Crystal structure of plant legumain reveals a unique two-chain state with pH-dependent activity regulation

Florian B Zauner, Elfriede Dall, Christof Regl, Luigi Grassi, Chrisitian G Huber, Chiara Cabrele, Hans Brandstetter

Plant Cell. Advance Publication February 16, 2018; doi:10.1105/tpc.17.00963

Corresponding author: Hans Brandstetter (hans.brandstetter@sbg.ac.at).

Review timeline:

TPC2017-00264-RA	Submission received:	Apr. 2, 2017
	1st Decision:	May 29, 2017 *manuscript declined*
TPC2017-00582-RA	Submission received:	Jul. 21, 2017
	1st Decision:	Aug. 23, 2017 *manuscript declined*
TPC2017-00963-RA	Submission received:	Dec. 13, 2017
	1st Decision:	Jan. 30, 2017 *acceptance pending, sent to sci editor*
	Final acceptance:	Feb. 13, 2018
	Advance publication:	Feb. 16, 2018

REPORT: (The report shows the major requests for revision and author responses. Minor comments for revision and miscellaneous correspondence are not included. The original format may not be reflected in this compilation, but the reviewer comments and author responses are not edited, except to correct minor typographical or spelling errors that could be a source of ambiguity.)

TPC2017-00264-RA 1st Editorial decision – *declined* **May 29, 2017**

Thank you for choosing to send your manuscript entitled "Crystal structure of plant legumain reveals a unique two-chain state with pH-dependent activity regulation" for consideration at The Plant Cell. Your submission has been evaluated by members of the editorial board as well as expert reviewers in your field, and we regret to inform you that we are not able to recommend publication of this manuscript. We have not made this decision lightly. We have had input from multiple scientists, and have solicited post-review comments as well. Our present policy is to offer streamlined decisions and to not advise on the direction of the work by requesting extensive modifications or substantial additional experiments.

As you will see, both reviewers consider that the structure and ligation activity of this plant legumain are interesting. However their reviews also converge on the fact that a major weakness is that the biology of this finding has not been explored. As such, the manuscript remains too specialized. Note also that a structure for *Oldenlandia affinis* legumain/AEP/VPE has already been solved by Yang et al. and this would need to be cited. Moreover, there are also issues regarding the pH-dependent dimerization of the enzyme and other technical concerns.

Therefore we cannot recommend resubmission of this work. The reviewers point out a number of areas in which the work could be strengthened, which may be helpful to you as you continue your project or revise your manuscript for submission elsewhere. If you decide to complete the story and resubmit to The Plant Cell, it will be evaluated as a new submission subject to full assessment by the editorial board, including pre-review and editor selection, and if sent for external review, a new set of reviewers is likely to be chosen.

-- Reviewer comments:

[Reviewer comments shown below along with author responses]

TPC2017-00582-RA Submission received **July 21, 2017**

Reviewer comments on previously declined manuscript and **author responses:**

Reviewer #1:

The authors have solved the structure for a protease (asparaginyl endopeptidase/AEP, aka vacuolar processing enzyme/VPE, aka legumain and one AEP has been called Butelase 1). Specifically they have solved the crystal structure of AEP3 or gamma-VPE encoded by At4g32940. They observed a ligation activity for it at neutral pH.

What was interesting to me was the observation that an AEP not shown or thought previously to ligate can do so under neutral pH. The structure is clearly a good one and the description of it seems solid.

Point 1. It does not cite some recent literature, in particular the first plant AEP structure published in JACS. The 'two chain' crystal structure of AtLEGγ is highly similar to that of the recently published *Oldenlandia affinis* AEP structure described by Yang et al. in JACS. They solved the structure for immature OaAEP1 which was purified at neutral pH and their data lacked electron density for the linker between the cap and core domain. It's not clear if it was removed from OaAEP1 during the crystallisation conditions or remains present, but is too flexible to give electron density. Moreover, the OaAEP1 structure also exhibited a dimeric structure that is stabilised by intermolecular interactions between the cap domains. The OaAEP1 structure is not discussed or referenced in this article and it should be.

RESPONSE: We have now cited the Yang et al. paper.

As becomes clear from our manuscript, the dimerization will also be relevant for the activation and activity regulation of OaAEP1. Apparently, Yang et al. did not recognize nor study the relevance of dimerization, underscoring the importance of our findings. We are thankful to include the Yang paper as it further underlines the biological relevance of our findings which quite apparently are not restricted to Arabidopsis, but apply to many different plants.

Point 2. The introduction does not prepare you for the work that follows. There is little about AEP biochemical prior art. The intro does not explain all the terms about to be used such as AP, LSAM etc. The paper also dives very quickly into structure with no rationale for acquiring the structure of this particular protein or starting with any specific question to answer?

RESPONSE: We will improve the introduction as suggested by the reviewer.

Point 3. The connection of this protein to programmed cell death is dubious. It is worth noting that since the 2004 Science paper and a review in 2013, there is very little on the subject of AEPs involved in senescence. The prevailing view is a role for AEPs in senescence is at best controversial and worst, incorrect. A recent study by Pružinská et al. (BMC Plant Biol. 2017 Jan 6;17(1):4. doi: 10.1186/s12870-016-0955-5) was entitled "Major Cys protease activities are not essential for senescence in individually darkened Arabidopsis leaves". It said that despite being "amongst the most highly expressed proteases during leaf senescence in Arabidopsis", by using a quadruple aep null as well as a line overexpressing AEP (the same one this structure was done for - AEP3 or gamma-VPE) in dark-induced senescence experiments the authors found no change from wild type senescence. They also noted that despite increased VPE transcript levels, active VPE (using activity based probes for AEP) decreased in senescing leaves.

RESPONSE: We agree that this connection is controversially discussed and we provided a more complete overview on this topic, which is very complex indeed. Also the recent Pružinská et al. paper comes with a series of limitations. Given the controversy in this important question, it is all the more important to point out the biological conclusions which we can draw from our structural and functional investigations, because they represent testable hypotheses and should contribute to clarify the role of AEPs in plant cell biology, including senescence.

Point 4. The substrates used to show the ligation reaction are an unusual choice. They are not actual substrates of any AEP. They are partial linear pentapeptides that come from the ends of a globular, S-S knotted peptide called kalata B1 that's nearly 30 amino acids. The kalata B1 trailing sequence His-Val is similarly not native to the kalata B1 sequence or most cyclotide precursors. It means the ligation reaction is curious, but not a biologically relevant one?

RESPONSE: The substrates were carefully selected because they unambiguously show the ligase activity of AtLEGγ in a situation that is not simply proximity-driven. It is undoubtedly clear that the intermolecular ligation of two independent peptides is kinetically more challenging than the intramolecular ligation of an S-S knotted peptide like kalata B1, where ligation is strongly proximity-driven.

The identification of physiologically relevant substrates is among the most challenging tasks in enzymology, and it is naïve to assume all plant legumain substrates are discovered already. By contrast, by designing a substrate chimera of kalata B1 and the butelase1-optimized tail sequence, we demonstrate that a much broader variety of substrates need to be considered.

Point 5. The mentions of significant conservation based on a handful of sequences (n = 9) (Page 4 Line 94) and mention of something that "independently evolved" (Page 16 Line 260) are similarly based on a very small sample size.

RESPONSE: We used a more careful wording in these instances.

Reviewer #2:

Zauner et al. present the first crystal structure and mechanistic insights of a plant legumain, a vacuolar cysteine protease with potential roles in programmed cell death. Work on legumain is of particular interest for general plant biology as well as for potential future applications due to the special features of legumains which can selectively cleave proteins and ligate peptide fragments. Activity of the legumain is pH-regulated like with most of the cysteine proteases which are strongly pH-dependent which is their way to shift from active to inactive state e.g. for self-protection.
The authors use very clear and structured visualization of data and concepts in Fig. 1 and 4 (structures) and Fig. 7 and Supp. Fig. S7 (models). In my impression, the results are correctly designed and explained. The authors have performed a lot of biochemical work that is well written and explained but often the message gets diluted.

The manuscript seems to be of interest to a specialist community and the experiments look well performed and interpreted.

However, the paper contains very detailed descriptions of biochemical, especially enzymological and structural work which represents one, albeit an important, aspect of legumains. The "biology" of legumains is referred to at some instances but falls short in general. Examples are their role in programmed cell death, use in biotech applications to form cyclic peptides and other potential uses of peptide ligase activities. The authors state that they "discuss the relevance of our findings in the context of plant programmed cell death and peptide ligation" but this aspect falls short. At the same time, I am not fully convinced that "This study will not only broaden the understanding of plant (patho)physiology and crop design but also enables the design of efficient peptide ligases and will bring such a technology closer to broad application". In my opinion, the discussion remains very vague and "practical" examples e.g. shown in Fig. 6 are a bit thin (or relevance not sufficiently clearly highlighted). It might improve this outspoken mechanistic manuscript to reduce complexity and refocus on the whole picture, the plant cell.

Point 1. They tried to show a pH-dependent dimerization of the enzyme (monomer at neutral pH and dimer at acidic pH). In my opinion, to prove that this is a two-state structure that is pH-controlled, intermediate pH values (such as 5, 5.5, and 6) have to be tested as well.

RESPONSE: We have carried out experiments for these intermediate pH values (figure 3).

Point 2. I think that this is because they claim that "Dimerization and electrostatic anchoring of the α6-helix suppress peptidase activity at neutral pH" page 13, line 220 and lines 227-229 "Conversely, our structural model further suggests an opening of the active site at acidic pH and a closure at the more neutral pH with the α6-helix as central lock/key. To test this model, we determined the pH-dependent proteolytic activities of the two-chain and protease-only forms."

This and their other structural discussions (about charge) would expect a higher proteolytic activity at acidic pH (like pH 4 and 4.5), a moderate one at pH 5.5 (here we'd expect a transition or both states to exist equally) and a weaker one at near neutral pH (like pH 6.5-7, here dimer dominates). However, Figure 5a shows that we have a maximum at pH 5.5 and weaker activity at pH 4 and 7 for both two-chain and protease only. Well, they explain that there's an abrupt change at pH 5.5 for two-chain, but it is not, at pH 5.5 is still very high activity and goes sudden at pH 6, this is true only for Tm values. Otherwise, how do they explain this low activity at pH 4? The claim that Cys nucleophilicity cannot stand that much, this is quite the same in that pH range and TC stability cannot explain 75% activity fold.

RESPONSE: Reviewer 2 points at the important interference of three key factors which all combine in the resulting proteolytic activity. We have schematically summarized the major contributions in figure 5, 7a.

(1) The <u>nucleophilicity of the catalytic cysteine</u> increases with pH, which is also indicated in fig. 5.

(2) The <u>conformational stability</u> of TC and protease are pH-dependent, referred to as “TC stability” (cyan) and “CD stability” (black) in fig 5b.

(3) The <u>accessibility of the active site</u> is controlled by the α6 helix; this relation is represented in figure 5, 7a by the lines “TC active site accessibility” in cyan (TC: Two-Chain) and “CD active site accessibility” in black (CD: Catalytic Domain).

Although Reviewer 2 appreciates the relevance of these three factors, (s)he questions the resulting activity profile (fig. 5a).

We would like to point out that it is easier to understand the activity of the protease-only profile because the first factor (accessibility of the active site) does not change with pH (this is more complicated in the two-chain form).

The protease-only profile (grey) is overall bi-phasic, increasing from pH 3.5 to 5 and decreasing from pH 6 to 7, with an intermediate regime from pH 5 to 6. The explanation is given by the superposition of the effects (1) and (2). Such bi-phasic activity profiles have been reported for several lysosomal proteases, simply because their conformational stability is adapted to the acidic pH of their home organelle. Lysosomal/vacuolar proteases leaking into the cytosol cannot do much harm, because they will be conformationally inactivated by the neutral pH. Similarly, at pH < 4 the enzyme becomes instable (see fig. 5b)

The nucleophilicity of the catalytic cysteine does change with pH considerably; increasing the pH for one unit (from say 4 to 5) increases the likelihood of having a thiolate (S-) by a factor of 10. Of course, the pH will affect the protonation of other structure elements as well.

The activity profile of the two-chain form can be understood on the basis of the protease form with the additional damping effect which results from the α6 helix blocking the active site at pH > 5.5. In summary, the combination of the three key effects can explain the observed activity profile very well.

Point 3. Therefore, point 1) above is even more important since dimerization/activity and pH are correlated in the schematic model of Fig. 7. I think that it needs to be shown that at ~ pH 5.5 both forms exist, so gelfiltration chromatography at more pH intermediate values is needed, not only pH 4 and 7 (Fig 3a).

RESPONSE: We appreciate the Reviewer's suggestion and have added a gel filtration chromatogram at a more intermediate pH, see figure 3. This chromatogram indeed reveals a shifted retention volume, as anticipated by reviewer 2!

TPC2017-00582-RA 1st Editorial decision – *declined*	August 23, 2017

Thank you for choosing to send your manuscript entitled "Crystal structure of plant legumain reveals a unique two-chain state with pH-dependent activity regulation" for consideration at The Plant Cell. Your submission has been evaluated by members of the editorial board as well as expert reviewers in your field, and we regret to inform you that we are not able to recommend publication of this manuscript. We have not made this decision lightly. We have had input from multiple scientists, and have solicited post-review comments as well.

Although all reviewers consider your work as potentially interesting, this second version of the manuscript still appears to require serious improvements regarding its clarity (see Rev2, Rev3 and Rev4 comments). In particular Rev3 and 4, who have expertise in structural biology, provide a long list of points to clarify. Even more problematic is the fact that some statements are wrong or overinterpreted and would need to be corrected or toned down and/or addressed experimentally, such as thermodynamics violations, the electrostatic interactions in controlling the activation mechanism (see Rev3 and Rev4 specific comments) and the effect of dimerization on legumain activity (see Rev4 specific comments).

During the post-review consultation session, we agreed that if you could address the major points raised by the reviewers, we would welcome a resubmission. This would be treated as a new submission, but we would attempt to use the same reviewers.

-- Reviewer comments:

[Reviewer comments shown below along with author responses]

TPC2017-00963-RA Submission received	**July 21, 2017**

Reviewer comments on previously declined manuscript and **author responses:**

Reviewer #2:

I have compared the previous and current versions of the manuscript and studied the responses to the questions. It looks as if the comments of reviewer 1 and 2 have been adequately addressed.
The paper is still very specialized and has a strong focus on biochemical and structural details and has only few links to the "whole picture" as I mentioned in the very end of my previous comment sto the author already. However, I think that, taken into account that not much is known about this specific subset of proteases, the class of plant legumains, it will be of interest to the readership of The Plant Cell.

Reviewer #3:

The authors describe structural and biochemical studies on a plant legumain enzyme, which catalyzes peptide hydrolysis and cyclization reactions in diverse cellular contexts. The crystallographic work appears to be in order and the biochemical studies seem sound. The work reveals some new and interesting enzyme features distinct from mammalian homologues, particularly regarding oligomeric state and the details of autoactivation and its pH dependence, which the authors relate to specific plant processes of storage and activation. On those grounds, the work is highly deserving of publication and could have substantial impact. On the other hand, the paper suffers from a number of serious weaknesses that need to be addressed on topics related to clarity of presentation and structural and biophysical analysis. Those issues are detailed below.

The authors' expertise on the system under study comes through immediately, yet the writing presents numerous gaps -- failures to explain names or abbreviations, statements about features that haven't been introduced yet, omissions of essential points -- that make it frustrating for the uninitiated reader to follow.

Point 1. In broad strokes, the key idea behind the authors' main hypothesis or model - that electrostatic interactions are important in controlling the activation mechanism -- is sensible and supported at least in general aspects by experimental data. But the electrostatics are overinterpreted in terms of what can be reliably concluded about exactly which amino acid interactions and protonation states are key to the mechanism. If it was sufficiently important, specific elements of an atomic level model might have been addressed experimentally by mutagenesis. In the absence of that the authors need to be much more circumspect about their details. The authors state in several places what groups will be protonated at what specific pH values as if the pKa values of particular side chains are not much affected by whether or not they are near counterions, or whether or not they are exposed to solvent (where the dielectric is high) or buried in the protein interior (where the dielectric is lower). Again, mutagenesis experiments would be needed to test such things. The authors mention using various programs for calculations, but such calculations are notoriously difficult to validate. What values (or approximate models) to use for the dielectric in different regions is constantly argued, and this strongly affects calculations of potentials and interaction energies. The authors seem to be putting a lot of faith in these calculations.

RESPONSE: We understand the critical remarks of the reviewer and we acknowledge that the wording was in parts too optimistic, in particular where it left the impression that it was possible to assign accurate pKa values to individual residues within its protein environment – which is not reliably possible. We further agree that multiple factors, such as mentioned by the reviewer, contribute in a complex way to the protonation of an amino acid.

Accordingly, we have corrected our wording in the manuscript where the impression of overinterpretation could have resulted. In one instance we were apparently tempted to emphasize the correlation of our findings with the computationally determined pKa values. We have not done ourselves any favour, because with this emphasis we have veiled the fact that the presented model results from robust electrostatic analysis and reasoning, e.g. (line

178f), "*We could identify two prominent positively charged clusters (Arg, Lys, His) at either side of the α6- helix, which are neutralized by negatively charged glutamate and aspartate patches (Fig. 4 a-c).*"

We now clarify that no black-box electrostatics calculations with specific pKa assignments are required to arrive at the activity regulation model. All electrostatic considerations and conclusions can be deduced by fundamental text book knowledge such as (i) Arg, Lys are positively charged at neutral and acidic pH; (ii) Asp, Glu are negatively charged at neutral pH and become protonated (i.e. neutral) at acidic pH; (iii) His and Cys are also able to undergo protonation in the physiological pH regime.

We emphasize that all text providing quantitative electrostatic calculations solely serves to illustrate the general qualitative considerations, thus avoiding the risk of overinterpretation.

Finally, we have also taken the reviewer's request for mutagenesis data seriously and invested significant effort in producing relevant mutants to test the critical relevance of the charge distribution around the α6 cluster, e.g., charge reversal mutations (E220K), charge neutralization (D356N), or short deletions (Δ354NQR). We have gone the full cycle of cloning, expression, upscaling etc. Unfortunately, despite the month-long efforts we could not get sufficient protein for enzymatic testing. This is disappointing but can be taken as indirect evidence that the identified charged amino acids at and around the α6 helix are indeed critical for the stability of the protein.

We have therefore introduced an E371D mutation mimicking the mammalian situation, which should result in the release of the α6 helix at acidic pH. Additionally and to complement the mutagenesis studies, we have carried out peptide competition experiments where we investigated the effect of the peptide of the α6 helix peptides on enzymatic activity. All these experiments are consistent with our proposed model. Please see our reply to Reviewer 4, point 6 where we summarize the results of these experiments.

Point 2. The idea of a quadrupolar arrangement comes up at least a couple of times; I don't feel that this is a very careful or clarifying way to describe the electrostatic forces at play here.

RESPONSE: We agree and are thankful for this comment. With this term we have not done ourselves any favour, it is confusing, see above. We have corrected this.

Point 3. The helix dipole is included as another contributing factor for the conformational transition. The use of a helix dipole as a simplifying description has, on occasion, provided a useful intuition about protein electrostatics, but it's really not a very accurate way to think about electrostatic interactions. And the magnitude of the interaction might not be physically important; this is not addressed.

RESPONSE: We largely agree. The helix dipole consideration is not necessary, and its effect is likely a minor one. The two dominating factors are the side chain of Arg355 and the neo N-terminus of Gln354, both contributing a full positive charge to the entrance of the α6 helix. Only as a third and subordinate factor the helix dipole comes into play, with an (unspecified, yet minor) partial charge δ+.

Then again, we cannot see any reason to completely black out the helix dipole. Consequently and in accordance with the reviewer's suggestion, we made clear that the physical relevance of the helix dipole partial charge distribution is subordinate and in quantitative terms unclear.

Point 4. I'm not sure whether the authors' idea makes sense that having a few positively charged side chains in a helix would destabilize it (when nearby negatively charged side chains get neutralized). Many proteins have exposed alpha helices with a lot of charge. I'm not persuaded that the idea of helix destabilizing by charge would hold up to scrutiny based on electrostatics or on surveys of natural protein amino acid sequences. [It wasn't easy to know how many helix charges were being discussed.]

RESPONSE: This comment is another valuable pointer how our wording was misunderstood. We did not want to indicate that the α6 helix gets destabilized as a secondary structure element due to unbalanced charges. Instead, our point is that upon protonation of the acidic residues (Asp, Glu; cf. Fig. 4) the anchoring of the helix gets loosened, thereby regulating the access to the active site. With other words, not the secondary structure element, but rather its relative positioning gets more flexible (destabilized).

We have changed the wording to avoid this misunderstanding and explicitly stated that the secondary structure element as such will likely remain unaffected by the pH change.

Point 5. In a few places, especially lines 284 and figure 5b, the authors seem to be committing outright thermodynamic heresy. The authors explain that certain features, like alpha helix 6, 'allows ligation to happen but preventing peptide hydrolysis' (line 284). Obviously nothing about the enzyme, in fact nothing besides a change in the standard Gibbs free energy values, can have an effect on the relative tendency toward ligation vs hydrolysis of any given substrate molecule. That would violate the principle of microscopic reversibility. The authors need to purge the manuscript of this thinking, including in Fig. 5d.

RESPONSE: By the use of the wording, the Reviewer leaves no doubt about the seriousness of this particular statement and concern. We are thankful for the comment; it is of utmost importance to us, too. It proved most valuable to improve the clarity and correctness of important aspects of our manuscript.

Firstly, we absolutely and without restrictions agree with the Reviewer's statement that enzymes do not change reaction equilibria (thermodynamics) – unless a second reaction is coupled which would need to be considered in the net balance. Enzymes are powerful at changing the kinetics of reactions, combined with an often amazing substrate selectivity and specificity. We appreciate the reviewer's concern that these fundamentals of enzymology may be unclear or get confused.

Taking this perspective, we can see that we used wording such as preferred ligation (or proteolysis) regime (Fig. 5d), which is sloppy and when taken literally incorrect. If nothing else, we should speak of "apparent ligation regime".

Otherwise readers could be left with the impression as if substrates would be ligated by legumain which would and could never be ligated alone; or substrates which are prone to undergo hydrolysis will not do so in the presence of the enzyme. This is clearly NOT the case.

What really happens when legumain switches from an "apparent proteolysis regime" to an "apparent ligation regime" with increasing pH is a change in legumain's substrate preference. At low pH legumain preferentially turns over substrates with a tendency for hydrolysis whereas this preference changes towards ligation-prone substrate at near neutral pH.

However, important key questions remain unanswered which should be addressed by structural enzymology, including: How is the substrate selection and the rate acceleration accomplished in mechanistic terms? Or: Why

have only very few enzymes a preference for substrates which exhibit the inherent tendency for peptide bond formation? Here, structural elements like the α6 helix play essential roles as they can provide a rationale how the enzyme assists the (ligation) substrates to meet more frequently and more productive. Answering these questions can provide a rationale to why certain, selected peptides can be efficiently ligated. But again: It's about kinetics, substrate selectivity and specificity, but not about equilibria (thermodynamics).

Finally, we wish to emphasize that one and the same substrates may change its tendency for cleavage vs re-ligation with pH. Classic examples are the Laskowski-mechanism protein inhibitors of proteases, i.e. inhibitors that bind like a substrate to the enzyme's active site. Ozawa & Laskowski have reported an analogous pH dependent tendency of these inhibitors for hydrolysis versus religation in their seminal 1966 paper (JBC 247:3955-3961). They explain the preference for ligation at near neutral pH as the optimum regime to find a nucleophilic amino terminus (-NH2) and a partially protonated carboxy terminus (-COOH), enabling a condensation reaction, whereas hydrolysis was found at pH 3.5 to 4.5. With other words: Ozawa & Laskowski deduced the pH regimes in figure 5d as a substrate inherent feature already 50 years ago!

There is no magic and no heresy with figure 5d. We only report of legumain as an enzyme which is able to exploit substrate-inherent properties.

We have emphasized the substrate-selection aspects and kinetic interpretation of the catalysed reaction in the legend to Fig. 5d and also put all main text through the "thermodynamics purgatory".

Reviewer #4:

The manuscript by Florian Zauner et al. describes the crystal structure of the Arabidopsis legumain gamma-isoform. The structure reveals a previously uncharacterized dimeric form of the enzyme in an apparent resting state. Because a significant portion of the dimer interface is formed by the so-called LSAM and not by the peptidase domain, the authors further investigate dimerization and propose a regulatory function of LSAM. Exposure to low pH, as in vivo in the vacuole, would destabilize the interface, thereby activating the enzyme.

The overall conformation of the Arabidopsis legumain is consistent with previously described structures, yet the regulation of enzyme activity through dimerization is interesting and worth communicating.

Point 1. I think the text is quite hard to follow at times. For example, the authors should better explain what is referred to as 'two-chain model', 'peptidase domain only', and 'zymogenic protein'. In particular, the difference between the two-chain and zymogenic state should be explained.

RESPONSE: We appreciate these comments, which in part were also raised by Reviewer 3. With respect to the term zymogenic state: The zymogen is the enzymatically inactive precursor of an enzyme (https://en.wikipedia.org/wiki/Zymogen). The zymogen (or proenzyme) form is the full-length form of legumain. When speaking of zymogen form, we wish to emphasize the enzymatic latency, contrasting the situation of cleaved and activated forms, e.g., the two-chain form.

Point 2. Crystallization: a) It is not clear to me what construct has been crystallized, is it the two-chain or zymogenic state? b) Also, the authors should state clearly the content of the asymmetric unit. Is it a legumain monomer or dimer? c) It would also be informative to show densities for the various disulfide bonds that are mentioned in the text to be important for dimer stability. d) Further, how different is the MR search model from the determined structure and what regions were manually built?

RESPONSE: a) The two-chain form has been crystallized. (Cleavage of the single-chain zymogen was accomplished during preparation and controlled by SDS-PAGE.)

b) The asymmetric unit contained two dimers, i.e. four independent protein molecules.

c) A stereo representation of the relevant electron density, including the disulphide bond Cys252-Cys266, is displayed in supplemental figure 2b.

d) The initial model could be generated by molecular replacement using human prolegumain (PDB ID entry: 4FGU) as a search model. The Arabidopsis AtLEGγ shares with human LEG (the search model) 47 % sequence identity, whereby the two proteins differ most at the C- terminal activation peptide (AP) and LSAM (legumain stabilization and activity modulation) domain with only 14.4% identity. These segments consequently required more manual rebuilding, in particular in loop regions. We added the missing information in the Material and Methods section.

Point 3. Glycosylation: The authors stress that glycosylation is important for the stability/folding of the protein, yet the text does not address this at all. Is there any density visible for the sugars? I think this should be stated early on when introducing the structure.

RESPONSE: Mass spectrometric analysis (Supplementary Fig. 3) revealed a sugar moiety consisting of an N-glycan

core, very likely on its only N-glycosylation site at Asn336. Due to its location at the relatively disordered AP, only an undefined electron density blob was found, which prevented unambiguous modelling of the sugar in the coordinates.

As a self-critical side remark: As crystallographers we are sometimes biased to think important features must be well-defined in the electron density, or conversely, a flexible feature with blurred electron density is functionally less important. This notion is obviously not generally valid. We added the information to the main text.

Point 4. Dimerization: The effect of dimerization on legumain activity is probably the most interesting observation. Thus, I think the dimer interface could be explored/described in more detail. a) First, how big is the interface and are there any conserved residues (other than the positive and negative charges)? Perhaps a surface representation of one protomer colored by sequence conservation could stress this point. b) Second, what is the evidence that the crystallized dimer also exists in solution? Or could it be that multiple dimers exists that all stabilize the protein at neutral pH? Perhaps a disulfide cross-linking approach could verify the existence of the observed dimer in solution? c) In addition, the authors briefly mention the plant-specific extension of the C341 loop but its contribution to or function in dimerization is not described.

RESPONSE: a) The dimer interface is large with 2891 Å2 buried surface area between the two monomers within the dimer (calculated with the PISA server application, http://www.ebi.ac.uk/msd- srv/prot_int/cgi-bin/piserver). We added this information to the main text.

- Yes, there are conserved residues, in particular a hydrophobic core in the centre of the four helix bundle formed by α6, α7, α6', α7'. The detailed interaction interface is provided in Supplementary Fig. 7b, with the relevant (non-charged) residues being labelled. The figure (S7b) together with the sequence alignment (suppl. Fig. S1) further emphasize that the hydrophobic core of the four helix bundle is conserved within certain isoforms. The sequence comparison, therefore, allows for a prediction of the dimerization tendency of a given isoform.

b) We have an array of (I) structural and (II) functional data all supporting the existence of a unique dimer interface.

(I) With respect to the structural data:

- We have crystallized AtLEGγ under a second, alternative crystallization condition. Albeit less ordered, the second crystal form reveals the same dimer interface, i.e. the conserved four- helix bundle α6, α7, α6', α7'.

- Furthermore and as discussed in the manuscript, a recent structure of dimeric OaAEP1 from the plant Oldenlandia affinis was published; it displayed the same four-helix-bundle based dimerization mode as AtLEGγ (Yang et al. 2017), as we discuss in our manuscript. Yang and colleagues also found a dimer in solution. However, the authors did not mention ("overlooked") any regulatory role of dimerization, which however is strongly suggested by our manuscript.

- Additionally, the crystal structure of mouse pro-legumain revealed an analogous dimer which was also mediated by the α6, α7, α6', α7' four helix bundle, which was also confirmed in solution (Zhao et al. 2014).

- Finally, also the crystal lattice of human prolegumain revealed a dimeric arrangement mediated by the α6, α7, α6', α7' four helix bundle (Dall et al. 2013).

(II) These structural data are supported by strong functional evidence in the case of plant legumains, which appears to be a critical difference to mammals, where the α6 helix is proteolytically released upon activation. No functional relevance has been reported or suggested for mammalian legumains. Also the results of the additionally conducted α6 helix peptide competition experiments confirm the functional relevance of the four-helix-bundle mediated dimer) below. Taken together, there is no doubt about the presence and relevance of the crystallized AtLEGγ dimer in solution.

c) The reviewer possibly refers to line 106-109, where we wrote: "The dimer was further stabilized by a plant-specific disulphide-linked (Cys252-Cys266) proline rich insertion (Fig. 1a) that extends the c341 specificity loop and resembles a typical cyclic Protein Recognition Motif (cPRM)".

This statement relates structurally to the c341 loop's contribution to the dimer interface; functionally, the redox sensitivity of AtLEGγ's auto-activation and enzymatic activity. We clarified the role of the cPRM in the main text.

Point 5. Figure 2a: Shown is a decrease in enzymatic activity at higher protein concentrations, indicating that there is a monomer/dimer equilibrium with the dimeric form being inactive. Since the gel filtration profile shown in Fig. 3a shows a monomer at pH 4.0, I am wondering whether dimerization at a higher protein concentration at pH 4.0 can be observed (gel filtration, light scattering, etc.). Also, the error bars were calculated from duplicate measurements, which is insufficient.

RESPONSE: This is a very interesting question. The reviewer correctly interprets the reported dimer- monomer transition, which is reflected for example by pH dependent gel filtration runs and pH-dependent activity assays, but

obviously also noted the sharp transition with pH, as evidenced by figure 3a. While the protein migrates significantly as a dimer at pH 4.2, it is completely monomeric at pH 4.0, at all protein concentrations tested (10 mg/ml). The dimer- monomer transition thus resembles a phase transition-like behaviour.

Duplicate measurements are indeed insufficient, and in fact the graph shown in figure 2a represented triplicate measurements, but the legend was unfortunately not accordingly updated. The data were confirmed by additional measurements. Figure was updated.

TPC2017-00963-RAR1 2nd Editorial decision – *acceptance pending* **January 30, 2018**

We are pleased to inform you that your paper entitled "Crystal structure of plant legumain reveals a unique two-chain state with pH-dependent activity regulation" has been accepted for publication in The Plant Cell, pending a final minor editorial review by journal staff. At this stage, your manuscript will be evaluated by a Science Editor with respect to scientific content presentation, compliance with journal policies, and presentation for a broad readership.

Final acceptance from Science Editor **February 13, 2018**