A family of papain-like fungal chimerolectins with distinct Ca^{2+} -dependent activation mechanism

Gabriele Cordara^{1,†,‡,}, Dipankar Manna^{1,†,‡} and Ute Krengel^{1,*}*

¹Department of Chemistry, University of Oslo, PO Box 1033 Blindern, 0315 Oslo, Norway

Keywords: calcium-binding protein, crystal structure, cysteine protease, gating tyrosine, protease inhibitor

Graphical abstract

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

ABSTRACT

An important function of fungal lectins is to protect their host. *Marasmius oreades* agglutinin (MOA) is toxic to nematodes and exerts its protective effect through protease activity. Its proteolytic function is associated with a papain-like dimerization domain. The closest homolog of MOA is *Polyporus squamosus* lectin 1a (PSL1a). Here we probed PSL1a for catalytic activity and confirmed that it is a calcium-dependent cysteine protease, like MOA. The X-ray crystal structures of PSL1a (1.5 Å) and MOA (1.3 Å) in complex with calcium and the irreversible cysteine protease inhibitor E-64 revealed the structural basis for their mechanism of action. The comparison with other calcium-dependent proteases (calpains, LapG) reveals a unique metal-dependent activation mechanism relying on a calcium-induced backbone shift and intra-dimer cooperation. Intriguingly, the enzymes appear to use a tyrosine-gating mechanism instead of pro-peptide processing. A search for potential MOA orthologs suggests the existence of a whole new family of fungal-specific chimerolectins with these unique features.

INTRODUCTION

Lectins are defined as proteins that “possess at least one non-catalytic domain that binds reversibly to a specific mono- or oligosaccharide”.¹ They are classified into *merolectins* (single carbohydrate-binding domain), *hololectins* (two or more carbohydrate-binding domains) and *chimerolectins* (carbohydrate-binding module fused to one or more differently purposed domains).¹

Fungi are a well-known and mostly unexploited source of chimerolectins.² Their bioactivity is often associated with the presence of a catalytic domain.^{3, 4} The *Marasmius oreades* agglutinin (MOA) is a histo-blood group B-specific chimerolectin extracted from the fruiting bodies of the fairy ring mushroom *Marasmius oreades*.⁵ The crystal structure of MOA shows a ricin B chain-like / papain-like two-domain partition.⁶ Its papain-like domain exhibits calcium- or manganese (II)-dependent proteolytic activity.⁷ MOA exerts a cytotoxic activity, similarly to other proteins carrying the ricin B chain-like fold.^{8, 9} The catalytic domain is a key element in mediating toxicity against nematodes or NIH/3T3 cells.^{10, 11} MOA has been crystallized in either a calcium-free (PDB ID: 2IHO)⁶ or calcium-bound form (PDB ID: 3EF2)¹². Calcium binding triggers a conformational change¹² that is essential for the lectin’s proteolytic activity.^{7, 10, 13}

The closest homolog of MOA is the *Polyporus squamosus* lectin 1a (PSL1a; 38% sequence identity)^{14, 15}, which like MOA exhibits cytotoxic activity.¹⁶ Compared to MOA, this 31.2 kDa homodimeric lectin has different ligand specificity, to terminal sialic acid-containing glycotopes.^{15, 17} The only available crystal structure of PSL1a (PDB ID: 3PHZ)¹⁸ closely mimics the tertiary structure of MOA in its calcium-free state.¹⁸ In particular, the dimerization domain of PSL1a retains the papain-like fold and its hallmark L(efl)/R(ight)-domain partition (Figure S1).¹⁹ The good conservation of the papain-like domain (r.m.s.d. = 1.0 Å) extends to the components of

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

the putative catalytic machinery. The catalytic triad of MOA (Cys215, His257, Glu274) perfectly aligns with three equivalent residues of PSL1a (Cys208, His248, Glu266). Likewise, each MOA residue involved in calcium coordination has a counterpart in PSL1a. The good conservation of structural features, and their importance for cytotoxic activity,^{10, 11, 16} suggests a conserved enzymatic function. However, the catalytic activity of PSL1a has not been probed to date.

Here we present a thorough characterization of PSL1a catalytic activity, its structural underpinnings, and compare PSL1a to MOA. Both lectins were crystallized in complex with the E-64 cysteine protease inhibitor and calcium. The structures suggest an unusual metal-dependent activation mechanism, involving the interplay with the symmetry-related protomer, which appears to be common to a new family of fungal chimerlectins.

MATERIALS AND METHODS

Expression and purification

For PSL1a production (UniProt reference: Q75WT9), ArcticExpress cells (DE3) cells (Agilent Technologies) were transformed with the pET43.1a-PSL1a vector, and cultivated according to the protocol suggested by the supplier. Gene expression was induced at 11°C for 24h with 0.1 mM IPTG. Subsequently, *E. coli* pellets were collected by centrifugation and stored at -80°C overnight before lysis. After thawing, the bacterial pellets were resuspended in a lysis buffer containing 50 mM Tris-HCl pH 8.0, 0.15 M NaCl, 2 mM EDTA, 1x concentrated cOmplete protease inhibitor cocktail EDTA free (Roche Diagnostics Ltd), 1 µl/ml Benzonase nuclease (Thermo Scientific) and 4 mg/ml hen egg white lysozyme. After incubation on a shaker for two hours at RT, the insoluble fraction was removed by two rounds of centrifugation (20000 rcf, 45 min). The clarified cell lysate was passed through a D-Gal-sepharose affinity column (Thermo Scientific), followed by extensive washing with 20 mM imidazole pH 8.0 and elution of the protein using a single step 1.0 M D-Gal gradient. The fraction containing the eluted PSL1a was concentrated using a 10000 MWCO PES membrane (Vivaspin, Sartorius AG) to a volume of ≈500 µl. Polishing of the protein preparation was carried out by size-exclusion chromatography (SEC) using a Superdex 75 10/300 GL gel-filtration column (Tricorn, GE Healthcare Life Sciences) with 20 mM imidazole pH 8.0, 2 mM EDTA, 0.2M D-Gal, 0.15M NaCl and DTT 2 mM. The fractions containing the purified protein were pooled, concentrated to a final protein concentration of ≈10-15 mg/ml using concentrator tubes with a 10000 MWCO PES membrane (Vivaspin, Sartorius AG) and underwent three rounds of buffer exchange against 20 mM imidazole/HCl pH 8.0, 2 mM EDTA, 2 mM DTT.

MOA (UniProt reference: Q8X123) was expressed in *E. coli* BL21 (DE3) cells transformed with the pT7-LO-MOA expression vector, and purified following an established protocol.^{7, 11} The protein was stored at -80°C at ≈10-15 mg/ml in 10 mM Tris-HCl pH 8.0, 5 mM DTT. The structure described here was obtained from the Cys63Ala variant of MOA, which was generated by site-directed mutagenesis through the QuikChange II kit (Stratagene) following the protocol provided by the manufacturer. This variant was produced as a preventive measure against a possible cross-reactivity with thiol-modifying compounds, since Cys63 was found to be chemically modified by *N*-ethylmaleimide in an earlier experiment. Parallel experiments with wild-type MOA yielded equivalent results, although with marginally lower resolution.

Activity assay

The activity of PSL1a or MOA was tested using native α1-antitrypsin (α1-AT) from human plasma (Sigma-Aldrich) as the target substrate, following an established protocol.⁷ A mixture containing 0.4 mg/ml α1-AT, 50 mM Na-HEPES pH 7.5, 10 mM DTT, 10 mM CaCl₂ and 0.2 mg/ml PSL1a or MOA was incubated at 37°C for 24 h. Variations of the protocol to profile the enzymatic activity of PSL1a included the use of a 50 mM concentration of a different buffer than Na-HEPES pH 7.5 or a 10 mM concentration of a different chloride salt than CaCl₂. The reaction was stopped by adding 5 μl of SDS-PAGE sample buffer 4x to 15 μl of the reaction mixture and boiling at 100°C for 10 min. Samples were loaded alongside with SeeBlue Plus2 Pre-Stained Standard (Invitrogen) molecular weight markers on a NuPAGE SDS-PAGE 4-12% gel for the electrophoretic run; the gels were stained with Coomassie Brilliant Blue dye (Sigma-Aldrich).. Double digestion with a peptide:*N*-glycanase was carried out by pre-incubating α1-AT 24 h at 37°C in presence of 2 UN of PNGase F from *Elizabethkingia miricola* in 50 mM Na-HEPES pH

7.5. The reaction mixture containing the deglycosylated α 1-AT was complemented with 10 mM DTT, 10 mM CaCl_2 , 0.2 mg/ml PSL1a or MOA and further incubated as described for the untreated substrate. Gel images were analyzed using the Fiji distribution of the ImageJ manipulation software.²⁰

Crystallization

MOA crystals were obtained for the MOA Cys63Ala variant (concentrated to 5 mg/ml). The protein solution contained the $\text{Gal}\alpha 1,3(\text{Fuc}\alpha 1,2)\text{Gal}$ trisaccharide (Dextra Laboratories) in a 1:20 MOA:sugar molar ratio and the proteolytic activity inhibitor E-64 in a MOA:E-64 1:3 molar ratio. Crystals grew after a two-week incubation period at 20°C from a crystallization mixture containing 0.1 M imidazole pH 8.0, 15% PEG 8000, 7.5% DMSO and 0.2 M calcium acetate, pre-mixed with the protein solution in a 1:1 ratio (drop volumes).

PSL1a crystals were obtained from a solution containing 5 mg/ml PSL1a, 20 mM imidazole pH 8.0, 5 mM calcium chloride and the DMSO-dissolved E-64 inhibitor in a PSL1a:E-64 1:3 molar ratio. The protein-inhibitor mixture was pre-incubated for 24h before mixing in a 1:1 ratio (drop volumes) with the reservoir solution. Hanging drops were transferred after 1h incubation from initial conditions containing 0.1 M citrate pH 5.6, 0.2 NaSCN and 50% PEG 3350 to wells containing the same components, but a lower PEG 3350 concentration (40% w/v). All crystals were cryoprotected using the crystallization solution supplemented with 15% ethylene glycol, successively frozen in liquid nitrogen for data collection.

Data collection, processing, scaling and structure determination

Diffraction data were collected at the European Synchrotron Radiation Facility (Grenoble, France) at beamlines ID29 and ID23-2. The images were processed and scaled using the XDS software package.²¹ Scaling statistics are given in Table 1. MOA and PSL1a crystals belong to different space groups, containing either a single protomer (MOA, $P6_322$) or the biological dimer (PSL1a, $P2_12_12_1$) in the asymmetric unit. The structures were solved by molecular replacement with *PHASER*,²² using the calcium-bound structure of MOA (PDB ID: 3EF2)¹² or the calcium-free structure of PSL1a (PDB ID: 3PHZ)¹⁸ as search models, respectively. The MOA search model was modified by stripping off atoms with HETATM record (including the calcium ions), deleting the Pro54-Val56 loop, and by mutating residues showing flexible or generally poorly defined side chains to Ala. The *mFo-DFc* difference electron density maps showed well-defined, positive density peaks at the metal binding sites, the three sugar binding sites and in the active site cleft. In PSL1a, additional positive and negative density peaks suggest a conformational change involving residues Ala165-Gly177.

Model building and refinement

Model building and refinement were carried out in cycles, using Coot²³ and REFMAC5²⁴, respectively. In an initial step, ill-defined side chains and loops were removed from the original *PHASER* output. The models were subsequently rebuilt by adding the missing structural elements in a step-wise fashion as the quality of the electron density map improved, including (in this order) the metal ions, the sugar ligands, water molecules and additional buffer components (*e.g.* acetate, ethylene glycol). The metal ions were modeled as calcium based on previous structural data¹², the presence of 0.2 M Ca^{2+} in the crystallization solution and a compatible ligand coordination^{25, 26}. The E-64 inhibitor was modeled at the end of the refinement process,

when the difference electron density at the active site allowed unambiguous tracing of the inhibitor. Ligand occupancy was determined by minimizing the residual difference electron density for the inhibitor molecule and taking the B-factors of nearby interacting atoms into account. A final refinement step included anisotropic refinement of the MOA-E-64 model.

PSL1a diffraction data beyond an $R_{\text{meas}} > 0.6$ and $\bar{I}/\sigma(\bar{I}) < 2$ were included in the refinement process by assessing the value of the $CC_{1/2}$ parameter, as suggested by Diederichs and Karplus.²⁷⁻

²⁹ A $CC_{1/2} \approx 0.3$ was taken as the cut-off limit, based on the suggestion by Evans and Murshudov that data below a $CC_{1/2}$ of 0.2-0.4 contain negligible residual information.³⁰ The final model was refined against additional diffraction data with a $CC_{1/2}$ below the 0.3 limit, settling for a final resolution of 1.5 Å ($CC_{1/2} \approx 0.3$). The final resolution limits were chosen based on a visual inspection of the electron density at different resolution cut-offs.

Validation of the refined models was carried out using the MolProbity server (<http://molprobity.biochem.duke.edu>)³¹, and additional parameters were calculated with phenix.validate.³² R.m.s.d. values were calculated using the PDBeFold server.³³ Total composite OMIT maps were calculated with COMIT, which is part of the CCP4 software suite for macromolecular crystallography.³⁴ All the figures displaying structural data were generated using PyMOL, version 1.5.0.4 (Schrödinger LLC). R-domain rotation angle measurements were performed within PyMOL with specific scripts from the ‘psico’ module, an Open Software tool developed by Thomas Holder (<http://github.com/speleo3/pymol-psico>; Max Planck Institute for Developmental Biology, Germany).

The atomic coordinates have been submitted to the Protein Data Bank (<http://www.pdb.org>)³⁵ with accession codes 5MU9 and 5MUA.

RESULTS AND DISCUSSION

PSL1a is an active protease

Primed by the significant similarities of PSL1a and MOA (regarding sequence, structure and toxicity)^{15, 16, 18}, we tested the hypothesis that also PSL1a is an active protease. α 1-antitrypsin from human plasma (α 1-AT) was selected as a model substrate, since it allows discrimination between proteolytic and *N*-glycanase activity and has been used in previous work.⁷ Its three *N*-linked glycan chains allow it to be processed as a substrate by enzymes of the PNGase family,³⁶ at the same time, the presence of a long, flexible loop constitutes an easy target for proteolytic cleavage.³⁷

The incubation with PSL1a led to a shift in the molecular weight of the α 1-AT SDS-PAGE band of about 7 kDa, comparable to the proteolytic removal of a 6 kDa fragment, as observed for MOA (Figure 1a). The cleavage pattern is consistent with proteolytic activity and rules out *N*-glycolytic activity. Mass-spectrometric analysis of the digested products confirmed proteolytic cleavage (data not shown). The difference in molecular mass of α 1-AT digestion products upon incubation with PSL1a or MOA can be explained by different substrate specificities. Similar to MOA,⁷ PSL1a is enzymatically active between pH 5.5 and 10 (Figure S2a) and requires the presence of calcium in the reaction environment (Figure 1b; note that calcium is not required for protein stability). For both proteins, manganese (II) can act as a functional substitute for calcium

(Figure 1b). As for other enzymes of the cysteine protease family, the proteolytic activity of PSL1a can be inhibited by thiol-modifying agents like iodoacetamide and *N*-ethylmaleimide (Figure S2b) and by specific thiol-reactive compounds, such as the Cys-protease inhibitor E-64 (Figure 1b), whereas pepstatin A (aspartyl proteases) and PMSF (serine proteases) did not show any inhibitory effect (not shown).⁷

Figure 1. Proteolytic activity of PSL1a. (a) SDS-PAGE analysis of α 1-antitrypsin (α 1-AT) digested with PNGase F, MOA or PSL1a in the presence of calcium. Incubation of α 1-AT with MOA or PSL1a alone (lanes 2 and 3) yields single-band shifts toward lower molecular weight, as expected for defined proteolytic cleavage. In contrast, sequential digestion of α 1-AT with PNGase F and either PSL1a or MOA (lanes 5 and 6) shifts the entire band pattern observed in lane 4 towards lower molecular weight (deglycosylated products, see Hirsch *et al.*).³⁶ As previously established for MOA,⁷ this result is consistent with PSL1a proteolytic rather than *N*-glycolytic activity. (b) Effect of different divalent cations on the enzymatic activity of PSL1a. Activity was only observed when digesting α 1-AT in the presence of calcium or manganese (II) (5 mM) (compare lanes 2, 3 and 8). Other divalent cations, tested at the same concentration, could not functionally replace Ca²⁺ or Mn²⁺; E-64 inhibits PSL1a activity (lane 11). Metal ion-mediated protein degradation can be held accountable for the decrease in intensity of the α 1-AT band in the presence of Cu²⁺ or Fe²⁺ (lanes 6 and 9; see Stadtman³⁸).

Calcium-bound PSL1a and MOA form a covalent complex with E-64

The crystal structures of MOA and PSL1a in complex with calcium and the irreversible protease inhibitor E-64 were solved to a resolution of 1.5 Å and 1.3 Å, respectively. Data collection and refinement statistics are summarized in Table 1. In the MOA-E64 structure, the asymmetric unit contains a single protomer, whereas PSL1a crystals contain the biological dimer (Figure 2a). The biologically relevant homodimer of MOA is reconstituted through crystallographic symmetry. A comparison of MOA bound to E-64 with the coordinates of the calcium-bound protein in complex with Z-VAD-fmk (PDB ID: 5D61¹³) or unliganded (PDB ID: 3EF2¹²) shows very little variation (r.m.s.d. = 0.1 Å). The two protomers in PSL1a are essentially identical (r.m.s.d. = 0.3 Å).

A galactose molecule, one of the buffer components throughout purification, was found to bind to the ricin B chain-like lectin module of both PSL1a subunits (Figure 2a). The monosaccharide matches the orientation of the galactose moiety of the human-type influenza receptor epitope Neu5Aca2,6Galβ1,4GlcNAc, characterized previously in complex with PSL1a (PDB ID: 3PHZ).¹⁸ As already noted in the case of MOA¹², each PSL1a protomer binds to two calcium ions in a binuclear cluster, with each of the ions exhibiting distinct ligand coordination (sites A, pentagonal bipyramidal and B, octahedral, respectively). In the biological dimer, the two binuclear calcium clusters are symmetrically positioned at the dimerization interface (Figure 2a).

Figure 2. E-64 inhibitor bound to PSL1a and MOA. (a) Structure of PSL1a homodimer (light/dark blue) in complex with two E-64 inhibitor molecules (green sticks), four calcium ions (magenta spheres) and two galactose molecules (yellow sticks). (b-d) Structures and composite OMIT maps of E-64 (contoured at 1σ) bound to PSL1a (b,c) or MOA (d). In one of the two PSL1a protomers (c), the arginine-like (agmatine) tail of E-64 adopts only a single conformation, stabilized by crystal contacts. (e) Chemical structure of the E-64 inhibitor, with the atom numbering suggested by Leung-Toung *et al.*³⁹ (f) Reaction mechanism of E-64 with the nucleophilic cysteine of PLCPs, as proposed by Matsumoto *et al.*⁴⁰ (g) Scheme of the PLCP catalytic cleft, first shown by Cordara *et al.*¹³ and adapted from the representation proposed by Brömme.⁴¹ Subsites S3 and S2' are drawn as dotted lines to stress their nature of 'binding areas', whereas subsites S4 and S3' are represented as shallow grooves to stress their very low conservation among PLCP members. The S2 subsite is shown as a deep pocket to highlight its prominent role in determining the substrate preference.

E-64 exhibits the same essential enzyme-substrate interactions in PSL1a and MOA

According to the literature, the catalytic cysteine of papain-like cysteine proteases (PLCPs) performs a nucleophilic attack on the oxirane ring of E-64 (Figure 2e,f), leading to the formation of a covalent adduct and the irreversible inhibition of the enzyme (Figure 2f).⁴⁰ In both PSL1a and MOA, E-64 forms a covalent linkage between the C2 carbon of the inhibitor molecule and the S^γ of the catalytic cysteine (PSL1a: Cys208, MOA: Cys215; Figure 2b-d). The inhibitor is positioned at the dimerization interface, interacting with the active site of PSL1a and MOA through polar and van der Waals contacts, as detailed in Figure 3a,b.

Figure 3. Active site interactions. (a) PSL1a. Left panel: Detailed interactions between the E-64 inhibitor (green) and the catalytic cleft of PSL1a (blue). Hydrogen bonding interactions are highlighted as yellow dotted lines, calcium ions represented by magenta spheres. Right panel: Schematic representation of PSL1a active site. (b) MOA. Left panel: E-64 interactions with the catalytic cleft of MOA (orange/wheat). Right panel: Schematic representation of the MOA active site. (c) Comparison of PSL1a, MOA and papain. Left panel: S2 binding pocket (superimposition of PSL1a, blue, and MOA, orange/wheat). Middle panel: active site (superimposition includes papain (grey/black); PDB ID: 1PPN).⁴² Right panel: Schematic view of the catalytic cleft of papain (adapted from Turk *et al.*).⁴³ The oxyanion hole is marked.

The three most important interactions between enzyme and inhibitor involve the carboxyl group, the C⁴=O⁴ carbonyl group and the leucyl moiety of E-64. The carboxylate “head” points to a cavity on the active site of PSL1a lined by the NH group of a tryptophan side chain (PSL1a: Trp201, MOA: Trp208) and the backbone NH group of the catalytic Cys. This feature, known in proteases as the *oxyanion hole*^{44, 45}, plays a fundamental role in stabilizing a negatively charged intermediate during the proteolytic reaction.⁴⁴ The C⁴=O⁴ group forms a strong interaction with the calcium ion occupying the octahedral cavity. The carbonyl oxygen completes the calcium ion coordination on the axial plane and keeps the E64 molecule tightly bound to the catalytic cleft. The leucyl side chain of E-64 points inward, toward the surface of the active cleft, fitting into a shallow cavity lined by the side chains of two residues, each contributed by a different protomer (PSL1a: Ser175# and Phe240, MOA: Leu182# and Leu247; Figure 3). This structural feature, found in all PLCPs and referred to as the *S2 binding pocket*⁴⁶ (Figure 2g), is a strong determinant of the substrate preference of proteases.⁴⁷

1
2
3 The conformation of the arginine-like (agmatine) tail of E-64 differs between MOA and
4 PSL1a, and between the two PSL1a protomers (Figure 2b-d). When bound to chain A of the
5 PSL1a dimer, the residual electron density at the active site supports the assignment of the
6 agmatine tail to three different conformations (α , β , γ), modeled with partial occupancy (0.4,
7 0.35, 0.25, respectively; Figure 2b), whereas it exhibits only a single conformation in PSL1a
8 chain B, facilitated by a crystal contact (Figure 2c; modeled at full occupancy). In MOA, the
9 agmatine tail of the inhibitor is poorly defined (occupancies α/β of 0.55/0.45, respectively;
10 Figure 2d).

Metal-mediated activation mechanism of MOA and PSL1a

21
22
23
24
25
26
27 The high-resolution crystal structures of MOA and PSL1a in complexes with calcium and the
28 irreversible E-64 inhibitor enable a direct comparison between the active sites of the two
29 enzymes. In MOA, calcium binding triggers a conformational change,¹² which is conserved in
30 PSL1a. At the domain level, the sixth strand of the R-domain barrel motif is pushed away and
31 forced into a different conformation (PSL1a: Ala165-Gly177, MOA: Ile173-Gly184), (Figure
32 4a). This sliding motion results in a partial rotation of the R-domain, which is more pronounced
33 in MOA than in PSL1a (5.2°, compared to the 1.6° measured in PSL1a; Figure 4a). The
34 combination of the sliding motion and the R-domain rotation opens the catalytic cleft for
35 substrate binding.

Figure 4. Calcium-based activation mechanism. (a) Comparison between the calcium-free (dark colors)^{6, 18} and the calcium-bound forms (light colors; this work and ¹²) of PSL1a (upper panels) and MOA (lower panels). Calcium binding triggers a conformational change of a stretch of residues of the R-domain antiparallel β -barrel (PSL1a: Ala165-Gly177, MOA: Ile173-Gly184). The change leads to a tilt of the R-domain relative to the L-domain (lighter colors; for R/L domain partitioning, see Figure S1). At the molecular level, the conformational change results in a $\approx 6\text{\AA}$ backbone shift of the affected residues. The hydroxyl group of a tyrosine residue (PSL1a: Tyr173, MOA: Tyr180) shifts $\approx 15\text{\AA}$, freeing the catalytic cleft of the symmetry-related protomer (transparent). (b) The conformational change leads to the proper alignment of the catalytic machinery. The tyrosine shown in (a) acts as a gating residue, moving away from the

catalytic cysteine and out of the S2 subsite (solvent exposed surface shown in either light blue, for PSL1a, or wheat, for MOA). A key aspartate (PSL1a: Asp207; MOA: Asp214) is also realigned. PDB files: PSL1a (PDB ID: 5MUA, Ca^{2+} -bound, PDB ID: 3PHZ,¹⁸ Ca^{2+} -free); MOA (PDB ID: 2IHO,⁶ Ca^{2+} -free; PDB ID: 3EF2,¹² Ca^{2+} -bound).

At the molecular level, calcium binding initiates a series of smaller conformational changes (Figure 4), which can be broken down into four molecular events: i) the realignment of the catalytic residues to match the geometric arrangement observed in other PLCPs; ii) binding of the calcium ions in site B tilts an aspartate residue away from the oxyanion hole (Asp207 in PSL1a, Asp214 in MOA; Figure 4b); iii) the conformational change leads to the formation of the S2 specificity pocket, by the correct placement of the residues lining its bottom (Ser175# in PSL1a, Leu182# in MOA); and iv) a tyrosine residue (Tyr173# in PSL1a, Tyr180# in MOA) undergoes a large shift ($\approx 15\text{\AA}$; Figure 4a). The conformational shift of this ‘gating tyrosine’, which would otherwise block the entrance to the S2 pocket, removes a constraint on the conformational freedom of the catalytic cysteine (Figure 4b).

Unique activation mechanism among PLCPs

There are only two other known classes of Ca^{2+} -dependent PLCPs: calpains and LapG. Calpains are a group of multi-domain eukaryotic proteases playing an important role in physiology and disease,⁴⁸⁻⁵¹ LapG is a bacterial protease involved in c-di-GMP signal transduction.⁵² The core event in calpain activation is a calcium-induced conformational change in its papain-like domain.^{49, 53} Calcium binds to two non-EF hand sites, leading to a 25° rotation between the L- and R-domain and the rearrangement of two gating loops.⁵⁴ This in turn realigns

1
2
3 the catalytic machinery, positioning the catalytic His in the optimal orientation to activate the
4 nucleophilic Cys of the catalytic thiolate-imidazolium pair, and removes a tryptophan residue
5 wedging the catalytic cleft.⁵⁵ While the calcium binding sites of calpains do not superimpose
6 with those in MOA,⁵⁶ the single calcium binding site of LapG matches the calcium ion A in
7 MOA/PSL1a. Calcium binding in LapG does not significantly affect the catalytic triad, but rather
8 leads to a structural change of the lower portion of the catalytic cleft for substrate interaction.⁵⁷
9

10
11 The calcium-dependent activation mechanism of PSL1a and MOA relies on both principles,
12 including an R-domain rotation (calpains) and the formation/opening of the catalytic cleft
13 (calpains, LapG). Unique among PLCPs, is that calcium binding to MOA and PSL1a triggers the
14 conformational change of key residues. Upon calcium binding, an aspartate residue tilts to allow
15 the substrate access to the oxyanion hole, while a tyrosine, provided by the second protomer,
16 behaves as a gating element on the S2 pocket and catalytic cysteine (Figure 4b). While partially
17 similar to the removal of the active site-occluding Trp298 in calpains, the gating tyrosine
18 observed in MOA and PSL1a exerts a direct inhibitory effect on the catalytic cysteine. Most
19 PLCPs are synthesized as proenzymes and subsequently cleaved, thus removing the propeptide
20 and freeing the catalytic cleft.⁵⁸ Propeptide-based inhibition relies on both keeping the
21 nucleophilic cysteine in check by engaging it through an incorrectly oriented peptide bond, and
22 blocking access to the catalytic center. A calcium-induced backbone shift resulting in tyrosine
23 gating, as observed in MOA and PSL1a, could represent a novel alternative to the propeptide-
24 based activation mechanism common in other PLCPs and the mobile loop-based mechanism of
25 calpains.⁵⁴ The role of tyrosine as a gating residue has been observed in other biological systems,
26 *e.g.* aquaporin,⁵⁹ RNA polymerase,⁶⁰ CIC chloride channels,⁶¹ and even, presumably as a “key in
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

lock”, in proteosomal activation;^{62, 63} however, tyrosine gating has to our knowledge not earlier been observed in small proteases.

Active site of MOA and PSL1a compared to other PLCPs

Enzymes belonging to the PLCP superfamily share a common active site architecture (Figure 2g), originally defined by Schechter and Berger in 1967.⁶⁴ When bound to the active site of a cysteine protease, the E-64 molecule can be likened to a dipeptide with a carboxyl-carrying tail attached to its N-terminus, a Leu residue as the P2 moiety and a pseudo-arginine as the P3 moiety (Figure 2e). Compared to a standard PLCP substrate, E-64 binds to sites S1' and S1-S3 of the PLCP active cleft in a reversed backbone orientation (compare Figure 2e,g).^{40, 65}

When bound to MOA or PSL1a, E-64 retains the inverted backbone orientation observed in other PLCPs.⁴⁰ The PSL1a- and MOA-E-64 complexes confirm Cys208 and Cys215 the nucleophilic residues of PSL1a and MOA, respectively, in line with published cytotoxicity experiments.¹⁶ The Cys(S^γ)-E-64(C²) distance observed in different PDB-deposited PCLP-E-64 complexes ranges between 1.8 Å and 2.1 Å, in good agreement with the value of 1.9-2.0 Å determined for MOA and PSL1a (Table 3; see Supplementary Table 1 for an extended comparison with all the known PLCP-E-64 complexes).

The oxyanion hole of most PLCPs is lined by the backbone NH group of the catalytic cysteine and the side chain amide of a glutamine residue (Gln19 in papain; Figure 3c). While the NH group of cysteine shoulders most of the stabilizing effect, a substitution of the glutamine residue leads to variations ranging from a change of the catalytic efficiency^{44, 66} to a change in reactivity (e.g. papain to a nitrile hydratase).⁶⁷ In PSL1a and MOA, the glutamine residue is replaced by

tryptophan. This could explain the low measured *in vitro* catalytic efficiency of both enzymes, or potentially indicate the presence of an alternative catalytic activity coexisting with the proteolytic one.

At the S1' subsite of papain and other PLCPs, E-64 directly interacts with the catalytic histidine and coordinates a tryptophan residue through a water molecule (not shown).⁶⁸⁻⁷⁴ Both residues are well conserved throughout the PLCP family, acting as the general acid-base catalyst (His159, papain numbering) and a docking point for the substrate backbone (Trp177; Figure 3c). In MOA and PSL1a, the E-64-His interaction is conserved, whereas the water-mediated interaction is to a glutamine and not a tryptophan residue (Figure 3a,b). The role of glutamine in substrate coordination is further reinforced by previous structural data of MOA in complex with Z-VAD-fmk.¹³

PLCPs coordinate their substrates through backbone-backbone interactions mediated by L- and R-domain residues. In particular, the L-domain binds the substrate through the backbone carbonyl and NH groups of a highly conserved glycine (Gly66 in papain) (Figure 3c). In MOA and PSL1a, the two glycine-mediated interactions are replaced by a single contact to the site B metal ion with octahedral configuration (Figure 3a,b).

An emerging family of chimerolectins

In order to investigate if other proteins contain the newly identified structural features, we performed a BLAST search using the sequence of the proteolytic domain of MOA. We identified several putative protein orthologs from newly sequenced fungal genomes, *e.g. Moniliophthora perniciosa*, *Serpula lacrymans* and *Singulisphaera acidiphila*, as well as bacterial genomes (*e.g. Xenorhabdus khoisanæ* and *Dehalococcoides mccartyi*; Figure S3).

A Promals3D⁷⁵ structure- and sequence-based alignment of MOA against known and potential homologs shows the conservation of several key elements (Table 2). Residues involved in metal binding through their side chain are generally well conserved, whereas residues mediating metal interactions *via* their backbone carbonyl are more prone to variation. The ‘gating tyrosine’ of MOA (Tyr180) seems to be conserved in some members (Table 2), while in others it is replaced by an amino acid of approximately similar size (*e.g.* Phe, Arg, Glu, His). In contrast, the catalytic Cys-His-Glu triad shared by MOA and PSL1a is not universally conserved. The cysteine is replaced by a serine residue in PSL1b and AGL1, by (*e.g.*) an isoleucine and an aspartate in the putative members from *Moniliophthora perniciosa*, or a glutamine residue in all the putative *Serpula lacrymans* members. The histidine residue is subject to variation as well, whereas the acidic residue (Glu or Asp) of the catalytic triad is rather well conserved.

MOA and PSL1a can hence be considered as founding members of an emerging family of fungal chimerolectins carrying a ricin B-chain-like fold and an unusual dimerization domain. Conservation of residues involved in metal binding and gating hint at a possible shared activation mechanism. In contrast, the catalytic triad is subject to more variation. Some PLCPs tolerate the substitution of cysteine by serine⁷⁶; however, in MOA this substitution results in the loss of proteolytic activity (data not shown). The loss of catalytic activity and gain of a new function by mutation of the catalytic cysteine is not unprecedented for PLCPs.⁷⁷ Examples of the latter are *e.g.* P34 from *Glycine max* and related proteins, such as SPE31 from *Pachyrhizus erosus*. These plant proteins have either gained or retained a fundamental role in signaling despite a mutation of their catalytic cysteine to a glycine.^{78, 79} Overall, conserved and diverging structural features suggest that the conformational changes associated with metal binding, including tyrosine gating,

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

could play an important functional role in this group of proteins, even in the absence or with a different kind of enzymatic activity.

Conclusions and perspectives

By joining a sugar binding module with a differently purposed domain, chimerolectins can deliver a specific function at a precise cellular location. The chimerolectins MOA and PSL1a carry a papain-like proteolytic domain. The domains provide for an unusual metal-dependent activation mechanism. Unlike any other enzyme with variations of the papain fold, the metal ions directly interact with the substrate. The likely consequence is to add multiple layers of regulation to the biological activity of the two chimerolectins. As the proteolytic activities of MOA and PSL1a are the main source of their cytotoxicity,^{10, 11, 16, 80} the role and nature of the available metal ions could determine the behavior in the original or the target host.⁸¹

Prospective MOA and PSL1a orthologs show a high degree of variability among putative components of the catalytic machinery, whereas residues responsible for metal binding are better conserved. This suggests that the MOA/PSL1a papain-like domain may serve as a flexible scaffold, hosting a range of different, metal-regulated enzymatic activities. The functional and structural characterization of more known and potential members of the MOA/PSL1a family will lead to a better understanding of its papain-like domain, and the different enzymatic activities associated with it.

TABLES

Table 1. Data collection and refinement statistics

	PSL1a-E-64	MOA-E-64
A. Data collection		
Beamline	ESRF ID23-2	ESRF ID23-2
Wavelength (Å)	0.8726	0.8726
Space group	$P2_12_12_1$	$P6_322$
Cell parameters: a, b, c (Å)	75.7 94.4 102.8	120.9 120.9 99.9
Resolution (Å) ^a	47.2-1.5 (1.57-1.54)	46.4-1.3 (1.32-1.30)
R_{merge} (%) ^{ab}	17.8 (>100.0)	7.0 (56.4)
R_{meas} ^{ac}	20.3 (>100.0)	7.8 (68.7)
$R_{\text{p.i.m.}}$ ^{ad}	9.8 (>100)	3.3 (38.2)
$CC_{1/2}$ ^{ae}	99.1 (33.3)	99.9 (61.6)
Mean $I/\sigma(I)$ ^a	6.2 (0.6)	14.0 (1.6)
Completeness (%) ^a	99.0 (93.8)	94.2 (61.2)
Multiplicity ^a	4.1 (3.8)	5.2 (2.5)
Unique reflections ^a	107830 (4995)	99395 (3104)
B. Refinement		
Resolution (Å) ^a	47.2-1.5	46.4-1.3
$R_{\text{work}}/R_{\text{free}}$ (%) ^f	22.1 / 24.8	16.0/19.2
Macromolecules / a.s.u.	2	1
<i>No. atoms</i>		
Protein	4372	2386
Water	451	331
Ligands	102	147
<i>B-factor (Å²)</i>		
Protein	14.9	11.4
Water	23.3	24.2
Ligands	19.2	15.1
<i>r.m.s.d. from ideal values</i>		
Bond lengths (Å)	0.02	0.02
Bond angles (deg.)	2.0	2.3
<i>Ramachandran plot</i>		
Core region (%)	97.3	97.4
Outliers (%)	0.0	0.0
PDB ID	5MUA	5MU9

^aValues in parentheses refer to highest resolution shell

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

^b $R_{\text{merge}} = \sum_h \sum_j |I_{hj} - \langle I_h \rangle| / \sum_h \sum_j I_{hj}$, where $\langle I_h \rangle$ is the mean intensity of symmetry-related reflections I_h

^c $R_{\text{meas}} = \sum_h [N_h/(N_h-1)]^{1/2} \sum_i |I_{hi} - \langle I_h \rangle| / \sum_h \sum_i I_{hi}$, where N is the redundancy of reflection **h** ⁸²

^d $R_{\text{p.i.m.}} = \sum_h [1/(N_h-1)]^{1/2} \sum_j |I_{hj} - \langle I_h \rangle| / \sum_h \sum_j I_{hj}$ ⁸³

^eThe high resolution cut-off was chosen despite the low $CC_{1/2}$ ensuring the presence of a low signal-to-noise ratio by visual inspection of the electron density map

^f R_{free} was calculated from 5% of randomly selected data for each data set

Table 2. Conserved catalytic residues in the MOA-PSL1a family

Protein/Organism	PDB/UniProt	MOA %ID	cat. triad			oxy. hole	subst. coord.		Metal binding					gating	S2 pocket		
									bb		side chain						
Known members																	
MOA	2IHO (PDB)	-	C *	H	E	W	Q	A	L	Q	D	D	D	D	Y	L	L
MOA residue number	-	-	215	257	274	208	276	256	182	211	183	214	216	217	180	182	247
PSL1a	3PHZ (PDB)	38	C	H	E	W	Q	A	S	E	D	D	D	D	Y	S	F
PSL1b	BAC87876	38	S	H	E	R	Q	R	Y	G	D	D	D	E	Y	Y	F
SCL/AGL1	XP_003030975	32	S	I	D	P	E	K	P	P	D	Q	E	A	E	P	S
Proposed members																	
<i>M. pernicioso</i>	XP_002398917	43	I	?	?	S	?	?	A	E	D	D	D	D	Y	A	?
<i>M. pernicioso</i>	XP_002393886	71	D	?	?	R	?	?	A	P	D	N	G	E	Y	A	?
<i>S. acidiphila</i>	ZP_09568250	33	C	H	E	Y	Q	S	L	E	D	D	D	D	R	L	L
<i>S. lacrimans</i>	EGN95167	38	Q	Q	D	L	Q	G	L	E	D	D	D	D	Y	L	L
<i>S. lacrimans</i>	EGO20674	37	Q	Q	D	L	Q	G	M	E	N	D	D	D	Y	M	L
<i>S. lacrimans</i>	EGO20678	38	Q	Q	D	L	Q	G	L	E	D	D	D	D	Y	L	L
<i>D. mccartyi</i>	WP_041331075	27	C	H	E	Y	Q	Y	I	D	D	D	D	N	I	I	R
<i>D. mccartyi</i>	WP_072555782	27	C	H	E	Y	Q	Y	I	D	D	D	D	N	I	I	R
<i>D. mccartyi</i>	OBW61112	29	C	H	E	Y	Q	Y	I	D	D	D	D	N	I	I	R
<i>Peniophora</i>	KZV64829	44	C	H	E	W	Q	A	S	E	D	D	D	D	F	S	F
<i>Peniophora</i>	KZV73639	42	Y	S	E	W	Q	R	S	E	D	D	E	D	Y	S	F
<i>Peniophora</i>	KZV65628	32	E	C	D	R	S	R	P	D	N	D	D	A	F	P	Y
<i>Pleurocapsa</i>	WP_019503841	33	C	H	E	W	Q	P	P	E	D	D	D	D	H	P	W
<i>P. ostreatus</i>	KDQ29914	35	I	T	R	S	L	S	A	E	D	D	D	D	F	A	F
<i>P. crispa</i>	KII87094	41	N	H	E	W	Q	T	F	E	D	D	Q	D	Y	F	F
<i>P. umbellatus</i>	ANC28063	50	C	H	E	W	Q	A	T	E	D	D	D	D	Y	T	L
<i>P. guariconensis</i>	SDC03422	30	C	H	E	Y	Q	A	V	E	D	D	D	D	T	V	F
<i>P. monteili</i>	WP_070091156	33	C	H	E	Y	Q	A	V	E	D	D	D	D	T	V	F
<i>P. putida</i>	WP_043215011	29	C	H	E	Y	Q	A	V	E	D	D	D	D	T	V	F
<i>P. sp. 25</i>	WP_050705034	35	C	H	E	W	Q	A	A	E	D	D	D	D	C	A	A
<i>P. sp. HMSC08G10</i>	WP_070572439	30	C	H	E	Y	Q	A	V	E	D	D	D	D	T	V	F
<i>S. vermifera</i>	KIM22138	34	C	H	E	W	Q	V	T	E	D	D	D	D	Y	T	F
<i>S. vermifera</i>	KIM25353	34	S	H	E	S	Q	L	V	E	E	D	D	Y	Y	V	F
<i>S. stellatus</i>	KIJ30840	44	C	H	E	W	Q	A	S	N	D	D	D	D	Y	S	F
<i>S. stellatus</i>	KIJ39841	43	C	H	E	W	Q	A	S	D	D	D	D	D	Y	S	F
<i>S. stellatus</i>	KIJ35645	43	C	H	E	W	Q	A	S	D	D	D	D	D	Y	S	F
<i>T. cinnabarina</i>	CDO76517	43	C	H	E	W	Q	G	S	E	D	D	D	D	Y	S	F
<i>T. cinnabarina</i>	CDO76070	44	C	H	E	W	Q	G	S	E	D	D	D	D	Y	S	F
<i>T. calospora</i>	KIO23019	33	Y	R	D	S	Q	P	D	Y	D	D	D	D	Y	D	Y
<i>T. calospora</i>	KIO23020	31	Y	K	D	V	Q	P	D	H	D	D	D	N	Y	D	Y
<i>T. calospora</i>	KIO23021	31	Y	K	D	V	Q	S	D	G	N	D	D	D	Y	D	Y
<i>T. calospora</i>	KIO23033	34	D	R	D	L	Q	P	D	H	S	D	D	S	H	D	Y
<i>T. calospora</i>	KIO25705	35	Y	K	D	V	Q	V	D	G	D	D	D	D	Y	D	Y
<i>T. calospora</i>	KIO17420	31	Y	R	D	I	Q	P	D	H	S	D	D	D	Y	D	Y
<i>T. calospora</i>	KIO17421	34	Y	K	D	I	Q	P	E	N	S	D	D	D	Y	E	F
<i>T. calospora</i>	KIO16956	28	Y	K	D	V	Q	R	T	G	Q	D	N	D	H	T	Y

<i>T. calospora</i>	KIO20974	32	Y	R	D	K	Q	Y	A	G	D	D	I	D	I	A	E
<i>X. khoisanae</i>	WP_047961396	33	C	H	E	Y	Q	A	V	Q	D	D	D	D	R	V	L

Notes: MOA %ID. = % identity to the MOA sequence
cat. triad = catalytic triad
oxy. hole = oxyanion hole
subst. coord. = residues involved in substrate coordination
bb = backbone-metal ion interaction
* = column containing the catalytic nucleophile

Table 3. Conserved E-64-PLCPs interactions

Feature	E64 Atom	PSL1a (chain A/B) (PDB ID: 5MU9; 1.5Å)		MOA (PDB ID: 5MU9; 1.3Å)		Papain ^{40, 73} (PDB ID: -; 2.4Å)		Mexicain (chain A/B/C/D) ⁶⁸ (PDB ID: 2BDZ; 2.10Å)		Cathepsin K (chain A/B) ⁷² (PDB ID: 3CE9; 1.8Å)		Staphopain ⁸⁴ (PDB ID: 1CV8; 1.75Å)	
		Residue(atom)	Dist.	Residue(atom)	Dist.	Residue(atom)	Dist.	Residue(atom)	Dist.	Residue(atom)	Dist.	Residue(atom)	Dist.
Catalytic Cys	C ²	Cys208(S ^γ)	1.9/2.0	Cys215(S ^γ)	1.9	Cys25(S ^γ)	1.8	Cys25(S ^γ)	2.3/2.1/2.3/2.1	Cys25(S ^γ)	1.8	Cys24(S ^γ)	1.9
Oxyanion hole	O ²	Cys208(N)	2.9/3.0	Cys215(N)	2.8	Cys25(N)	3.0	Cys-25(N)	3.1/3.0/3.0/3.0	Cys25(N)	3.0	Cys24(N)	2.9
	O ²	Trp201(N ^{ε1})	2.7/2.7	Trp208(N ^{ε1})	2.7	Gln19(N ^{ε2})	2.9	Gln-19(N ^{ε2})	2.9/2.7/2.9/2.6	Gln19(N ^{ε2})	2.8	Gln18(N ^{ε2})	2.9
S1'	O ¹	His248(N ^{δ1})	2.9/2.9	His257(N ^{δ1})	3.6	His159(N ^{δ1})	2.9	His159(N ^{δ1})	3.3/3.5/3.3/3.6	His162(N ^{δ1})	2.7	His120(N ^{δ1})	2.7
	O ¹	HOH(Gln276-N ^{ε1}), Trp208-N ^{ε2}	2.7/2.6	HOH(Gln276-N ^{ε2}), Trp208-N ^{ε1}	2.6	HOH(Trp177-N ^{ε1})	2.7	HOH(Trp177-N ^{ε1})	2.6/2.5/-/-	HOH(Trp184-N ^{ε1}), Gln19-N ^{ε2}	2.8	HOH(Trp143-N ^{ε1})	2.6
Substrate coordination	O ⁴	Ca ²⁺	2.2/2.3	Ca ²⁺	2.2	Gly66(N)	2.9	Gly66(N)	2.9/2.9/2.9/2.9	Gly66(N)	3.0	Leu65(N)	3.0
	N ²	Ser175#(O)	-/3.0	-	-	Gly66(O)	3.0	Gly66(O)	2.7/2.9/2.9/2.8	Gly66(O)	3.4	Leu65(O)	3.0
	N ¹	Ala247(O)	3.4/3.8	Ala256(O)	3.5	Asp158(O)	3.4	Asp158(O)	3.2/2.9/2.9/3.0	Asn161(O)	3.1	Gly119(O)	3.5

ASSOCIATED CONTENT

Supporting Information

The following files are available free of charge: overall structural comparison of PSL1a, MOA and papain (Figure S1), role of different agents on PSL1a enzymatic activity (Figure S2), Promals3D mixed structure- and sequence-based alignment of MOA against known and proposed homologs (Figure S3) and E-64-protein atomic distances for different PLCPs in complex with the E-64 inhibitor (Table S1, Microsoft Excel file).

AUTHOR INFORMATION

Corresponding Authors

*Gabriele Cordara - E-mail: gabriele.cordara@kemi.uio.no. Phone: +47 22 85 53 85. Ute Krengel - E-mail: ute.krengel@kjemi.uio.no. Phone: +47 22 85 54 61.

Present Addresses

†Gabriele Cordara: Biocenter Oulu and Faculty of Biochemistry and Molecular Medicine, University of Oulu, Aapistie 7A, 90220 Oulu, Finland. Phone: +358 (0)294 481 179.

†Dipankar Manna: Department of Molecular Medicine, Institute of Basic Medical Sciences, University of Oslo, 0372 Oslo, Norway. Phone: +47 228 51 464.

Author Contributions

The work was conceived by GC and UK, experiments performed by GC and DM, and analyzed by all authors. The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript. ‡These authors contributed equally.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Funding Sources

This work was funded by the University of Oslo, the Norwegian Research Council (grant no. 216625), the Norwegian National graduate School in Structural Biology (BioStruct) and BioStruct-X.

ACKNOWLEDGMENT

We would like to thank Dr. Harry C. Winter and Prof. Irwin J. Goldstein (University of Michigan, Ann Arbor, U.S.A.) for a very fruitful long-term collaboration on this project and Markus Künzler (ETH, Zürich, Switzerland) for providing us with the expression plasmid carrying the full length PSL1a gene. We further acknowledge the European Synchrotron Radiation Facility staff for assistance and support in using the beamlines ID29 and ID23-2. This research project was carried out within the frame of the GlycoNor consortium.

ABBREVIATIONS

Abbreviations: α 1-AT: α 1-antitrypsin from human plasma; MalNEt: *N*-ethylmaleimide; MOA: *Marasmius oreades* agglutinin; MR: molecular replacement; PLCPs: papain-like cysteine proteases; PNGase: peptide:*N*-glycanase; PSL1a: *Polyporus squamosus* lectin 1a; PSL1b: *Polyporus squamosus* lectin 1b; SCL: *Schizophyllum commune* lectin; r.m.s.d.: root mean square deviation/difference.

REFERENCES

- (1) Peumans, W. J., and Van Damme, E. J. M. (1995) Lectins as plant defense proteins, *Plant. Physiol.* 109, 347-352.
- (2) Hassan, M. A. A., Rouf, R., Tiralongo, E., May, T. W., and Tiralongo, J. (2015) Mushroom lectins: specificity, structure and bioactivity relevant to human disease, *Int. J. Mol. Sci.* 16, 7802-7838.
- (3) Khan, F., and Khan, M. I. (2011) Fungal lectins: current molecular and biochemical perspectives, *Int. J. Biol. Chem.* 5, 1-20.
- (4) Xu, X., Yan, H., Chen, J., and Zhang, X. (2011) Bioactive proteins from mushrooms, *Biotechnol. Adv.* 29, 667-674.
- (5) Winter, H. C., Mostafapour, K., and Goldstein, I. J. (2002) The mushroom *Marasmius oreades* lectin is a blood group type B agglutinin that recognizes the Gala1,3Gal and Gala1,3Galb1,4GlcNAc porcine xenotransplantation epitopes with high affinity, *J. Biol. Chem.* 277, 14996-15001.
- (6) Grahn, E., Askarieh, G., Holmner, Å., Tateno, H., Winter, H. C., Goldstein, I. J., and Krenzel, U. (2007) Crystal structure of the *Marasmius oreades* mushroom lectin in complex with a xenotransplantation epitope, *J. Mol. Biol.* 369, 710-721.
- (7) Cordara, G., Egge-Jacobsen, W., Johansen, H. T., Winter, H. C., Goldstein, I. J., Sandvig, K., and Krenzel, U. (2011) *Marasmius oreades* agglutinin (MOA) is a chimerolectin with proteolytic activity, *Biochem. Biophys. Res. Commun.* 408, 405-410.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

(8) Bleuler-Martínez, S., Butschi, A., Garbani, M., Wälti, M. A., Wohlschlager, T., Potthoff, E., Sabotîc, J., Pohleven, J., Lüthy, P., Hengartner, M. O., Aebi, M., and Künzler, M. (2011) A lectin-mediated resistance of higher fungi against predators and parasites, *Mol. Ecol.* 20, 3056-3070.

(9) Bleuler-Martinez, S., Stutz, K., Sieber, R., Collot, M., Mallet, J.-M., Hengartner, M., Schubert, M., Varrot, A., and Künzler, M. (2016) Dimerization of the fungal defense lectin CCL2 is essential for its toxicity against nematodes, *Glycobiology* 27, 486-500.

(10) Wohlschlager, T., Butschi, A., Zurfluh, K., Vonesch, S. C., auf dem Keller, U., Gehrig, P., Bleuler-Martinez, S., Hengartner, M. O., Aebi, M., and Künzler, M. (2011) Nematotoxicity of *Marasmius oreades* agglutinin (MOA) depends on glycolipid binding and cysteine protease activity, *J. Biol. Chem.* 286, 30337-30343.

(11) Cordara, G., Winter, H. C., Goldstein, I. J., Krengel, U., and Sandvig, K. (2014) The fungal chimerolectin MOA inhibits protein and DNA synthesis in NIH/3T3 cells and may induce BAX-mediated apoptosis, *Biochem. Biophys. Res. Commun.* 447, 586-589.

(12) Grahn, E. M., Winter, H. C., Tateno, H., Goldstein, I. J., and Krengel, U. (2009) Structural characterization of a lectin from the mushroom *Marasmius oreades* in complex with the blood group B trisaccharide and Calcium, *J. Mol. Biol.* 390, 457-466.

(13) Cordara, G., van Eerde, A., Grahn, E. M., Winter, H. C., Goldstein, I. J., and Krengel, U. (2016) An unusual member of the papain superfamily: mapping the catalytic cleft of the *Marasmius oreades* agglutinin (MOA) with a caspase inhibitor, *PLoS One* 11, e0149407.

(14) Kruger, R. P., Winter, H. C., Simonson-Leff, N., Stuckey, J. A., Goldstein, I. J., and Dixon, J. E. (2002) Cloning, expression, and characterization of the Gal α 1,3Gal high affinity lectin from the mushroom *Marasmius oreades*, *J. Biol. Chem.* 277, 15002-15005.

(15) Tateno, H., Winter, H. C., and Goldstein, I. J. (2004) Cloning, expression in *Escherichia coli* and characterization of the recombinant Neu5Aca2,6Gal β 1,4GlcNAc-specific high-affinity lectin and its mutants from the mushroom *Polyporus squamosus*, *Biochem. J.* 382, 667-675.

(16) Manna, D., Pust, S., Torgersen, M. L., Cordara, G., Künzler, M., Krengel, U., and Sandvig, K. (2017) *Polyporus squamosus* lectin 1a (PSL1a) exhibits cytotoxicity in mammalian cells by disruption of focal adhesions, inhibition of protein synthesis and induction of apoptosis, *PLoS One* 12, e0170716.

(17) Mo, H., Winter, H. C., and Goldstein, I. J. (2000) Purification and characterization of a Neu5Aca2-6Gal β 1-4Glc/GlcNAc-specific lectin from the fruiting body of the polypore mushroom *Polyporus squamosus*, *J. Biol. Chem.* 275, 10623-10629.

(18) Kadirvelraj, R., Grant, O. C., Goldstein, I. J., Winter, H. C., Tateno, H., Fadda, E., and Woods, R. J. (2011) Structure and binding analysis of *Polyporus squamosus* lectin in complex with the Neu5Aca2-6Gal β 1-4GlcNAc human-type influenza receptor, *Glycobiology* 21, 973-984.

(19) Heinemann, U., Pal, G. P., Hilgenfeld, R., and Saenger, W. (1982) Crystal and molecular structure of the sulfhydryl protease calotropin DI at 3.2 Å resolution, *J. Mol. Biol.* 161, 591-606.

(20) Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J. Y., White, D. J., Hartenstein, V.,

Eliceiri, K., Tomancak, P., and Cardona, A. (2012) Fiji: an open-source platform for biological-image analysis, *Nat. Methods* 9, 676-682.

(21) Kabsch, W. (2010) XDS, *Acta Crystallogr. D Biol. Crystallogr.* 66, 125-132.

(22) McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) *Phaser* crystallographic software, *J. Appl. Crystallogr.* 40, 658-674.

(23) Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of *Coot*, *Acta Crystallogr. D Biol. Crystallogr.* 66, 486-501.

(24) Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., and Vagin, A. A. (2011) *REFMAC5* for the refinement of macromolecular crystal structures, *Acta Crystallogr. D Biol. Crystallogr.* 67, 355-367.

(25) Katz, A. K., Glusker, J. P., Beebe, S. A., and Bock, C. W. (1996) Calcium ion coordination: a comparison with that of beryllium, magnesium, and zinc, *J. Am. Chem. Soc.* 118, 5752-5763.

(26) McPhalen, C. A., Strynadka, N. C. J., and James, M. N. G. (1991) Calcium-binding sites in proteins: a structural perspective, *Adv. Protein Chem.* 42, 77-144.

(27) Karplus, P. A., and Diederichs, K. (2012) Linking crystallographic model and data quality, *Science* 336, 1030-1033.

(28) Diederichs, K., and Karplus, P. A. (2013) Better models by discarding data?, *Acta Crystallogr. D Biol. Crystallogr.* 69, 1215-1222.

(29) Karplus, P. A., and Diederichs, K. (2015) Assessing and maximizing data quality in macromolecular crystallography, *Curr. Opin. Struct. Biol.* 34, 60-68.

(30) Evans, P. R., and Murshudov, G. N. (2013) How good are my data and what is the resolution?, *Acta Crystallogr. D Biol. Crystallogr.* 69, 1204-1214.

(31) Chen, V. B., Arendall, W. B. I., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) *MolProbity*: all-atom structure validation for macromolecular crystallography, *Acta Crystallogr. D Biol. Crystallogr.* 66, 12-21.

(32) Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) *PHENIX*: a comprehensive Python-based system for macromolecular structure solution, *Acta Crystallogr. D Biol. Crystallogr.* 66, 213-221.

(33) Krissinel, E., and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions, *Acta Crystallogr. D Biol. Crystallogr.* 60, 2256-2268.

(34) 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., and Wilson, K. S. (2011) Overview of the *CCP4* suite and current developments, *Acta Crystallogr. D Biol. Crystallogr.* 67, 235-242.

- (35) Berman, H. M. (2008) The Protein Data Bank: a historical perspective, *Acta Crystallogr. A Found. Crystallogr.* 64, 88-95.
- (36) Hirsch, C., Misaghi, S., Blom, D., Pacold, M. E., and Ploegh, H. L. (2004) Yeast *N*-glycanase distinguishes between native and non-native glycoproteins, *EMBO Rep.* 5, 201-206.
- (37) Kalsheker, N. (1989) Alpha₁-antitrypsin: structure, function and molecular biology of the gene, *Biosci. Rep.* 9, 129-138.
- (38) Stadtman, E. R. (1993) Oxidation of free amino acids and amino acid residues in proteins by radiolysis and by metal-catalyzed reactions, *Annu. Rev. Biochem.* 62, 797-821.
- (39) Leung-Toung, R., Li, W., Tam, T. F., and Karimian, K. (2002) Thiol-dependent enzymes and their inhibitors: a review, *Curr. Med. Chem.* 9, 979-1002.
- (40) Matsumoto, K., Mizoue, K., Kitamura, K., Tse, W.-C., Huber, C. P., and Ishida, T. (1999) Structural basis of inhibition of cysteine proteases by E-64 and its derivatives, *Biopolymers* 51, 99-107.
- (41) Brömme, D. (2001) Papain-like cysteine proteases, *Curr. Protoc. Protein Sci. Chapter 21*.
- (42) Pickersgill, R. W., Harris, G. W., and Garman, E. (1992) Structure of monoclinic papain at 1.60 Å resolution, *Acta Crystallogr. B Struct. Sci.* 48, 59-67.
- (43) Turk, B., Turk, V., and Turk, D. (1997) Structural and functional aspects of papain-like cysteine proteinases and their protein inhibitors, *Biol. Chem.* 378, 141-150.

(44) Ménard, R., Plouffe, C., Laflamme, P., Vernet, T., Tessier, D. C., Thomas, D. Y., and Storer, A. C. (1995) Modification of the electrostatic environment is tolerated in the oxyanion hole of the cysteine protease papain, *Biochemistry* 34, 464-471.

(45) Robertus, J. D., Kraut, J., Alden, R. A., and Birktoft, J. J. (1972) Subtilisin; a stereochemical mechanism involving transition-state stabilization, *Biochemistry* 11, 4293-4303.

(46) Kamphuis, I. G., Drenth, J., and Baker, E. N. (1985) Thiol proteases. Comparative studies based on the high-resolution structures of papain and actinidin, and on amino acid sequence information for cathepsins B and H, and stem bromelain, *J. Mol. Biol.* 182, 317-329.

(47) Turk, D., Gunčar, G., Podobnik, M., and Turk, B. (1998) Revised definition of substrate binding sites of papain-like cysteine proteases, *Biol. Chem.* 379, 137-147.

(48) Suzuki, K., Hata, S., Kawabata, Y., and Sorimachi, H. (2004) Structure, activation, and biology of calpain, *Diabetes* 53 Suppl. 1, S12-S18.

(49) Campbell, R. L., and Davies, P. L. (2012) Structure-function relationships in calpains, *Biochem. J.* 447, 335-351.

(50) Low, K. E., Partha, S. K., Davies, P. L., and Campbell, R. L. (2014) Allosteric inhibitors of calpains: reevaluating inhibition by PD150606 and LSEAL, *Biochim. Biophys. Acta* 1840, 3367-3373.

(51) Ono, Y., Saido, T. C., and Sorimachi, H. (2016) Calpain research for drug discovery: challenges and potential, *Nat. Rev. Drug Discov.* 15, 854-876.

(52) Newell, P. D., Boyd, C. D., Sondermann, H., and O'Toole, G. A. (2011) A c-di-GMP effector system controls cell adhesion by inside-out signaling and surface protein cleavage, *PLoS Biol.* 9, e1000587.

(53) Moldoveanu, T., Hosfield, C. M., Lim, D., Elce, J. S., Jia, Z., and Davies, P. L. (2002) A Ca^{2+} switch aligns the active site of calpain, *Cell* 108, 649-660.

(54) Moldoveanu, T., Campbell, R. L., Cuerrier, D., and Davies, P. L. (2004) Crystal structures of calpain-E64 and -leupeptin inhibitor complexes reveal mobile loops gating the active site, *J. Mol. Biol.* 343, 1313-1326.

(55) Hosfield, C. M., Elce, J. S., Davies, P. L., and Jia, Z. (1999) Crystal structure of calpain reveals the structural basis for Ca^{2+} -dependent protease activity and a novel mode of enzyme activation, *EMBO J.* 18, 6880-6889.

(56) Partha, S. K., Ravulapalli, R., Allingham, J. S., Campbell, R. L., and Davies, P. L. (2014) Crystal structure of calpain-3 penta-EF-hand (PEF) domain - a homodimerized PEF family member with calcium bound at the fifth EF-hand, *FEBS J.* 281, 3138-3149.

(57) Chatterjee, D., Boyd, C. D., O'Toole, G. A., and Sondermann, H. (2012) Structural characterization of a conserved, calcium-dependent periplasmic protease from *Legionella pneumophila*, *J. Bacteriol.* 194, 4415-4425.

(58) Cygler, M., and Mort, J. S. (1997) Proregion structure of members of the papain superfamily. Mode of inhibition of enzymatic activity, *Biochimie* 79, 645-652.

(59) Fischer, G., Kosinska-Eriksson, U., Aponte-Santamaría, C., Palmgren, M., Geijer, C., Hedfalk, K., Hohmann, S., de Groot, B. L., Neutze, R., and Lindkvist-Petersson, K. (2009)

Crystal structure of a yeast aquaporin at 1.15 Å reveals a novel gating mechanism, *PLoS Biol.* 7, e1000130.

(60) Cheung, A. C. M., and Cramer, P. (2011) Structural basis of RNA polymerase II backtracking, arrest and reactivation, *Nature* 471, 249-253.

(61) Basilio, D., Noack, K., Picollo, A., and Accardi, A. (2014) Conformational changes required for H⁺/Cl⁻ exchange mediated by a CLC transporter, *Nat. Struct. Mol. Biol.* 21, 456-463.

(62) Smith, D. M., Chang, S.-C., Park, S., Finley, D., Cheng, Y., and Goldberg, A. L. (2007) Docking of the proteasomal ATPases' carboxyl termini in the 20S proteasome's a ring opens the gate for substrate entry, *Mol. Cell.* 27, 731-744.

(63) Dange, T., Smith, D., Noy, T., Rommel, P. C., Jurzitza, L., Cordero, R. J. B., Legendre, A., Finley, D., Goldberg, A. L., and Schmidt, M. (2011) Bln10 protein promotes proteasomal substrate turnover by an active gating mechanism, *J. Biol. Chem.* 286, 42830-42839.

(64) Schechter, I., and Berger, A. (1967) On the size of the active site in proteases. I. Papain, *Biochem. Biophys. Res. Commun.* 27, 157-162.

(65) Yamamoto, D., Ishida, T., and Inoue, M. (1990) A comparison between the binding modes of a substrate and inhibitor to papain as observed in complex crystal structures, *Biochem. Biophys. Res. Commun.* 171, 711-716.

(66) Ménard, R., Carrière, J., Laflamme, P., Plouffe, C., Khouri, H. E., Vernet, T., Tessier, D. C., Thomas, D. Y., and Storer, A. C. (1991) Contribution of the glutamine 19 side chain to transition-state stabilization in the oxyanion hole of papain, *Biochemistry* 30, 8924-8928.

(67) Dufour, E., Storer, A. C., and Ménard, R. (1995) Engineering nitrile hydratase activity into a cysteine protease by a single mutation, *Biochemistry* 34, 16382-16388.

(68) Gavira, J. A., González-Ramírez, L. A., Oliver-Salvador, M. C., Soriano-García, M., and García-Ruiz, J. M. (2007) Structure of the mexicain-E-64 complex and comparison with other cysteine proteases of the papain family, *Acta Crystallogr. D Biol. Crystallogr.* 63, 555-563.

(69) Gomes, M. T. R., Teixeira, R. D., Lopes, M. T. P., Nagem, R. A. P., and Salas, C. E. (2012) X-ray crystal structure of CMS1MS2: a high proteolytic activity cysteine proteinase from *Carica candamarcensis*, *Amino Acids* 43, 2381-2391.

(70) González-Páez, G. E., and Wolan, D. W. (2012) Ultrahigh and high resolution structures and mutational analysis of monomeric *Streptococcus pyogenes* SpeB reveal a functional role for the glycine-rich C-terminal loop, *J. Biol. Chem.* 287, 24412-24426.

(71) Katerelos, N. A., Taylor, M. A. J., Scott, M., Goodenough, P. W., and Pickersgill, R. W. (1996) Crystal structure of a caricain D158E mutant in complex with E-64, *FEBS Lett.* 392, 35-39.

(72) Li, Z., Kienetz, M., Cherney, M. M., James, M. N. G., and Brömme, D. (2008) The crystal and molecular structures of a cathepsin K:chondroitin sulfate complex, *J. Mol. Biol.* 383, 78-91.

(73) Varughese, K. I., Ahmed, F. R., Carey, P. R., Hasnain, S., Huber, C. P., and Storer, A. C. (1989) Crystal structure of a papain-E-64 complex, *Biochemistry* 28, 1330-1332.

(74) Varughese, K. I., Su, Y., Cromwell, D., Hasnain, S., and Xuong, N.-H. (1992) Crystal structure of an actinidin-E-64 complex, *Biochemistry* 31, 5172-5176.

(75) Pei, J., Kim, B. H., and Grishin, N. V. (2008) PROMALS3D: a tool for multiple protein sequence and structure alignments, *Nucleic Acids Res.* 36, 2295-2300.

(76) Hodder, A. N., Drew, D. R., Epa, V. C., Delorenzi, M., Bourgon, R., Miller, S. K., Moritz, R. L., Frecklington, D. F., Simpson, R. J., Speed, T. P., Pike, R. N., and Crabb, B. S. (2003) Enzymic, phylogenetic, and structural characterization of the unusual papain-like protease domain of *Plasmodium falciparum* SERA5, *J. Biol. Chem.* 278, 48169-48177.

(77) Berti, P. J., and Storer, A. C. (1995) Alignment/phylogeny of the papain superfamily of cysteine proteases, *J. Mol. Biol.* 246, 273-283.

(78) Li, Q.-G., and Zhang, Y.-M. (2013) The origin and functional transition of P34, *Heredity (Edinb.)* 110, 259-266.

(79) Zhang, M., Wei, Z., Chang, S., Teng, M., and Gong, W. (2006) Crystal structure of a papain-fold protein without the catalytic residue: a novel member in the cysteine proteinase family, *J. Mol. Biol.* 358, 97-105.

(80) Juillot, S., Cott, C., Madl, J., Claudinon, J., van der Velden, N. S. J., Künzler, M., Thuenauer, R., and Römer, W. (2016) Uptake of *Marasmius oreades* agglutinin disrupts integrin-dependent cell adhesion, *Biochim Biophys. Acta* 1860, 392-401.

(81) Sueldo, D. J., and van der Hoorn, R. A. L. (2017) Plant life needs cell death, but does plant cell death need Cys proteases?, *FEBS J.* (in press).

(82) Diederichs, K., and Karplus, P. A. (1997) Improved *R*-factors for diffraction data analysis in macromolecular crystallography, *Nat. Struct. Biol.* 4, 269-275.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

(83) Evans, P. (2006) Scaling and assessment of data quality, *Acta Crystallogr. D Biol. Crystallogr.* 62, 72-82.

(84) Hofmann, B., Schomburg, D., and Hecht, H. J. (1993) Crystal structure of a thiol proteinase from *Staphylococcus aureus* V-8 in the E-64 inhibitor complex, *Acta Crystallogr. A Biol. Crystallogr.* 2, c102.

Proteolytic activity of PSL1a. (a) SDS-PAGE analysis of α 1-antitrypsin (α 1-AT) digested with PNGase F, MOA or PSL1a in the presence of calcium. Incubation of α 1-AT with MOA or PSL1a alone (lanes 2 and 3) yields single-band shifts toward lower molecular weight, as expected for defined proteolytic cleavage. In contrast, sequential digestion of α 1-AT with PNGase F and either PSL1a or MOA (lanes 5 and 6) shifts the entire band pattern observed in lane 4 towards lower molecular weight (deglycosylated products, see Hirsch et al.).³⁶ As previously established for MOA,⁷ this result is consistent with PSL1a proteolytic rather than N-glycolytic activity. (b) Effect of different divalent cations on the enzymatic activity of PSL1a. Activity was only observed when digesting α 1-AT in the presence of calcium or manganese (II) (5 mM) (compare lanes 2, 3 and 8). Other divalent cations, tested at the same concentration, could not functionally replace Ca²⁺ or Mn²⁺; E-64 inhibits PSL1a activity (lane 11). Metal ion-mediated protein degradation can be held accountable for the decrease in intensity of the α 1-AT band in the presence of Cu²⁺ or Fe²⁺ (lanes 6 and 9; see Stadtman³⁸).

1057x251mm (72 x 72 DPI)

Figure 2. E-64 inhibitor bound to PSL1a and MOA. (a) Structure of PSL1a homodimer (light/dark blue) in complex with two E-64 inhibitor molecules (green sticks), four calcium ions (magenta spheres) and two galactose molecules (yellow sticks). (b-d) Structures and composite OMIT maps of E-64 (contoured at 1σ) bound to PSL1a (b,c) or MOA (d). In one of the two PSL1a protomers (c), the arginine-like (agmatine) tail of E-64 adopts only a single conformation, stabilized by crystal contacts. (e) Chemical structure of the E-64 inhibitor, with the atom numbering suggested by Leung-Toung et al..38 (f) Reaction mechanism of E-64 with the nucleophilic cysteine of PLCPs, as proposed by Matsumoto et al..39 (g) Scheme of the PLCP catalytic cleft, first shown by Cordara et al.13 and adapted from the representation proposed by Brömme.40 Subsites S3 and S2' are drawn as dotted lines to stress their nature of 'binding areas', whereas subsites S4 and S3' are represented as shallow grooves to stress their very low conservation among PLCP members. The S2 subsite is shown as a deep pocket to highlight its prominent role in determining the substrate preference.

558x767mm (72 x 72 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure 3. Active site interactions. (a) PSL1a. Left panel: Detailed interactions between the E-64 inhibitor (green) and the catalytic cleft of PSL1a (blue). Hydrogen bonding interactions are highlighted as yellow dotted lines, calcium ions represented by magenta spheres. Right panel: Schematic representation of PSL1a active site. (b) MOA. Left panel: E-64 interactions with the catalytic cleft of MOA (orange/wheat). Right panel: Schematic representation of the MOA active site. (c) Comparison of PSL1a, MOA and papain. Left panel: S2 binding pocket (superimposition of PSL1a, blue, and MOA, orange/wheat). Middle panel: active site (superimposition includes papain (grey/black); PDB ID: 1PPN).⁴¹ Right panel: Schematic view of the catalytic cleft of papain (adapted from Turk et al.).⁴² The oxyanion hole is marked.

1057x1527mm (72 x 72 DPI)

Figure 4. Calcium-based activation mechanism. (a) Comparison between the calcium-free (dark colors)6, 18 and the calcium-bound forms (light colors; this work and 12) of PSL1a (upper panels) and MOA (lower panels). Calcium binding triggers a conformational change of a stretch of residues of the R-domain antiparallel β -barrel (PSL1a: Ala165-Gly177, MOA: Ile173-Gly184). The change leads to a tilt of the R-domain relative to the L-domain (lighter colors; for R/L domain partitioning, see Figure S1). At the molecular level, the conformational change results in a $\approx 6\text{\AA}$ backbone shift of the affected residues. The hydroxyl group of a tyrosine residue (PSL1a: Tyr173, MOA: Tyr180) shifts $\approx 15\text{\AA}$, freeing the catalytic cleft of the symmetry-related protomer (transparent). (b) The conformational change leads to the proper alignment of the catalytic machinery. The tyrosine shown in (a) acts as a gating residue, moving away from the catalytic cysteine and out of the S2 subsite (solvent exposed surface shown in either light blue, for PSL1a, or wheat, for MOA). A key aspartate (PSL1a: Asp207; MOA: Asp214) is also realigned. PDB files: PSL1a (PDB ID: 5MUA, Ca²⁺-bound, PDB ID: 3PHZ,18 Ca²⁺-free); MOA (PDB ID: 2IHO,6 Ca²⁺-free; PDB ID: 3EF2,12 Ca²⁺-bound).

1017x732mm (72 x 72 DPI)

SUPPLEMENTARY INFORMATION

A family of papain-like fungal chimerolectins with distinct Ca^{2+} -dependent activation mechanism

Gabriele Cordara^{1,†,‡,}, Dipankar Manna^{1,†,‡} and Ute Krengel^{1,*}*

¹Department of Chemistry, University of Oslo, PO Box 1033 Blindern, 0315 Oslo, Norway

Supplementary Figure S1

Figure S1. Overall structural comparison of PSL1a, MOA and papain. (a) Superimposition of the homodimers of MOA (orange, PDB ID: 2IHO)¹ and PSL1a (blue, PDB ID: 3PHZ)². (b) Superimposition of the papain-like domain of PSL1a and MOA with papain (PDB ID: 1PPN)³ aligned according to the 'standard view'.^{4, 5} The canonical L- ('left') and R-domain ('right') partition found in all PLCPs is well preserved in the dimerization domains of MOA and PSL1a.

Supplementary Figure S2

Figure S2. Role of different agents on PSL1a enzymatic activities. (a) The active pH region of PSL1a was probed by digesting α 1-AT at a range of different pHs (from 3 to 11), using different buffering agents for each pH point. Those include citrate (pH 3), formate (pH 3.5), succinate (pH 4), acetate (pH 4.5), propionate (pH 5), citrate (pH 5.5), MES (pH 6), bistris (pH 6.5), bistris propane (pH 7), HEPES (pH 7.5), Tris (pH 8), bicine (pH 8.5), bistris propane (pH 9), CHES (pH 9.5), piperazine (pH 10), 1,3-diaminopropane (pH 11). (b) Thiol-modifying agents such as iodoacetamide or *N*-ethylmaleimide alkylate the catalytic cysteine (Cys208) inhibiting proteolytic activity.

Supplementary Figure S3

MOA	1	M-----	1	M-----	1
PSL-1a (38%)	1	M-----	1	M-----	1
PSL-1b (38%)	1	M-----	1	M-----	1
SCL-AGLL (32%)	1	M-----	1	M-----	1
XP_00239917 (43%)	1	M-----	1	M-----	1
XP_002393886 (71%)	1	M-----	1	M-----	1
ZP_09568250 (33%)	1	MDRLVLRNNVAQ-----	53	PRPARVVEPWDRKDRQVPEPTT-----	53
EGN95167.1 (38%)	1	M-----	1	M-----	1
EGO20674 (37%)	1	M-----	1	M-----	1
EGO20678 (38%)	1	M-----	1	M-----	1
WP_041331075 (27%)	1	M-----	1	M-----	1
WP_072555782 (27%)	1	M-----	1	M-----	1
OBW61112 (29%)	1	M-----	1	M-----	1
KZV64829 (44%)	1	M-----	1	M-----	1
KZV73639 (32%)	1	M-----	1	M-----	1
KZV65628 (42%)	1	M-----	1	M-----	1
WP_019503841 (33%)	1	M-----	1	M-----	1
KQZ29914 (35%)	1	M-----	1	M-----	1
KI187094 (41%)	1	M-----	1	M-----	1
ANC28063 (50%)	1	M-----	1	M-----	1
SDO3422 (30%)	1	MPHKLVPGQPDANSPLV-----	52	GQILPRPEVEPPPEDDTAPD-----	52
WP_070091156 (33%)	1	MPHKLVPGQPDANSPLV-----	52	GQILPRPEVEPPPEDDTAPD-----	52
WP_043215011 (29%)	1	M-----	1	M-----	1
WP_050705034 (35%)	1	MSL-----	6	DTL-----	6
WP_070572439 (30%)	1	MPHKLVPGQPDANSPLV-----	52	GQILPRPEVEPPPEDDTAPD-----	52
KIM22138 (34%)	1	M-----	1	M-----	1
KIM25353 (34%)	1	MSSTVYSQFVKVVDGEPNSKPSNVI	126	TDVSGNGEDINKYAGRYVNI	126
KI130840 (44%)	1	M-----	1	M-----	1
KI139841 (43%)	1	M-----	1	M-----	1
KI135645 (43%)	1	M-----	1	M-----	1
CDO76517 (44%)	1	M-----	1	M-----	1
CDO76070 (43%)	1	M-----	1	M-----	1
KIO23019 (33%)	1	MPAYSA-----	31	TTSRNRKTKGY-----	31
KIO23020 (31%)	1	M-----	1	M-----	1
KIO23021 (31%)	1	MAAAT-----	29	TNMQSKNTDES-----	29
KIO23033 (34%)	1	MEIE-----	15	GLSRS-----	15
KIO25705 (35%)	1	MQSN-----	9	TAEN-----	9
KIO17420 (31%)	1	M-----	1	M-----	1
KIO17421 (32%)	1	MDVGGTENGEPVPCSAEACQQLMDL	68	ERASRTTLEIKTMPMVS	68
KIO16956 (28%)	1	M-----	1	M-----	1
KIO20974 (32%)	1	M-----	1	M-----	1
WP_047961396 (33%)	1	MEKDKXTIKI-----	53	FSVSDG--KSTLPKSENRRSED-----	53
				PISLQNDSSFFPMENDPIPE-----	
MOA	2	-----	76	-----	76
PSL-1a (38%)	2	-----	76	-----	76
PSL-1b (38%)	2	-----	76	-----	76
SCL-AGLL (32%)	2	-----	72	-----	72
XP_00239917 (43%)	2	-----	81	-----	81
XP_002393886 (71%)	2	-----	118	-----	118
ZP_09568250 (33%)	54	-----	79	-----	79
EGN95167.1 (38%)	2	-----	30	-----	30
EGO20674 (37%)	2	-----	75	-----	75
EGO20678 (38%)	2	-----	76	-----	76
WP_041331075 (27%)	2	-----	30	-----	30
WP_072555782 (27%)	2	-----	78	-----	78
OBW61112 (29%)	2	-----	77	-----	77
KZV64829 (44%)	1	-----	78	-----	78
KZV73639 (32%)	2	-----	77	-----	77
KZV65628 (42%)	2	-----	78	-----	78
WP_019503841 (33%)	1	-----	77	-----	77
KQZ29914 (35%)	2	-----	127	-----	127
KI187094 (41%)	2	-----	32	-----	32
ANC28063 (50%)	2	-----	66	-----	66
SDO3422 (30%)	53	-----	127	-----	127
WP_070091156 (33%)	3	-----	74	-----	74
WP_043215011 (29%)	1	-----	75	-----	75
WP_050705034 (35%)	7	-----	223	-----	223
WP_070572439 (30%)	53	-----	51	-----	51
KIM22138 (34%)	3	-----	74	-----	74
KIM25353 (34%)	127	-----	75	-----	75
KI130840 (44%)	2	-----	107	-----	107
KI139841 (43%)	2	-----	40	-----	40
KI135645 (43%)	2	-----	105	-----	105
CDO76517 (44%)	2	-----	88	-----	88
CDO76070 (43%)	2	-----	85	-----	85
KIO23019 (33%)	32	-----	144	-----	144
KIO23020 (31%)	1	-----	24	-----	24
KIO23021 (31%)	30	-----	126	-----	126
KIO23033 (34%)	16	-----		-----	
KIO25705 (35%)	10	-----		-----	
KIO17420 (31%)	69	-----		-----	
KIO17421 (32%)	1	-----		-----	
KIO16956 (28%)	2	-----		-----	
KIO20974 (32%)	2	-----		-----	
WP_047961396 (33%)	54	-----		-----	
MOA	77	SS-----EAG-----	136	-----	136
PSL-1a (38%)	77	AA-----TDS-----	132	-----	132
PSL-1b (38%)	77	AV-----ADN-----	136	-----	136
SCL-AGLL (32%)	73	LA-----DRG-----	129	-----	129
XP_00239917 (43%)	82	MA-----ASS-----	137	-----	137
XP_002393886 (71%)	119	SS-----DGS-----	185	-----	185
ZP_09568250 (33%)	80	KS-----TNG-----	133	-----	133
EGN95167.1 (38%)	78	SS-----ANG-----	154	-----	154
EGO20674 (37%)	80	KS-----TNG-----	133	-----	133
EGO20678 (38%)	78	SS-----ANG-----	20	-----	20
WP_041331075 (27%)	2	-----	25	-----	25
WP_072555782 (27%)	1	-----	15	-----	15
OBW61112 (29%)	1	-----	87	-----	87
KZV64829 (44%)	31	A-----TNG-----	134	-----	134
KZV73639 (32%)	76	SA-----ATE-----	133	-----	133
KZV65628 (42%)	77	SS-----DGS-----	133	-----	133
WP_019503841 (33%)	31	VO-----SDG-----	91	-----	91
KQZ29914 (35%)	79	NH-----TNG-----	136	-----	136
KI187094 (41%)	78	SS-----DNQ-----	133	-----	133
ANC28063 (50%)	128	ST-----AGD-----	191	-----	191
SDO3422 (30%)	128	ST-----DGD-----	191	-----	191
WP_070091156 (33%)	33	ST-----DGD-----	191	-----	191
WP_043215011 (29%)	67	DRPWLFLANGPIGRDPNLFASGQNN	135	PGVPRKLCPVYS-----	135
WP_050705034 (35%)	67	DRPWLFLANGPIGRDPNLFASGQNN	135	PGVPRKLCPVYS-----	135
WP_070572439 (30%)	128	ST-----AGD-----	191	-----	191
KIM22138 (34%)	38	VG-----ANG-----	93	-----	93
KIM25353 (34%)	224	SA-----VDS-----	279	-----	279
KI130840 (44%)	52	-----	99	-----	99
KI139841 (43%)	75	-----	100	-----	100
KI135645 (43%)	75	-----	165	-----	165
CDO76517 (44%)	1	-----	148	-----	148
CDO76070 (43%)	1	-----	24	-----	24
KIO23019 (33%)	108	LF-----SILPQ-----	173	-----	173
KIO23020 (31%)	41	SS-----SAG-----	100	-----	100
KIO23021 (31%)	106	KT-----ABR-----	165	-----	165
KIO23033 (34%)	89	LR-----NKG-----	148	-----	148
KIO25705 (35%)	86	KK-----DPK-----	10	-----	10
KIO17420 (31%)	2	-----	2	-----	2
KIO17421 (32%)	142	KT-----ABR-----	204	-----	204
KIO16956 (28%)	2	-----	70	-----	70
KIO20974 (32%)	25	-----	187	-----	187
WP_047961396 (33%)	127	DT-----SEE-----		-----	

[illegible]

5

Supplementary References

- (1) Grahn, E., Askarieh, G., Holmner, Å., Tateno, H., Winter, H. C., Goldstein, I. J., and Krenzel, U. (2007) Crystal structure of the *Marasmius oreades* mushroom lectin in complex with a xenotransplantation epitope, *J. Mol. Biol.* **369**, 710-721.
- (2) Kadirvelraj, R., Grant, O. C., Goldstein, I. J., Winter, H. C., Tateno, H., Fadda, E., and Woods, R. J. (2011) Structure and binding analysis of *Polyporus squamosus* lectin in complex with the Neu5Aca2-6Galb1-4GlcNAc human-type influenza receptor, *Glycobiology* **21**, 973-984.
- (3) Pickersgill, R. W., Harris, G. W., and Garman, E. (1992) Structure of monoclinic papain at 1.60 Å resolution, *Acta Crystallogr. B Struct. Sci.* **48**, 59-67.
- (4) Heinemann, U., Pal, G. P., Hilgenfeld, R., and Saenger, W. (1982) Crystal and molecular structure of the sulfhydryl protease calotropin DI at 3.2 Å resolution, *J. Mol. Biol.* **161**, 591-606.
- (5) Kamphuis, I. G., Drenth, J., and Baker, E. N. (1985) Thiol proteases. Comparative studies based on the high-resolution structures of papain and actinidin, and on amino acid sequence information for cathepsins B and H, and stem bromelain, *J. Mol. Biol.* **182**, 317-329.
- (6) Pei, J., Kim, B. H., and Grishin, N. V. (2008) PROMALS3D: a tool for multiple protein sequence and structure alignments, *Nucleic Acids Res.* **36**, 2295-2300.