u>K</u> <u>V</u> <u>G</u> <u>F</u> <u>T</u> <u>E</u> <u>A</u> <u>Y</u> <u>L</u> <u>E</u>	060				
BsClpC	001	<u>M</u> <u>F</u> <u>G</u> <u>R</u> <u>F</u> <u>T</u> <u>E</u> <u>R</u> <u>A</u> <u>Q</u> <u>V</u> <u>L</u> <u>A</u> <u>L</u> <u>A</u> <u>Q</u> <u>E</u> <u>E</u> <u>A</u> <u>L</u> <u>R</u> <u>L</u> <u>G</u> <u>H</u> <u>N</u> <u>I</u> <u>G</u> <u>T</u> <u>E</u> <u>H</u> <u>I</u> <u>L</u> <u>L</u> <u>G</u> <u>L</u> <u>V</u> <u>R</u> <u>E</u> <u>G</u> <u>E</u> <u>G</u> <u>I</u> <u>A</u> <u>A</u> <u>K</u> <u>A</u> <u>L</u> <u>Q</u> <u>A</u> <u>L</u> <u>G</u> <u>L</u> <u>G</u> <u>S</u> <u>E</u> <u>K</u> <u>I</u> <u>Q</u>	060				
BlClpC	001	<u>M</u> <u>F</u> <u>G</u> <u>R</u> <u>F</u> <u>S</u> <u>E</u> <u>R</u> <u>A</u> <u>Q</u> <u>V</u> <u>L</u> <u>A</u> <u>L</u> <u>S</u> <u>Q</u> <u>E</u> <u>E</u> <u>A</u> <u>I</u> <u>R</u> <u>L</u> <u>S</u> <u>H</u> <u>H</u> <u>N</u> <u>I</u> <u>G</u> <u>T</u> <u>E</u> <u>H</u> <u>I</u> <u>L</u> <u>L</u> <u>G</u> <u>L</u> <u>I</u> <u>R</u> <u>E</u> <u>G</u> <u>E</u> <u>G</u> <u>I</u> <u>A</u> <u>A</u> <u>K</u> <u>A</u> <u>L</u> <u>Q</u> <u>Q</u> <u>L</u> <u>G</u> <u>L</u> <u>S</u> <u>D</u> <u>K</u> <u>L</u> <u>Q</u>	060				
AtClpD	079	<u>V</u> <u>F</u> <u>E</u> <u>R</u> <u>F</u> <u>T</u> <u>E</u> <u>R</u> <u>A</u> <u>I</u> <u>R</u> <u>A</u> <u>I</u> <u>F</u> <u>S</u> <u>Q</u> <u>K</u> <u>E</u> <u>A</u> <u>K</u> <u>S</u> <u>L</u> <u>G</u> <u>K</u> <u>D</u> <u>M</u> <u>V</u> <u>Y</u> <u>T</u> <u>Q</u> <u>H</u> <u>L</u> <u>L</u> <u>L</u> <u>G</u> <u>L</u> <u>I</u> <u>A</u> <u>E</u> <u>D</u> <u>R</u> <u>D</u> <u>--</u> <u>P</u> <u>Q</u> <u>G</u> <u>F</u> <u>L</u> <u>G</u> <u>S</u> <u>G</u> <u>I</u> <u>T</u> <u>I</u> <u>D</u> <u>K</u> <u>A</u> <u>R</u>	136				
		$\alpha 4$		$\alpha 5$			
AtClpC1	154	<u>V</u> <u>E</u> <u>V</u> <u>E</u> <u>K</u> <u>I</u> <u>I</u> <u>G</u> <u>R</u> <u>S</u> <u>G</u> <u>F</u> <u>V</u> <u>A</u> <u>V</u> <u>E</u> <u>-----</u> <u>I</u> <u>P</u> <u>F</u> <u>T</u> <u>P</u> <u>R</u> <u>A</u> <u>K</u> <u>R</u> <u>V</u> <u>L</u> <u>E</u> <u>L</u> <u>S</u> <u>L</u> <u>E</u> <u>E</u> <u>A</u> <u>R</u> <u>Q</u> <u>L</u> <u>G</u> <u>H</u> <u>N</u> <u>Y</u> <u>I</u> <u>G</u> <u>S</u> <u>E</u> <u>H</u> <u>L</u>	201				
MtClpC1	061	<u>S</u> <u>Q</u> <u>V</u> <u>E</u> <u>E</u> <u>I</u> <u>I</u> <u>G</u> <u>Q</u> <u>Q</u> <u>A</u> <u>P</u> <u>S</u> <u>G</u> <u>H</u> <u>-----</u> <u>I</u> <u>P</u> <u>F</u> <u>T</u> <u>P</u> <u>R</u> <u>A</u> <u>K</u> <u>K</u> <u>V</u> <u>L</u> <u>E</u> <u>L</u> <u>S</u> <u>L</u> <u>R</u> <u>E</u> <u>A</u> <u>L</u> <u>Q</u> <u>L</u> <u>G</u> <u>H</u> <u>N</u> <u>Y</u> <u>I</u> <u>G</u> <u>T</u> <u>E</u> <u>H</u> <u>I</u>	108				
CgClpC	061	<u>Q</u> <u>E</u> <u>V</u> <u>E</u> <u>E</u> <u>I</u> <u>I</u> <u>G</u> <u>Q</u> <u>S</u> <u>Q</u> <u>P</u> <u>T</u> <u>T</u> <u>G</u> <u>H</u> <u>-----</u> <u>I</u> <u>P</u> <u>F</u> <u>T</u> <u>P</u> <u>R</u> <u>A</u> <u>K</u> <u>K</u> <u>V</u> <u>L</u> <u>E</u> <u>L</u> <u>S</u> <u>L</u> <u>R</u> <u>E</u> <u>G</u> <u>L</u> <u>Q</u> <u>M</u> <u>G</u> <u>H</u> <u>K</u> <u>Y</u> <u>I</u> <u>G</u> <u>T</u> <u>E</u> <u>F</u> <u>L</u>	108				
CdClpC	061	<u>G</u> <u>K</u> <u>I</u> <u>V</u> <u>D</u> <u>M</u> <u>E</u> <u>G</u> <u>K</u> <u>G</u> <u>E</u> <u>E</u> <u>I</u> <u>S</u> <u>E</u> <u>D</u> <u>-----</u> <u>I</u> <u>V</u> <u>L</u> <u>S</u> <u>P</u> <u>R</u> <u>S</u> <u>K</u> <u>Q</u> <u>I</u> <u>L</u> <u>E</u> <u>L</u> <u>S</u> <u>G</u> <u>M</u> <u>F</u> <u>A</u> <u>N</u> <u>K</u> <u>L</u> <u>K</u> <u>T</u> <u>N</u> <u>Y</u> <u>I</u> <u>G</u> <u>T</u> <u>E</u> <u>H</u> <u>I</u>	107				
BsClpC	061	<u>K</u> <u>E</u> <u>V</u> <u>E</u> <u>S</u> <u>L</u> <u>I</u> <u>G</u> <u>R</u> <u>A</u> <u>Q</u> <u>E</u> <u>M</u> <u>S</u> <u>Q</u> <u>T</u> <u>-----</u> <u>I</u> <u>H</u> <u>Y</u> <u>T</u> <u>P</u> <u>R</u> <u>A</u> <u>K</u> <u>K</u> <u>V</u> <u>I</u> <u>E</u> <u>L</u> <u>S</u> <u>M</u> <u>D</u> <u>E</u> <u>A</u> <u>R</u> <u>K</u> <u>L</u> <u>G</u> <u>H</u> <u>S</u> <u>Y</u> <u>V</u> <u>G</u> <u>T</u> <u>E</u> <u>H</u> <u>I</u>	107				
BlClpC	061	<u>K</u> <u>E</u> <u>V</u> <u>E</u> <u>G</u> <u>L</u> <u>V</u> <u>G</u> <u>K</u> <u>Q</u> <u>E</u> <u>G</u> <u>Q</u> <u>S</u> <u>I</u> <u>H</u> <u>S</u> <u>T</u> <u>-----</u> <u>P</u> <u>H</u> <u>Y</u> <u>T</u> <u>P</u> <u>R</u> <u>A</u> <u>K</u> <u>K</u> <u>V</u> <u>I</u> <u>E</u> <u>L</u> <u>S</u> <u>M</u> <u>D</u> <u>E</u> <u>A</u> <u>R</u> <u>K</u> <u>L</u> <u>G</u> <u>H</u> <u>S</u> <u>Y</u> <u>V</u> <u>G</u> <u>T</u> <u>E</u> <u>H</u> <u>I</u>	111				
AtClpD	137	<u>E</u> <u>A</u> <u>W</u> <u>S</u> <u>I</u> <u>W</u> <u>D</u> <u>E</u> <u>A</u> <u>N</u> <u>S</u> <u>D</u> <u>S</u> <u>K</u> <u>Q</u> <u>E</u> <u>A</u> <u>S</u> <u>S</u> <u>T</u> <u>S</u> <u>Y</u> <u>S</u> <u>K</u> <u>S</u> <u>T</u> <u>D</u> <u>M</u> <u>P</u> <u>F</u> <u>S</u> <u>I</u> <u>S</u> <u>T</u> <u>R</u> <u>V</u> <u>F</u> <u>E</u> <u>A</u> <u>A</u> <u>V</u> <u>E</u> <u>Y</u> <u>S</u> <u>R</u> <u>T</u> <u>M</u> <u>D</u> <u>C</u> <u>Q</u> <u>Y</u> <u>I</u> <u>A</u> <u>P</u> <u>E</u> <u>H</u> <u>I</u>	196				
		$\alpha 6$	$\alpha 7$	$\alpha 8$			
AtClpC1	202	<u>L</u> <u>L</u> <u>G</u> <u>L</u> <u>I</u> <u>R</u> <u>E</u> <u>G</u> <u>E</u> <u>G</u> <u>V</u> <u>A</u> <u>A</u> <u>R</u> <u>V</u> <u>D</u> <u>E</u> <u>N</u> <u>L</u> <u>G</u> <u>A</u> <u>D</u> <u>P</u> <u>S</u> <u>N</u> <u>I</u> <u>T</u> <u>Q</u> <u>V</u> <u>I</u> <u>R</u> <u>M</u> <u>V</u> <u>G</u> <u>E</u>	237				
MtClpC1	109	<u>L</u> <u>L</u> <u>G</u> <u>L</u> <u>I</u> <u>R</u> <u>E</u> <u>G</u> <u>E</u> <u>G</u> <u>V</u> <u>A</u> <u>A</u> <u>Q</u> <u>V</u> <u>L</u> <u>V</u> <u>K</u> <u>L</u> <u>G</u> <u>A</u> <u>E</u> <u>L</u> <u>T</u> <u>R</u> <u>V</u> <u>R</u> <u>Q</u> <u>O</u> <u>V</u> <u>I</u> <u>Q</u> <u>L</u> <u>L</u> <u>S</u> <u>G</u>	144				
CgClpC	109	<u>L</u> <u>L</u> <u>G</u> <u>L</u> <u>I</u> <u>R</u> <u>E</u> <u>G</u> <u>E</u> <u>G</u> <u>V</u> <u>A</u> <u>A</u> <u>Q</u> <u>V</u> <u>L</u> <u>V</u> <u>K</u> <u>L</u> <u>G</u> <u>A</u> <u>D</u> <u>L</u> <u>P</u> <u>R</u> <u>V</u> <u>R</u> <u>Q</u> <u>O</u> <u>V</u> <u>I</u> <u>Q</u> <u>L</u> <u>L</u> <u>S</u> <u>G</u>	144				
CdClpC	108	<u>L</u> <u>L</u> <u>A</u> <u>I</u> <u>Q</u> <u>E</u> <u>G</u> <u>E</u> <u>G</u> <u>I</u> <u>A</u> <u>N</u> <u>K</u> <u>I</u> <u>L</u> <u>N</u> <u>Y</u> <u>A</u> <u>G</u> <u>V</u> <u>N</u> <u>D</u> <u>R</u> <u>T</u> <u>L</u> <u>A</u> <u>Q</u> <u>L</u> <u>T</u> <u>I</u> <u>D</u> <u>M</u> <u>M</u> <u>G</u> <u>I</u>	143				
BsClpC	108	<u>L</u> <u>L</u> <u>G</u> <u>L</u> <u>I</u> <u>R</u> <u>E</u> <u>G</u> <u>E</u> <u>G</u> <u>V</u> <u>A</u> <u>A</u> <u>R</u> <u>V</u> <u>L</u> <u>N</u> <u>N</u> <u>L</u> <u>G</u> <u>V</u> <u>S</u> <u>L</u> <u>N</u> <u>K</u> <u>A</u> <u>R</u> <u>Q</u> <u>O</u> <u>V</u> <u>L</u> <u>Q</u> <u>L</u> <u>L</u> <u>G</u> <u>N</u>	143				
BlClpC	112	<u>L</u> <u>L</u> <u>G</u> <u>L</u> <u>I</u> <u>R</u> <u>E</u> <u>G</u> <u>E</u> <u>G</u> <u>V</u> <u>A</u> <u>A</u> <u>R</u> <u>V</u> <u>L</u> <u>N</u> <u>N</u> <u>L</u> <u>G</u> <u>V</u> <u>S</u> <u>L</u> <u>N</u> <u>K</u> <u>A</u> <u>R</u> <u>Q</u> <u>O</u> <u>V</u> <u>L</u> <u>Q</u> <u>L</u> <u>L</u> <u>G</u> <u>S</u>	147				
AtClpD	197	<u>A</u> <u>V</u> <u>G</u> <u>L</u> <u>F</u> <u>T</u> <u>V</u> <u>D</u> <u>D</u> <u>S</u> <u>S</u> <u>A</u> <u>G</u> <u>R</u> <u>V</u> <u>L</u> <u>K</u> <u>R</u> <u>L</u> <u>G</u> <u>A</u> <u>N</u> <u>M</u> <u>N</u> <u>L</u> <u>L</u> <u>T</u> <u>A</u> <u>A</u> <u>A</u> <u>L</u> <u>T</u> <u>R</u> <u>L</u> <u>K</u> <u>G</u>	232				

**Figure 7. Structure-based sequence alignment of AtClpC1 NTD and AtClpD NTD with bacterial ClpC NTDs.** The sequences have been aligned using ClustalW. The residues forming helices are underlined and the helices are labelled  $\alpha 1$  to  $\alpha 8$  (ClpC numbering). The residues conserved among all seven sequences are highlighted in red, those conserved among five or six sequences are highlighted in green and those conserved in four sequences are highlighted in grey. The residues of regions that are substantially different in AtClpD structure have been given in blue. Mt, Cg, Cd, Bs and Bl corresponds to *Mycobacterium tuberculosis*, *Corynebacterium glutamicum*, *Clostridium difficile*, *Bacillus subtilis* and *Bacillus lehensis*, respectively.

with an r.m.s. deviation of 1.03 Å (for 88 C $\alpha$  atoms) and 1.31 Å (for 84 C $\alpha$  atoms), respectively and with AtClpD with an r.m.s. deviation of 1.35 Å (for 92 C $\alpha$  atoms) and 1.44 Å (for 90 C $\alpha$  atoms), respectively. Structure-based sequence alignment of the NTD of AtClpC1 and AtClpD with AtClpT1, AtClpT2 as well the NTD of *E. coli* ClpA, *E. coli* ClpB and malarial ClpB2 is presented in Fig. 8.

**AtClpC1/D NTD and interaction partners.** The NTD of *E. coli* ClpA protein is only about 23% identical to AtClpC1 NTD in its sequence; however, the two proteins seem to have a very similar structural organization. *E. coli* ClpA NTD residues involved in interaction with ClpS are conserved in AtClpC1 NTD as well<sup>16</sup>. *E. coli* ClpS protein is only about 27% identical to AtClpS1 protein with respect to its sequence. The interaction between AtClpC1 and AtClpS1 is thought to be quite similar to that of *E. coli* ClpA and ClpS proteins<sup>16</sup>. After having the AtClpC1 NTD structure, we also presume the mode of interaction between AtClpC1 NTD and AtClpS1 to be similar to that between *E. coli* ClpA and *E. coli* ClpS. The discovery of AtClpF protein as a third partner of the complex<sup>18</sup> has added another level of complexity and with the current level of understanding, predictions looking at AtClpC1 and AtClpS1 in isolation may not be prudent. In fact, our attempts to obtain a complex of AtClpS1 with AtClpC1 NTD by pull down assays and co-purification of AtClpS1 with monomeric as well as hexameric forms of full-length AtClpC1 were unsuccessful, suggesting the need to have even AtClpF in the complex (Supplemental data). Future studies should focus on the structural characterization of a complex made up of the interacting domains of AtClpC1, AtClpS1 and AtClpF. This would shed light on the exact mode of substrate recognition by the plant Clp machinery.

Unlike AtClpC1 and AtClpC2, AtClpD is not known to have an adaptor protein. AtClpS1 and AtClpF do not interact with AtClpD<sup>16,18</sup>. AtClpD not interacting with AtClpS1 has primarily been attributed to the absence of residues present in AtClpC1 that support its interaction with AtClpS1. When the structure of AtClpD NTD is aligned with AtClpC1 NTD, we realize that the long middle loop comes directly over the region which may otherwise contribute to ClpS1 interaction. Therefore, AtClpD not binding to AtClpS1 or AtClpF could at least partly be due to the presence of the long middle loop in AtClpD NTD which would sterically prevent any such interaction. In the functional and hexameric context of AtClpD, the absence of a helix ( $\alpha 3$ ) in the face opposing the longer loop region might also attribute to the inability to interact with the adaptor proteins AtClpF/S1. It is not clear whether the absence of an adaptor protein is compensated by these structural variations.

While looking for sequence similarity between AtClpC1/C2 NTD and *Mycobacterium tuberculosis* ClpC1 NTD, we found the conserved nature of residues in AtClpC1/C2 that are involved in the interaction of MtClpC1 NTD with the cyclic peptide Cyclomarin A (CymA). Residues Phe2, Phe80, Lys85 and Glu89 of MtClpC1 NTD are responsible for CymA interaction as has been reported in the crystal structure of the complex<sup>27</sup>. A comparison of the crystal structures of MtClpC1-CymA complex and AtClpC1 shows the CymA-interacting residues to

		$\alpha 1$	$\alpha 2$	$\alpha 3$	
EcClpA	001	---MLNQELELSINMAFARAREHRHEFMTVEHL	LALISNP--	SAREAEACSV	DLVALR 055
EcClpB	002	RLDRLTNKFOALALADAQSLALGHDNQFIEPLH	MSALLNQEGGSV	SPLT	SAGINAGQLR 061
PfClpB2	008	EQGKYLNRTEINILNAGKNIKSYGHNKPKPIHI	SALAKSD--	YGSTL	FKENNVNAANLK 065
AtClpT1	082	KIPKWSARAIKSLAMGELEARKIKYPSTGTEAI	LMGILVEGTSTVAKF	IRGNGV	TFLFKVR 141
AtClpT2	090	KKPKWSWRAIKSFAMGELEARKIKYPNTGTEAL	LMGILIEGTSTFTSKF	IRANKIM	LYKVR 149
AtClpC1	094	MFERFTEKAIKVIMLAQEEARRLGHNQVTEQI	LLGLIGEGTGIAAKV	LKSMGIN	LKDPAR 153
AtClpD	079	VFERFTEARAIRAIIFSQKEAKSLGKDMVYTQ	HLHLGLIAERDRD--	PQGF	LGSGITIDKAR 136
		$\alpha 4$	$\alpha 5$		
EcClpA	056	QLEAFIEQTFVLPASFEERD-----	TPPTLSFORVLR	VFHVQSSGRN	-EVTGAN 107
EcClpB	062	TDINQALNRLPQVEGTGGD-----	VQPSQDLVRVNL	CDKLAQKRGDN	-FISSSEL 110
PfClpB2	066	EYIDIALEQTRAGAPLDNKS-----	IVNSAEVKETLAL	EAAANKYKSP	-KVDVEH 116
AtClpT1	142	DETLSSLGKSDMYFFSPEH-----	PPLTEPAOKAIAWA	IDENKNSD	VDGELTAY 191
AtClpT2	150	EETVKLLGKADMYFFSPEH-----	PPLTEDAQRALDS	ALDQNLKAGG	GIGVMPAH 199
AtClpC1	154	VEVEKIIIGRSG-FVAVEI-----	PFTPRAKRVLELS	SLEEARQLGHN	-YIGSEH 200
AtClpD	137	EAVWSIWDEANSDSKQEEASSTSYSKSTDM	PFSISTKR	VFEAVEYSRT	MDCQ-YIAPEH 195
		$\alpha 6$	$\alpha 7$	$\alpha 8$	
EcClpA	108	VLVAIFSEQESQAAYL	LRKHEVSR	LDVNVNF	ISHGTRKDEP 147
EcClpB	111	FVLAALSRGTLADILKAAGATTANITQAI	EQMRGGSV		
PfClpB2	117	LLSGLSN--DELVNEIFNEVYLTDEAIKAIL	KRKFECTLV		
AtClpT1	192	LLLGVWSQKDSAGROI	LEKLGFNEDKAKEVEKSMNEDVDL		
AtClpT2	200	LLLGWSEVESPGHKILATLGFTDEKSKELES	FASESGFL		
AtClpC1	201	LLGLLREGEGVAARV	LENLGADPSNIR	TQVIRMV	GENNE 240
AtClpD	196	IAVGLFTVDDGSAGRV	LKRLGANMNL	LTAALTRL	KGEIA 235

**Figure 8.** Structure-based sequence alignment of AtClpC1 NTD and AtClpD NTD with ClpA/B NTD and ClpT1/T2 proteins. The sequences have been aligned using ClustalW and the residues forming helices are underlined and helices labelled  $\alpha 1$  to  $\alpha 8$  (ClpA/B/C/T numbering). The residues conserved among all seven sequences are highlighted in red, those conserved among five or six sequences are highlighted in green and those conserved among four sequences are highlighted in grey. The loop region of AtClpD corresponding to  $\alpha 3$  of other structures is given in blue. Ec and Pf correspond to *Escherichia coli* and *Plasmodium falciparum*, respectively.

be conserved. Phe95, Phe173, Lys178 and Glu182 of AtClpC1 occupy the same position as the residues involved in MtClpC1 NTD interaction with CymA (Fig. 9). So we presume CymA would bind readily to AtClpC1 NTD as well. CymA has been studied functionally and structurally as an anti-TB compound targeting mycobacterial ClpC1 and not active against other Gram-positive bacteria where not all these residues are conserved in the ClpC protein<sup>27,28</sup>. It has been proposed that the lack of activity of CymA on Gram-positive bacteria could be due to weaker CymA binding as a result of sequence variation, as well as the presence of the adaptor protein MecA that binds close to the CymA-binding site which is absent in the case of mycobacteria<sup>27</sup>. In the context of plant Clp machinery, though this compound would be of no value as a drug, it could serve as a reagent useful in functionally characterizing the chaperone; more as an activator to the machinery. Interestingly, AtClpD NTD also has all the four CymA-interacting residues conserved with MtClpC1: AtClpD residues Phe80, Phe168, Lys173 and Glu177 (Figs 9 and 10). Due to non-availability of CymA from commercial sources, we could not perform binding or co-crystallization experiments of CymA with AtClpC1/D proteins. MtClpC1 has more recently become an attractive target for anti-TB drug research with peptides of actinomycete origin such as Cyclomarin A, Lassomycin and Ecumicin reported to target the protein's NTD<sup>28-30</sup>. However, the residues predicted to be involved in MtClpC1 interaction with Lassomycin and Ecumicin based on genome-wide mutational studies with resistant mycobacterial strains<sup>29,30</sup> are not conserved in AtClpC1/C2 as well as AtClpD.

In summary, we have structurally characterized the NTD of AtClpC1 and AtClpD. The structures have also been compared with other known Clp chaperone NTD structures. AtClpC1 resembles all other ClpC chaperone protein NTDs whereas, AtClpD turned out to be a divergent one lacking the typical 4-helix repeat motif that is present in all known Clp chaperone structures. The longer middle loop in its NTD also makes AtClpD different from other known Clp chaperone families. The structural difference in the N-terminus could perhaps be important for AtClpD in specifically recognizing misfolded substrates produced during stress conditions and for subsequent unfolding and proteolysis. An understanding about the structural variations in the NTD of AtClpC1 and AtClpD from this work paves way to further studies involving their interaction with adaptor proteins and/or substrates within the plant chloroplast. We propose Cyclomarin A to be a suitable reagent that would activate plant ClpC/D family proteins and thereby aiding their *in vitro* functional characterization.

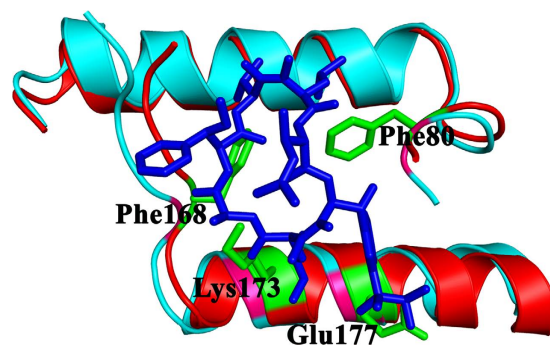
## Methods

**Construction of *Escherichia coli* expression plasmids and protein purification.** Genes coding for AtClpC1 and AtClpD, optimized for overexpression in *E. coli* were obtained from Genscript (NJ, USA) in pUC57 vector. The DNA sequence coding for the N-terminal domain of AtClpC1 spanning residues 94–238 was amplified by PCR and cloned in frame into a pET22b (Novagen) vector between NdeI and XhoI restriction sites for expression with a non-cleavable carboxyl-terminal hexahistidine tag. The DNA sequence coding for AtClpD



MtClpC1	001	<u>M</u> <u>E</u> <u>R</u> <u>F</u> <u>T</u> <u>D</u> <u>R</u> <u>A</u> <u>R</u> <u>R</u> <u>V</u> <u>V</u> <u>L</u> <u>A</u> <u>Q</u> <u>E</u> <u>E</u> <u>A</u> <u>R</u> <u>M</u> <u>I</u> <u>N</u> <u>H</u> <u>N</u> <u>Y</u> <u>I</u> <u>G</u> <u>T</u> <u>E</u> <u>H</u> <u>I</u> <u>L</u> <u>L</u> <u>G</u> <u>L</u> <u>I</u> <u>H</u> <u>E</u> <u>G</u> <u>E</u> <u>G</u> <u>V</u> <u>A</u> <u>A</u> <u>K</u> <u>S</u> <u>L</u> <u>E</u> <u>S</u> <u>L</u> <u>G</u> <u>I</u> <u>S</u> <u>L</u> <u>E</u> <u>G</u> <u>V</u> <u>R</u>	060
AtClpC1	094	<u>M</u> <u>E</u> <u>R</u> <u>F</u> <u>T</u> <u>E</u> <u>K</u> <u>A</u> <u>K</u> <u>V</u> <u>I</u> <u>M</u> <u>L</u> <u>A</u> <u>Q</u> <u>E</u> <u>E</u> <u>A</u> <u>R</u> <u>R</u> <u>L</u> <u>G</u> <u>H</u> <u>N</u> <u>F</u> <u>V</u> <u>G</u> <u>T</u> <u>E</u> <u>Q</u> <u>I</u> <u>L</u> <u>L</u> <u>G</u> <u>L</u> <u>I</u> <u>G</u> <u>E</u> <u>G</u> <u>T</u> <u>G</u> <u>I</u> <u>A</u> <u>A</u> <u>K</u> <u>V</u> <u>L</u> <u>K</u> <u>S</u> <u>M</u> <u>G</u> <u>I</u> <u>N</u> <u>L</u> <u>K</u> <u>D</u> <u>A</u> <u>R</u>	153
AtClpD	079	<u>V</u> <u>E</u> <u>R</u> <u>F</u> <u>T</u> <u>E</u> <u>R</u> <u>A</u> <u>I</u> <u>R</u> <u>A</u> <u>I</u> <u>F</u> <u>S</u> <u>K</u> <u>E</u> <u>A</u> <u>K</u> <u>S</u> <u>L</u> <u>G</u> <u>K</u> <u>D</u> <u>M</u> <u>V</u> <u>T</u> <u>Q</u> <u>H</u> <u>L</u> <u>L</u> <u>L</u> <u>G</u> <u>L</u> <u>I</u> <u>A</u> <u>E</u> <u>D</u> <u>R</u> <u>D</u> <u>--</u> <u>P</u> <u>Q</u> <u>G</u> <u>F</u> <u>L</u> <u>G</u> <u>S</u> <u>G</u> <u>I</u> <u>T</u> <u>I</u> <u>D</u> <u>K</u> <u>A</u> <u>R</u>	136
MtClpC1	061	<u>S</u> <u>Q</u> <u>V</u> <u>E</u> <u>I</u> <u>I</u> <u>G</u> <u>Q</u> <u>E</u> <u>Q</u> <u>Q</u> <u>A</u> <u>P</u> <u>S</u> <u>G</u> <u>H</u> <u>-----</u> <u>I</u> <u>P</u> <u>T</u> <u>P</u> <u>R</u> <u>A</u> <u>K</u> <u>V</u> <u>L</u> <u>S</u> <u>L</u> <u>R</u> <u>E</u> <u>A</u> <u>L</u> <u>Q</u> <u>L</u> <u>G</u> <u>H</u> <u>N</u> <u>Y</u> <u>I</u> <u>G</u> <u>T</u> <u>E</u> <u>H</u> <u>I</u>	108
AtClpC1	154	<u>V</u> <u>E</u> <u>V</u> <u>E</u> <u>K</u> <u>I</u> <u>I</u> <u>G</u> <u>R</u> <u>S</u> <u>G</u> <u>F</u> <u>V</u> <u>A</u> <u>V</u> <u>E</u> <u>-----</u> <u>I</u> <u>P</u> <u>T</u> <u>P</u> <u>R</u> <u>A</u> <u>R</u> <u>V</u> <u>L</u> <u>S</u> <u>L</u> <u>E</u> <u>E</u> <u>A</u> <u>R</u> <u>Q</u> <u>L</u> <u>G</u> <u>H</u> <u>N</u> <u>Y</u> <u>I</u> <u>G</u> <u>S</u> <u>E</u> <u>H</u> <u>L</u>	201
AtClpD	137	<u>E</u> <u>A</u> <u>W</u> <u>S</u> <u>I</u> <u>W</u> <u>D</u> <u>E</u> <u>A</u> <u>N</u> <u>S</u> <u>D</u> <u>S</u> <u>K</u> <u>Q</u> <u>E</u> <u>A</u> <u>S</u> <u>S</u> <u>T</u> <u>S</u> <u>Y</u> <u>S</u> <u>K</u> <u>S</u> <u>T</u> <u>D</u> <u>M</u> <u>P</u> <u>S</u> <u>I</u> <u>S</u> <u>T</u> <u>R</u> <u>V</u> <u>F</u> <u>T</u> <u>A</u> <u>A</u> <u>V</u> <u>E</u> <u>Y</u> <u>S</u> <u>R</u> <u>T</u> <u>M</u> <u>D</u> <u>C</u> <u>Q</u> <u>Y</u> <u>T</u> <u>A</u> <u>P</u> <u>E</u> <u>H</u> <u>I</u>	196
MtClpC1	109	<u>L</u> <u>L</u> <u>G</u> <u>L</u> <u>I</u> <u>R</u> <u>E</u> <u>G</u> <u>E</u> <u>G</u> <u>V</u> <u>A</u> <u>A</u> <u>Q</u> <u>V</u> <u>L</u> <u>V</u> <u>K</u> <u>L</u> <u>G</u> <u>A</u> <u>E</u> <u>L</u> <u>T</u> <u>R</u> <u>V</u> <u>R</u> <u>Q</u> <u>O</u> <u>V</u> <u>I</u> <u>Q</u> <u>L</u> <u>L</u> <u>S</u> <u>G</u>	144
AtClpC1	202	<u>L</u> <u>L</u> <u>G</u> <u>L</u> <u>I</u> <u>R</u> <u>E</u> <u>G</u> <u>E</u> <u>G</u> <u>V</u> <u>A</u> <u>R</u> <u>V</u> <u>L</u> <u>E</u> <u>N</u> <u>L</u> <u>G</u> <u>A</u> <u>D</u> <u>P</u> <u>S</u> <u>N</u> <u>I</u> <u>R</u> <u>T</u> <u>O</u> <u>V</u> <u>I</u> <u>R</u> <u>M</u> <u>V</u> <u>G</u> <u>E</u>	237
AtClpD	197	<u>A</u> <u>V</u> <u>S</u> <u>L</u> <u>F</u> <u>T</u> <u>V</u> <u>D</u> <u>D</u> <u>S</u> <u>A</u> <u>G</u> <u>R</u> <u>V</u> <u>L</u> <u>K</u> <u>R</u> <u>L</u> <u>G</u> <u>A</u> <u>N</u> <u>M</u> <u>N</u> <u>L</u> <u>L</u> <u>T</u> <u>A</u> <u>A</u> <u>A</u> <u>L</u> <u>T</u> <u>R</u> <u>L</u> <u>K</u> <u>G</u>	232

**Figure 9. Structure-based sequence alignment of AtClpC1 NTD and AtClpD NTD with MtClpC1 NTD, highlighting the residues involved in Cyclomarlin A interaction.** The sequences have been aligned using ClustalW and the residues forming helices are underlined. The residues of MtClpC1 NTD involved in Cyclomarlin A interaction are shown in blue and with a star on top. The corresponding residues of AtClpC1 NTD and AtClpD NTD are shown in blue. The residues conserved among all three sequences are highlighted in red and those conserved among two sequences are highlighted in green.



**Figure 10. Structural alignment of AtClpD NTD with MtClpC1 NTD in complex with Cyclomarlin A.** MtClpC1 NTD is shown in cyan and residues involved in Cyclomarlin A (CymA) interaction are shown in magenta (coordinates from PDB id: 3WDC). CymA is shown in stick model and in blue. AtClpD NTD is shown in red and the residues which could interact with CymA (Phe80, Phe168, Lys173 and Glu177) are shown as sticks in green and labelled. The region of NTDs coming in the proximity of CymA alone is shown for figure clarity.

NTD spanning residues 79–233 was amplified by PCR and cloned into a pGEX-6P-1 (GE Healthcare) vector between BamHI and XhoI restriction sites for expression with a cleavable N-terminal GST tag. The recombinant proteins were expressed in *E. coli* (BL21/DE3 strain). The expression of AtClpC1 NTD was induced at OD<sub>600</sub> of 0.4 with 0.5 mM isopropyl β-D-thiogalactopyranoside (IPTG) at 37 °C in 2xYT medium and allowed to proceed for 4 hours. The cell lysate for AtClpC1 NTD was clarified by centrifugation and the supernatant was passed through a HisTrap FF nickel affinity column (GE Healthcare), followed by a HiLoad 16/600 Superdex 75 prep grade column (GE Healthcare). The expression of recombinant AtClpD NTD was induced at OD<sub>600</sub> of 0.6 with 0.2 mM IPTG at 16 °C in 2xYT medium and allowed to proceed overnight. The cell lysate for AtClpD NTD was clarified by centrifugation, the supernatant passed through a GSTrap FF column (GE Healthcare) and the bound protein eluted from the column for cleavage with PreScission protease (GE Healthcare). In order to separate AtClpD NTD from any excess of the uncleaved GST fusion protein and cleaved off GST, the protein was again passed through GSTrap FF column. The protein was then passed through a HiLoad 16/600 Superdex 75 prep grade column. The eluant from the column was in a buffer containing 20 mM Tris pH 7.5, 150 mM NaCl, 1 mM DTT, and 1 mM PMSE.

**Crystallization and data collection.** Crystals of AtClpC1 NTD with carboxyl-terminal hexahistidine tag appeared in sitting drop vapor diffusion plate in 3–4 days in a condition having 1 M ammonium phosphate and 100 mM sodium citrate, pH 5.6. The crystals were transferred into a cryoprotectant solution containing the reservoir solution supplemented with 25% glycerol and then flash cooled in liquid nitrogen. Preliminary diffraction data collection was carried out at the x-ray diffraction facility of Indian Institute of Chemical Technology (IICT), Hyderabad, India on an R-axis IV++ machine equipped with an image plate detector. High resolution dataset was later collected using the Indian synchrotron radiation source at the RRCAT beamline BL21 and recorded with a MarCCD detector. A total of 120 frames of 1° oscillation were collected at a wavelength of 0.9794 Å. The

diffraction data were processed using iMOSFLM<sup>31</sup> and AIMLESS<sup>32</sup> from the CCP4 suite of programs<sup>33</sup>. The crystal belonged to the orthorhombic space group  $P2_12_12_1$  with unit cell dimensions  $a = 40.10 \text{ \AA}$ ,  $b = 44.29 \text{ \AA}$ ,  $c = 97.56 \text{ \AA}$ , and  $\alpha = \beta = \gamma = 90^\circ$ . One molecule was found per asymmetric unit with a solvent content of ~54%.

AtClpD NTD with carboxyl-terminal hexahistidine tag was highly soluble and did not yield any crystals even after extensive screening. However, crystals of AtClpD NTD (obtained from GST-tagged construct) appeared in sitting drop vapor diffusion plate in 1–2 days in a condition having 20% v/v 2-propanol, 20% w/v PEG MME 2000 and 100 mM MES monohydrate, pH 6.0. The drop containing crystals was overlaid with mineral oil to avoid excessive evaporation of 2-propanol and a cryoprotectant solution containing the reservoir solution supplemented with 20% ethylene glycol was added to the drop and a single crystal flash cooled in liquid nitrogen. High resolution dataset was collected at beamline BM14 of ESRF and recorded with a MarCCD detector. The diffraction data of AtClpD NTD crystal were also processed using iMOSFLM and AIMLESS from the CCP4 suite of programs. The crystal belonged to the monoclinic space group  $P2_1$  with unit cell dimensions  $a = 38.08 \text{ \AA}$ ,  $b = 37.31 \text{ \AA}$ ,  $c = 99.98 \text{ \AA}$ ,  $\alpha = \gamma = 90^\circ$ , and  $\beta = 96.08^\circ$ . Two molecules were found per asymmetric unit with a solvent content of ~39%.

**Structure determination.** The crystal structure of AtClpC1 was solved by molecular replacement method using the program Molrep<sup>34</sup> from CCP4 suite of programs, using the coordinates of *M. tuberculosis* ClpC1 NTD crystal structure in complex with Cyclomarin A (PDB id: 3WDC)<sup>27</sup>, but using only the NTD coordinates as a search model. Model building and structure refinement were carried out using COOT<sup>35</sup> and Refmac5<sup>36</sup> from CCP4 suite of programs. The crystal structure of AtClpD NTD was solved by molecular replacement method with Molrep program, using the coordinates of AtClpC1 NTD crystal structure (PDB id: 5GUI) as a search model. Model building and structure refinement of AtClpD NTD were carried out similar to AtClpC1 NTD. The structures were analyzed for stereochemical quality with the help of Ramachandran Plot from the program PROCHECK<sup>37</sup>. The structure figures were prepared and structural superpositions were prepared and calculated using PyMOL Molecular Graphics System (Schrodinger, LLC).

**Accession numbers.** Atomic coordinates and structure factors have been deposited with the Protein Data Bank under the accession codes 5GUI (AtClpC1 NTD) and 5GKM (AtClpD NTD).

## References

- Peltier, J. B. *et al.* 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, doi: 10.1074/jbc.M309212200 (2004).
- Peltier, J. B., Ytterberg, J., Liberles, D. A., Roepstorff, P. & van Wijk, K. J. Identification of a 350-kDa ClpP protease complex with 10 different Clp isoforms in chloroplasts of *Arabidopsis thaliana*. *J Biol Chem.* **276**, 16318–16327, doi: 10.1074/jbc.M010503200 (2001).
- Adam, Z. & Clarke, A. K. Cutting edge of chloroplast proteolysis. *Trends Plant Sci.* **7**, 451–456, doi: S1360-1385(02)02326-9 (2002).
- Lee, U. *et al.* The *Arabidopsis* ClpB/Hsp100 family of proteins: chaperones for stress and chloroplast development. *Plant J.* **49**, 115–127, doi: 10.1111/j.1365-313X.2006.02940.x (2007).
- Mishra, R. C. & Grover, A. ClpB/Hsp100 proteins and heat stress tolerance in plants. *Crit Rev Biotechnol.* **1**–13, doi: 10.3109/07388551.2015.1051942 (2015).
- Constan, D., Froehlich, J. E., Rangarajan, S. & Keegstra, K. A stromal Hsp100 protein is required for normal chloroplast development and function in *Arabidopsis*. *Plant Physiol.* **136**, 3605–3615, doi: 10.1104/pp.104.052928 (2004).
- Simpson, S. D. *et al.* Two different novel cis-acting elements of *erd1*, a *clpA* homologous *Arabidopsis* gene function in induction by dehydration stress and dark-induced senescence. *Plant J.* **33**, 259–270, doi: 10.1046/j.1365-313X.2003.01624.x (2003).
- Nakashima, K., Kiyosue, T., Yamaguchi-Shinozaki, K. & Shinozaki, K. A nuclear gene, *erd1*, encoding a chloroplast-targeted Clp protease regulatory subunit homolog is not only induced by water stress but also developmentally up-regulated during senescence in *Arabidopsis thaliana*. *Plant J.* **12**, 851–861, doi: 10.1046/j.1365-313X.1997.12040851.x (1997).
- Weaver, L. M., Froehlich, J. E. & Amasino, R. M. Chloroplast-targeted ERD1 protein declines but its mRNA increases during senescence in *Arabidopsis*. *Plant Physiol.* **119**, 1209–1216, doi: 10.1104/pp.119.4.1209 (1999).
- Zheng, B., Halperin, T., Hruskova-Heidingsfeldova, O., Adam, Z. & Clarke, A. K. Characterization of Chloroplast Clp proteins in *Arabidopsis*: Localization, tissue specificity and stress responses. *Physiol Plant.* **114**, 92–101, doi: 10.1034/j.1399-3054.2002.1140113.x (2002).
- Mishra, R. C., Richa & Grover, A. Constitutive over-expression of rice ClpD1 protein enhances tolerance to salt and desiccation stresses in transgenic *Arabidopsis* plants. *Plant Sci.* **250**, 69–78, doi: 10.1016/j.plantsci.2016.06.004 (2016).
- Nishimura, K. & van Wijk, K. J. Organization, function and substrates of the essential Clp protease system in plastids. *Biochim Biophys Acta.* **1847**, 915–930, doi: 10.1016/j.bbabi.2014.11.012 (2015).
- Schirmer, E. C., Glover, J. R., Singer, M. A. & Lindquist, S. HSP100/Clp proteins: a common mechanism explains diverse functions. *Trends Biochem Sci.* **21**, 289–296, doi: 10.1016/S0968-0004(96)10038-4 (1996).
- Rosano, G. L., Bruch, E. M. & Ceccarelli, E. A. Insights into the Clp/HSP100 chaperone system from chloroplasts of *Arabidopsis thaliana*. *J Biol Chem.* **286**, 29671–29680, doi: 10.1074/jbc.M110.211946 (2011).
- Turgay, K., Persuh, M., Hahn, J. & Dubnau, D. Roles of the two ClpC ATP binding sites in the regulation of competence and the stress response. *Mol Microbiol.* **42**, 717–727, doi: 10.1046/j.1365-2958.2001.02623.x (2001).
- Nishimura, K. *et al.* ClpS1 is a conserved substrate selector for the chloroplast Clp protease system in *Arabidopsis*. *Plant Cell.* **25**, 2276–2301, doi: 10.1105/tpc.113.112557 (2013).
- Tapken, W., Kim, J., Nishimura, K., van Wijk, K. J. & Pilon, M. The Clp protease system is required for copper ion-dependent turnover of the PAA2/HMA8 copper transporter in chloroplasts. *New Phytol.* **205**, 511–517, doi: 10.1111/nph.13093 (2015).
- Nishimura, K. *et al.* 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, doi: 10.1105/tpc.15.00574 (2015).
- Guo, F., Esser, L., Singh, S. K., Maurizi, M. R. & Xia, D. 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, doi: 10.1074/jbc.M208104200 (2002).
- Holm, L. & Rosenstrom, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545–549, doi: 10.1093/nar/gkq366 (2010).
- Krissinel, E. & Henrick, K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr.* **60**, 2256–2268, doi: 10.1107/S0907444904026460 (2004).
- Erbse, A. H. *et al.* Conserved residues in the N-domain of the AAA+ chaperone ClpA regulate substrate recognition and unfolding. *FEBS J.* **275**, 1400–1410, doi: 10.1111/j.1742-4658.2008.06304.x (2008).
- Kirstein, J., Moliere, N., Dougan, D. A. & Turgay, K. Adapting the machine: adaptor proteins for Hsp100/Clp and AAA+ proteases. *Nat Rev Microbiol.* **7**, 589–599, doi: 10.1038/nrmicro2185 (2009).

24. Kress, W., Maglica, Z. & Weber-Ban, E. Clp chaperone-proteases: structure and function. *Res Microbiol.* **160**, 618–628, doi: 10.1016/j.resmic.2009.08.006 (2009).
25. Striebel, F., Kress, W. & Weber-Ban, E. Controlled destruction: AAA+ ATPases in protein degradation from bacteria to eukaryotes. *Curr Opin Struct Biol.* **19**, 209–217, doi: 10.1016/j.sbi.2009.02.006 (2009).
26. Kim, J. *et al.* Structures, Functions, and Interactions of ClpT1 and ClpT2 in the Clp Protease System of Arabidopsis Chloroplasts. *Plant Cell.* **27**, 1477–1496, doi: 10.1105/tpc.15.00106 (2015).
27. Vasudevan, D., Rao, S. P. & Noble, C. G. Structural basis of mycobacterial inhibition by cyclomarin A. *J Biol Chem.* **288**, 30883–30891, doi: 10.1074/jbc.M113.493767 (2013).
28. Schmitt, E. K. *et al.* The natural product cyclomarin kills Mycobacterium tuberculosis by targeting the ClpC1 subunit of the caseinolytic protease. *Angew Chem Int Ed Engl.* **50**, 5889–5891, doi: 10.1002/anie.201101740 (2011).
29. Gavriš, E. *et al.* Lassomycin, a ribosomally synthesized cyclic peptide, kills mycobacterium tuberculosis by targeting the ATP-dependent protease ClpC1P2. *Chem Biol.* **21**, 509–518, doi: 10.1016/j.chembiol.2014.01.014 (2014).
30. Gao, W. *et al.* The cyclic peptide ecumicin targeting ClpC1 is active against Mycobacterium tuberculosis *in vivo*. *Antimicrob Agents Chemother.* **59**, 880–889, doi: 10.1128/AAC.04054-14 (2015).
31. Battye, T. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr D Biol Crystallogr.* **67**, 271–281, doi: 10.1107/S0907444910048675 (2011).
32. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr.* **69**, 1204–1214, doi: 10.1107/S0907444913000061 (2013).
33. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr.* **67**, 235–242, doi: 10.1107/S0907444910045749 (2011).
34. Vagin, A. & Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallogr D Biol Crystallogr.* **66**, 22–25, doi: 10.1107/S0907444909042589 (2010).
35. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr.* **66**, 486–501, doi: 10.1107/S0907444910007493 (2010).
36. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr.* **53**, 240–255, doi: 10.1107/S0907444996012255 (1997).
37. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structure. *J. Appl. Crystallogr.* **26**, 283–291, doi: 10.1107/S0021889892009944 (1993).

## Acknowledgements

We acknowledge the intramural support provided to D.V. by the Institute of Life Sciences, Bhubaneswar and DBT-RA fellowship to C.M. by the Department of Biotechnology, Ministry of Science and Technology, Government of India. We thank Ravindra D. Makde, Biplob Ghosh and Ashwani Kumar of BL21/Indus II in RRCAT (Indore, India) for their help with AtClpC1 NTD crystal data collection. We thank Babu A. Manjasetty of BM14 in ESRF (Grenoble, France) for his help with AtClpD NTD crystal data collection. Help of Anthony Addlagatta of CSIR-IICT, Hyderabad by providing in-house diffraction facility for our preliminary diffraction experiments on AtClpC1 NTD is also greatly acknowledged. We also thank Manikkoth B. Asha, Christian G. Noble, Jobichen Chacko and Narottam Acharya for their valuable suggestions during the preparation of this manuscript.

## Author Contributions

C.M. did all experiments on AtClpD. M.K.J. did all experiments on AtClpC1. D.V. conceived the study, participated in the design of experiments and wrote the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing Interests:** The authors declare no competing financial interests.

**How to cite this article:** Mohapatra, C. *et al.* Crystal structures reveal N-terminal Domain of *Arabidopsis thaliana* ClpD to be highly divergent from that of ClpC1. *Sci. Rep.* **7**, 44366; doi: 10.1038/srep44366 (2017).

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017