



## OPEN Targeting *Aspergillus* allergen oryzin with a chemical probe at atomic precision

Olivia N. Pattelli<sup>1,2</sup>, Dinh Dinh Ly Diec<sup>1,2</sup>, Wanting Guo<sup>1,2</sup>, Silvia Russi<sup>3</sup> & Daniel Fernandez<sup>1,2</sup>✉

We report the molecular basis of *Aspergillus fumigatus* oryzin, allergen Asp f 13, or alkaline proteinase ALP1, containing the sequence motif His–Asp–Ser of the subtilisin family, structure, and function at atomic detail. Given the resolution of the data (1.06 Å), we use fragment molecular replacement with ideal polyalanine  $\alpha$ -helices to determine the first crystal structure of oryzin. We probe the catalytic serine through formation of an irreversible bond to a small molecule compound, specifically labeling it, describing the amino acid residues performing the catalytic function. Defined by a self-processed pro-peptide, the active site architecture shapes up pocket-like subsites that bind to and unveil the S1–S4 substrate binding preferences. We use molecular modeling to dock a model of the pro-peptide in the S1–S4 region and to dock collagen along the active site cleft. Opposite to the face harboring the catalytic serine, the enzyme binds to a calcium ion in a binding site created by backbone flipping. We use thermal unfolding to show that this metal ion provides structural stability. With no known host inhibitor identified thus far, this structure may hasten the progress of developing new therapeutic agents for diseases caused by pathogenic fungi.

*Aspergillus* spp. are microorganisms commonly found in the environment that produce airborne fungal spores. Several hundred spores can be inhaled and eliminated per day without harm to the host; however, in individuals predisposed to infection through lung disease and in patients undergoing cancer chemotherapy or immunosuppressive therapy following organ transplants, aspergilli can cause invasive pulmonary aspergillosis, which can be a life-threatening condition<sup>1,2</sup>. Several studies have shown that patients with severe COVID-19 had a high risk of developing invasive fungal infections, including aspergillosis<sup>3</sup>. Among aspergilli, the World Health Organization has designated *Aspergillus fumigatus* as a critical priority pathogen<sup>4</sup>. Azoles are the first choice of drug for the treatment of aspergillosis; however, reports of antifungal resistance are increasing in the literature<sup>5,6</sup> such that up to 20% of *Aspergillus* isolates exhibit de novo resistance to commonly used antifungal agents<sup>7</sup>. Aspergillosis is an increasingly difficult condition to diagnose and treat<sup>8,9</sup>. Recent data suggests that the pathogen's genetic pool for key metabolic and catabolic enzymes in human infection is more versatile than originally thought<sup>10</sup>. More research into available molecular targets will help us understand the complexities of this organism and its relation to host virulence, drug resistance, and target characterization with the goal of developing new therapeutics.

Fungal colonization of the lungs requires a vast supply of hydrolytic enzymes. *Aspergillus* spp. degradome includes 111 peptidase genes (about 1% of the genome) coding for 14 different types of peptidases<sup>11</sup>. The most studied of such peptidases is the alkaline serine peptidase oryzin (also known in the medical literature as allergen Asp f 13, alkaline proteinase Alp1, and elastinolytic serine proteinase). By a conserved sequence motif, and the Asp–His–Ser triad arrangement, oryzin belongs to the SB clan, S8A subfamily of subtilisins<sup>12</sup>. Initially, oryzin clinical isolates were found to hydrolyze elastin, a key protein component of the connective tissue, and fibrinogen, a blood-circulating glycoprotein, hence mechanistically this enzyme was characterized as elastinolytic or fibrinogenolytic<sup>13,14</sup>. Different *Aspergillus* strains including *A. fumigatus*, *A. flavus*, *A. nidulans*, and *A. terreus* showed a strong digestion of collagen mediated by abundantly secreted oryzin when grown in a collagen medium<sup>15,16</sup>. Another study showed oryzin efficiently cleaved complement components C3, C4, C5, and C1q, as well as immunoglobulin G, and that this was the only serine protease in culture supernatant performing this activity, such that it could be prevented by abrogating the oryzin gene<sup>17</sup>. Additionally, oryzin isolated from fungal spores was shown to detach cultured Vero cells suggesting oryzin's ability to remove attachment without

<sup>1</sup>Sarafan ChEM-H, Stanford University, Stanford, CA 94305, USA. <sup>2</sup>Macromolecular Structure Knowledge Center, Stanford University, Stanford, CA 93405, USA. <sup>3</sup>Structural Molecular Biology Group, Stanford Synchrotron Radiation Lightsource (SSRL), SLAC National Accelerator Laboratory, Menlo Park, CA 94205, USA. ✉email: danilo@stanford.edu

impairing the cells<sup>18</sup>. In patients with asthma, *Aspergillus* may cause a spectrum of conditions from allergic bronchopulmonary aspergillosis to severe asthma with fungal sensitization. In murine models of fungal asthma, oryzin was found as an essential mediator for the recruitment of inflammatory cells and remodeling of the airways<sup>19</sup>, disrupting airway smooth muscle cell-extracellular matrix interactions, and eroding epithelial cell junctions<sup>20–22</sup>. Using a deletion mutant of the oryzin gene, Namvar et al.<sup>19</sup> found that oryzin was indispensable for the initiation and progression of fungal asthma. In summary, oryzin activity on host proteins during infection serves as a promising route for the development of diagnostic tools and inhibitors to target this peptidase. However, the molecular basis for substrate recognition is largely unknown and is difficult to rationalize based on primary sequence alone. To provide a deeper insight into the function and structure of the *Aspergillus* secreted serine peptidase oryzin, we determined its first three-dimensional structure by X-ray crystallography. We obtained complexes with its pro-peptide segment, defined the substrate binding pockets, a tetrahedral irreversible adduct with the catalytic serine, captured the interactions between the key residues and calcium in catalysis, defined the metal ion binding site that provides structural stability.

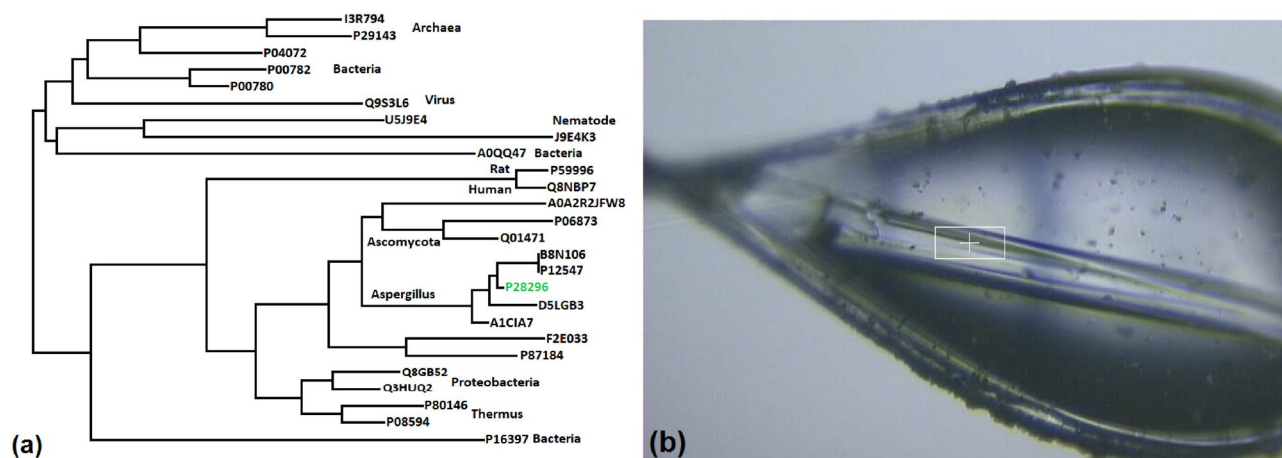
## Results

### Oryzin folds as an $\alpha/\beta$ -hydrolase enzyme used by prokaryotic and eukaryotic cells

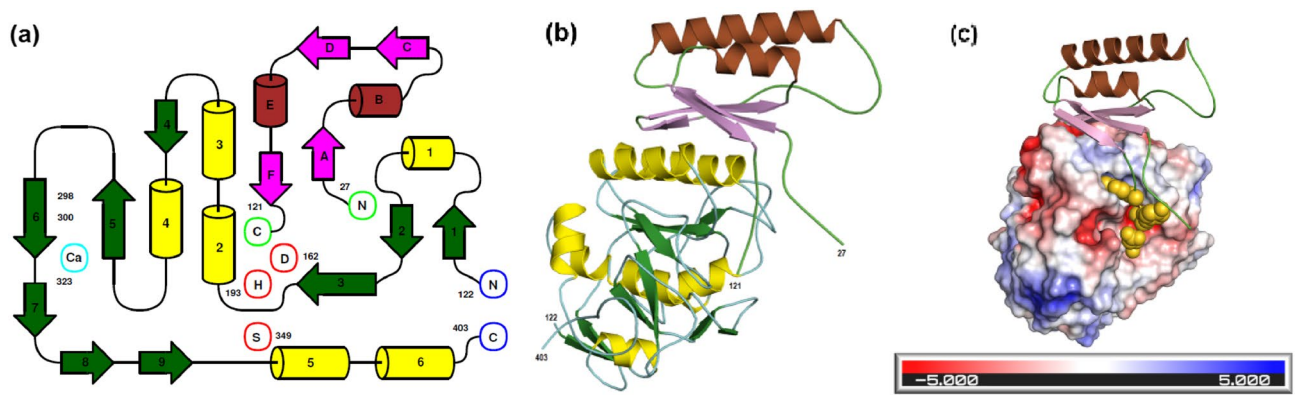
Proteolysis is the mechanism by which peptide bonds in proteins are broken, creating new peptides and aminoacids. This splitting reaction is facilitated by peptide-degrading enzymes called proteases or peptidases or peptide hydrolases. Peptidases differ by the nature of the active group that drives the hydrolytic reaction, with serine peptidases as the most abundant in nature, followed by metallo, cysteine, aspartate and threonine proteases, respectively<sup>23</sup>. Typically, serine peptidases are synthesized as large multi-segmented polypeptides (zymogens) that are activated auto-catalytically and/or exogenously. Previous work with clinical isolates of oryzin-producing spores identified the Asp–His–Ser catalytic signature typical for serine peptidases of the subtilisin type<sup>24–26</sup>. Given the importance of serine proteases in key biological reactions carried out both inside and outside the cell, they are present among all taxonomic kingdoms, from archaea to animals (Fig. 1a). Specifically, the extracellular activity of oryzin is being investigated for its involvement in aspergillosis and fungal-associated diseases. The reactive sites and mode of substrate binding of oryzin are largely unknown and are difficult to predict based on sequence alone. To structurally characterize oryzin, we recombinantly produced it in bacterial cells and purified it to homogeneity. We then obtained crystals of oryzin (Fig. 1b), performed X-ray diffraction, and determined the three-dimensional structure.

The crystal structure of oryzin consists of two polypeptide chains, the pro-peptide (residues 27–121) and the catalytic domain (residues 122–403), originating from a single polypeptide precursor (Fig. 2a). Cleavage at the junction between Asp121–Ala122 releases a new N-terminal segment that displaces to a new position in the mature structure. Upon bond breaking the initially contiguous amino acids are separated by 36 Å. The segment containing residues 122–132 pivots on Gly132 to its final position away from the catalytic site and rebuilds to form the new  $\beta$ -strand 1. This process is similar to the maturation of *Aspergillus fumigatus* secreted zinc-metalloprotease<sup>28</sup>, where the P1'-peptide refolds to a new secondary structure element in a new location. The oryzin pro-peptide fold consists of two  $\alpha$ -helices packed against a  $\beta$ -sheet with a C-terminal coil partially occupying the catalytic cleft (Fig. 2b). The oryzin pro-peptide is attached to the catalytic domain by non-covalent interactions mediated by C-terminal residues, Trp118–Leu119–Tyr120–Asp121 (this will be discussed in a later section). Oryzin displays one face crossed by a long cleft that is predominantly negatively charged at the entry of the catalytic site (Fig. 2c). As shown below, these charges and surface properties will help position oryzin to bind to its substrate.

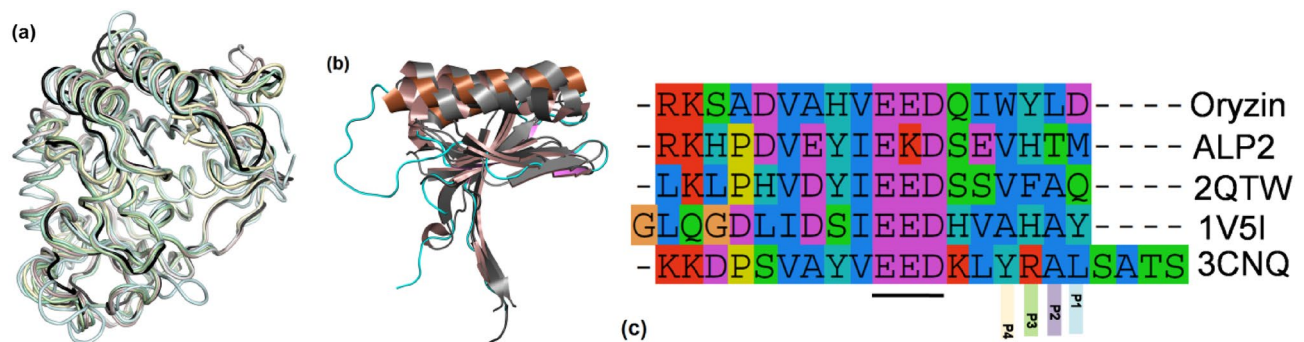
Oryzin has a globular fold that conforms to the  $\alpha/\beta$  three-layered sandwich type resembling eukaryotic and prokaryotic subtilisin-like peptidase enzymes (Fig. 3a). A 3D structural comparison using the DALI server<sup>31</sup>



**Figure 1.** Distribution of subtilisins among taxonomic kingdoms. (a), phylogenetic distribution. Codes are from UniProtKB<sup>27</sup>. P28296, *Aspergillus fumigatus* oryzin (this work, in green). (b), oryzin crystallized as a long rod (dimensions  $10 \times 20 \times 400 \mu\text{m}$ ) that was used for a helical data collection strategy with microfocused X-ray beam along the longest dimension of the crystal. Box size dimensions:  $50 \times 15 \mu\text{m}$ .



**Figure 2.** Oryzin is a binary complex resulting from proteolysis and trapping of cleavage product. (a), topological diagram of oryzin indicating secondary structure elements and key residues. Catalytic domain (green and yellow), pro-domain (magenta and brown). N- and C-termini, catalytic triad, and divalent metal binding site are marked as circles (blue, green, red, sky-blue). Figure prepared with Topdraw<sup>29</sup>. (b), cartoon depicting oryzin with secondary structure elements colored as in panel (a). (c), electrostatic potential surface of oryzin. Pro-domain shown as cartoon, with residues 118–121 depicted as space-filling model. Orientation shifted 45 degrees to facilitate view to the catalytic site (center of figure). Figures in panels (b) and (c) prepared with Pymol<sup>30</sup>.



**Figure 3.** Structure of oryzin catalytic domain and pro-peptide show fold similarity to enzymes from different origins. (a), ribbon representation of the catalytic domain of oryzin (black) and subtilisin-like enzymes (tones of green; shown PDB IDs: 5z6o, 3f7o, 6kkf, 2b6n, 4dzt and 2qtw (except that the C-terminal 250 residues in human kexin were omitted for clarity). Same orientation as in Fig. 2, c. Ca atoms aligned: 5z6o (276 residues, rmsd 0.96 Å); 3f7o (273 residues, rmsd 1.01 Å); 6kkf (266 residues, rmsd 1.06 Å); 2b6n (264 residues, rmsd 0.85 Å); 4dzt (268 residues, rmsd 0.97 Å); and 2qtw (251 residues, rmsd 1.41 Å). (b), cartoon representation of the pro-peptide (colored brown, magenta and cyan) overlay to equivalent regions (grey) in human proprotein convertase subtilisin/kexin type 9 [PDB ID: 2qtw; (80 residues, rmsd 1.84 Å)] and the bacterial subtilisin from *Bacillus amyloliquefaciens* [PDB ID: 3cnq; (62 residues, rmsd 1.71 Å)] and a natural subtilisin inhibitor from the fungus *Pleurotus ostreatus* [PDB ID: 1v5i; (72 residues, rmsd 2.07 Å)]. (c), detail of alignment of oryzin C-terminal tail to subtilisin-like pro-domains and a peptidase inhibitor (from top to bottom: oryzin, ALP2 (Uniprot P81784, *A. fumigatus* intracellular alkaline serine protease), 2qtw, 1v5i, and 3cnq). Underlined, conserved negatively charged residues. P1–P4 residues N-terminal to the scissile bond. Alignment made with Seeview<sup>38</sup>.

top-score hits were with a subtilisin protease from *Penicillium cyclopium*, a plant-infecting fungus [PDB ID: 5z6o;<sup>32</sup>]; a cuticle-degrading protease from *Paecilomyces lilacinum*, an opportunistic fungus commonly found in the soil [PDB ID: 3f7o;<sup>33</sup>]; and, a proteinase K-like enzyme from *Parengyodontium album* (also known as *Tritirachium album*, *Beauveria alba*, and *Engyodontium album*), a causative agent of mycoses [PDB ID: 6kkf;<sup>34</sup>]. Further, bacterial enzymes such as a proteinase K-like from *Serratia* spp. [PDB ID: 2b6n;<sup>35</sup>] and the aqualysin from *Thermus aquaticus* [PDB ID: 4dzt;<sup>36</sup>] are highly similar. In addition, the structure aligns to the human proprotein convertase/subtilisin kexin type 9 [PDB ID: 2qtw;<sup>37</sup>] at a rmsd 1.4 Å over 251 Ca atoms.

There is also significant structural homology in the pro-domain regions of human and bacterial subtilisin despite low sequence similarity (<20%) [Fig. 3b]. These include the pro-domain of the proprotein convertase subtilisin/kexin type 9 (PDB ID: 2qtw) and a bacterial subtilisin from *Bacillus amyloliquefaciens* [PDB ID: 3cnq;<sup>39</sup>]. Interestingly, it also is similar to a natural subtilisin inhibitor from the fungus *Pleurotus ostreatus* [PDB ID: 1v5i;<sup>40</sup>]. Contained in the C-terminal tail of oryzin, there is a common motif involving a stretch of negatively charged residues around position 113 (Fig. 3c). Within this acidic region, His111 stacks to Tyr262, while the carboxyl of Glu113 hydrogen bonds to the main chain N of Tyr262 and the O $\gamma$  of Ser261. Similarly, Gln116 is



hydrogen bonded to the main chain carbonyl atoms of Gly259 and Tyr260. In summary, these interactions anchor and orient the C-terminal tail toward the catalytic site. Next, to precisely delineate the catalytic active site cleft architecture, we use this crystal structure in combination with molecular modeling to determine the structural basis for substrate recognition.

### Oryzin substrate binding specificity

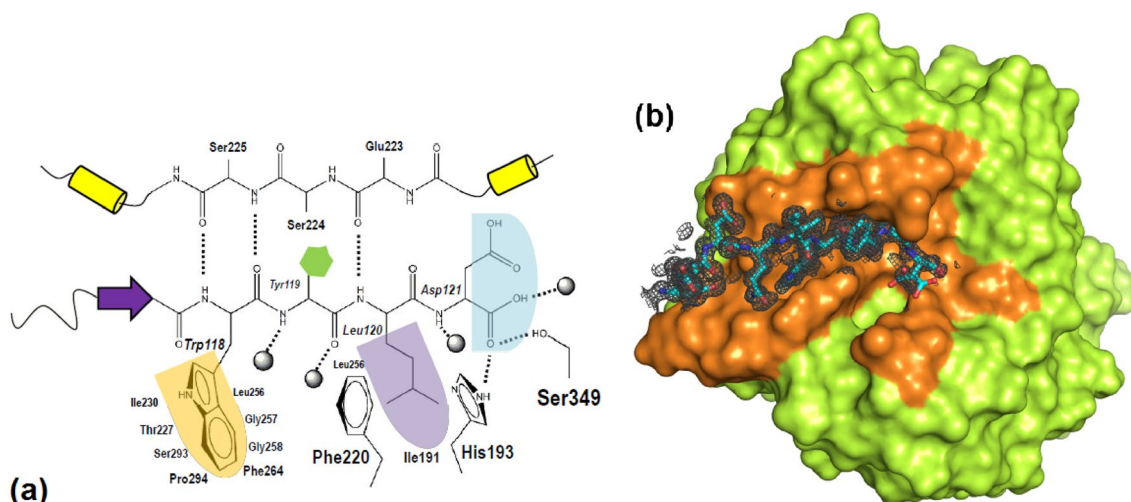
About 12% of the accessible surface area of oryzin catalytic domain ( $\sim 1220 \text{ \AA}^2$  out of total  $\sim 10,080 \text{ \AA}^2$ ) is occupied by pro-domain atoms. Although both the N- and C-termini are unstructured segments, the C-terminal residues Trp118-Leu119-Tyr120-Asp121 from the pro-peptide remain bound in the catalytic cleft. They reveal the S1 to S4 substrate binding pockets in oryzin (Fig. 4a).

Trp118 is buried deep in a hydrophobic pocket lined by several small amino acid side chains where strong hydrogen bonds are established between the main chain atoms of this residue and Ser225 (Fig. 4b). The bulky aromatic moiety of Trp118 forms hydrophobic interactions with Leu256, Ser293, Pro294, Phe264, Ile230, Thr227, Gly257 and Gly258. The Tyr119 side chain makes a hydrophobic contact to Ser224 but mostly faces outward towards the solvent and further side chains in the pro-peptide. Here, the bulky side chain of Phe220 protrudes, further constraining contacts in S3, while most of interactions are through water molecules to the main chain N and O of Gly257. Leu120 is sandwiched between Phe220 and His193 with extra hydrophobic interactions to Leu256 and Ile191. In addition, the N of the amide bond of Leu120 hydrogen bonds to the main chain O of Glu223. Finally, the side chain of P1, Asp121, faces outwards towards the solvent while its main chain O (now the new COOH terminus) is hydrogen bonded to both Ser349 and His193. A further hydrogen bond is mediated through a water molecule to the main chain O of Ser346. An overlay with the human proprotein convertase subtilisin/kexin type 9 (PDB 2qtw; sequence ID 19%, 80 aligned C $\alpha$ s) shows that the equivalent side chains occupying the S4–S1 subsites in the human pro-peptide are valine, phenylalanine, alanine, and glutamine. The bacterial serine proteinase inhibitor from the mushroom *P. ostreatus* (PDB 1v5i; sequence ID 20%, 70 aligned C $\alpha$ s) shows alanine, histidine, alanine, and tyrosine. In the model serine proteinase from *B. amyloliquefaciens* (PDB 3cnq; sequence ID 28%, 67 aligned C $\alpha$ s), the P4–P1 residues are tyrosine, arginine, alanine, and leucine. The comparison to structurally characterized counterparts indicate that substrate binding preferences are for subsites: (S4), bulky aromatic/aliphatic; (S3), large side chains, aromatic or otherwise; (S2), small aliphatic side chain; and (S1), large side chains, aliphatic or not. Interestingly, subsites S4 and S2 show a preference to bury aromatic/aliphatic side chains while also promoting strong hydrogen bonding interactions; in contrast, subsites S3 and S1 seem open to accommodate side chains with lower selectivity. In summary, these binding interactions illustrate a distinct pattern of substrate-binding preferences among different subtilisins. Next, we use molecular modeling to explore the C-terminal subsites and the binding of an oryzin protein substrate.

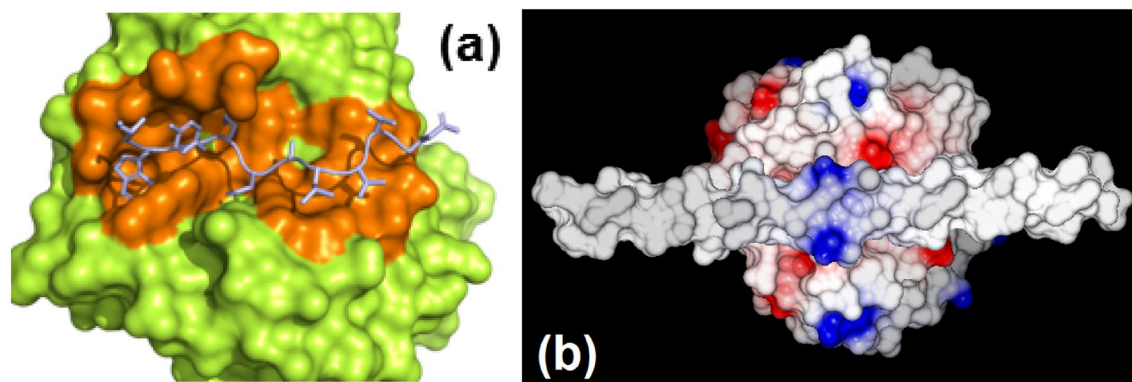
### Molecular modeling—a structural rationale for substrate recognition

To delineate oryzin's binding pockets beyond the S1 site, we performed molecular modeling using a peptide fragment with adjacent P1'–P4' residues extending the pro-domain from C-terminal residue Asp121.

Our model shows a transition from pocket-like on the non-prime subsites to a shallow plain on the prime subsites (Fig. 5a). Covering one face of the enzyme, the catalytic cleft is an elongated groove spanning  $\sim 32 \text{ \AA}$  from S4 to S4' ( $\sim 9$  peptide bonds). The P1–P1' scissile bond (Asp121–Ala122) is  $3.8 \text{ \AA}$  from the O $\gamma$  oxygen of the catalytic serine (Ser349). The P1' and P2' side chains, Ala122 and Leu123, make hydrophobic contacts with the side chains of Ile345 and Trp335 and form hydrogen bonds to main chain atoms anchoring them. These small



**Figure 4.** Oryzin pro-peptide C-terminal tail binding pocket. (a), schematic diagram showing pockets S4 to S1 from left to right (colored yellow, green, violet and sky blue). (b), pro-peptide segment 113–121 (sticks) harboring P4–P1 residues Ile117–Trp118–Tyr119–Leu120–Asp121. At center of figure, Asp121 is in two split orientations. Oryzin shown as a surface (lemon), with pro-peptide-interacting residues (under  $4 \text{ \AA}$  distance) in orange. Electron density map contoured at the  $1.0\text{-}\sigma$  level (mesh in gray).

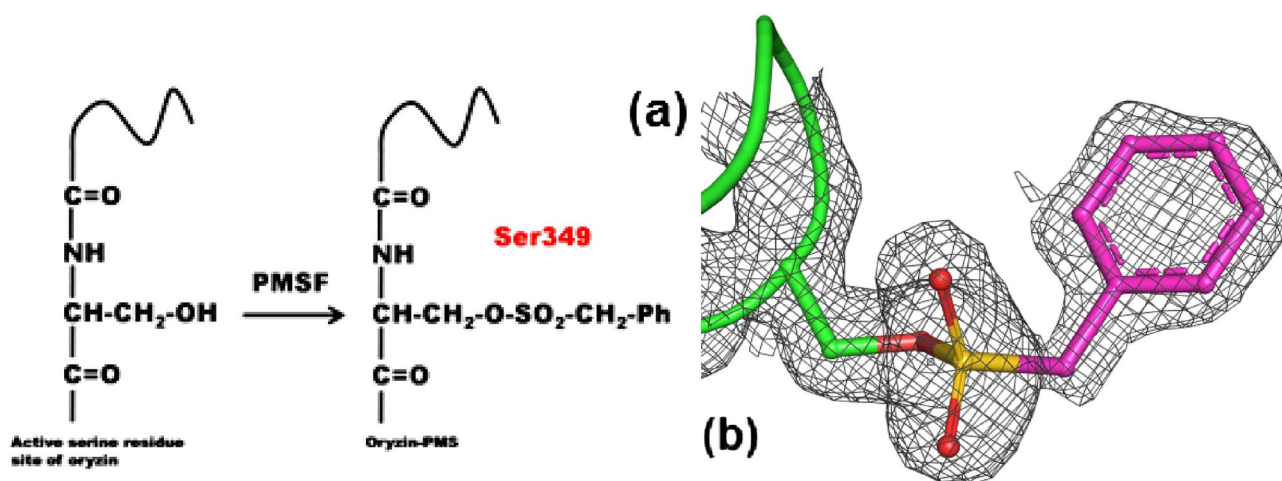


**Figure 5.** The catalytic cleft extends 32 Å or about a 9-aminoacids-long stretch. (a), model of a short peptide encompassing residues 118–126 (sticks representation in gray) binding to oryzin (surface in lime, residues interacting with the peptide highlighted orange). (b), model of collagen docked to oryzin. Oryzin electrostatic potential surface as in Fig. 2.

side chains would be small enough to fit the scissile bond closer to the catalytic serine. Similarly, two further small residues in the sequence, Thr124 (P3') and Thr125 (P4'), would be buried in a depression at Ser346 and Gly347, promoting hydrogen bonding interactions between main-chain atoms. The prime residues are exposed to the solvent more than the non-prime ones, explaining why the new C-terminal segment at residue 122 springs out of the catalytic site after cleavage while the N-terminal residue 121 remains bound. To study the binding between oryzin and its substrate collagen, we used molecular modeling to dock collagen to an oryzin structure that does not contain the pro-domain. Collagen is a neutrally charged polypeptide, except for a segment of positively charged arginines (Fig. 5b, center of figure in blue) that would interact with the negatively charged catalytic pocket in oryzin. In summary, the architecture of oryzin catalytic cleft facilitates binding of extended protein substrates along the cleft's axis, exposing the scissile peptide bond to the catalytic machinery. We next use a chemical probe to confirm the identity of the catalytic active site residue and the oxyanion binding site in oryzin.

#### **The hydrogen bonding in the Asp–His–Ser catalytic triad and the oxyanion binding site**

The mechanism of the catalytic triad in serine peptidases is highly studied and well understood. In serine peptidases, the primary hydroxyl group of the active site serine residue acts as the nucleophile. The histidine activates serine by acting as a base and accepting the serine's single proton. The polarized hydroxyl group attacks the carbonyl carbon resulting in a negatively charged tetrahedral intermediate that is stabilized by a hydrogen bond from a nearby asparagine that constitutes the oxyanion binding site. Finally, the amine group leaves and the tetrahedral intermediate is broken by a histidine-activated water molecule<sup>41</sup>. We use phenylmethanesulfonyl fluoride (PMSF), a chemical compound that sulfonylates serine to form a stable adduct, to introduce an oryzin-PMS crystal structure (Fig. 6a). The covalent attachment between the inhibitor and the O<sub>γ</sub> of Ser349 shows that the enzyme is fully active and confirms the identity of Ser349 as the oryzin catalytic reactive serine (Fig. 6b).



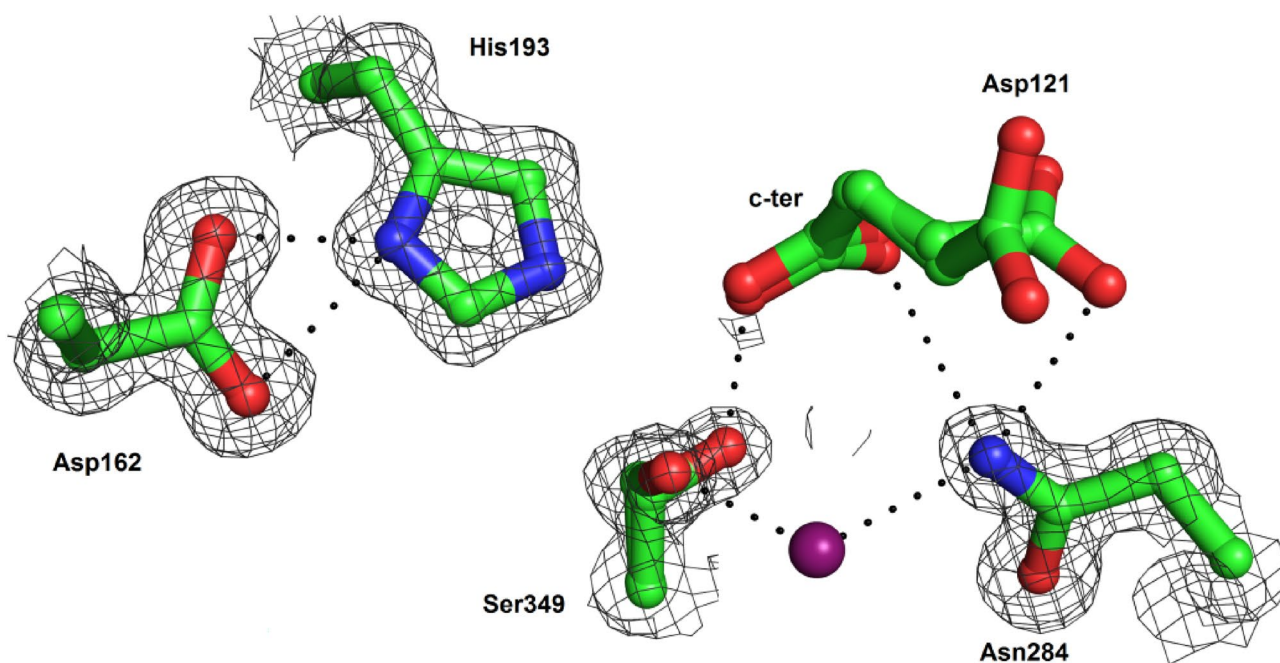
**Figure 6.** Oryzin-PMS covalent bonding crystal structure probes Ser349. (a), chemical scheme for reacting a serine to the irreversible inhibitor PMSF. (b), Ser349 in oryzin-PMS crystal structure. The covalent bonding to Ser349 is verified. Atoms rendered as sticks in green (carbon, protein), magenta (carbon, PMS), gold (sulfur), and red (oxygen). Omit-calculated electron density map (black mesh) contoured at the 1.0-σ level.

Next, we analyze the oryzin Asp–His–Ser triad hydrogen bonding interactions (Fig. 7). The pair His193 and Asp162 are at hydrogen bonding distance sharing the N $\delta$ 1 hydrogen. We hypothesize that histidine may be in its ionized form. For this to occur, the N $\epsilon$ 2 atom of His193 would have to abstract a hydrogen atom from Ser349. These atoms are 3.3 Å apart, at hydrogen bonding distance, but neither of the two split orientations of the hydroxyl group of Ser349 (this side chain is disordered and modeled in two sites) points in the direction of the histidine. One of the two disordered oxygens of Ser349 is hydrogen bonded to the C-terminal carboxyl of Asp121 (residue disordered and modeled at two alternative sites). The second disordered oxygen of Ser349 is hydrogen bonded through a water molecule to the nearby Asn284. Asn284 hydrogen bonds to the C-terminal carboxyl and the side chain of Asp121. Asn284, a conserved residue in the subtilisin family, is the oxyanion binding site in oryzin.

The side chain hydroxyl group of Ser349 builds a hydrogen binding network with the C-terminal carboxyl group of Asp121 and the side chain amide nitrogen of Asn284 (through a water molecule). In summary, the geometry around base His193, hydrogen bonded to Asp162 farther from the catalytic Ser349 O $\gamma$ , and the latter hydroxyl group pointing away His193 in the direction of Asn284 and the new C-terminal carboxyl, would suggest that the enzyme's structure represents an end-product state. We then use a peptide substrate to analyze the enzyme's activity after its activation through chemical unfolding.

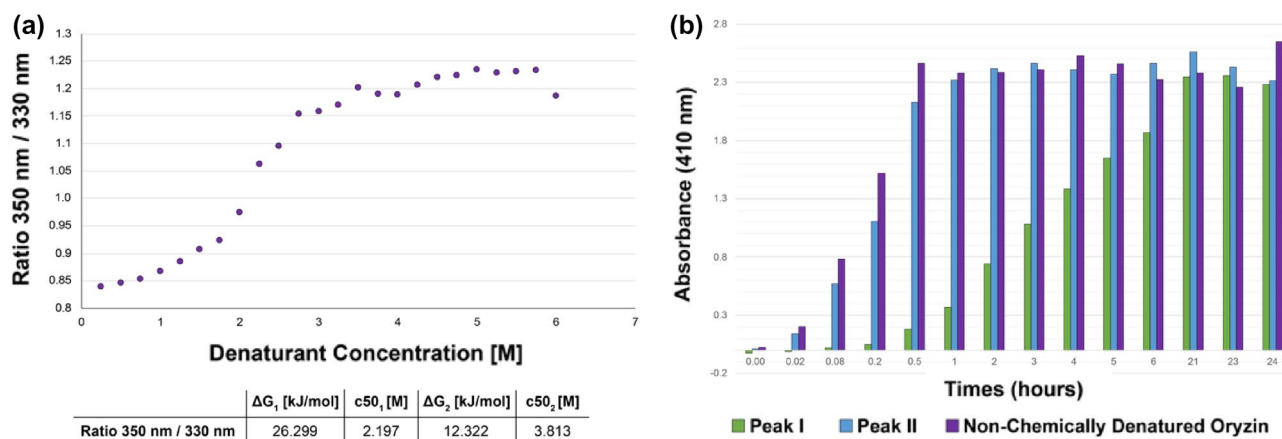
### Enzyme activity

Oryzin is expressed as a single polypeptide that can self-cleave the pro-peptide portion of the protein. The cleaved pro-peptide remains bound, partially occupying the catalytic cleft and hindering access to the catalytic Ser349. Previous work established that chemical denaturants, such as urea, can be used to dissociate an enzyme's catalytic domain from its pro-peptide<sup>42</sup>. Though urea is a commonly used chemical denaturant, solutions of urea are not stable and are subject to decomposition. To circumvent this, we use guanidine hydrochloride (GuHCl). To determine the ideal denaturant concentrations for purification, we perform chemical unfolding of oryzin with differential scanning fluorimetry (DSF) monitoring changes in fluorescence as a function of GuHCl concentration. We determined that oryzin has two major unfolding events at 2.2 M and 3.8 M GuHCl (Fig. 8a). Due to the shorter length of the pro-peptide sequence and the nature of its tertiary structure (about 50% of its fold is loops or unstructured regions; Fig. 2b), we hypothesized that this portion of the zymogen will unfold first, suggesting that denaturing at this concentration of GuHCl is sufficient for the unfolding and dissociation of the pro-peptide and the catalytic domain of oryzin. Oryzin was expressed, the lysate was incubated with 2.2 M GuHCl, and then purified using immobilized metal affinity chromatography (IMAC) and size exclusion chromatography (SEC) to isolate two distinct species as shown by SDS-PAGE chromatography (Supplemental Figure S2). Peak I contained a major band with an estimated molecular weight of 28 kDa, and a 21 kDa band suggesting one or more events of hydrolytic cleavage. Peak II contained a major band with an apparent molecular weight of 33 kDa, indicating a distinctly cleaved oryzin containing the two minor bands of 28 and 21 kDa. This is in agreement with several previous works that isolated oryzin from hospital samples predominantly as 32–33 kDa products and degradation products at 28 and 23 kDa and smaller<sup>13–15,26</sup>. We then use a previously identified synthetic substrate such



**Figure 7.** Hydrogen bonding interactions in oryzin catalytic site. Atoms rendered as sticks in green (carbon), blue (nitrogen), and magenta (oxygen). Hydrogen bonds shown as dotted lines (black). Electron density map (gray mesh) for catalytic site residues contoured at the 1.0- $\sigma$  level.





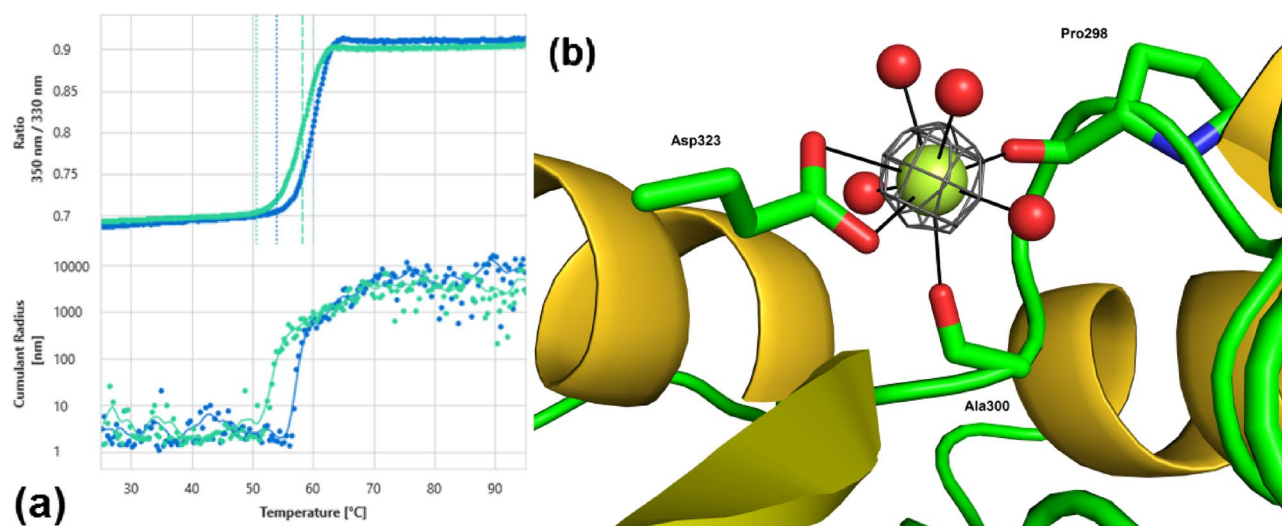
**Figure 8.** Chemical unfolding of oryzin leads to cleavage of the pro-peptide and slowed enzymatic activity. (a) Chemical unfolding of oryzin using increasing concentrations of GuHCl as shown as a fluorescence ratio 350 nm/330 nm. (b) Enzyme–substrate activity assay using peak I, peak II, and non-chemically denatured oryzin samples. The substrate used was Suc-Ala-Ala-Pro-Phe-pNA. Activity of each sample was measured at 410 nm.

as Suc-Ala-Ala-Pro-Phe-pNA (Suc-AAPF-pNA)<sup>13</sup> for the determination of oryzin enzymatic activity of peaks I and II and a non-chemically-denatured sample (33 KDa, see Supplemental Figure S1).

We found that while all three samples reach similar hydrolytic activity after 24 h (Fig. 8b), the peak I sample does so more slowly than the Peak II and the non-denatured sample (green vs. blue/purple bars). This suggests that chemical denaturation of the protein results in the dissociation of the pro-peptide with concurrent enzyme cleaving itself to lower molecular weight products, weakening its enzymatic activity. An alternative could be that degradation products could remain bound and block enzymatic activity.

### Calcium-binding stabilizes the enzyme through backbone remodeling

Earlier work demonstrated that ethylenediaminetetraacetic acid (EDTA), a compound typically used for sequestering metal ions, could render oryzin inactive<sup>14</sup>. However, other researchers found EDTA did not inhibit activity, but divalent metal ions, like calcium, had an activity-modifying effect<sup>15</sup>. To better understand the role of divalent metals on oryzin, we use differential scanning fluorimetry (DSF) to measure the unfolding of oryzin in the presence of calcium, by monitoring changes in fluorescence as a function of temperature. We found that oryzin unfolds at a lower  $T_m$  (the  $T_m$ s are 58.38 (0.07) °C and 60.09 (0.02) °C in the absence and presence of calcium, respectively) compared to oryzin in the presence of calcium (Fig. 9a). Calcium at less than 10  $\mu$ M

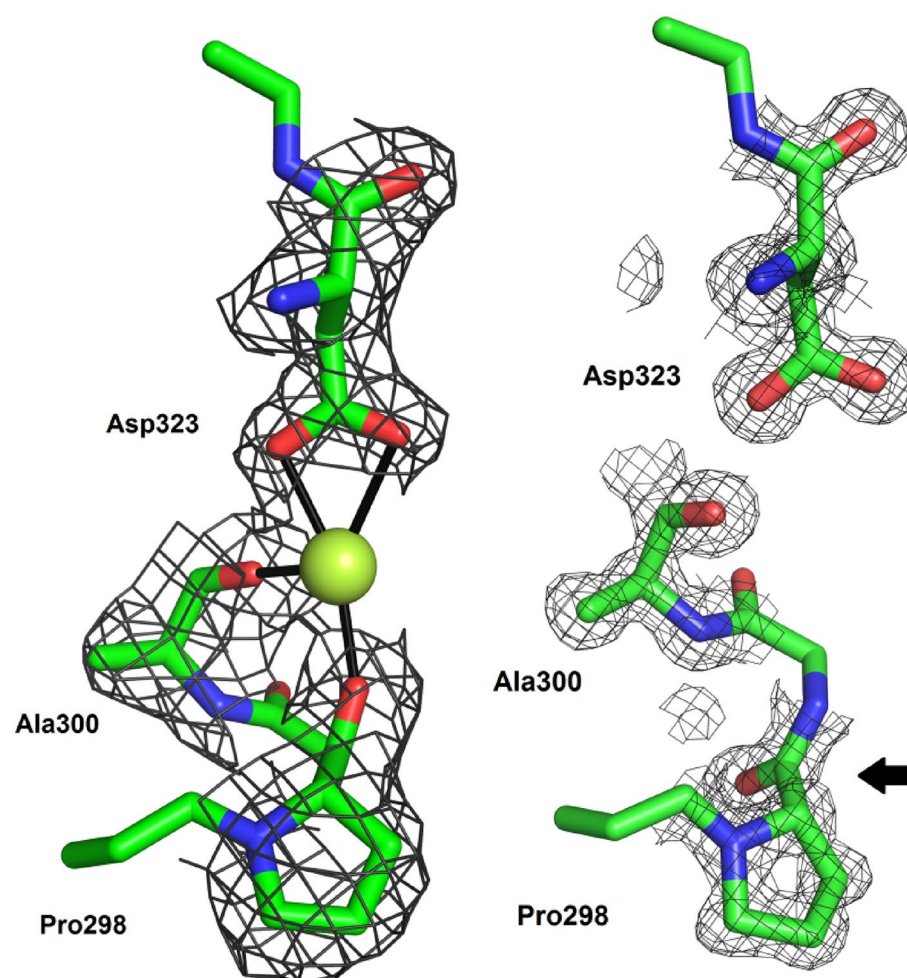


**Figure 9.** Oryzin thermal unfolding is slowed by calcium binding to a pocket created by backbone flipping. (a), nanoDSF trace of oryzin melting. On increasing temperature, in the absence of calcium (green) oryzin melts (*top panel*) and aggregates (*bottom*) at a lower temperature than in the presence of calcium (blue). (b), calcium binding site in oryzin. Shown as sticks Pro298, Ala300, and Asp323 (carbon atoms, green; nitrogen, blue; oxygen, red). Calcium is as a sphere (lemon) with bonding interactions (solid lines in black). Anomalous electron density map (grey mesh) contoured at the 4.0- $\sigma$  level.

caused no variation of the transition temperature, but the  $T_m$  steadily increases with increasing calcium from 20  $\mu\text{M}$ . A  $\Delta T_m = 1^\circ\text{C}$  is achieved at 312  $\mu\text{M}$  calcium. Since oryzin is secreted it may be exposed to an as high as 2 mM extracellular calcium environment<sup>43</sup> suggesting its physiological significance. While the difference in  $T_m$  is small ( $\Delta T_m = 2^\circ\text{C}$ ), these changes in  $T_m$  in the presence of calcium compares to what has been previously observed in subtilisin with similar magnitudes<sup>44</sup>. Similarly, calcium increases the transition temperature at which the protein aggregates resulting in an increase in the unfolding  $\Delta T_m$  of the protein. Oryzin begins to aggregate at 57.40 (0.12)  $^\circ\text{C}$  and 59.39 (0.06)  $^\circ\text{C}$  in the absence and presence of calcium, respectively, following the trend observed in the thermal unfolding experiment. As the effect of calcium is to structurally stabilize the molecule, the questions as to where the calcium binding site locates and what its binding means to the protein fold, are to be solved. To this end, we analyzed a third crystal form, oryzin-calcium, obtained with calcium as an additive in the crystallization cocktail.

The calcium ion is placed at the center of an almost regular octahedron formed by oxygen ligands (Fig. 9b). Three of these ligands belong to the protein: carbonyl groups of Pro298 and Ala300 and the carboxylate of Asp323. Four other oxygens are provided by water molecules indicating that the calcium site is fully exposed to the solvent. The apex of the octahedron is defined by the Pro298–Asp323 axis while the rest of binders lie on a plane equatorially. The oxygen–calcium distances are all about 2.3–2.5  $\text{\AA}$ , typical for calcium–oxygen contacts.

The oryzin–calcium crystal shows that a calcium site is in a turn between sheets 6 and 7 located opposite to the catalytic cleft face (Fig. 10). An overlay between oryzin and oryzin–calcium reveals that calcium binding has a minimal effect on overall fold (rmsd alignment of 0.35  $\text{\AA}$  along 282 C $\alpha$ s). However, the largest offset is at position 298 (rmsd of 2.1  $\text{\AA}$  between Pro298 C $\alpha$ ). This calcium binding site is created by backbone rearrangements in oryzin at Pro298 and stabilized by hydrogen bonding interactions between Glu269 and the amide N atom of Pro298, and the amide N and side chain of Asn299. Interestingly, while Asp323 is conserved among subtilisins, Glu269 is not. For example, in *P. cyclopium* subtilisin, an alanine occupies the equivalent place of Glu269, the proline is further from Asp323, and its carbonyl O atom is replaced by a water molecule. The calcium ions are



**Figure 10.** Backbone remodeling at Pro298–calcium binding site. Detail of binding interactions for calcium (water molecules omitted for clarity). Pro298 carbonyl oxygen provides one of oxygen ligands to the calcium, but its carbonyl is flipped (arrow) when unbound. Atoms rendered as sticks in green (carbon), blue (nitrogen), and red (oxygen). Calcium (sphere, lemon). Electron density map (grey mesh) contoured at the 1.0- $\sigma$  level.



in the same position in both structures. Moreover, in human pro-protein convertase, the calcium site relates to the oryzin calcium site (they are further away to 2.5 Å). In the unbound form (Fig. 10, arrow), the backbone at Pro298–Asn299 bond is flipped outward of the turn and cannot form a binding site for calcium.

## Discussion

We report on the first crystal structure of *Aspergillus fumigatus* oryzin. This structure shows a product-trapped catalytic domain with its self-cleaved pro-domain peptide segment marking the enzyme's substrate prime binding pocket preferences. Analysis of the complex with the pro-peptide and modeling illustrates oryzin's S4'–S4 substrate binding pockets differ from the typical subtilisin-like  $\alpha$ , $\beta$ -hydrolases. In complex to an irreversible inhibitor, we label and confirm the identity of the catalytic machinery around the catalytic serine (Ser349). Combining crystallography and in vitro assays, our work confirms that oryzin binds calcium, that calcium stabilizes the enzyme, that the metal binding site is rather flexible and solvent exposed further from of the catalytic machinery. Thermal unfolding analysis combined with crystallographic data shows that the binding of calcium results in local structural changes, without detriment for the overall fold. Together with earlier observations our work suggests that enzyme activity changes are due to extra cleavage events that would result in changes in the protein integrity or because of products of hydrolysis directly modulating the catalytic mechanism by physical hindrance. Further work will be needed to address the complexities of this serine hydrolase substrate recognition properties and progression to its mature state. Diseases caused by *Aspergillus* spp. remain a challenging problem<sup>45</sup> and oryzin is well suited for target-based drug development<sup>46</sup>. In the human genome, serine proteases are a predominant class of enzymes<sup>47</sup> making the design of *Aspergillus*-specific inhibitors a difficult endeavor. Structure-based screening in exploratory research of new inhibitors is well established<sup>48–50</sup> and may assist in the design effort. Combined with the present structural work it will help progress research into new agents to block progression of aspergillosis.

## Experimental procedures

### Oryzin cloning, expression, and purification

A plasmid for expression of the *A. fumigatus* sequence [UniProtKB<sup>27</sup> ID: P28296; *Neosartorya fumigata* (strain ATCC MYA-4609/Af293/CBS 101355/FGSC A1100)] encoding a truncated protein lacking the 20 amino acids from the secretion signal and containing a C-terminal His<sub>6</sub>-tag cloned into a pD444 vector with ampicillin resistance and IPTG inducible T5 promoter was purchased as a synthetic gene (ATUM, Newark, CA, USA). After transformation of protease deficient *E. coli* strain BL21(DE3) [New England BioLabs, Ipswich, MA, USA], transformants were grown in lysogeny broth (Thermo Fisher Scientific, Waltham, MA, USA) containing ampicillin (100 µg/ml) at 37 °C in a shaking incubator at 225 rpm. The culture was grown until an optical density of 0.6–0.8 was reached and then transferred to a shaking incubator at 16 °C for 1 h before induction with 1 mM IPTG (G-Biosciences, St. Louis, MO, USA). After 3–4 h post induction, the culture was spun down at 10,000×g for 60 min at 4 °C, resuspended in 1/50 culture volume of buffer (150 mM NaCl, 16 mM Na<sub>2</sub>HPO<sub>4</sub>, 4mM NaH<sub>2</sub>PO<sub>4</sub> [pH 7.3]), and, after addition of lysozyme and Triton X-100, incubated in a rotator shaker at 4 °C for 10 min. Cells were lysed by sonication in a Fisherbrand Model 705 Sonic Dismembrator (ThermoFisher Scientific) three or four times for 30 s each on ice, and the lysate was cleared by centrifugation at 12,000×g for 60 min at 4 °C. The supernatant was mixed with 1 ml of Ni-NTA resin (Qiagen, MD, USA), incubated in a rotator shaker at 4 °C for 30 min, and the resin was applied on a plastic column. After washing with 10 bed volumes of 20 mM Tris–HCl, 500 mM NaCl, 20 mM imidazole, pH 8.5, elution of the His<sub>6</sub>-tagged protein was performed with an imidazole gradient, 1 bed volume of 10–500 mM imidazole, stepwise by gravity. The purity of the eluted fusion protein was analyzed by SDS-PAGE (NuPAGE, Thermo Fisher Scientific). We attempted to capture the intact uncleaved zymogen by using acidic pH buffers, but we consistently obtained the separate polypeptides (judged by SDS-PAGE) suggesting that the cleavage products must have appeared very early during protein expression. Fractions containing pure protein as determined by SDS-PAGE analysis were loaded onto a Superdex 75 column for size-exclusion chromatography using a Akta Pure (Cytiva, Marlborough, MA, USA) system. The protein was concentrated to 30 mg/ml using Amicon Ultra 10 kDa MWCO spin filters (Millipore Sigma), in buffer 0.1 M MES, 0.05 M NaCl, pH 5.6, and aliquots stored at –80 °C. For chemically denatured protein purification, oryzin lysate was incubated with 2.2 M GuHCl (as determined by chemical unfolding experiments) for 1 h at room temperature to allow for chemical unfolding of the protein. Protein was further purified following the protocol from above.

### Crystallization, data collection, structure solution and refinement

Crystals were obtained using sitting-drop vapor diffusion for screening and sitting- or hanging-drop vapor diffusion for optimization at 6 °C or 16 °C. Crystallization experiments were setup using a Douglas Oryx8 Nanodrop dispensing robot (Douglas Instruments Ltd, East Garston, United Kingdom). Diffraction-quality crystals were obtained from a solution containing 20% (w/v) polyethyleneglycol 550 monomethylether, 10% (w/v) polyethyleneglycol 20,000, sodium nitrate, sodium dihydrogen phosphate, ammonium sulfate (each 0.03 M), in 0.1 M MES/imidazole buffer at pH 6.5. Crystals were harvested and cryocooled by plunging in liquid nitrogen with or without addition of cryoprotectant (polyethyleneglycol 200 or 1,2-ethanediol). Crystals were screened for their diffracting power at 100 K at beamlines BL14-1 and BL12-2<sup>51</sup> at the Stanford Synchrotron Radiation Lightsource (SSRL) at SLAC National Accelerator Laboratory (Menlo Park, CA, USA). The best diffracting sample was up to a Bragg spacing of 1 Å and data was collected taking advantage of the macroscopic properties of the crystal. To mitigate radiation damage, we employed a helical data collection strategy on a rod-shaped crystal using fine slicing by translating the microfocussed beam along the longest edge of the crystal. The crystal is in space group P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> with cell dimensions a: 60.60 Å, b: 70.76 Å, c: 78.55 Å;  $\alpha$ ,  $\beta$ ,  $\gamma$ : 90° with one polypeptide chain in the

asymmetric unit. Solving the phase problem using high-quality diffraction data to find one single polypeptide chain consisting of 382 aminoacids at ~ 1 Å resolution proved to be practicable using a method to place idealized peptide fragments via molecular replacement iterated with automated cycles of density modification. Phasing of the 1.06 Å data was with fragment location with ideal polyalanine  $\alpha$ -helices as search fragments. Calculations were performed with ARCIMBOLDO\_LITE<sup>52</sup>, a method which has previously been shown to successfully phase a 385-residues target protein at a lower resolution. This method uses Phaser<sup>53</sup> for searching combined with density modification and autotracing with SHELXE<sup>54</sup>. ARCIMBOLDO\_LITE runs in two macrocycles: first, Phaser sequentially places all requested copies of the search model(s) and partial structures are ranked and optimized with SHELXE. Similar solutions are clustered; then, top-ranked substructure solutions (according to top LLG and CC; a CC > 25% indicates a possibly solved structure) are selected and subjected to density modification and autotracing with SHELXE. A promising substructure comprising 8  $\alpha$ -helices was expanded to trace a polyalanine-only model encompassing 367 residues (out of 382 from sequence). This solution gave a final CC of 49.89 and was subsequently chosen for refinement. This partial model was completed by successive rounds of manual adjustment and model building with COOT<sup>55</sup> and refinement with REFMAC5<sup>56</sup>. The electron density map was of excellent quality and allowed for unambiguously tracing the pro-domain, residues Arg27–Asp121 (21–26 missing), while the catalytic domain could be traced fully from Ala122–Ala403. Refinement progressed to convergence and reached an excellent agreement to the experimental data (see Table S1 for crystal data quality parameters and refinement statistics). Additional molecules, namely formate, imidazole, and polyethyleneglycol were added in later stages of refinement. To test bond geometry hypotheses, we used SHELXL<sup>57</sup> at various stages of refinement. Solvent water molecules were first assigned based on their hydrogen bonding properties. To prepare other samples for crystallization, either 10 mM PMSF or CaCl<sub>2</sub> were added to the protein buffer before concentration. Oryzin-calcium was obtained from a crystallization condition containing 12.5% (w/v) PEG 1000, 12.5% (w/v) PEG 3350, 12.5% (v/v) MPD; L-glutamate, DL-alanine, glycine, DL-lysine, and DL-serine (each 0.02 M), in 0.1 M MOPS/HEPES, pH 7.5, with CaCl<sub>2</sub> as an additive. The crystal is in space group P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> with cell dimensions a: 59.25 Å, b: 66.07 Å, c: 86.79 Å;  $\alpha$ ,  $\beta$ ,  $\gamma$ : 90° with one polypeptide chain in the asymmetric unit. Data at 1.65 Å Bragg spacing was collected at SSRL beamline BL12-2. The condition for the 1.06 Å crystal was repeated to obtain oryzin-PMS containing PMSF as an additive. The crystal is in space group P2<sub>1</sub> with cell dimensions a: 51.09 Å, b: 75.21 Å, c: 87.92 Å;  $\alpha$ ,  $\beta$ ,  $\gamma$ : 90°, 101.30° with two polypeptide chains in the asymmetric unit. Data at 1.55 Å Bragg spacing was collected at SSRL beamline BL12-2. Further crystals were grown using CaCl<sub>2</sub> as an additive to collect anomalous data at 7 keV (1.7715 Å). The 1.06 Å crystal structure was used as the search model for molecular replacement of these derivatives with Phaser. See Tables S1–4 for crystal data quality parameters and refinement statistics. Data was reduced with XDS<sup>58</sup>, scaled with SCALA<sup>59</sup>, and analyzed with different software routines within the CCP4 suite<sup>60</sup>. Graphics were rendered with Pymol<sup>30</sup>.

### Molecular alignment, modeling, and docking

UniProtKB entries used for alignment: I3R794, Halolysin, *Haloferax mediterranei* (strain ATCC 33500/DSM 1411/JCM 8866/NBRC 14739/NCIMB 2177/R-4) (*Halobacterium mediterranei*); P29143, Halolysin, *Natrialba asiatica* (strain ATCC 700177/DSM 12278/JCM 9576/FERM P-10747/NBRC 102637/172P1); P04072, Thermitase, *Thermoactinomyces vulgaris*; P00782, Subtilisin BPN', *Bacillus amyloliquefaciens* (*Bacillus velezensis*); P00780, Subtilisin Carlsberg, *Bacillus licheniformis*; Q3S3L6, Putative 36kDa protease, *Lysinibacillus sphaericus* (*Bacillus sphaericus*); U5J9E4, Peptidase\_S8 domain-containing protein, Bacillus phage vB\_BanS-Tsamsa; J9E4K3, Peptidase\_S8 domain-containing protein, *Wuchereria bancrofti*; A0QQ47, Subtilase family protein, *Mycobacterium smegmatis* (strain ATCC 700084/mc(2)155) (*Mycobacterium smegmatis*); P59996, Proprotein convertase subtilisin/kexin type 9, *Rattus norvegicus* (Rat); Q8NBP7, Proprotein convertase subtilisin/kexin type 9, *Homo sapiens* (Human); A0A2R2JFW8, Protease, *Penicillium cyclopium*; P06873, Proteinase K, *Paronygodontium album* (*Tritirachium album*); Q01471, Serine protease, *Purpureocillium lilacinum* (*Paecilomyces lilacinus*); B8N106, Alkaline protease 1, *Aspergillus flavus* (strain ATCC 200026/FGSC A1120/IAM 13836/NRRL 3357/JCM 12722/SRRC 167); P12547, Alkaline protease 1, *Aspergillus oryzae* (strain ATCC 42149/RIB 40) (Yellow koji mold); P28296, Alkaline protease 1, *Neosartorya fumigata* (strain ATCC MYA-4609/Af293/CBS 101355/FGSC A1100) (*Aspergillus fumigatus*); D5LGB3, Alkaline protease 1, *Aspergillus versicolor*; A1CIA7, Alkaline protease 1, *Aspergillus clavatus* (strain ATCC 1007/CBS 513.65/DSM 816/NCTC 3887/NRRL 1/QM 1276/107); F2E033, Predicted protein, *Hordeum vulgare* subsp. vulgare (Domesticated barley); P87184, Alkaline protease 2, *Neosartorya fumigata* (strain ATCC MYA-4609/Af293/CBS 101355/FGSC A1100) (*Aspergillus fumigatus*); Q8GB52, Extracellular subtilisin-like serine proteinase, *Vibrio* sp. PA-44; Q3HUQ2, Proteinase K, *Serratia* sp. GF96; P80146, Extracellular serine proteinase, *Thermus* sp. (strain Rt41A); P08594, Aqualysin-1, *Thermus aquaticus*; P16397, Bacillopeptidase F, *Bacillus subtilis* (strain 168). Alignment was done with Seaview<sup>38</sup>. Structure comparisons were done with DALI server<sup>31</sup>. All molecular modeling was performed using Coot and the model coordinates refined against reference parameters with REFMAC5 (version 5.8.0135) until deviation from ideal bond length, bond angle, planar restraints, chiral volume, reached convergence and Ramachandran outliers were minimized (1.7%, 5 out of 347, of the total residues were outliers). Extension of the pro-domain beyond residue 121 was made exploiting the location of water molecules that hydrogen bond the protease to guide the tracing of the hydroxyl group of residues Thr124 and Thr125 and the main chain carbonyl of Gln126. The docking model was based on a synthetic triple-helix peptide containing a region from human collagen type III<sup>61</sup>. Molecular protein docking was made with Frodock version 2.0<sup>62</sup>. Docking was validated by using the pro-domain for a docking prediction and comparison to the experimentally determined structure. As a test, the pro-domain's coordinates were randomly oriented and then submitted to docking. The comparison with the experimentally determined binary complex showed an excellent agreement.

### Differential scanning fluorimetry and light scattering

Thermal unfolding measurements were performed using a Prometheus Panta instrument (NanoTemper Inc, South San Francisco, CA, USA). The thermal unfolding was on a linear thermal ramp from 25 to 95 °C to the sample while reading out fluorescence and turbidity to analyze thermal stability and aggregation. To assess protein aggregation and precipitation backreflection was used for measuring light intensity loss due to scattering. Data recorded by the backreflection optics is termed turbidity. The unfolding profile plots the 350nm/330 nm fluorescence ratio as a function of temperature. The ratio is a measure for a spectral shift in the fluorescence emission profile of tryptophan residues. Non-aggregated oryzin exhibits a turbidity signal of under 100 mAU and then sharply goes up on melting. The overall degree of aggregation is calculated according to the Beer–Lambert law and expressed in attenuation units (AU), plotted as milli-AU. Turbidity data was measured in parallel to the thermal unfolding experiment on the same capillaries. The cumulant analysis was used for dynamic light scattering (DLS) data analysis and to assess sample homogeneity. It models the autocorrelation function (ACF) using an average diffusion coefficient to obtain a single averaged hydrodynamic radius  $r_H$ . Heating rate was with a temperature slope of 0.5 °C/min. Experiments performed by triplicate. Oryzin stock concentration was at 0.4 mg/ml in 0.05 M Hepes buffer, pH 7.5 with 0.1M NaCl.  $CaCl_2$  was used in serial dilutions between 50 mM and 1.2  $\mu$ M. Additionally,  $MgCl_2$  was tested at concentrations between 6.25 and 50 mM, but the  $T_m$  for this metal ion was indistinguishable from protein control without metal. Tabulated data are the average value and standard deviation (calculated from all DLS acquisitions).

### Chemical unfolding

Chemical unfolding measurements were performed using Prometheus Panta (NanoTemper Inc., South San Francisco, USA). Guanidine hydrochloride (GuHCl) was used as the chemical denaturant with 30  $\mu$ L dilutions from 6 to 0.25 M prepared in Tris–HCl buffer. A 0.5 mg/mL stock of oryzin in 0.2 M Tris–HCl buffer, pH 8.0, was added to each sample and incubated for 1 h at room temperature. The unfolding profile plots the 350 nm/330 nm fluorescence ratio as a function of denaturant concentration. The ratio is a measure for a spectral shift in the fluorescence emission profile of tryptophan residues. The  $c_{50}$ , denaturant concentration in which 50% of the protein is unfolded, is calculated.

### Substrate-enzyme activity assay

The substrate used to determine enzymatic activity of oryzin was Suc-Ala-Ala-Pro-Phe-pNA (Suc-AAPF-pNA; Bachem Americas Inc., Torrance, CA, USA). Suc-AAPF-pNA 10 mM stock solutions were prepared in 0.2 M Tris–HCl buffer, pH 8.0, aliquoted, and stored at –20 °C. Protein stocks were all diluted to the lowest concentration stock (0.2 mg/mL) using Tris–HCl, pH 8.0. Samples were prepared in a cuvette as 1 mL total volume as follows: 100  $\mu$ L protein of interest, 50  $\mu$ L of Suc-AAPF-pNA substrate, 850  $\mu$ L Tris–HCl, pH 8.0 buffer. A blank was prepared using 50  $\mu$ L of Suc-AAPF-pNA substrate and 950  $\mu$ L Tris–HCl, pH 8.0 buffer. Samples were measured at 410 nm using a Nanodrop 2000c spectrophotometer (ThermoFisher Scientific, Wilmington, DE, USA) at 0, 0.02, 0.08, 0.2, 0.5, 1, 2, 3, 4, 5, 6, 21, 23, 24 h, making sure to blank before each measurement. Substrate-enzyme activity could also be visualized by a color change from clear to yellow.

### Data availability

The atomic coordinates and structure factors generated during the current study are available in the Protein Data Bank<sup>63</sup> with accession codes: oryzin (PDB: 8GKO), oryzin-PMS (PDB: 8GKP), and oryzin-calcium (PDBs: 8GKQ and 8U45). All other data are contained in this manuscript.

Received: 28 February 2023; Accepted: 14 October 2023

Published online: 20 October 2023

### References

- Henneberg, S. *et al.* Antibody-guided in vivo imaging of *Aspergillus fumigatus* lung infections during antifungal azole treatment. *Nat. Commun.* **12**, 1707 (2021).
- Dobiáš, R., Stevens, D. A. & Havlíček, V. Current and future pathways in *Aspergillus* diagnosis. *Antibiotics* **12**, 385 (2023).
- Singh, A., Kaur, A. & Chowdhary, A. Fungal pathogens and COVID-19. *Curr. Opin. Microbiol.* **75**, 102365 (2023).
- Burki, T. WHO publish fungal priority pathogens list. *Lancet Microbe* **4**, e74 (2023).
- van de Veerdonk, F. L., Gresnigt, M. S., Romani, L., Netea, M. G. & Latgé, J.-P. *Aspergillus fumigatus* morphology and dynamic host interactions. *Nat. Rev. Microbiol.* **15**, 661–674 (2017).
- Lockhart, S. R., Chowdhary, A. & Gold, J. A. W. The rapid emergence of antifungal-resistant human-pathogenic fungi. *Nat. Rev. Microbiol.* <https://doi.org/10.1038/s41579-023-00960-9> (2023).
- Cadena, J., Thompson, G. R. & Patterson, T. F. Aspergillosis. *Infect. Dis. Clin. N. Am.* **35**, 415–434 (2021).
- Echeverria-Esnal, D. *et al.* Pharmacological management of antifungal agents in pulmonary aspergillosis: An updated review. *Expert Rev. Anti-Infect. Ther.* **20**, 179–197 (2022).
- Salmanton-García, J. *et al.* The current state of laboratory mycology and access to antifungal treatment in Europe: A European confederation of medical mycology survey. *Lancet Microbe* **4**, e47–e56 (2023).
- Barber, A. E. *et al.* *Aspergillus fumigatus* pan-genome analysis identifies genetic variants associated with human infection. *Nat. Microbiol.* **6**, 1526–1536 (2021).
- Monod, M., Jousson, O. & Reichard, U. *Aspergillus fumigatus* Secreted Proteases. in *Aspergillus fumigatus and Aspergillosis*, 87–106 (ASM Press, 2009).
- Dong, Z., Yang, S. & Lee, B. H. Bioinformatic mapping of a more precise *Aspergillus niger* degradome. *Sci. Rep.* **11**, 693 (2021).
- Larcher, G. *et al.* Purification and characterization of a fibrinogenolytic serine proteinase from *Aspergillus fumigatus* culture filtrate. *FEBS Lett.* **308**, 65–69 (1992).
- Frosco, M., Chase, T. & Macmillan, J. D. Purification and properties of the elastase from *Aspergillus fumigatus*. *Infect. Immun.* **60**, 728–734 (1992).



15. Monod, M., Togni, G., Rahalison, L. & Frenk, E. Isolation and characterisation of an extracellular alkaline protease of *Aspergillus fumigatus*. *J. Med. Microbiol.* **35**, 23–28 (1991).
16. Hanzi, M., Shimizu, M., Hearn, V. M. & Monod, M. A study of the alkaline proteases secreted by different *Aspergillus* species. *Mycoses* **36**, 351–356 (1993).
17. Behnsen, J. *et al.* Secreted *Aspergillus fumigatus* protease Alp1 degrades human complement proteins C3, C4, and C5. *Infect. Immun.* **78**, 3585–3594 (2010).
18. Reichard, U., Buttner, S., Eiffert, H., Staib, F. & Ruchel, R. Purification and characterisation of an extracellular serine proteinase from *Aspergillus fumigatus* and its detection in tissue. *J. Med. Microbiol.* **33**, 243–251 (1990).
19. Namvar, S. *et al.* *Aspergillus fumigatus* proteases, Asp f 5 and Asp f 13, are essential for airway inflammation and remodelling in a murine inhalation model. *Clin. Exp. Allergy* **45**, 982–993 (2015).
20. Balenga, N. A. *et al.* A fungal protease allergen provokes airway hyper-responsiveness in asthma. *Nat. Commun.* **6**, 6763 (2015).
21. Basu, T. *et al.* *Aspergillus fumigatus* alkaline protease 1 (Alp1/Asp f13) in the airways correlates with asthma severity. *J. Allergy Clin. Immunol.* **141**, 423–425.e7 (2018).
22. Wiesner, D. L. *et al.* Club cell TRPV4 serves as a damage sensor driving lung allergic inflammation. *Cell Host & Microbe* **27**, 614–628.e6 (2020).
23. Rawlings, N. D., Barrett, A. J. & Bateman, A. MEROPS: The peptidase database. *Nucleic Acids Res.* **38**, D227–D233 (2010).
24. Jatón-Ogay, K. *et al.* Nucleotide sequence of a genomic and a cDNA clone encoding an extracellular alkaline protease of *Aspergillus fumigatus*. *FEMS Microbiol. Lett.* **92**, 163–168 (1992).
25. Tang, C. M., Cohen, J. & Holden, D. W. An *Aspergillus fumigatus* alkaline protease mutant constructed by gene disruption is deficient in extracellular elastase activity. *Mol. Microbiol.* **6**, 1663–1671 (1992).
26. Kollattukudy, P. E. *et al.* Evidence for possible involvement of an elastolytic serine protease in aspergillosis. *Infect. Immun.* **61**, 2357–2368 (1993).
27. Breuza, L. *et al.* The UniProtKB guide to the human proteome. *Database* **2016**, bav120 (2016).
28. Fernández, D., Russi, S., Vendrell, J., Monod, M. & Pallarès, I. A functional and structural study of the major metalloprotease secreted by the pathogenic fungus *Aspergillus fumigatus*. *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1946–1957 (2013).
29. Bond, C. S. TopDraw: A sketchpad for protein structure topology cartoons. *Bioinformatics* **19**, 311–312 (2003).
30. DeLano, W. L. Pymol: An open-source molecular graphics tool. *CCP4 Newsl. Protein Crystallogr.* **40**, 82–92 (2002).
31. Holm, L. DALI and the persistence of protein shape. *Protein Sci.* **29**, 128–140 (2020).
32. Koszelak, S. *et al.* The crystallographic structure of the subtilisin protease from *Penicillium cyclopium*. *Biochemistry* **36**, 6597–6604 (1997).
33. Liang, L. *et al.* The crystal structures of two cuticle-degrading proteases from nematophagous fungi and their contribution to infection against nematodes. *FASEB J.* **24**, 1391–1400 (2010).
34. Nam, S. Triglycine-based approach for identifying the substrate recognition site of an enzyme. *Crystals* **9**, 444 (2019).
35. Helland, R., Larsen, A. N., Smalås, A. O. & Willassen, N. P. The 1.8 Å crystal structure of a proteinase K-like enzyme from a psychrophilic *Serratia* species. *FEBS J.* **273**, 61–71 (2006).
36. Barnett, B. L., Green, P. R., Strickland, L. C., Oliver, J. D., Rydel, T. & Sullivan, J. F. Aqualysin I: The crystal structure of a serine protease from an extreme thermophile, *Thermus aquaticus* YT-1. Unpublished (2012).
37. Hampton, E. N. *et al.* The self-inhibited structure of full-length PCSK9 at 1.9 Å reveals structural homology with resistin within the C-terminal domain. *Proc. Natl. Acad. Sci.* **104**, 14604–14609 (2007).
38. Gouy, M., Guindon, S. & Gascuel, O. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* **27**, 221–224 (2010).
39. Ruan, B., London, V., Fisher, K. E., Gallagher, D. T. & Bryan, P. N. Engineering substrate preference in subtilisin: Structural and kinetic analysis of a specificity mutant. *Biochemistry* **47**, 6628–6636 (2008).
40. Lee, W. C., Kikkawa, M., Kojima, S., Miura, K. & Tanokura, M. Crystal structure of serine protease inhibitor POIA1 in complex with subtilisin BPN'. Unpublished (2005).
41. Polgár, L. The catalytic triad of serine peptidases. *Cell. Mol. Life Sci.* **62**, 2161–2172 (2005).
42. San Segundo, B., Martínez, M. C., Vilanova, M., Cuchillo, C. M. & Avilés, F. X. The severed activation segment of porcine pancreatic procarboxypeptidase is a powerful inhibitor of the active enzyme Isolation and characterisation of the activation peptide. *Biochim. Biophys. Acta (BBA) Protein Struct. Mol. Enzymol.* **707**, 74–80 (1982).
43. Valentine, M. L., Cardenas, A. E., Elber, R. & Baiz, C. R. Physiological calcium concentrations slow dynamics at the lipid-water interface. *Biophys. J.* **115**, 1541–1551 (2018).
44. Narhi, L. O. *et al.* Enhanced stability of subtilisin by three point mutations. *Biotechnol. Appl. Biochem.* **13**, 12–24 (1991).
45. Fisher, M. C. *et al.* Threats posed by the fungal kingdom to humans, wildlife, and agriculture. *mBio* **11**, e00449–20 (2020).
46. Druey, K. M., McCullough, M. & Krishnan, R. *Aspergillus fumigatus* protease alkaline protease 1 (Alp1): A new therapeutic target for fungal asthma. *JoF* **6**, 88 (2020).
47. Simon, G. M. & Cravatt, B. F. Activity-based proteomics of enzyme superfamilies: Serine hydrolases as a case study. *J. Biol. Chem.* **285**, 11051–11055 (2010).
48. Zhang, L. *et al.* Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved  $\alpha$ -ketoamide inhibitors. *Science* **368**, 409–412 (2020).
49. Turner, L. D. *et al.* Use of crystallography and molecular modeling for the inhibition of the botulinum neurotoxin A protease. *ACS Med. Chem. Lett.* **12**, 1318–1324 (2021).
50. Burchacka, E., Pięta, P. & Łupicka-Słowik, A. Recent advances in fungal serine protease inhibitors. *Biomed. Pharmacother.* **146**, 112523 (2022).
51. Russi, S., Song, J., McPhillips, S. E. & Cohen, A. E. The Stanford Automated Mounter: Pushing the limits of sample exchange at the SSRL macromolecular crystallography beamlines. *J. Appl. Cryst.* **49**, 622–626 (2016).
52. Sammito, M. *et al.* ARCIMBOLDO\_LITE: Single-workstation implementation and use. *Acta Crystallogr. D Biol. Crystallogr.* **71**, 1921–1930 (2015).
53. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
54. Uson, I. & Sheldrick, G. M. An introduction to experimental phasing of macromolecules illustrated by SHELX; new autotracing features. *Acta Crystallogr. D Struct. Biol.* **74**, 106–116 (2018).
55. Emsley, P. & Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
56. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240–255 (1997).
57. Sheldrick, G. M. Crystal structure refinement with SHELXL. *Acta Crystallogr. C Struct. Chem.* **71**, 3–8 (2015).
58. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
59. Evans, P. R. An introduction to data reduction: Space-group determination, scaling and intensity statistics. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 282–292 (2011).
60. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242 (2011).
61. Kramer, R. Z., Bella, J., Mayville, P., Brodsky, B. & Berman, H. M. Sequence dependent conformational variations of collagen triple-helical structure. *Nat. Struct. Biol.* **6**, 454–457 (1999).

62. Ramírez-Aportela, E., López-Blanco, J. R. & Chacón, P. FRODOCK 2.0: Fast protein–protein docking server. *Bioinformatics* **32**, 2386–2388 (2016).
63. Berman, H. M. The protein data bank. *Nucleic Acids Res.* **28**, 235–242 (2000).

## Acknowledgements

W. G. wishes to thank Foothill College Science Learning Institute program for a student internship. The Science Learning Institute is a nonacademic program at Foothill College whose mission is advancing diversity and equity in STEM by supporting students from underrepresented groups in their STEM pathways. The authors wish to thank Drs. Jacqueline Carozza, Natalia Gomez-Ospina, and Laura M. K. Dassama for constructive discussions. We thank H. Perry at the Stanford Environmental Health & Safety Surplus Chemical Program for some of the reagents used in this study. This work utilized the NanoTemper Prometheus Panta dynamic light scattering (DLS) and nanoDSF system that was purchased with funding from Stanford c-SHARP Program. Use of the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research, and by the National Institutes of Health, National Institute of General Medical Sciences (including P41GM103393). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of NIGMS or NIH.

## Author contributions

D.F. conceptualized experiments, wrote the main manuscript text, analyzed data, and prepared figures and tables. O.N.P. expressed and purified the protein and performed thermal melting and chemical denaturing experiments, in vitro activity assay measurements, analyzed data, prepared figures, and wrote text. D.D.L.D. expressed, purified, and crystallized the protein. W.G. expressed and purified the protein and performed thermal melting experiments. D.F. and S.R. collected X-ray diffraction data and analyzed the data. All authors reviewed the manuscript.

## Funding

This work was partly supported by Stanford Sarafan ChEM-H.

## 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/s41598-023-45028-z>.

**Correspondence** and requests for materials should be addressed to D.F.

**Reprints and permissions information** is available at [www.nature.com/reprints](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 4.0 International License, which permits use, sharing, adaptation, 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 changes were made. 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/4.0/>.

© The Author(s) 2023