Zymogenic latency in an 250-million-year-old astacin metallopeptidase

Tibisay Guevara,a,‡ Arturo Rodriguez-Banqueri,a,‡ Walter Stöcker, b Christoph Becker-Paulyc and F. Xavier Gomis-Rutha*

aProteolysis Laboratory, Department of Structural Biology, Molecular Biology Institute of Barcelona (IBMB), Higher Scientific Research Council (CSIC), Barcelona Science Park, Baldirí Reixac 15–21, Helix Building, 08028 Barcelona, Catalonia, Spain,
‡Institut für Molekulare Physiologie (IMP), Johannes-Gutenberg Universität Mainz (JGU), Johann-Joachim-Becher-Weg 7, 55128 Mainz, Germany, and
Biochemical Institute, Christian-Albrechts-Universität zu Kiel, Otto-Hahn-Platz 9, 24118 Kiel, Germany. *Correspondence e-mail: xgrcri@ibmb.csic.es

The horseshoe crab Limulus polyphemus is one of few extant Limulus species, which date back to 250 million years ago under the conservation of a common Bauplan documented by fossil records. It possesses the only proteolytic blood-coagulation and innate immunity system outside vertebrates and is a model organism for the study of the evolution and function of peptidases. The astacins are a family of metallopeptidases that share a central 200-residue catalytic domain (CD), which is found in >1000 species across holozoans and, sporadically, bacteria. Here, the zymogen of an astacin from L. polyphemus was crystallized and its structure was solved. A 34-residue, mostly unstructured pro-peptide (PP) traverses, and thus blocks, the active-site cleft of the CD in the opposite direction to a substrate. A central ‘PP motif’ (F35-E-G-D-I39) adopts a loop structure which positions Asp38 to bind the catalytic metal, replacing the solvent molecule required for catalysis in the mature enzyme according to an ‘aspartate-switch’ mechanism. Maturation cleavage of the PP liberates the cleft and causes the rearrangement of an ‘activation segment’. Moreover, the mature N-terminus is repositioned to penetrate the CD moiety and is anchored to a buried ‘family-specific’ glutamate. Overall, this mechanism of latency is reminiscent of that of the other three astacins with known zymogenic and mature structures, namely crayfish astacin, human meprin β and bacterial myroilysin, but each shows specific structural characteristics. Remarkably, myroilysin lacks the PP motif and employs a cysteine instead of the aspartate to block the catalytic metal.

1. Introduction

The Atlantic horseshoe crab Limulus polyphemus (Linnaeus, 1758) is a unique marine merostomatus decapod that is endemic to North America (Shuster, 1982; Walls et al., 2002). It is one of four closely related extant species of horseshoe crabs together with Tachypleus tridentatus, Tachypleus gigas and Carcinoscorpius rotundicaudia, which are found in Asia (Sekiguchi & Shuster, 2009). They are the only survivors of the order Xiphosurida (Bicknell & Pates, 2020) and are the closest living relatives of trilobites (Shuster, 1982). Indeed, Limulus spp. go back to 250 million years ago (Mya) (Bicknell & Pates, 2020) and the Limulidae family has existed since the Carboniferous period (~360 Mya; Bicknell & Pates, 2019, 2020). Xiphosurida, which share a highly conserved horseshoe-crab-like Bauplan as inferred from an exceptionally extensive fossil record (Bicknell & Pates, 2020), date as far back as the Late Ordovician (~445 Mya; Rudkin et al., 2008; Bicknell & Pates, 2020) or Cambrian (~540 Mya; Störmer, 1952). Thus, these animals have survived all five great mass extinctions and are sometimes considered to be ‘living fossils’.
a term introduced by Charles Darwin (Darwin, 1859), or ‘stabilomorphs’ (Kin & Blażejowski, 2014) and are an example of ‘evolutionary stasis’ (Rudkin et al., 2008).

Despite their name, horsehoe crabs are actually not crustaceans but chelicerates that are phylogenetically closer to spiders, ticks and scorpions than to crabs (Lankester, 1881; Ballesteros & Sharma, 2019). *L. polyphemus* has a remarkable estimated life expectancy of up to 20 years (Walls et al., 2002) and is frequently used as a laboratory animal model to study its compound eyes, its simple nervous system and marine invertebrate embryology in general (Smith, 2022). Moreover, it possesses an ancient and primitive proteolytic blood-coagulation and innate immunity system, which is the only one found outside vertebrates (Rowley et al., 1984; Doolittle, 2010; Schmid et al., 2019; Winter et al., 2020; Eleftherianos et al., 2021). Thus, *L. polyphemus* is an important organism for study of the evolution and function of peptidases (Becker-Pauly et al., 2009).

The astacins are a family of zinc-dependent metallopeptidases (MPs; Stöcker et al., 1993; Gomis-Rúth, Trillo-Muyó et al., 2012; Stöcker & Gomis-Rúth, 2013; Bond, 2019) named after the archetypal digestive enzyme astacin from the European freshwater crayfish *Astacus astacus* L., which was first described in 1967 (Pleiderer et al., 1967; Stöcker et al., 1988, 1992; Stöcker & Yiallourou, 2013). Astacins are characterized by a central ~200-residue zinc-dependent catalytic domain (CD), which occurs in >12 000 sequences from >1000 species of identified and putative family members grouped into family PF01400 within the PFAM database (Mistry et al., 2021). Sequences are found consistently with Darwinian vertical descent throughout metazoans and, sporadically, up to the root of holozoans. They are absent from plants and viruses (Semenova & Rudenskaia, 2008) and are found to be scattered across bacteria, which suggests that they are xenologues resulting from horizontal gene transfer from euarkyotes (Koonin et al., 2001; Keeling & Palmer, 2008). The structural characteristics of astacin CDs further place the family within the metzincin clan of MPs (Bode et al., 1993; Stöcker et al., 1995; Gomis-Rúth, Trillo-Muyó et al., 2012; Cerdá-Costa & Gomis-Rúth, 2014) and family M12A of the MEROPS database (Rawlings & Bateman, 2021).

Astacins share a basic domain architecture consisting of an N-terminal signal peptide for secretion, a pro-peptide (PP) of variable length (from 34 residues in astacin to 486 residues in *Drosophila melanogaster* tolkin; Finelli et al., 1995; Gomis-Rúth, Trillo-Muyó et al., 2012; Arolas et al., 2018) for zymogenic latency and the CD (Gomis-Rúth, Trillo-Muyó et al., 2012). This core may be C-terminally extended by disparate modules, among which are linkers (LNK), CUB domains (found in the complement component C1r/1s, the embryonic sea urchin Uegf and bone morphogenetic protein 1; Bork & Beckmann, 1993; PF00431) and MAM domains (common to meprins, A5 receptor protein and tyrosine phosphatase μ; Cismasu et al., 2004; PF00629). Two astacins, namely a short 240-residue protein (astl gene; UniProt accession B4F319) and a long 403-residue protein (astl-mam gene; UniProt B4F320), were identified in *L. polyphemus*, recombinantly expressed and biochemically characterized (Becker-Pauly et al., 2009, 2011). The short form was predominantly found in the eyes and the brain, which suggests a function in the nervous system, while the long form was ubiquitous (Becker-Pauly et al., 2009). The short paralogue has the basic domain architecture of the family, while the long paralogue further contains an LNK and an MAM domain. Both astacins share 46% sequence identity within the PP and the CD, and their trypsin-activated forms showed proteolytic activity in gelatin zymography and in solution against azocasein and the extracellular matrix proteins fibronectin, type IV collagen, gelatin and laminin, but not triple-helical collagen (Becker-Pauly et al., 2009). Finally, consistent with the horsehoe crab being a chelicerate, these astacins were found to be closer to an orthologue from the brown spider *Loxoceles intermedia* in a phylogenetic analysis than to the crustacean orthologs from the crayfish *A. astacus* and the shrimp *Panaeus vannamei* (Becker-Pauly et al., 2009).

Here, we crystallized the zymogen of the long paralogue, hereafter referred to as pLAST-MAM, and solved its crystal structure. Our results provide structural and molecular insight into the latency mechanism of the currently evolutionarily oldest holozoan astacin.

2. Methods

2.1. Protein crystallization

The pLAST-MAM zymogen was obtained by recombinant expression in *Trichoplusia ni* High Five insect cells, purified as described in Becker-Pauly et al. (2009) and subsequently concentrated in a Vivaspin device using a polyethersulfone membrane with 10 kDa cutoff (Vivaproduits). We screened for crystallization conditions using the sitting-drop vapour-diffusion method at the joint IBMB/IRB Automated Crystallography Platform (https://www.ibmb.csic.es/en/facilities/automated-crystallographic-platform). Reservoir solutions were prepared using a Tecan Freedom EVO robot and were dispensed into 96 × 2-well MRC plates (Innovadyne Technologies). A Phoenix/RE robot (Art Robbins) administered crystallization nanodrops consisting of 100 nl each of protein and reservoir solution. Crystallization plates were subsequently incubated at 4 or 20°C in Bruker steady-temperature crystal farms. Successful initial conditions were refined and scaled up to the microlitre range in 24-well Cryscem crystallization dishes (Hampton Research) whenever possible. Optimal crystals of the protein at ~7 mg ml⁻¹ in 50 mM HEPES pH 7.0 were obtained at 20°C using 0.1 M bicine pH 9.0, 10% polyethylene glycol (PEG) 40 000, 2% dioxane as the reservoir solution. Crystals were thin and fragile rectangular plates, which were harvested using cryo-loops (Molecular Dimensions), rapidly passed through a cryo-buffer consisting of reservoir solution plus 20% (v/v) glycerol and flash-vitrified in liquid nitrogen for transport and data collection.

2.2. Diffraction data collection and processing

X-ray diffraction data were collected on 18 April 2010 using an ADSC Quantum 315r detector on beamline ID29 of the
Table 1
Crystallographic data.

Abbreviations: AU, crystallographic asymmetric unit; GOL, glycerol; PEG, diethylene glycol; PGE, triethylene glycol; RSRZ, real-space R-value Z-score. Values in parentheses are for the outermost resolution shell.

| Beamline   | ID29, ESRF | Space group | C2 |
|------------|------------|-------------|----|
| Protomers per AU | 2          |             |    |
| a, b, c (Å) | 115.83, 47.57, 236.40 | 90, 102.91, 90 |
| Wavelength (Å) | 0.9724 |             |    |
| No. of measurements | 316453 |             |    |
| No. of unique reflections | 49732 |             |    |
| Resolution range (Å) | 76.8–2.40 (2.54–2.40) | 99.6 (98.6) |
| Rmerge† | 0.151 (1.431) |             |    |
| Rmerge‡ | 0.164 (1.564) |             |    |
| CC1/2‡ | 0.995 (0.594) |             |    |
| Average intensity‡ | 9.8 (1.9) |             |    |
| Wilson B factor (Å²) | 53.7 |             |    |
| Average multiplicity | 6.4 (6.1) |             |    |
| Resolution range used for refinement (Å) | 46.7–2.40 |             |    |
| Reflections used (total/test set) | 489887 (741) |             |    |
| Crystallographic R/Rmerge† | 0.253/0.289 |             |    |
| Contents of AU | Non-H protein atoms | 4801 | 2 Zn²⁺, 1 Mg²⁺, 229 |
| Ionic ligands | 2 |             |    |
| Waters | 229 |             |    |
| Non-ionic ligands | 1 PGE, 1 PEG, 2 GOL | 1 PEG, 1 PEG, 2 GOL |
| R.m.s.d. from target values | Bond lengths (Å) | 0.014 | 1.26 |
| Angles (°) | 77.6 |             |    |
| Average B factor (Å²) | 550 [91%]/10603 |             |    |
| Analysis of protein contacts and geometrinya | Ramachandran favoured/outliers/all analyzed | 550 [91%]/10603 | 43 |
| Bond-length/bond-angle/chirality/planarity outliers | 0/0/0 | 29 [5.8%] |
| Side-chain outliers | 43 |             |    |
| All-atom clashes | 4.5 |             |    |
| Clashscore | 131 [21.6%] |             |    |
| Fc–Fo correlation | 0.89 (0.88) |             |    |
| PDB code | 8a28 |             |    |

† For definitions, see Einspruch & Weiss (2012). ‡ Average intensity is the ⟨|I(0)/I(t)|⟩ of unique reflections after merging according to XSCELE (Kabsch, 2010). § According to the wwPDB Validation Service (https://validate-rcsb-1.wwpdb.org/validservice).

ESRF synchrotron, Grenoble, France. Diffraction data were processed using XDS (Kabsch, 2010) and XSCALE, and were transformed to MTZ format using XDSCONV for use with the Phenix (Liebschner et al., 2019) and CCP4 (Winn et al., 2011) suites. Analysis with phenix.xtriage within Phenix revealed an absence of translational noncrystallographic symmetry (NCS) and no significant twinning according to the L-test. The crystals contained two monomers in the asymmetric unit and Table 1 provides essential statistics on data collection and processing.

2.3. Structure solution and refinement

The structure of pLAST-MAM was solved by molecular replacement using the Phaser crystallographic software (McCoy et al., 2007) and a homology model for the CD and MAM domain predicted with AlphaFold (Jumper et al., 2021). After several trials, we could only obtain correct solutions by searching with the domains separately, i.e. two for the CD but only one for the MAM domain. Those for the CD corresponded to Eulerian angles of ω = 54.2, β = 54.0, γ = 116.3 and cell-fracton translation values of x = 0.106, y = 0.002, z = 0.210 for one protomer and α = 261.5, β = 125.5, γ = 297.0, x = 0.419, y = 0.884, z = 0.303 for the second protomer. The corresponding values for the MAM moiety were α = 64.8, β = 106.0, γ = 172.9, x = 0.285, y = 0.751, z = 0.991. These solutions had a final translation-function Z-score of 17.1 and a global log-likelihood gain after refinement of 782.

The suitably rotated and translated molecules were subjected to the phenix.autobuild protocol (Terwilliger et al., 2008) within Phenix, which yielded a greatly improved Fourier map for manual model building with Coot (Casanal et al., 2020). The latter alternated with crystallographic refinement using the phenix.refine protocol (van Zundert et al., 2021) and BUSTER (Smart et al., 2012), which both included translation/liberation/screw motion and NCS restraints, until completion of the model. The latter comprised residues Glu22–Cys403 of protomer A and Glu22–Gly246 of protomer B, each with a catalytic zinc ion plus one tentatively assigned magnesium cation, one diethylene glycol molecule, one triethylene glycol molecule, two glycerol molecules and 229 solvent molecules. The occupancy of LNK and MAM of protomer A refined to 87%. Table 1 provides essential statistics on the final refined model, which was validated through the wwPDB validation service (https://validate-rcsb-1.wwpdb.org/validservice). The coordinates can be retrieved from the Protein Data Bank (https://www.wwpdb.org/) as entry 8a28.

2.4. Miscellaneous

Structure superpositions were performed with SSM (Krissinel & Henrick, 2004) within Coot. Figures were prepared using UCSF Chimera (Goddard et al., 2018). Protein interfaces and intermolecular interactions were analysed using PDB-PIASA (https://www.ebi.ac.uk/pdbe/pisa; Krissinel & Henrick, 2007) and verified by visual inspection. For this, the interacting surface of a complex was taken as half of the sum of the buried surface areas of either molecule.

3. Results and discussion

3.1. Overall crystal arrangement

To prevent autolysis, pLAST-MAM was recombinantly expressed in insect cells as a point mutant in which the general base/acid glutamate for catalysis (Arolas et al., 2018; E¹⁴⁰, residues are given as single-letter codes with numbering in superscript according to UniProt B4F320; other proteins are numbered in subscript) was replaced by alanine to create a catalytically impaired variant. This strategy has often been employed in the past to prevent autolysis when crystallizing MPzymogens (see Table 1 in Arolas et al., 2018). pLAST-MAM crystals with two protomers (A and B) in the crystallographic asymmetric unit were obtained in 2010 (Table 1) but the structure was only solved very recently using a homology model predicted by AlphaFold (Jumper et al., 2021) for molecular replacement. After extensive calculations with the whole molecule and separate domains, the two CDs (N⁹⁰–C²⁴⁴) could confidently be placed, rebuilt and refined, as well
as the respective PPs (defined for E_{22}–K_{48}). In contrast, the LNK (F_{245}–D_{257}) and MAM (F_{358}–C_{403}) moieties were flexible and only those of protomer A could be placed in the structure. Moreover, crystallographic refinement revealed that the final Fourier map was discontinuous in several places in the MAM domain owing to this flexibility (Fig. 1a). Indeed, while the

Figure 1
Crystal features. (a) Experimental structure of pLAST-MAM protomer A in C^* representation superposed with the final (2mF_oobs – DF_calcd)-type Fourier map contoured at 0.5σ above the threshold. The CD is in yellow with the zinc ion as a magenta sphere, the LNK is in purple and the MAM domain is in orange. (b) Crystal packing viewed down the x axis of the crystal in cross-eyed stereo. The CDs of protomers A (orange C^* trace) and B (green C^* trace), together with their symmetry mates (purple C^* traces), are in planes (1 and 2; pink background) that parallel the xy plane of the crystal. The two CDs are in a relative upside-down conformation, so their respective C-termini point either downwards (protomer A, orange arrow) or upwards (protomer B, green arrow). The MAM domains of protomer A and its symmetry mates occupy the space between the CD sections (yellow background). (c) Same as (b) but viewed down the crystal y axis, i.e. after a vertical 90° rotation. Note that each LNK interacts in extended conformation with a symmetric MAM domain to build up the crystal in the section with the yellow background between the CD planes. Further contacts are observed between symmetric MAM domains. (d) Stereoview down the y axis as in (c) showing four CD sections (1–4) as they occur in the crystal. No atoms are found in the space between sections 2 and 3 as the LNK and MAM domain of the B protomers (CD in green C^* trace) is disordered in the final model. (e) Same as (d) after superposing protomers A and B using their respective CDs. The LNK and MAM domain of protomer B and its symmetry mates (in purplish colours) would establish the crystal contacts required to build up the crystal in the space between sections 2 and 3. Thus, these moieties must be present in the crystal but are disordered. (f) Superposition in stereo of the MAM domain of pLAST-MAM predicted with *AlphaFold* and refined against experimental diffraction data (LNK in purple, MAM domain in orange) and the experimental MAM domain of human meprin β (chartreuse; PDB entry 4gwm; Arolas et al., 2012).
CDs showed average thermal displacement parameters (B factors) of 60 and 73 Å² for protomers A and B, respectively, the segment spanning LNK and MAM of protomer A had an average B factor of 116 Å² after occupancy refinement to 87%.

Inspection of the crystal packing revealed that the two CDs form tight layers parallel to the xy plane of the crystal with their respective crystallographic symmetry mates (1 and 2 in Figs. 1b and 1c). They are in a relative upside-down conformation, so that the C-termini protrude either above or below the CD layer. In the case of the A protomers, LNK and MAM project into the space between CD sections and make interactions with symmetric MAM and LNK moieties from the CD layer beneath, respectively, which are required to form the crystal (Figs. 1b and 1c). In contrast, the space between CD sections into which the C-termini of the B-protomer CDs point (sections 2 and 3 in Fig. 1d) does not contain any atoms and thus lacks crystal contacts owing to the missing LNKs and MAMs. However, when superposing the full-length protomer A on protomer B by their respective CDs, the LNK and MAM moieties adopt a very similar arrangement in the space between the two CD layers to that seen in the A protomers (sections 2 and 3 in Fig. 1e). Thus, LNK and MAM of the B protomers must also be present in the crystal to establish the intermolecular contacts necessary to build the crystal. Overall, we conclude that while both LNK-MAM moieties are very flexible and adopt several slightly different orientations that are able to assemble the crystal, those of protomer A are somewhat more rigid, so they are grossly defined in the final Fourier maps. In contrast, those of protomer B are so flexible that the density is too poor to confidently place them.

Thus, given the poor definition of the MAM domains, we will concentrate the discussion hereafter on the PP and CD moieties of the zymogen (referred to here as pLAST) and the mature CD (LAST) of protomer A, and the mechanism of latency in the context of other structurally characterized astacin zymogens. Suffice to say that the predicted structure of the MAM domain of pLAST-MAM is very similar to that of the human astacin-family member meprin β except for some loops (Fig. 1f). For a discussion of the architecture and features of these domains, please refer to Cismasiu et al. (2004), Aricescu et al. (2006, 2007), Arolas et al. (2012), Yelland & Djordjevic (2016) and Eckhard et al. (2021).

3.2. Structure of the zymogen

The pLAST moiety subdivides into three segments when viewed in the standard orientation of MPs (Gomis-Ruth,
Botelho et al., 2012): the N-terminal PP (E^{22–K^{48}}), an upper N-terminal subdomain (NTS) of the CD (N^{49–G^{146}}) and a lower C-terminal subdomain (CTS) of the CD (F^{147–C^{244}}) (Fig. 2a). The PP runs along the front surface of pLAST from right to left and features helix α1 on the primed side of the cleft (substrate and active-site subsite terminology based on Schechter & Berger, 1967; Gomis-Ruth, Botelho et al., 2012). It adopts a wide loop structure protruding from the cleft between L^{29} and D^{38} (Fig. 3a), which is stabilized by two intramain-chain hydrogen bonds (H^{31} N−G^{37} O and F^{35} O−I^{39} N). The intervening residues are included in a ‘PP motif’ found in astacins (F−E/O−D−I; Gomis-Ruth, Trillo-Muyo et al., 2012), F^{35}−G^{31}−E^{30}−I^{29} in pLAST (Becker-Pauly et al., 2009). For F^{39}−G^{41}, the polypeptide adopts an extended conformation along the nonprimed side of the cleft before turning 90° downwards for V^{42}−Y^{45} and then leftwards for Y^{45}−D^{47}. Thereafter, the peptide containing the primary activation cleavage site (K^{48}−H^{54}) enters into the CD, which adopts a helical conformation for K^{48}−H^{54} (α2; Figs. 2a and 3a).

As in other astacins, the 195-residue CD divides into an NTS and a CTS of approximately equal size (Fig. 2a). The NTS is rich in regular secondary structure and consists of a five-stranded arched and twisted β-sheet (β1−β5), the strands of which parallel the active-site cleft except for the lowermost (β4), which is antiparallel and frames the upper rim of the cleft. The concave face of the sheet accommodates three helices (α3−α5), among which are a ‘backing helix’ (α4) and an ‘active-site helix’ (α5) that are characteristic of astacins and metzincins in general (Bode et al., 1993; Stöcker et al., 1993; Stöcker & Bode, 1995; Gomis-Ruth, 2009; Gomis-Ruth, Trillo-Muyo et al., 2012; Cerda-Costa & Gomis-Ruth, 2014; Arolas et al., 2018). The active-site helix encompasses the first two-thirds of a conserved zinc-binding motif (H^{139}−E−X−X−H−X−X−X−H^{149} in pLAST) found in astacins and other metzincins, which features three metal-binding histidines and the general base/acid glutamate, here replaced with an alanine (see above and Fig. 2c). At the glycine of the motif (G^{146}), the polypeptide undergoes a sharp downwards turn to enter the CTS, which in contrast to the NTS is more irregular. It contains two short helices (α6 and α7) and the short β-ribbon β6β7 in addition to a ‘C-terminal helix’ (α8), which again is characteristic of metzincins. Of note is another conserved structural element of metzincins, the ‘Met-turn’, which is a tight 1,4-turn (S^{194}−L^{197}) encompassing the strictly conserved M^{196} (Fig. 2c). Its side chain provides a hydrophobic pillow for the metal-binding site that is essential for the stability and function of metzincins (Tallant, Garcia-Castellanos et al., 2010). Immediately downstream of this methionine, Y^{198} provides the fourth zinc ligand of the CD through its somewhat more distant O^δ atom. In other astacins, this residue is swung out upon substrate binding following a ‘tyrosine switch’ and its O^δ atom participates in stabilization of the reaction intermediate during catalysis (Stöcker & Yiallouros, 2013). Finally, a disulfide bond links the back of the NTS with the C-terminal helix α8 of the CTS (C^{210}−C^{244}) and a second one links strand β4 with the loop connecting β5 and α5 (L^{β5α5}) (C^{112}−C^{131}) (Fig. 2a).

3.3. Mechanism of latency

Latency is achieved in pLAST by blocking access of substrates through the PP, which runs across the active-site cleft of the CD moiety in the opposite direction to a substrate (Figs. 2a, 2b and 3a). This is a strategy to prevent untimely autolytic cleavage in cis (Khan & James, 1998; Arolas et al., 2018). In addition, the polypeptide chain does not adopt an extended conformation as required for substrates to be cleaved.
(Tylndall et al., 2005) but rather the aforementioned loop structure protrudes from the cleft (Fig. 2b). This prevents a scissile bond from extending across cleft subsites S1 and S2 (Fig. 3a), which is another mechanism to prevent undesired cleavage (Arolas et al., 2018). The surface occluded by the PP–CD interaction spans 1207 Å², which is in the range reported for protein–protein complexes (~380–3390 Å²; Chen et al., 2013), and has a solvation free-energy gain upon interface formation (∆G) of ~16.9 kcal mol⁻¹ (Krissinel & Henrick, 2007), indicating a strong interaction. Participating structural elements include the entire PP and segments N49–V52, D38–V12, Y173–H143, W148–N152, S170–M178, Y198–T208 and P223–K226 of the CD, with the establishment of 20 electrostatic interactions and hydrophobic contacts between 17 pairs of residues of either moiety (Table 2).

The primary activation site of pLAST (K48–N49) is inserted within short helix α2 and buried in the zymogen, thus preventing access by activating enzymes in a similar fashion as found in pro-astacin (Guevara et al., 2010). Moreover, K48 N makes strong interactions with Y173 O (2.7 Å apart) and N176 O (3.1 Å) of the CD and with E36 O°C (2.7 Å) of the PP motif, which likewise hinder activation. The latter interaction is reminiscent of the double salt bridge between an arginine and an aspartate in a PP motif found in matrix metallopeptidase (MMP) zymogens (P-R-C-G-X-P-D; van Wart & Birkedal-Hansen, 1990; Springman et al., 1990; Tallant, Marrero et al., 2010; Arolas et al., 2018). Moreover, the activation-scissile-bond N atom is bound to D47 O (2.8 Å) within the PP, so the activation site is additionally protected in the zymogen. All of these findings support the maturation of pLAST requiring partial unfolding of the segment flanking the activation site and/or preliminary cleavages, as described for crayfish astacin (Yiallouros et al., 2002; Guevara et al., 2010).

The most relevant element for latency is D38, which binds the catalytic zinc in a bidentate manner through its O41 (2.2 Å) and O62 (2.4 Å) atoms (Fig. 2c), thus replacing the catalytic solvent required for catalysis in mature MPs (Arolas et al., 2018). This aspartate is embedded in the PP motif and contributes to a distorted octahedral metal coordination sphere together with H143 N (2.1 Å) and H149 N (2.1 Å) in plane with the cation and with H143 N (2.1 Å) and Y198 O (3.3 Å) in the apical positions. Thus, D38 functions as an ‘aspartate switch’ for latency maintenance as described previously for crayfish astacin (Guevara et al., 2010) and human meprin β (Arolas et al., 2012) within the astacins (see below) and for fragilysin-3 (Goulas et al., 2011) and the bacterial MMP karilysin (Cerdà-Costa et al., 2011) within other metzincins (Arolas et al., 2018).

3.4. Proposed mechanism of activation

The archetypal astacin from crayfish, which like the horse-shoe crab is an arthropod, represents the evolutionarily closest orthologue of LAST with a known mature structure (Bode et al., 1992). Indeed, 157 Cα atoms from these proteins superpose with a core root-mean-square deviation (r.m.s.d.) of 1.3 Å (38% sequence identity). Moreover, a predicted homology model of LAST was obtained with AlphaFold (Jumper et al., 2021), which showed most of the common features in relevant segments described for mature astacin. It had an average predicted local distance difference test (pLDDT) value of >97, which is indicative of high reliability (Tunyasuvunakool et al., 2021). Thus, this model is taken hereafter as a working model of mature Limulus astacin.

Superposition of the pLAST structure and the LAST model (Fig. 3b) reveals that the CD moieties mostly coincide. In particular, the NTSs match best, with an r.m.s.d. of 0.93 Å for all 746 atoms of segment L1–H149. The metal-binding site and most of the active-site cleft would largely be preformed in the zymogen, as observed for other MP zymogens (Arolas et al., 2018). Within the CTS, good agreement is observed for the segment E188–G199, which includes the Met-turn, and the entire C-terminal stretch from G207 to C244. Loop G199–D206, which frames the lower rim of the cleft, slightly deviates, with a maximal displacement of ~2 Å that closes the cleft on the primed side upon activation. On the bottom of the nonprimed side of the cleft, E150–E179 would additionally undergo a

Table 2

| PP          | CD            | Distance (Å) |
|-------------|---------------|--------------|
| Hydrogen bonds (<3.5 Å) |               |              |
| E22 O       | W225 N31      | 2.8          |
| N23 N32     | T201 O41      | 3.0          |
| N23 N33     | G202 O        | 2.9          |
| D38 O31     | W225 N        | 3.1          |
| D38 O32     | V222 N        | 3.0          |
| P38 O       | C122 O        | 2.9          |
| E36 O       | Y198 O7       | 2.7          |
| E36 O41     | H149 N31      | 3.1          |
| E36 O42     | N156 N32      | 2.9          |
| D38 O       | S134 N        | 3.0          |
| D38 O33     | Y198 O7       | 3.0          |
| A40 N       | S134 O        | 3.2          |
| A46 O       | Y177 O7       | 2.8          |
| D47 O       | I51 N         | 3.2          |
| D47 O31     | N95 N         | 3.2          |
| D47 O32     | A56 N         | 3.5          |
| K48 N33     | Y177 O       | 2.7          |
| K48 N34     | N176 O        | 3.1          |
| Metallorganic bonds |               |              |
| D38 O31     | Zn999         | 2.2          |
| D38 O41     | Zn999         | 2.4          |
| Van der Waals interactions (<4 Å) |               |              |
| P28         | C112          |              |
| P28         | C131          |              |
| L29         | I136          |              |
| L29         | V198          |              |
| L29         | V224          |              |
| V30         | A202          |              |
| Y30         | F203          |              |
| F35         | I51           |              |
| F35         | W148          |              |
| F39         | W148          |              |
| F39         | W148          |              |
| A49         | M145          |              |
| A49         | M145          |              |
| A55         | F135          |              |
| A55         | F135          |              |

Acta Cryst. (2022). D78, 1347–1357

Tibisay Guevara et al.  •  Latency regulation of L. polyphemus astacin 1353
closing motion of maximally ~3 Å facilitated by a ~10° rotation around W298. The largest deviation, however, is observed for the segment P138–N187, which conforms to a flexible ‘activation domain’ and would become significantly rearranged (Fig. 3b), as described for other astacins (Guevara et al., 2010) and the otherwise unrelated trypsin-like serine endopeptidases (Huber & Bode, 1978). This rearrangement would result from the displacement of N49–L56, which upon maturation cleavage at K48–N49 would become rotated outwards around the Cα–C bond of L50. In this way, the seven preceding residues would be amply repositioned by up to ~11 Å and penetrate the mature enzyme moiety, so the first three residues (N49–A50–I51) would be completely inaccessible to solvent, as reported for meprin β (see Section 3.5). Next, N49 would bind the ‘family-specific residue’ immediately after the third zinc-binding histidine (E150, Bode et al., 1993; Gomis-Rüth, 2003), which in turn is held in place by internal salt bridges with R237 and R153 in the zymogen. This interaction could occur directly through the N49 Nδ2 atom, as observed in meprin β (Arolas et al., 2012). An alternative interaction through the α-amino group (N49 N) mediated by a solvent molecule, as observed in crayfish astacin (Bode et al., 1992), is also conceivable. Moreover, the N49 Oε1 atom might also bind the R237 side chain. Overall, this scenario of a deeply buried mature N-terminus is very similar to that found in other astacins, in which the maturation mechanism has been structurally verified (see Section 3.5). This, in turn, provides confidence in the reliability of the LAST homology model.

3.5. Comparison with other astacin latency mechanisms

To date, the crystal structures of crayfish pro-astacin (PDB entry 3lq0; Guevara et al., 2010), human pro-meprin β (PDB entry 4gwn; Arolas et al., 2012) and pro-myroilysin from two closely related bacterial species, Myroides profundi (PDB entry 5czw; Xu et al., 2017) and Myroides sp. CSLB8 (PDB entry 5gw; Xu et al., 2017), have been reported, as well as their respective mature forms astacin (PDB entry 1ast; Bode et al., 1992; Gomis-Rüth et al., 1993), meprin β (PDB entry 4gwn; Arolas et al., 2012) and myroilysin from Myroides sp. CSLB8 (PDB entry 5zjk; Ran et al., 2020). The two proteins from Myroides are 99.6% identical, so only that from Myroides sp. CSLB8 will be discussed here. Of all these structures, only pro-meprin β spans additional domains downstream of the CD, namely an MAM and a TRAF domain (Arolas et al., 2012). Pictures of the three zymogens superposed onto the mature forms, together with those of the pLAST structure and the LAST model, are provided in Figs. 4(a)–4(d).

In all cases, the mature N-terminus is buried inside the catalytic moiety and is bound to the family-specific glutamate of astacins either directly through an N-terminal asparagine (LAST and meprin β) or glycine (myroilysin) or mediated by a solvent molecule because the N-terminal segment is one residue shorter (astacin). The position of the new N-terminus in the zymogen and the mature moiety is very close in astacin (~2 Å; Fig. 4b), quite close in meprin β (~6 Å; Fig. 4c), farther apart in LAST (~11 Å; Fig. 4a) and farthest in myroilysin (~17 Å; Fig. 4d).

Detailed analysis of the four zymogen–mature enzyme pairs reveals that in all cases the PP is poor in regular secondary structure and adopts a mostly extended conformation that traverses the active-site cleft in the opposite direction to a substrate. In pro-myroilysin it is additionally elongated at the N-terminus and further extends along the front surface of the NTS (Fig. 4d), while in pro-meprin β (Fig. 4c) it runs in an extended conformation along a neighbouring TRAF domain on the right of the CD (not shown). In all cases, CTS regions framing the bottom of the active-site cleft on its nonprimed side constitute activation segments that undergo rearrangement upon maturation cleavage and repositioning of the new N-terminus. In astacin, only this activation segment (I130–E139, mature enzyme numbering according to PDB entry last; add 49 for full-gene numbering; see UniProt P07584) is reorganized, while the rest of the molecule is preformed in the zymogen (Guevara et al., 2010; Fig. 4b). Next, LAST is most likely to undergo slight rearrangement of two segments (G139–D260 and E150–E179) in addition to the major movement of the activation segment (P138–N187; see Section 3.4 and Fig. 4a). Meprin β, in turn, repositions most of its CTS (Q165–Y211 and L199–D233 according to UniProt Q16820; segment D194–L199 is disordered in the zymogen structure) in a concerted hinge motion that entirely closes the cleft at its bottom in response to maturation (Fig. 4c). Finally, the largest deviation is observed in myroilysin, which rearranges its entire CTS except for the Met-turn and the C-terminal helix (Fig. 4d). The segments affected are Q155–A201 and Y210–N225 (myroilysin numbering according to PDB entry 5gw; see also UniProt A0A0P0DZ84). A large flap (N167–S193), which encompasses two helices, is folded back on top of the active-site cleft and traps the PP in the zymogen. Upon maturation, this flap is rotated to the right with a maximal displacement of ~17 Å (measured at P176), thus liberating access to the cleft (Fig. 4d).

Differences are also found in the residues blocking the zinc ion in the zymogen. The three metazoan proteins contain an aspartate within the PP motif, which is structurally conserved (Fig. 4e), acting as an aspartate switch. In contrast, the bacterial enzyme lacks the PP motif and instead features a cysteine, which blocks the zinc according to a ‘cysteine-switch’ mechanism (Ran et al., 2020; Xu et al., 2017). Moreover, the polypeptide chain flanking the cysteine is in a canonical, extended conformation and does not adopt the loop of the PP motif. Overall, this is inversely reminiscent of MMPs, in which canonical vertebrate orthologues regulate latency according to a cysteine-switch mechanism (Springman et al., 1990; van Wart & Birkedal-Hansen, 1990; Rosenblum et al., 2007; Tallant, Marrero et al., 2010; Arolas et al., 2018), while the bacterial orthologue karilysin from the periodontopathogen Tannerella forsythia instead operates according to an asparagine switch. As in astacins, MMPs are only found dispersedly outside animals, and it has been proposed that karilysin is a xenologue co-opted from a mammalian host through horizontal gene transfer facilitated by intimate interaction between the host and the colonizing bacterium (Cerdà-Costa et al., 2011).
similar origin is conceivable for myroilysin within astacins given that *Myroides* spp. have been reported in several human body fluids and can trigger infection leading to soft-tissue infections in humans (Maraki *et al.*, 2012) and bacteraemia in a

Figure 4
Activation of astacins with reported zymogen structures and a conserved PP motif. (a)–(d) Superposition in cross-eyed stereo of the C' traces in standard orientation of the latent and mature forms of (a) *Limulus* astacin (latent, PDB entry 8a28; mature, *AlphaFold* model), (b) crayfish astacin (latent, PDB entry 3lj0; mature, PDB entry 1ast), (c) human meprin β (latent, PDB entry 4gwm; mature, PDB entry 5gwd), and (d) *Myroides* sp. CSLBB myroilysin (latent, PDB entry 5gwd; mature, PDB entry 5zk). The mature forms are in orange and the zymogens are in cyan (PP) and yellow (CD). The catalytic zinc ions are depicted as purple spheres. The PP of meprin β is N-terminally extended and runs across the front surface of a vicinal TRAF domain (not shown; Arolas *et al.*, 2012). The most relevant rearranged segments during maturation cleavage, the ‘activation segment’ and the mature N-terminal segment, are pinpointed by green and red stars in each structure, respectively. (e) Superposition of the segments encompassing the PP motif of astacins (F-E-G-D-I) in pLAST (C atoms in cyan), crayfish pro-astacin (C atoms in tan) and human pro-meprin β (C atoms in plum). Myroilysin lacks this motif.
diabetic patient (Endicott-Yazdani et al., 2015). Thus, as in MMPs, the latency mechanisms of holozoan orthologues and bacterial xenologues would also diverge in astacins.

4. Data availability
All data and reagents are freely available from the authors upon reasonable request.

Acknowledgements
We are grateful to Joan Pous from the joint IBMB/IRB Automated Crystallography Platform for assistance. Author contributions were as follows. FXG-R, CB-P and WS conceived and supervised the project, CB-P purified the protein, TG and AR-B performed experiments, FXG-R performed calculations and FXG-R wrote the manuscript with contributions from all authors. The authors declare no financial or nonfinancial conflicts of interest.

Funding information
This study was supported in part by grants from Spanish and Catalan public and private bodies (grant/fellowship references PID2019-107725RG-100 from MICIN/AEI/10.13039/501100011033 to FXG-R, TG and AR-B, 2017SGR3 and Fundació La Marató de TV3 201815 to FXG-R, TG and AR-B, 2017SGR3 and PID2019-107725RG-I00 from MICIN/AEI/10.13039/Catalan public and private bodies (grant/fellowship references F1356). Further support was obtained from German funding bodies (grant SFB877, Project A9: ‘Proteolysis as a regulatory AR-B). Further support was obtained from German funding bodies (grant SFB877, Project A9: ‘Proteolysis as a regulatory AR-B). Further support was obtained from German funding bodies (grant SFB877, Project A9: ‘Proteolysis as a regulatory AR-B)

References
Aricescu, A. R., Hon, W. C., Siebold, C., Lu, W., van der Merwe, P. A. & Jones, E. Y. (2006). EMBO J. 25, 701–712.

Becker-Pauly, C., Barre`re, O., Schilling, O., auf dem Keller, U., Ohler, A., Broder, C., Schütté, A., Kappelhoff, R., Stöcker, W. & Overall, C. M. (2011). J. Mol. Biol. 410, M111.009233.

Cerdà-Costa, N., Guevara, T., Karim, A. Y., Ksiazek, M., Nguyen, K. A., Arolas, J. L., Potempa, J. & Gomis-Rüth, F. X. (2011). Mol. Microbiol. 79, 119–132.
Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., Tosatto, S. C. E., Paladin, L., Raj, S., Richardson, L. J., Finn, R. D. & Bateman, A. (2021). *Nucleic Acids Res.* 49, D412–D419.

Pfeiderer, G., Zwilling, R. & Sonneborn, H. H. (1967). *Hoppe Seylers Z. Physiol. Chem.* 348, 1319–1331.

Ran, T., Li, W., Sun, B., Xu, M., Qiu, S., Xu, D. Q., He, J. & Wang, W. (2020). *Int. J. Biol. Macromol.* 156, 1556–1564.

Rawlings, N. D. & Bateman, A. (2021). *Protein Sci.* 30, 83–92.

Rosenblum, G., Meroueh, S., Toth, M., Fisher, J. F., Fridman, R., Mobashery, S. & Sagi, I. (2007). *J. Am. Chem. Soc.* 129, 13566–13574.

Rowley, A. F., Rhodes, C. P. & Ratcliffe, N. A. (1984). *Semenova, S. A. & Rudenskaia, G. N. (2008). Protein Sci.* 30, 83–92.

Rudkin, D. M., Young, G. A. & Nowlan, G. S. (2008). *Paleontology,* 51, 1–9.

Schechter, I. & Berger, A. (1967). *Biochem. Biophys. Res. Commun.* 27, 157–162.

Schmid, M. R., Dzedziech, A., Arefin, B., Kienzle, T., Wang, Z., Akhter, M., Berka, J. & Theopold, U. (2019). *Insect Biochem. Mol. Biol.* 109, 63–71.

Sekiguchi, K. & Shuster, C. N. Jr (2009). *Biological and Comparative Horseshoe Crab Research*. Edited by J. T. Tanacredi, M. L. Botton & D. R. Smith, pp. 5–24. Dordrecht: Springer.

Semenova, S. A. & Rudenskaia, G. N. (2008). *Biomed. Khim.* 54, 531–554.

Shuster, C. N. Jr (1982). *Physiology and Biology of Horseshoe Crabs: Studies on Normal and Environmentally Stressed Animals*. Edited by J. Bonaventura, C. Bonaventura & S. Tesh, pp. 1–52. New York: Alan R. Liss.

Smart, O. S., Womack, T. O., Flensburg, C., Keller, P., Paciorek, W., Shuster, C. N. Jr (1982). *Physiology and Biology of Horseshoe Crabs: Studies on Normal and Environmentally Stressed Animals*. Edited by J. T. Tanacredi, M. L. Botton & D. R. Smith, pp. 5–24. Dordrecht: Springer.

Smith, S. A. (2022). *Invertebrate Medicine*, 3rd ed., edited by G. A. Lewbart, pp. 283–300. Hoboken: John Wiley & Sons.

Springman, E. B., Angleton, E. L., Birkedal-Hansen, H. & Van Wart, H. E. (1990). *Proc. Natl Acad. Sci. USA,* 87, 364–368.

Stöcker, W. & Bode, W. (1995). *Curr. Opin. Struct. Biol.* 5, 383–390.

Stöcker, W. & Gomis-Rüth, F. X. (2013). *Proteases: Structure and Function*, edited by K. Brix & W. Stöcker, pp. 235–263. Vienna: Springer Verlag.

Stöcker, W., Gomis-Rüth, F. X., Bode, W. & Zwilling, R. (1993). *Eur. J. Biochem.* 214, 215–231.

Stöcker, W., Gomis-Rüth, F. X., Huber, R., Zwilling, R. & Bode, W. (1992). *Biol. Chem. Hoppe Seyler,* 373, 654.

Stöcker, W., Grams, F., Reinemer, P., Bode, W., Baumann, U., Gomis-Rüth, F. X. & McKay, D. B. (1995). *Protein Sci.* 4, 823–840.

Stöcker, W. & Yiallouros, I. (2013). *Handbook of Proteolytic Enzymes*, 3rd ed., edited by N. D. Rawlings & G. S. Salvesen, pp. 895–900. Oxford: Academic Press.

Stöcker, W., Wolz, R. L., Zwilling, R., Strydom, R. J. & Auld, D. S. (1988). *Biochemistry,* 27, 5026–5032.

Sturmer, L. (1952). *J. Paleontol.* 26, 630–639.

Tallant, C., García-Castellanos, R., Baumann, U. & Gomis-Rüth, F. X. (2010). *J. Biol. Chem.* 285, 13951–13957.

Tallant, C., Marrero, A. & Gomis-Rüth, F. X. (2010). *Biochim. Biophys. Acta,* 1803, 20–28.

Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Pardi, P. H., Hung, L.-W., Read, R. J. & Adams, P. D. (2008). *Acta Cryst.* D64, 61–69.

Tunyasuvunakool, K., Adler, J., Wu, Z., Green, T., Zielinski, M., Zidek, A., Bridgland, A., Cowie, A., Meyer, C., Laydon, A., Velankar, S., Kleywegt, G. J., Bateman, A., Evans, R., Pritez, A., Figurov, M., Ronneberger, O., Bates, R., Kohl, S. A. A., Potapenko, A., Ballard, A. J., Romera-Paredes, B., Nikolov, S., Jain, R., Clancy, E., Reiman, D., Petersen, S., Senior, A. W., Kavukcuoglu, K., Birney, E., Kohli, P., Jumper, J. & Hassabis, D. (2021). *Nature,* 596, 590–596.

Tyndall, J. D. A., Nall, T. & Fairlie, D. P. (2005). *Chem. Rev.* 105, 973–1000.

Van Wart, H. E. & Birkedal-Hansen, H. (1990). *Proc. Natl Acad. Sci. USA,* 87, 5578–5582.

Walls, E. A., Berksom, J. & Smith, S. A. (2002). *Rev. Fish. Sci.* 10, 39–73.

Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A. & Wilson, K. S. (2011). *Acta Cryst.* D67, 235–242.

Winter, W. E., Greene, D. N., Beal, S. G., Isom, J. A., Manning, H., Wilkerson, G. & Harris, N. (2020). *Adv. Clin. Chem.* 94, 31–84.

Xu, D., Zhou, J., Lou, X., He, J., Ran, T. & Wang, W. (2017). *J. Biol. Chem.* 292, 5195–5206.

Yealland, T. & Djordjevic, S. (2016). *Structure,* 24, 2008–2015.

Yiallouros, I., Kappelhoff, R., Schilling, O., Wegmann, F., Helms, M. W., Auge, A., Brachtendorf, G., Berkhoff, E. G., Beermann, B., Hinze, H. J., König, S., Peter-Katalinic, J. & Stöcker, W. (2002). *J. Mol. Biol.* 324, 237–246.

Zundert, G. C. P. van, Moriarty, N. W., Sobolev, O. V., Adams, P. D. & Borrelli, K. W. (2021). *Structure,* 29, 913–921.