

Structural determinants for pH-dependent activation of a plant metacaspase

Received: 25 January 2025

Accepted: 20 May 2025

Published online: 29 May 2025

Haijiao Liu^{1,2,5}, Max Henderson^{3,5}, Zhili Pang⁴, Qingfang Zhang³, Eric Lam⁴✉ & Qun Liu^{1,2,3}✉

Arabidopsis thaliana metacaspase 9 (*AtMC9*) plays roles in clearing dead cells, forming xylem vessels, and regulating immunity and programmed cell death in plants. The protease's activation is controlled by pH levels, but the exact structural mechanism behind this has not been elucidated. In this work, we report high-resolution crystal structures for *AtMC9* under both active (pH 5.5 and pH 4.2) and inactive (pH 7.5) conditions. The three structures are similar except for local conformations where their hydrogen bonding interactions with solvents are mediated through the protonation of specific titratable amino acid residues' side chains. By combining structural analysis, molecular dynamics simulations under constant pHs, and biochemical assays coupled with site-directed mutagenesis, we show that the regulation of *AtMC9* activation involves multiple titratable glutamate and histidine residues across the three domains of p20, linker, and p10. Specifically, deprotonated Glu112, His193, and His208 can suppress *AtMC9* proteolytic activity, while protonation of Glu255 and His307 at acidic pH may promote it. This study provides valuable insights into the pH-dependent activation of *AtMC9* and could potentially lead to improving crops with enhanced immunity and controlled cell death, ultimately increasing agricultural productivity.

Metacaspases (MCs) play a prominent role in plant immunity, biotic and abiotic stress-induced programmed cell death (PCD), aging, and cell proliferation during plant development^{1–4}. MCs are arginine/lysine-specific, caspase-like cysteine proteases that can be found in protists, fungi, and plants^{5,6}. They cleave their substrates after arginine or lysine residues instead of the aspartate specificity seen with caspases⁶. There are three major types of MCs. Type I MCs have a prodomain, which could contain a proline-rich repeat or a zinc finger motif, at the N-terminus. Type II MCs, found almost exclusively in the green lineage that includes plants and green algae, have an extended linker region between the p20 and p10 domains but without an obvious prodomain^{6–9}. Type III MCs, which are found in phytoplanktons such as diatoms and cryptophytes, have only p20 and p10 domains with the p10 domain preceding the p20 domain and a very short linker region^{7,10}.

Nine plant MCs were first identified in the *Arabidopsis thaliana* genome^{6,8}. Among those MCs, metacaspases 9 (*AtMC9*) belongs to type II MCs, of which there are 6 in *Arabidopsis*, with *AtMC4* and *AtMC9* being the most well-characterized^{5,6,9,11}. *AtMC9* and *AtMC4* are also renamed as *AtMCA-IIc* and *AtMCA-IIa*, respectively, based on the recommendation by the plant cell death research community¹². *AtMC9* is involved in the process of xylem development through its role in corpse clearing during the later stages of tracheary element differentiation¹¹. In addition, *AtMC9* is required for the cleavage of the Grim Reaper precursor (GRI) to produce the 11-residue GR1p elicitor peptide that can activate PCD under oxidative stress via the transmembrane receptor kinase PRK5¹³. More recently, *AtMC9* and other type II MCs of *Arabidopsis* were proposed as potential convertases for PROPEP1 based on transient expression studies to process the

¹Department of Materials Science and Chemical Engineering, Stony Brook University, Stony Brook, NY, USA. ²Biology Department, Brookhaven National Laboratory, Upton, NY, USA. ³Department of Biochemistry and Cell Biology, Stony Brook University, Stony Brook, NY, USA. ⁴Department of Plant Biology, Rutgers, The State University of New Jersey, New Brunswick, NJ, USA. ⁵These authors contributed equally: Haijiao Liu, Max Henderson.

✉ e-mail: eric.lam@rutgers.edu; qunliu@bnl.gov

precursor peptide PROPEP1 into the phytocytokine Pep1^{14,15}, which can play vital roles in governing essential processes in plant development and defense through its pattern recognition receptor (PRR) receptors PEPR1/2¹⁶.

Unlike other type II MCs that have been characterized thus far, activation of AtMC9 is uniquely dependent on an acidic pH, but not calcium, which is needed to activate *Arabidopsis thaliana* metacaspase 4 (AtMC4) and other type II MCs such as AtMC5 and AtMC8 that have pH optima at around 7.5^{5,17–20}. AtMC9 can be found expressed in the cytosol, which has a neutral pH under ambient conditions, while its activation likely occurs in the apoplast, where GRI resides, which has an acidic pH near 5¹³. In addition, there is evidence for possible nuclear-localized AtMC9 when expressed in transgenic plants as a GFP fusion²¹. A recent study reported the crystal structure of a non-functional AtMC9 mutant (C147A) crystalized at pH 6.3 and used it to characterize the mechanism of a specific inhibitor. The application of the AtMC9-specific inhibitor provides evidence for an additional role of AtMC9 in lateral root emergence²². However, it remains unclear how AtMC9's autolytic activation and subsequent substrate processing are regulated by the pH shift from 7.5 to 5.5.

Here, we report three structures of AtMC9 at its active and inactive pHs, along with molecular dynamics simulations under constant pH conditions, and biochemical assays with site-directed mutagenesis

at conserved residues throughout the protein. Our work reveals structural features that enable the combined use of titratable residues to suppress activation at neutral to basic pH, while other residues promote activation under acidic pH, resulting in tightly regulated activation of AtMC9 only at acidic pHs.

Results

pH-dependent AtMC9 activation for autolysis and structural changes

The AtMC9 zymogen is activated by protons, which lead to autolytic self-cleavage after Arg183 in the linker domain that relieves the auto-inhibitory function of the linker, similar to that found with AtMC4 upon calcium activation^{5,17,19}. After the cleavage and clearance of the linker from the active site, AtMC9 is activated to process substrates in an acidic environment⁵. Calcium-activated AtMC4 converts PROPEP1 to the Pep1 elicitor by cleavage at the site of Arg69¹⁴. We observed a similar cleavage pattern with a GST-tagged PROPEP1 substrate by AtMC4¹⁹ and propose that AtMC9 may likely cleave at the same target site in PROPEP1 (Fig. 1a). We thus produced recombinant AtMC9 and GST-fused PROPEP1 and performed pH-dependent AtMC9 activation (Fig. 1b) and substrate processing assays (Fig. 1c) in vitro. As expected, AtMC9 is inactive at a neutral or higher pH; while it is activated and autolyzed at an acidic pH starting around pH 5.6, and more obvious at

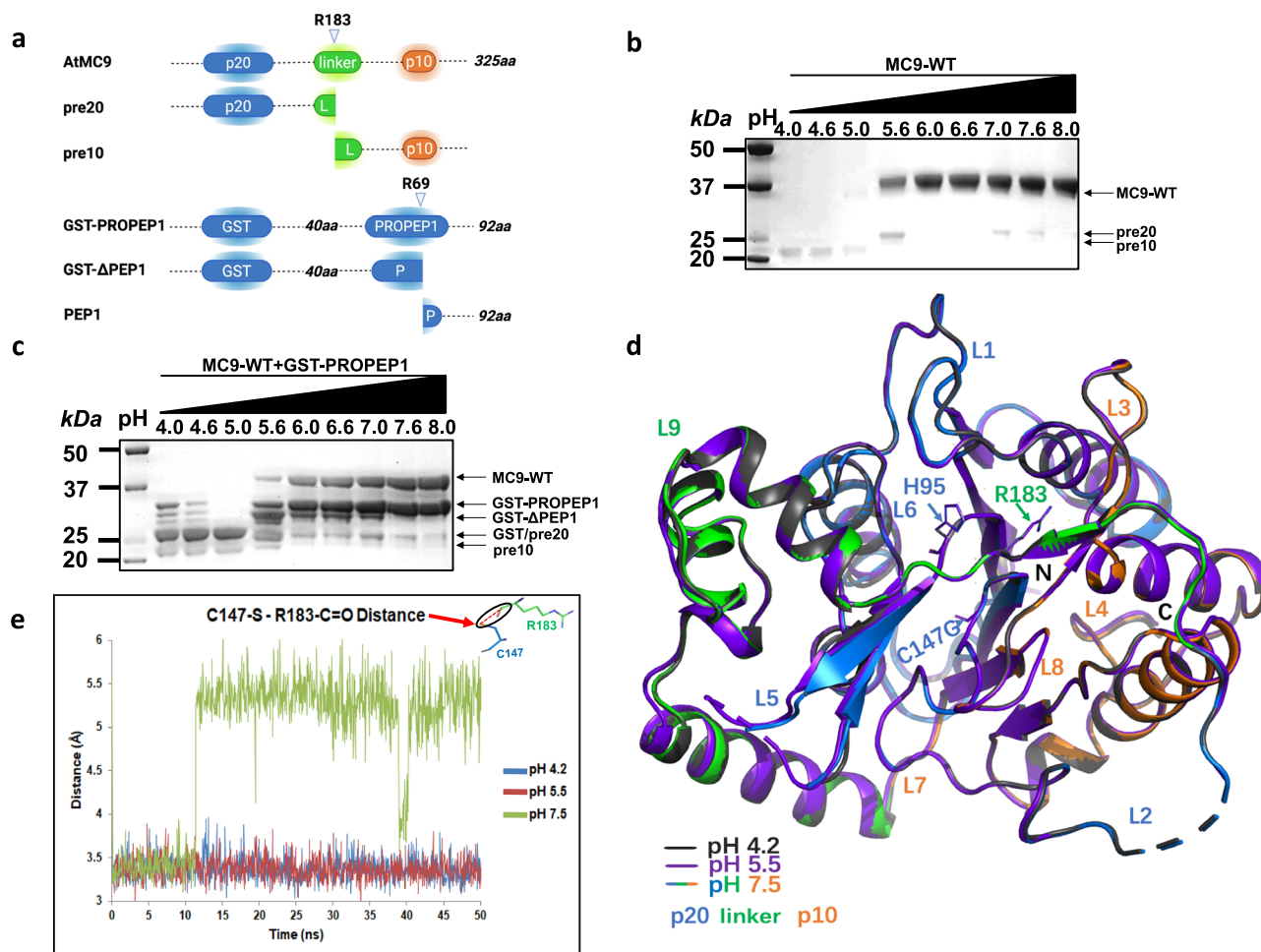


Fig. 1 | Structure and activity of AtMC9. a Schematics of major fragments produced in self-cleavage of AtMC9 and its cleavage of the GST-fused substrate PROPEP1 (residues 40–92) (GST-PROPEP1). Created in BioRender. Liu, H. (2025) <https://BioRender.com/x35j923>. **b** pH-dependent self-cleavage and activation of AtMC9. Data are representative of at least three independent experiments. **c** pH-dependent processing of GST-PROPEP1 by AtMC9. Data are representative of at least three

independent experiments. **d** AtMC9 C147G mutant structures at pH 4.2 (gray), pH 5.5 (purple), and pH 7.5 (colored differently by three domains). **e** Molecular dynamics trajectories showing the distance between the catalytic C147 sulfur and Arg183's carbonyl carbon over 50 ns at pH 7.5 (green), 5.5 (red), and 4.2 (blue). Source data are provided as a Source Data file.

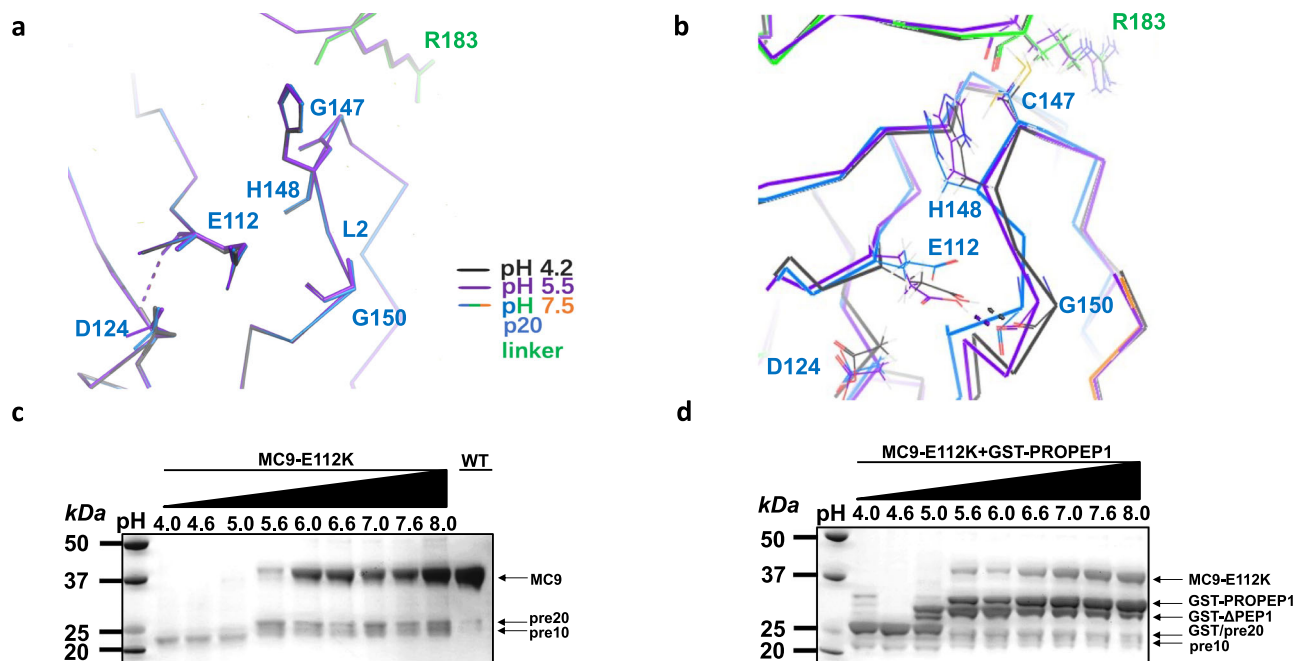


Fig. 2 | A conserved negatively charged residue, Glu112, in the p20 domain contributes to *AtMC9* activation and substrate processing. **a** Comparison of the interactions involving Glu112 at three different pHs. **b** Comparison of representative frames of the pH 4.2 (gray), 5.5 (purple), and 7.5 (green/marine/orange) simulations showing interactions involving Glu112. Interactions at each pH are

displayed as each pH's respective color but with pH 7.5 displayed as red. **c** Self-cleavage in E112K. For the wild-type (WT) sample, pH 8 was used in the incubation to compare with the mutant. Data are representative of two independent experiments. **d** Substrate GST-PROPEP1 processing by E112K. Data are representative of two independent experiments. Source data are provided as a Source Data file.

pH 5.0 or lower. Similarly, *AtMC9* starts to actively process its substrate GST-PROPEP1 at around pH 5.6, with the most PEPI peptides produced at pH 5.0. At a lower pH (pH < 4.6), *AtMC9* activity begins to decrease and is not as efficient in processing its substrate. It appeared that *AtMC9* is activated within a pH range between 4.6 and 5.6, similar to what was previously reported⁵.

To reveal the structural determinants for pH-dependent activation, we produced a Cys147 to Gly147 inactive mutant (C147G) of *AtMC9*, crystallized it at either acidic pHs (pH 4.2 and 5.5) where the zymogen is expected to be active or at higher pH where the zymogen stays inactive (pH 7.5), and determined their structures (Fig. 1d). The structures at different pHs are largely similar to each other with the largest root mean square deviation (RMSD) of 1.4 Å between pH 5.5 and pH 7.5. In the solved structures, Arg183 from the linker domain is inserted into the catalytic pocket and blocks its access by a substrate (Supplementary Fig. 1), analogous to Lys225 in the case of the calcium-dependent *AtMC4*^{17,19}. The Arg183 carbonyl carbon is sandwiched by the two catalytic residues of Cys147 (Gly147 in structure) and His95. For a type II MC, the *AtMC9* structure exhibits a relatively short linker domain positioned between its p20 and p10 domains compared to others^{6,20}. Among the structures at the three pHs, differences are found in the loops L2, L5, and L9 with part of L2 and L5 disordered (Fig. 1d).

Although the *AtMC9* structures at active and inactive pHs are rather similar, we found that the pH 5.5 structure is in space group P2₁ while the pH 4.2 and 7.5 structures are in space group I2, suggesting that pH may alter *AtMC9*'s molecular packing behavior during its crystallization process. To reveal the dynamics of the pH-dependent activation of *AtMC9*, we performed molecular dynamics simulations under constant pH conditions²³, with three replicates for each of the three structures. To understand the dynamic process of *AtMC9* activation, we measured the distance between the catalytic residue Cys147 thiol group and the carbonyl carbon of Arg183 with respect to the simulated trajectory throughout 50 ns (Fig. 1e). The smaller the distance between the two atoms, the more interactions between Cys147 and Arg183, and more efficient activation by cleavage would be

predicted. The simulations show that at the beginning (first 10 ns) of each simulation, the distances are between 3.3 Å and 3.6 Å. After that, the distances increased dramatically to around 5.5 Å at pH 7.5, suggesting pH 7.5 is not favorable for activation. For the simulation at pH 4.2 and 5.5, the distances remained the same, thus likely more favorable for activation. The results from constant pH simulations at the three pHs, which indicate that structural perturbations at more basic pHs resulted in unfavorable distance between Cys147 and Arg183 for catalysis, are consistent with those from *in vitro* autolytic activation (Fig. 1b) and substrate processing assays (Fig. 1c).

Our structural, biochemical, and computational analyses of *AtMC9* at inactive and active pHs provide the basis to further characterize the pH-dependent activation of this protease by protons. To delineate the specific structural determinants contributing to the pH-dependent activation and function of *AtMC9*, we performed sequence alignments of *AtMC9* orthologs from different plant phyla (Supplementary Fig. 2), but excluding other type II metacapses whose activation is calcium dependent^{17,20,24}, and identified among these conserved residues for which their protonation states and local structures may contribute to zymogen activation and catalytic activities. Based on pKa calculation and residue conservation within each of the *AtMC9* orthologs as criteria, we selected residues with titratable carboxylates or imidazoles for further characterization (Supplementary Table 1).

A conserved glutamate Glu112 in the p20 domain of *AtMC9* contributes to its activation

In our pKa calculations, we found that a conserved residue Glu112 has an increased pKa of 7.32 or higher at all three pHs (Supplementary Table 1). The theoretical pKa of a glutamate is 4.5. This increased pKa and its conservation across type II MCs suggest a potential role in activation. Under the three pHs, we observed Glu112 in *AtMC9* interacts with Asp124 via an H-bond, which stabilizes it (Fig. 2a, Supplementary Fig. 3). Our constant pH simulation shows that Glu112 is protonated at pH 5.5 and 4.2, and deprotonated at pH 7.5

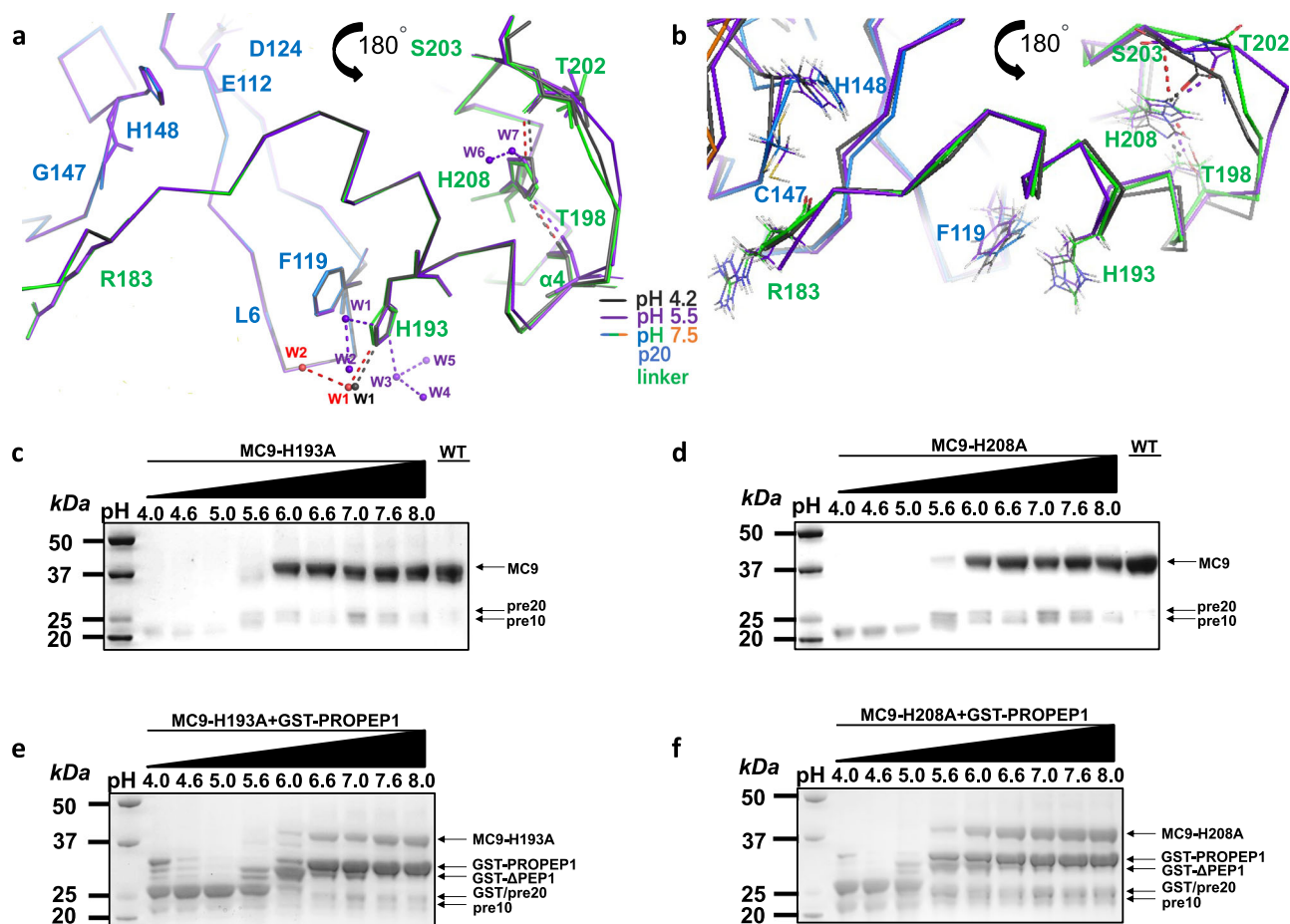


Fig. 3 | Two histidine residues in the linker domain regulate self-cleavage activation and substrate processing. **a** Comparison of the interactions involving H193 and H208 at pH 4.2, 5.5 and 7.5. Interactions and water molecules at each pH are displayed as each pH's respective color but with pH 7.5 displayed as red. **b** Comparison of representative frames of the pH 4.2 (gray), 5.5 (purple), and 7.5 (marine/green) simulations showing interactions found in and around the linker

domain. **c** Self-cleavage in H193A. **d** Self-cleavage in H208A. **e** Substrate processing of GST-PROPEP1 by H193A. **f** Substrate processing of GST-PROPEP1 by H208A. WT: For the wild-type *At*MC9 sample in panels **c** and **d**, pH 8 was used in the incubation to compare with the mutant. Data are representative of two independent experiments for (**c**–**f**). Source data are provided as a Source Data file.

(Supplementary Table 2), consistent with the pKa calculation. Interestingly, protonated Glu112 has an additional H-bond interaction with the Gly150 carbonyl oxygen, likely pushing His148 and catalytic residue Cys147 toward the self-cleavage site of Arg183 (Fig. 2b). In contrast, at pH 7.5, Glu112 is mostly deprotonated and does not have an interaction with Gly150. Consequently, His148 and Cys147 tend to be farther away from their active positions (Fig. 1e).

We hypothesize that the H-bond interaction between Glu112 and Gly150 carbonyl oxygen may contribute directly to *At*MC9 activation. To test this hypothesis, we mutated Glu112 to a positively charged residue Lys112 which might also form an H-bond with Gly150 for activation. As shown in Fig. 2c, d, E112K is indeed activated and can process the substrate GST-PROPEP1 to produce Pep1 peptide even at a basic pH of 8. Compared to the wild-type (WT), E112K is not only staying activated at a higher pH, but it could also be more active at low pHs for both self-activation and substrate processing.

Histidine residues in the linker domain suppress *At*MC9 activation

Type-II MCs have a linker domain that has been shown to regulate autolytic activation and substrate processing^{19,20}. Sequence alignment of *At*MC9 and its homologs from other plant phyla showed two conserved histidine residues His193 and His208 in its linker domain (Supplementary Fig. 2). The imidazole ring of histidine has two titratable groups, which are either protonated or deprotonated

depending on their local environment. Our crystal structures show that the His193 imidazole ring attracts several water molecules at pH 5.5 compared to pH 4.2 and pH 7.5 (Fig. 3a and Supplementary Fig. 4). Similarly, His208 attracts two water molecules at pH 5.5. At all three pHs, the imidazole ring of His193 interacts with the phenyl group of Phe119, which is located in L6 between Glu112 and Asp124 through π -stacking. In contrast, His208 is stabilized by interacting with small side-chain residues Thr198 and Thr202 (Fig. 3a). In addition, the His208 side chain rotates by 27.5° when pH decreases from pH 7.5 to 4.2 (Supplementary Fig. 5), suggesting its imidazole ring may have a different protonation state at different pHs (Supplementary Table 2). Both His193 and His208 are close to the inhibitory residue Arg183 in our resolved *At*MC9 structures. Based on the structure and location of the two histidine residues at the three different pHs, we propose that these two residues may be involved in the folding and stabilization of the linker domain to suppress *At*MC9 activation under neutral to higher pH conditions. Results from constant pH molecular dynamics simulations also suggest that His193 and His208 are involved in the folding and stabilization of the linker region to modulate *At*MC9 activity. At all three pHs, the deprotonated N δ of His208 interacts with Thr198 and the protonated N ϵ interacts with either Thr202 (pH 4.2 and 5.5) or Ser203 (pH 7.5). The π -stacking between His193 and Phe119 observed in the crystal structures appears to either become disrupted or undergo a change in conformation from π -shaped stacking (Fig. 3b and Supplementary Table 2).

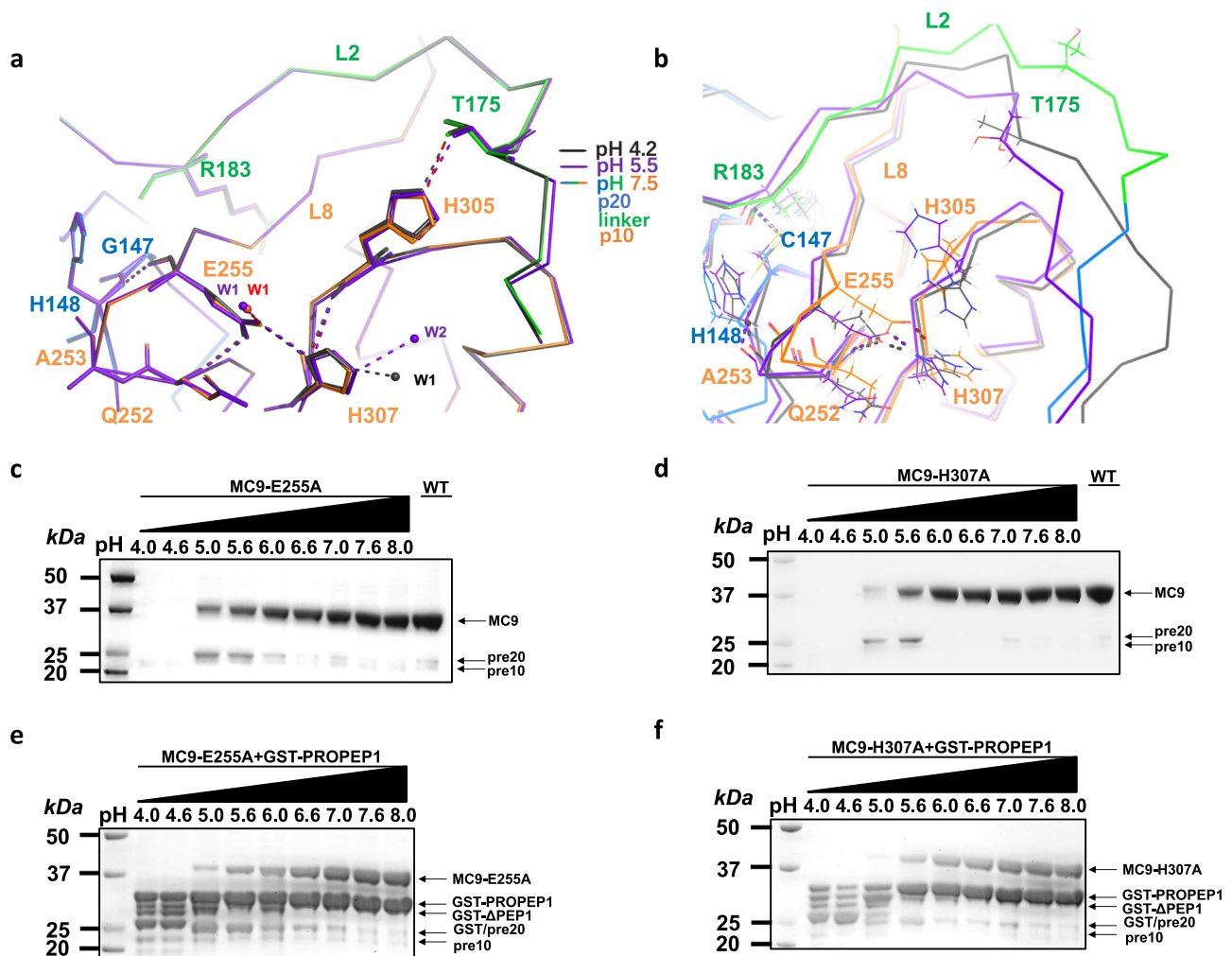


Fig. 4 | Residues Glu255 and His307 in the p10 domain regulate the self-cleavage activation and substrate processing. **a** Interactions involving Glu255 and His307 at pH 4.2, 5.5, and 7.5. Interactions and water at each pH are displayed as each pH's respective color, but with pH 7.5 displayed as red. **b** Comparison of representative frames of the pH 4.2 (gray), 5.5 (purple), and 7.5 (marine/orange) simulations showing interactions involving Glu255 and His307. **c** Self-cleavage in

E255A. **d** Self-cleavage in H307A. **e** Substrate processing of GST-PROPEP1 by E255A. **f** Substrate processing of GST-PROPEP1 by H307A. WT: For the wild-type *AtMC9* sample in (c, d), pH 8 was used in the incubation to compare with the mutant. Data are representative of two independent experiments for (c–f). Source data are provided as a Source Data file.

To further investigate the roles of these two histidine residues, we produced His193A and His208A mutants. Both mutants have increased activation and substrate processing activities at acidic pHs and remained active to a higher pH of 6.0 (Fig. 3c–f) compared to the WT (Fig. 1b, c). However, neither mutant appears to alter the repression of activity at pH 6.6 or higher, which implies that while the two mutants can enhance the activity of *AtMC9* at acidic pHs by removing their inhibitory interactions with other residues such as Phe119 and Thr202, unlike that of E112K, they are insufficient to overcome the remaining inhibitory interactions at pH 6.6 or higher. The two conserved, *AtMC9*-specific, histidine residues in the linker domain may thus help suppress the activation of *AtMC9* at a pH above 6.0 and minimize the background activity of the zymogen. They may not be directly involved in the repositioning of the critical Arg183 to the active site to drive autolytic cleavage at acidic pHs.

Residues Glu255 and His307 in the p10 domain mediate *AtMC9* activation at acidic pH

Based on sequence alignment between *AtMC9* homologs, titratable Glu255 and His307 from the p10 domain are also conserved²⁰ (Supplementary Fig. 2). pKa calculation shows that Glu255 is deprotonated in all three structures (Supplementary Table 1). At all three pHs, Glu255

interacts with Gln252 and the imidazole nitrogen of His307. His307 interacts with the carbonyl oxygen of His305, which in turn interacts with Thr175 in the linker domain. Further investigation reveals that at pH 4.2 and 5.5, His307 forms an H-bond with a nearby water molecule, while at pH 7.5, this interaction is not observed. This suggests that the protonation state of the imidazole ring of this residue is affected by pH (Fig. 4a and Supplementary Fig. 6). Through molecular dynamics simulations under constant pHs, it is predicted that the two titratable nitrogen groups of His307 can be either protonated or deprotonated at different pHs, potentially explaining its varying interactions with nearby water molecules (Fig. 4b). Glu255 on the L8 loop interacts with the active-site residue His148 via a main-chain H-bond. The interaction between Glu255 and His307 could push the Cys147 side chain toward the cleavage site of Arg183. In addition, His307 may regulate the flexibility of the L2 loop through its indirect interaction with Thr175, which could further enhance *AtMC9* activation by promoting the release of Arg183 from its inhibitory location at the catalytic site.

To validate the predicted roles of these two residues in *AtMC9* activation and substrate processing, we produced E255A and H307A mutants as well as an E255A-H307A double mutant. The activation assay shows that all three mutants are inhibited for autolytic activation compared to the WT (Fig. 4c, d and Supplementary Fig. 7a).

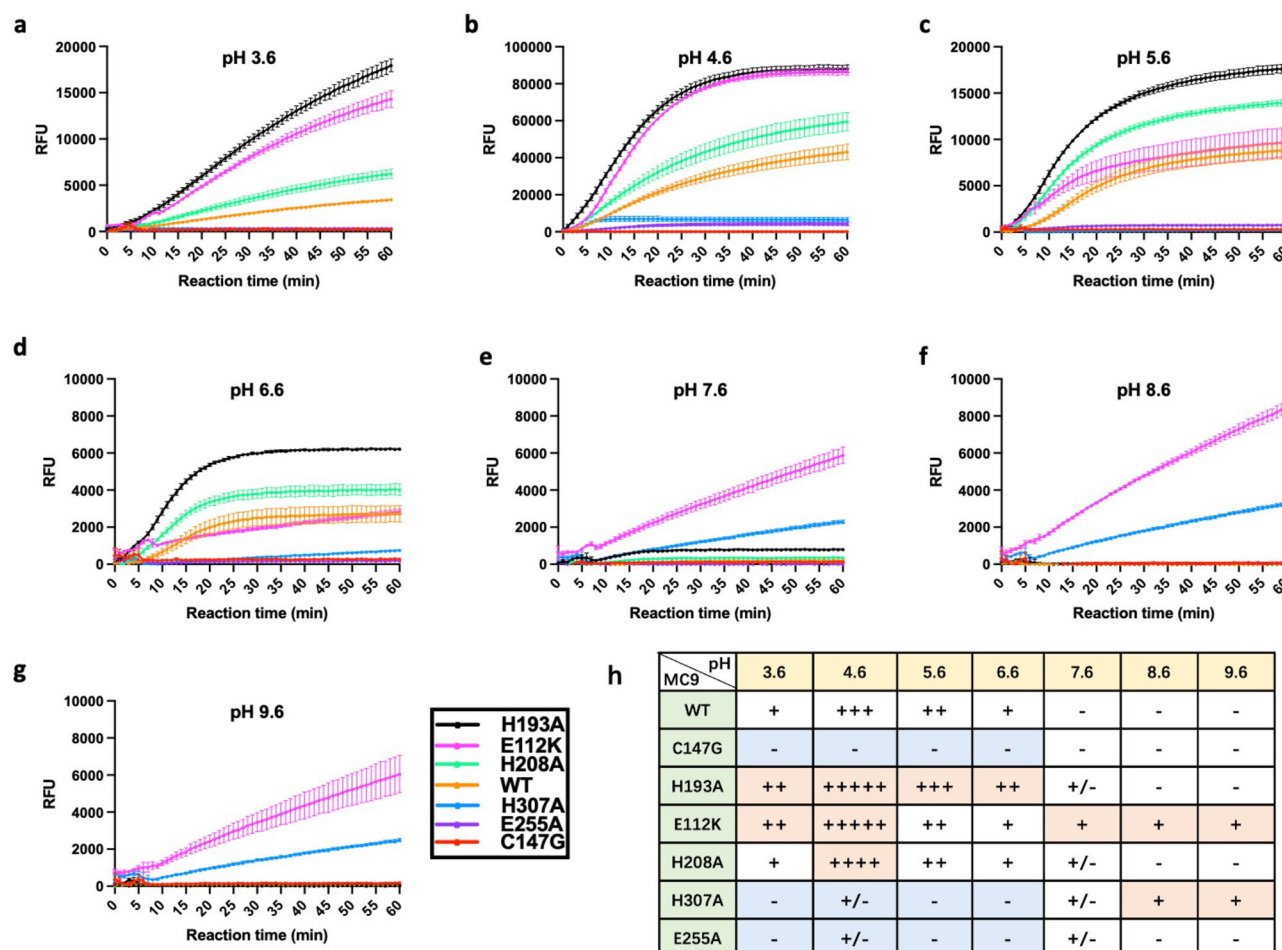


Fig. 5 | Functional characterization of AtMC9 and its mutants. The GRRase activity is shown as relative fluorescence units (RFUs) for 60 min with 1-min intervals for H193A (black), E112K (pink), H208A (green), WT (orange), H307A (blue), E255A (purple), and C147G (red) under indicated pHs for 60 min with 1-min intervals. Data are presented as mean $\pm$ standard deviation (SD). Error bars represent SD

($n = 3$ independent technical replicates). **a** 3.6; **b** 4.6; **c** 5.6; **d** 6.6; **e** 7.6; **f** 8.6; **g** 9.6. **h** The RFU range for AtMC9 and its mutants at 7 different pHs at the 20 min time point: +++++, >40,000; +++, >20,000; ++, >10,000; +, >5000; $\pm$, >200; -, 0. Activities lower than WT are highlighted in blue, and those higher in salmon. Source data are provided as a Source Data file.

Consequently, they are highly inefficient in processing the substrate GST-PROPEP1 (Fig. 4e, f and Supplementary Fig. 7b) in comparison to the WT AtMC9 protein (Fig. 1). Although the E225A-H307A double mutant can undergo some autolytic activation at a pH of 5.0 or lower, it has the lowest substrate processing activity among all mutants. Even at pH 4.0, the majority of the added GST-PROPEP1 remains unprocessed under our assay conditions (Supplementary Fig. 7b). These results indicate that these two residues in the p10 domain are critically important to promote the acidic pH-dependent AtMC9 activation, in contrast to the activation-suppressing residues in the linker and p20 domains described above.

Quantification of AtMC9 activation with a fluorogenic peptide

To quantify the activity of AtMC9 and its mutants, we measured the GRRase activity^{5,17,20} using a fluorogenic substrate Boc-Gly-Arg-Arg-MCA (GRR-MCA) under various pH conditions. We mixed GRR-MCA with buffers ranging from pH 3.6 to 9.6 and added WT AtMC9 or its mutants to initiate the activation, and the produced fluorescent signals after release of the peptide were measured using a plate reader as shown in Fig. 5. Except for the C147G active site mutant, which is inactive in all pHs tested, WT and other mutants showed varying GRRase activity between pH 3.6 and pH 9.6, extending our activation assays using SDS-PAGE analysis (Figs. 1–4). For the WT and most of the active mutants, with the exception of the H307A variant, the optimal

GRRase activity was measured at pH 4.6, which is lower than the previous report of pH 5.5⁵.

Among the mutants tested, the E112K mutant shows significantly increased GRRase activity compared with the WT at both the lower pHs of 4.6 and 3.6 while it also remains active at the high pHs of 7.6 to 9.6, albeit at a lower level, at which the WT enzyme is inactive (Fig. 5). The activation of the E112K variant at basic pHs is consistent with our SDS-PAGE self-cleavage assays and the interpretation that the deprotonated Glu112 residue plays an inhibitory role to suppress AtMC9 activation (Fig. 2c, d), while its significant enhancement of AtMC9 activity even at acidic pHs may result from the introduction of a lysine residue in place of Glu112. Linker domain histidine mutant H193A, and to a lesser extent H208A, show increased GRRase activity at pH 3.6–6.6, consistent with their increased activation and substrate processing activity at lower pH levels (Fig. 3c–f). Unlike the E112K mutant, however, these two AtMC9 variants do not show significant activity at the high pH range of 7.6–9.6 (Fig. 5). Among the gain-of-function mutants we tested, H193A exhibited the highest GRRase activity at pHs < 7.6 (Fig. 5). In contrast to these, the E255A and H307A mutants from the p10 domain showed reduced self-cleavage and GST-PROPEP1 processing activity compared to the AtMC9 wild type (Fig. 4c–f). Consistent with these results, both E255A and H307A displayed lower than WT GRRase activity between pH 3.6 and 6.6.

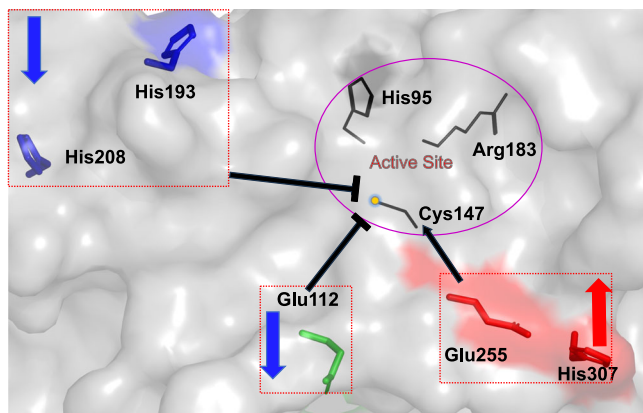


Fig. 6 | Proposed mechanism of AtMC9 activation and regulation. Surface model of the structures surrounding the active site region of AtMC9, with the key modulating residues outlined and labeled. Blue down-arrows signify decreased activation. The red up-arrow indicates increased activation. Black spindles represent suppressive functions as Glu112 keeps Cys147 away from cleaving the inhibitory Arg183 in the active site at a neutral pH or higher. In contrast, the Black arrow indicates the activating function of Glu255 and His307, as these residues may push Cys147 toward Arg183 under acidic pH conditions.

Interestingly, while the E255A mutant stays inactive at higher pHs like the WT, the H307A enzyme displayed a low level of activity in the basic pHs from 7.6 to 9.6. As our structural work suggested, deprotonation of His307 at higher pHs could be involved in restricting the Cys147 active site residue to approach Arg183. It's H307A variant, which has a small side chain, may remove this constraint under higher than neutral pH to enable the protease to attain some level of activation even under pH 9.6. In summary, our site-specific mutagenesis coupled with biochemical analysis of the selected residues in all three domains of AtMC9 demonstrated that they are critical residues for shaping the pH dependence of AtMC9 autoactivation.

Discussion

AtMC9 and its orthologs from other plant phyla require an acidic pH for their activation and function. Therefore, its pH-dependent autoactivation and substrate processing functions have to be precisely regulated. Based on the results presented in this work, we propose a regulatory model for its pH-dependent activation mechanism (Fig. 6). After AtMC9 is synthesized and accumulates in the cytosol as a zymogen, it is inactive at pH 7.5. To suppress spurious activation in neutral or basic pH of the cytosol under ambient conditions, AtMC9 deploys several strategies including deprotonated Glu112 in the p20 domain and the imidazole groups of two histidine residues (His193 and His208) from the linker domain to restraint the active site Cys147 from approaching Arg183, the target residue for autolysis. In contrast, the p10 domain provides the main driver upon transition to acidic pHs for the protease's autolytic activation through protonation of Glu255 and His307 under acidic pHs, while under these conditions, His193 and His208 will be re protonated and no longer exert their repressive functions. As predicted by this model, these residues are all highly conserved in orthologous positions in proteins of the AtMC9 clade (Supplementary Fig. 2) and correlate well with previously identified "Signature" residues that could define this subgroup of type II MCs²⁰. Consistent with our findings, recent site-specific mutagenesis of Glu246 residue in tomato SMC8, which is orthologous to Glu255 in AtMC9, also revealed its importance for autolytic activation as well as its role in defense against biotrophic pathogens²⁵. After exposure of AtMC9 to an acidic environment such as that found in the apoplast, AtMC9 is activated at a pH of 5.6 or lower, where it can process its targets, such as GRI. Alterations in the pH of the cellular milieu,

analogous to calcium spikes^{13,14}, also have the potential to trigger the rapid and transient activation of protease zymogens such as AtMC9 in the cytosol to process targets such as PROPEP1.

In this study, we identified several titratable residues that modulate AtMC9 activation and activity in a pH-dependent manner. For example, mutations of Glu112 and His193 were shown to increase the activation and shift the enzyme's activity profile toward a more neutral pH range. These findings suggest that pH sensitivity is modulated through the cooperative interaction of multiple titratable residues rather than being governed by a single residue or switch. Functionally, although AtMC9 is synthesized in the cytosol, its site of action is often associated with low-pH environments arising from biological processes such as xylem vessel differentiation and stress-triggered immune responses. These contexts involve cellular acidification through vacuolar rupture, oxidative bursts, or the hypersensitive response, providing the acidic conditions necessary for activation.

pH-dependent activation is a broadly conserved mechanism across biological systems. For example, human pepsin is activated in the stomach at pH 2–3²⁶; lysosomal cathepsins function optimally in the acidic lysosomal lumen (pH 4–5)²⁷; LeSBT1 which is a subtilase from tomato plants shows the highest activity at acidic pH²⁸; and parasite proteases, such as FheCLI from *Fasciola hepatica*, become active in the host's acidic gut²⁹. These examples support the notion that pH can serve as a regulatory cue to localize protease activity spatially and temporally.

The results from our molecular dynamics simulations can help rationalize the findings from our in vitro cleavage assays. The distance of the active site Cys147 sulfur atom from the carbonyl carbon of Arg183 was seen to increase at pH 7.5 (Fig. 1e). This is in line with the lower activity at pH 7.5 because the sulfur atom of the catalytic cysteine would likely need to be in proximity of Arg183's carbonyl carbon to initiate catalysis¹⁹. The simulations reveal the regulatory roles of titratable residues in the p20, linker, and p10 domains for AtMC9 activation as summarized in Supplementary Table 2 and Supplementary Fig. 8, and a comparison of the structures obtained under three different pHs as shown in Supplementary Fig. 9 and Supplementary Movies 1 and 2. While the simulations carried out in this study can be used to support and interpret our experimental findings, they do come with limitations. For example, the current study only allows for certain amino acids to be titratable, but it may be beneficial to include cysteine in future work to better understand the catalytic process.

Although previous fluorogenic peptide substrate cleavage assays indicated an optimal pH of around 5.5 for AtMC9 activity⁵, our GRRase activity assay of AtMC9 and its mutants shows an optimal pH of 4.6 (Fig. 5). This shift in optimal pH may be attributed to differences in how our recombinant AtMC9 was produced, purified or the exact buffers that were used. For example, in the previous study⁵, dithiothreitol (DTT, 10 mM) and a high substrate concentration (50 μ M) were used in a buffer composed of 50 mM acetic acid, 50 mM MES, and 100 mM Tris. In our work, we did not include DTT and used a lower concentration of substrate (10 μ M). Nevertheless, the pH-dependent activation profiles are similar, and AtMC9 is activated at a low pH.

Recently, the structure of an AtMC9-C147A mutant at pH 6.3 was reported²². By superimposing our obtained pH 7.5 AtMC9-C147G mutant structure with the pH 6.3 AtMC9-C147A mutant structure, we found that the linker domain does not align well between them, particularly the segment from Arg183 to His241. Notably, the His193 and His208 residues in the linker occupy different positions in these structures (Supplementary Fig. 10a). The activation of type II MCs is either calcium-dependent in AtMC4 or acidic pH-dependent in AtMC9. In AtMC4, a negatively charged loop L5 was proposed to have a critical role in AtMC4 activation and substrate processing. Interactions between calcium and the negatively charged L5 loop may destabilize the electrostatic interactions between L5 and the linker domain, leading to AtMC4 activation¹⁹. In contrast, AtMC9's L5 does not contain negatively charged glutamate or aspartate residues and is disordered

in crystal structures under all three pHs (Fig. 1d), while its linker domain is quite different and shorter than that of AtMC4 (Supplementary Fig. 10b). Therefore, AtMC9 does not utilize calcium for its activation. Instead, AtMC9 deploys titratable residues to precisely regulate its activation and substrate processing under acidic conditions through both positive and negative functions to tune its specific range of activity.

Our comprehensive investigation of the activation mechanism for AtMC9 sheds light on the structural foundation underlying pH-dependent MC activation, offering insights into the molecular details of its activation in response to both abiotic and biotic stresses, when pH changes in different subcellular compartments can play critical signaling functions^{20,30,31}. From our site-specific mutagenesis of titratable residues that are highly conserved between AtMC9 orthologs in diverse plant phyla, our work created several types of gain-of-function mutants of AtMC9 in addition to several regulatory mutants that have lost autolytic activity. These AtMC9 variants with novel activities and pH profiles could help future studies to dissect the function of this MC in plant defense and stress responses, while they may also provide tools to investigate potential strategies for crop protection through controlled manipulation of the downstream pathways from this phyto cytokine convertase in crop plants. This advance can hold promise for engineering more resilient crops tailored for enhanced food and biomass production.

Methods

Protein production for AtMC9

The full-length, WT AtMC9 (residues 1–325) was cloned into the pET-15b vector (Novagen) between the Thrombin and Bpu sites using the Gibson assembly cloning method and standard PCR-based protocols. Mutagenesis of AtMC9 was performed using a Q5 Site-Directed Mutagenesis kit following the manufacturer's protocol. Primers used for cloning and mutagenesis were listed in Supplementary Data 1.

Protein overexpression and purification of AtMC9 followed AtMC4 overexpression and purification conditions^{9,19}. Proteins were overexpressed in *Escherichia coli* C43 (DE3) cells at 15 °C for 16 h induced by the addition of 0.2 mM IPTG (final) to the cell culture at an OD₆₀₀ of 0.5–0.6. Harvested cells were resuspended in a lysis buffer that contains 25 mM Tris, pH 8.0, 250 mM NaCl, 0.5 mM TCEP (tris 2-carboxyethyl phosphine), 10% glycerol, and protease inhibitors. Cells were lysed by using an EmulsiFlex-C3 Homogenizer (Avestin, Ottawa, Canada). After centrifugation at 18,000 × g for 1 h, the supernatants were collected for a two-step purification by nickel-nitrilotriacetic acid (NTA) affinity chromatography (HisTrap FF column, GE Healthcare, Inc.) and size exclusion chromatography (Superdex-200 10/300 GL column, GE Healthcare, Inc.). Eluted protein was concentrated to 30 mg/ml by using a 10 kDa molecular weight cutoff Amicon Ultra-15 centrifugal filter (Millipore, Inc.).

Protein production for GST-PROPEP1

The coding sequence for AtPROPEP1 (residues 40–92 to remove the possible cleavage at the R6/R7 site)¹⁹ (from clone 06-11-N09 of RIKEN, Japan) was inserted into the BamHI and XhoI sites of the pGEX-4T-1 vector (GE Healthcare, Inc.) using standard PCR-based techniques. The GST-fusion protein was overexpressed in *Escherichia coli* BL21 (DE3) pLysS cells at 16 °C for 20 h upon induction with 0.2 mM IPTG (final) at an OD₆₀₀ of 0.5–0.6. Harvested cells were resuspended in a lysis buffer containing 25 mM Tris, pH 8.0, 150 mM NaCl, 10 mM DTT, 5% glycerol, and protease inhibitors. Cells were lysed by using an EmulsiFlex-C3 Homogenizer (Avestin, Ottawa, Canada), followed by the addition of Triton X-100 (Sigma) to a final concentration of 1% and stirring at 4 °C for 1 h. Subsequently, the lysates were centrifuged at 18,000 × g for 1 h, and the supernatant was subjected to purification using Glutathione Sepharose 4B resin and a PD10 column, following the manufacturer's protocols (GE Healthcare, Inc.). The resin was washed with a wash buffer containing 25 mM HEPES, pH 8.0, 150 mM NaCl, 10 mM DTT,

0.2 mM AEBSF, and 5% glycerol. The GST-PROPEP1 protein was eluted using a wash buffer supplemented with 10 mM glutathione and 0.1% Triton X-100.

Crystallization

Crystallization of AtMC9 C147G was performed by using the sitting-drop vapor diffusion method. For crystallization, 1 μL of 30 mg/ml protein was mixed with an equal volume of various crystallization reagents. After optimization, the best crystallization conditions are pH 7.5 (0.2 M ammonium sulfate, 0.1 M Tris and 20% w/v PEG 5000 MME), pH 4.2 (0.2 M Sodium chloride, 0.1 M Phosphate/citrate and 20 % w/v PEG 8000), and pH 5.5 (0.1 M sodium acetate, 0.2 M ammonium sulfate and 20% w/v PEG 5000 MME). For cryo-crystallography, 10% glycerol was added as a cryoprotectant before crystals were plunge-frozen into liquid nitrogen.

Diffraction data collection and processing

All crystallographic data were collected at the NSLS-II FMX and AMX beamlines, employing an Eiger 16 M detector for FMX and an Eiger 9 M detector for AMX, both at a cryogenic temperature of 100 K. The indexed and integrated data sets were processed using DIALS³², while the scaling and merging steps were carried out using the CCP4 programs (v8.0) POINTNESS and AIMLESS³³. Following data processing, a rejection of outlier crystals and frames was implemented using a developed procedure in PyMDA³⁴. Detailed statistics for data collection and reduction are shown in Supplementary Table 3.

Structure determination

The structure of the AtMC9 C147G mutant was determined using molecular replacement in PHENIX (v1.19)^{34,35} using a model predicted by AlphaFold2³⁶. The structures underwent iterative rebuilding and refinement using COOT (v0.9.8.91) and phenix.refine, respectively^{34,35}. Non-crystallographic symmetry was employed for restraints and TLS parameters were used to model anisotropy. The refined Rwork/Rfree values for the structures at pH 7.5, pH 5.5, and pH 4.2 are 0.197/0.215, 0.219/0.231, and 0.185/0.206, respectively. The final models were validated using the program MOLPROBITY³⁷. Detailed refinement statistics are shown in Supplementary Table 3. The structure figures were prepared using PyMOL³⁸.

The protonation states of titratable residues at pHs 7.5, 5.5, and 4.2 were calculated using the program Propka3 version 3.4.0³⁹. Water molecules were retained, and the pKa values were determined at the same pH used for crystallization.

In vitro pH-dependent cleavage assays

To assess AtMC9 self-cleavage activity, 5 μM of purified AtMC9 or its mutants were incubated for 15 min at room temperature in a 15 μL reaction buffer comprising varying pHs from pH 4.0 to 8.0, 150 mM NaCl, 10% glycerol, and 0.5 mM TCEP.

To assess the cleavage of substrate GST-PROPEP1 by AtMC9 and its mutants, 5 μM of purified protein and 10 μM of purified GST-PROPEP1 were incubated for 15 min at room temperature in a 15 μL reaction buffer containing a pH gradient reagent ranging from pH 4.0 to 8.0, 150 mM NaCl, 10% glycerol, and 0.5 mM TCEP. The reactions were stopped by adding an SDS-PAGE sample buffer. The proteins were then separated using 4–20% gradient precast PAGE gels (GenScript, Piscataway, NJ) and visualized through Coomassie blue staining.

Molecular dynamics simulations

The AtMC9 C147G crystal structures were prepared for molecular dynamics simulations. For structures at each pH, UCSF Chimera (v1.15)⁴⁰ was used to convert the structures to their WT sequences by introducing a G147C mutation. To fill in the residues (104–106 and 165–170) that were missing in the pH 5.5 structure, AlphaFold2 was used to model AtMC9 using the WT sequence³⁶. The top-ranked, relaxed structure produced

by AlphaFold2 was then aligned with the pH 5.5 crystal structure using PyMOL (v2.5.2)³⁸. All residues of the AlphaFold structure were removed except for the missing loops. PyMOL's Builder function was then used to create bonds between the AlphaFold2-built loops and the pH 5.5 crystal structure, followed by the implementation of the "Clean" command. This same procedure was used for missing residues 103–107 of the pH 4.2 and pH 7.5 structures.

All simulations were performed using a modified version of GROMACS (v2021)^{41,42}. Simulation systems setup is summarized in Supplementary Table 4. To ensure the systems were as accurate as possible, we chose all-atom models with explicit solvents for constant pH simulations to address the roles of protonation states of glutamate and histidine residues in AtMC9. The program pHbuilder (v1.2.1) was then used to setup constant pH simulations for each structure at their respective pH, using the workflow on the program's GitLab⁴³. The topology for each system was generated using the CHARMM36-mar2019-cphmd forcefield, the TIP3P water model⁴⁴, and a cubic periodic box with the distance flag set to 1.5 nm. For each system, Arg, Lys, Glu, Asp, and His residues were made titratable. The systems were then solvated after which the provided `clean_after_solvate` python script was used to remove poorly placed water molecules. Sodium chloride was then added to each system to neutralize the charge of the system as well as have a salt concentration of 150 mM. Energy minimization was performed using the steepest descent for 50 ns. Next, NVT ensemble relaxation was performed for 10 ps using PME electrostatics and the v-rescale thermostat with a reference temperature of 300 K⁴⁵. Bonds were constrained using the LINCS algorithm. The second round of relaxation lasted 10 ps using the NPT ensemble with the c-rescale barostat and a reference pressure of 1 bar⁴⁶. Finally, production simulations were run in triplicate for each pH using the equilibrated structure for each pH. Different seeds (ld-seed) were used for each replica. The production simulations used the previously mentioned electrostatics, thermostat, and barostat. Simulations were 50 ns in length with a step size of 2 fs and frames written every 50 ps, leading to 1000 frames for each trajectory. The timescale of 50 ns is chosen for the simulations because protonation/deprotonation events were observed for multiple titratable residues.

Molecular dynamics trajectory analysis

The triplicates at each pH were concatenated into a single trajectory and clusters of each concatenation (150 ns total time) were produced using the gmx cluster tool. For each pH every 30 frames were skipped, and the default clustering method (single linkage) was used to cluster frames based on the protein sidechain atoms, leading to 100 clustered frames. A cutoff distance of 0.2 nm was used for pH 4.2 and pH 5.5, whereas a cutoff distance of 0.22 nm was used for pH 7.5. At pH 4.2, twenty-nine clusters were produced with an average RMSD of 0.27 nm where the most populated cluster contained 29 frames (29% of the simulation). At pH 5.5, eighteen clusters were produced with an average RMSD of 0.28 nm where the most populated cluster contained 24 frames (24% of the simulation). At pH 7.5, three clusters were produced with an average RMSD of 0.28 nm, where the most populated cluster contained 88 frames (88% of the simulation). To produce a representative structure at each pH, the structure from the middle frame of the most populated cluster was used. For the creation of figures, hydrogens were either kept or removed for residues of interest, depending on average λ values (Supplementary Table 2). The supplemental movie was created using the third replica of each pH (Supplementary Movies 1 and 2). All trajectories were aligned to the first frame of the pH 7.5 trajectory using VMD (v1.9.3)⁴⁷. A smoothing size of 5 was used for visualization.

In vitro cleavage assays of a fluorogenic substrate

To quantify the activity of AtMC9 or its mutants in this study, we adapted the previously established enzyme assay protocol²⁰, with

several modifications. Specifically, 0.1 μ M of purified AtMC9 or each of its mutants were incubated in a 100 μ L reaction buffer with varying pHs from pH 3.6 to 9.6, 150 mM NaCl, and 10 μ M of the fluorogenic substrate Boc-Gly-Arg-Arg-MCA (Boc-GRR-MCA; Peptides International, KY) on 96-well flat bottom black plates (Greiner Bio-One, Monroe, NC) at 30 °C. The released 7-amino-4-methyl-coumarin (AMC) was continuously monitored for 60 min with 1-min intervals using a microtiter plate reader (Synergy HT, BIO-TEK, Santa Clara, CA) at 360 nm excitation wavelength and 460 nm emission wavelength. The relative fluorescence units in each well were measured after the AMC was released.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Atomic coordinates and structure factor files have been deposited in the RCSB Protein Data Bank (PDB) under the accession code [9N3F](#) (C147G mutant at pH7.5), [9N3E](#) (C147G mutant at pH 5.5), and [9N3D](#) (C147G mutant at pH 4.2). PDB codes of previously published structures used in this work are [8A53](#) (AtMC9 C147A at pH 6.3) and [6W8R](#) (AtMC4 C139A). Simulation input and output coordinates are provided as Supplementary Data 2. Source data are provided with this paper.

References

- Lam, E. Controlled cell death, plant survival and development. *Nat. Rev. Mol. Cell Biol.* **5**, 305–315 (2004).
- Piszczek, E. & Gutman, W. Caspase-like proteases and their role in programmed cell death in plants. *Acta Physiol. Plant* **29**, 391–398 (2007).
- Sundström, J. F. et al. Tudor staphylococcal nuclease is an evolutionarily conserved component of the programmed cell death degradome. *Nat. Cell Biol.* **11**, 1347–1354 (2009).
- Bozhkov, P. V. et al. Cysteine protease mcll-Pa executes programmed cell death during plant embryogenesis. *Proc. Natl. Acad. Sci. USA* **102**, 14463–14468 (2005).
- Vercammen, D. et al. Type II metacaspases Atmc4 and Atmc9 of *Arabidopsis thaliana* cleave substrates after arginine and lysine*. *J. Biol. Chem.* **279**, 45329–45336 (2004).
- Tsiatsiani, L. et al. Metacaspases. *Cell Death Differ.* **18**, 1279–1288 (2011).
- Choi, C. J. & Berges, J. A. New types of metacaspases in phytoplankton reveal diverse origins of cell death proteases. *Cell Death Dis.* **4**, e490 (2013).
- Klemenčič, M. & Funk, C. Structural and functional diversity of caspase homologues in non-metazoan organisms. *Protoplasma* **255**, 387–397 (2018).
- Liu, H., Zhu, P., Zhang, Q., Lam, E. & Liu, Q. Plant metacaspase: a case study of microcrystal structure determination and analysis. in *Methods in Enzymology* <https://doi.org/10.1016/bs.mie.2022.07.026> (Elsevier, 2022).
- Klemenčič, M. & Funk, C. Type III metacaspases: calcium-dependent activity proposes new function for the p10 domain. *N. Phytol.* **218**, 1179–1191 (2018).
- Bollhöner, B. et al. Post mortem function of AtMC9 in xylem vessel elements. *N. Phytol.* **200**, 498–510 (2013).
- Minina, E. A. et al. Classification and nomenclature of metacaspases and paracaspases: no more confusion with caspases. *Mol. Cell* **77**, 927–929 (2020).
- Wrzaczek, M. et al. GRIM REAPER peptide binds to receptor kinase PRK5 to trigger cell death in *Arabidopsis*. *EMBO J.* **34**, 55–66 (2015).
- Hander, T. et al. Damage on plants activates Ca²⁺-dependent metacaspases for release of immunomodulatory peptides. *Science* **363**, eaar7486 (2019).

15. Shen, W., Liu, J. & Li, J.-F. Type-II metacaspases mediate the processing of plant elicitor peptides in Arabidopsis. *Mol. Plant* **12**, 1524–1533 (2019).
16. Bartels, S. et al. The family of Peps and their precursors in Arabidopsis: differential expression and localization but similar induction of pattern-triggered immune responses. *J. Exp. Bot.* **64**, 5309–5321 (2013).
17. Watanabe, N. & Lam, E. Calcium-dependent activation and autolysis of Arabidopsis metacaspase 2d. *J. Biol. Chem.* **286**, 10027–10040 (2011).
18. He, R. et al. Metacaspase-8 modulates programmed cell death induced by ultraviolet light and H₂O₂ in Arabidopsis. *J. Biol. Chem.* **283**, 774–783 (2008).
19. Zhu, P. et al. Structural basis for Ca²⁺-dependent activation of a plant metacaspase. *Nat. Commun.* **11**, 2249 (2020).
20. Fortin, J. & Lam, E. Domain swap between two type-II metacaspases defines key elements for their biochemical properties. *Plant J.* **96**, 921–936 (2018).
21. Tsiatsiani, L. et al. The Arabidopsis METACASPASE9 degradome. *Plant Cell* **25**, 2831–2847 (2013).
22. Stael, S. et al. Structure-function study of a Ca²⁺-independent metacaspase involved in lateral root emergence. *Proc. Natl. Acad. Sci. USA* **120**, e2303480120 (2023).
23. Berendsen, H. J. C., van der Spoel, D. & van Drunen, R. GROMACS: a message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **91**, 43–56 (1995).
24. Lam, E. & Zhang, Y. Regulating the reapers: activating metacaspases for programmed cell death. *Trends Plant Sci.* **17**, 487–494 (2012).
25. Basak, S., Paul, D., Das, R., Dastidar, S. G. & Kundu, P. A novel acidic pH-dependent metacaspase governs defense-response against pathogens in tomato. *Plant Physiol. Biochem.* **213**, 108850 (2024).
26. Fujinaga, M., Chernaia, M. M., Tarasova, N. I., Mosimann, S. C. & James, M. N. Crystal structure of human pepsin and its complex with pepstatin. *Protein Sci.* **4**, 960–972 (1995).
27. Olson, O. C. & Joyce, J. A. Cysteine cathepsin proteases: regulators of cancer progression and therapeutic response. *Nat. Rev. Cancer* **15**, 712–729 (2015).
28. Janzik, I., Macheroux, P., Amrhein, N. & Schaller, A. LeSBT1, a subtilase from tomato plants. Overexpression in insect cells, purification, and characterization. *J. Biol. Chem.* **275**, 5193–5199 (2000).
29. Lowther, J. et al. The importance of pH in regulating the function of the *Fasciola hepatica* cathepsin L1 cysteine protease. *PLoS Negl. Trop. Dis.* **3**, e369 (2009).
30. Fernández-Fernández, Á. D., Stael, S. & Van Breusegem, F. Mechanisms controlling plant proteases and their substrates. *Cell Death Differ.* <https://doi.org/10.1038/s41418-023-01120-5> (2023).
31. Bosch, M. & Franklin-Tong, V. Regulating programmed cell death in plant cells: intracellular acidification plays a pivotal role together with calcium signaling. *Plant Cell* **36**, 4692–4702 (2024).
32. Winter, G. et al. DIALS: implementation and evaluation of a new integration package. *Acta Crystallogr. D Struct. Biol.* **74**, 85–97 (2018).
33. Winn, M. D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242 (2011).
34. Takemaru, L. et al. PyMDA: microcrystal data assembly using Python. *J. Appl. Crystallogr.* **53**, 277–281 (2020).
35. Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. D. Struct. Biol.* **75**, 861–877 (2019).
36. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
37. Williams, C. J. et al. MolProbity: more and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315 (2018).
38. DeLano, W. L. & Others Pymol: an open-source molecular graphics tool. *CCP4 Newsl. Protein Crystallogr.* **40**, 82–92 (2002).
39. Olsson, M. H. M., Søndergaard, C. R., Rostkowski, M. & Jensen, J. H. PROPKA3: consistent treatment of internal and surface residues in empirical pKa predictions. *J. Chem. Theory Comput.* **7**, 525–537 (2011).
40. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
41. Van Der Spoel, D. et al. GROMACS: fast, flexible, and free. *J. Comput. Chem.* **26**, 1701–1718 (2005).
42. Aho, N. et al. Scalable constant pH molecular dynamics in GRO-MACS. *J. Chem. Theory Comput.* **18**, 6148–6160 (2022).
43. phbuilder. *GitLab* <https://gitlab.com/gromacs-constantph/phbuilder>.
44. Mark, P. & Nilsson, L. Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *J. Phys. Chem. A* **105**, 9954–9960 (2001).
45. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101 (2007).
46. Bernetti, M. & Bussi, G. Pressure control using stochastic cell rescaling. *J. Chem. Phys.* **153**, 114107 (2020).
47. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 27–28 (1996).

Acknowledgements

We thank the staff at the NSLS-II beamlines FMX and AMX for their assistance in data collection. This work was supported by NSF grant IOS-2052997 (E.L. and Q.L.) and the U.S. DOE Office of Biological and Environmental Research (KP1601011) (Q.L.). The work used National Synchrotron Light Source II (NSLS-II), which is supported in part by the U.S. DOE Office of Basic Energy Sciences under contract number DE-SC0012704. Beamlines FMX and AMX are part of the Center for Bio-Molecular Structure (CBMS), which is primarily supported by the National Institutes of Health, National Institute of General Medical Sciences (NIGMS) through a Center Core P30 Grant (P30GM133893), and by the DOE Office of Biological and Environmental Research (KP1607011).

Author contributions

Q.L. and E.L. designed the study and experiments. H.L., M.H., Z.P., and Q.Z. performed the experiments. H.L., M.H., Z.P., E.L., and Q.L. analyzed the data. H.L., E.L., and Q.L. wrote the manuscript with help from other coauthors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-60253-y>.

Correspondence and requests for materials should be addressed to Eric Lam or Qun Liu.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025