A novel mechanism of latency in matrix metalloproteinases **

Mar López-Pelegrín 1, Miroslaw Ksiazek 2, Abdulkarim Y. Karim 2,4, Tibisay Guevara 1, Joan L. Arolas 1,*, Jan Potempa 2,3,6 & F. Xavier Gomis-Rüth 1,*

Running title: Structure of Tannerella forsythia prokarilysin

1 Proteolysis Lab; Department of Structural Biology; Molecular Biology Institute of Barcelona, CSIC; Barcelona Science Park; c/ Baldiri Reixac, 15-21; 08028 Barcelona, Catalonia (Spain).

2 Department of Microbiology, Faculty of Biochemistry; Biophysics and Biotechnology; Jagiellonian University; Ul. Gronostajowa 7; 30-387 Kraków (Poland).

3 Oral Immunology and Infectious Disease; University of Louisville School of Dentistry; Louisville, KY 40202 (USA).

4 Present address: Department of Biology; College of Science; University of Salahaddin; Erbil, Kurdistan (Irak).

* Correspondence: e-mail: fxgr@ibmb.csic.es; phone: (+34) 934 020 186; fax: (+34) 934 034 979; e-mail: jlacri@ibmb.csic.es; phone: (+34) 934 020 187; fax: (+34) 934 034 979; or e-mail: jan.potempa@louisville.edu; phone (+1) 502 852 5572; Fax (+1) 502 852 5572.

Keywords: matrixin; MMP; pro-domain; zymogen; X-ray crystal structure; microbial infection

BACKGROUND: Animal and plant MMPs are kept zymogenic through large pro-domains and a cysteine-switch mechanism.

RESULTS: Bacterial MMP karilysin has only a short N-terminal peptide upstream of the catalytic domain, which lacks cysteines. Here we determined the structure of a pro-karilysin fragment encompassing the pro-peptide and the catalytic domain, and found that the former runs across the cleft in the opposite direction to a bound substrate and inhibits the latter through an “aspartate-switch” mechanism. This finding is reminiscent of latency maintenance in the otherwise unrelated astacin and fragilysin metalloproteinase families. In addition, in vivo and biochemical assays showed that the pro-peptide contributes to protein folding and stability. Our analysis of pro-karilysin reveals a novel mechanism of latency and activation in MMPs. Finally, our findings support the view that the karilysin catalytic domain was co-opted by competent bacteria through horizontal gene transfer from a eukaryotic source, and later evolved in a specific bacterial environment.

ABSTRACT

The matrix metalloproteinases (MMPs) are a family of secreted soluble or membrane-anchored multi-modular peptidases regularly found in several paralogous copies in animals and plants, where they have multiple functions. The minimal consensus domain architecture comprises a signal peptide, a 60-to-90-residue globular pro-domain with a conserved sequence motif including a cysteine engaged in “cysteine-switch” or “Velcro” mediated latency, and a catalytic domain. Karilysin, from the human periodontopathogen Tannerella forsythia, is the only bacterial MMP to have been characterized biochemically to date. It shares with eukaryotic forms the catalytic domain but none of the flanking domains. Instead of the consensus MMP pro-domain, it features a 14-residue pro-peptide—the shortest reported for a metallopeptidase—, which lacks cysteines. In confirmation, we determined the structure of a pro-karilysin fragment encompassing the pro-peptide and the catalytic domain, and found that the former runs across the cleft in the opposite direction to a bound substrate and inhibits the latter through an “aspartate-switch” mechanism. This finding is reminiscent of latency maintenance in the otherwise unrelated astacin and fragilysin metalloproteinase families. In addition, in vivo and biochemical assays showed that the pro-peptide contributes to protein folding and stability. Our analysis of pro-karilysin reveals a novel mechanism of latency and activation in MMPs. Finally, our findings support the view that the karilysin catalytic domain was co-opted by competent bacteria through horizontal gene transfer from a eukaryotic source, and later evolved in a specific bacterial environment.

INTRODUCTION

The matrix metalloproteinases (MMPs) are a family of zinc- and calcium-dependent peptidases, which are grouped into the metzincin clan of metalloproteinases (MPs) together with other separate families such as the ADAMs/adamalysins, astacins, fragilysins, and serralysins (1-8). MMPs are found throughout animals and plants (9-12), where their distribution is consistent with a Darwinian tree-based pathway. In addition, polyplication has led to several paralogous MMP genes being present in the same organism: 24 in humans, 26 in sea urchin, 26 in zebra fish, seven in sea squirt and two in fruitfly (11). In contrast, only a patchy phylogenetic distribution of genes encoding hypothetical orthologs has been found in viruses, Bacteria,
Archaea and fungi. Earlier studies of the relationship between mammalian MMPs and supposed prokaryotic orthologs included—as we now know—bacterial members of other metzincin families such as serralysins, fragilysins, and astacins (13-16). Accordingly, it was suggested that a primordial MMP may have arisen from an ancestor that is common to vertebrates, invertebrates and plants but is not shared by earlier stages in evolution (11,16-18). This entails that the hypothetical prokaryotic, viral, and fungal MMPs are incongruent with the tree of life or, more accurately, xenologs, i.e. the result of direct or indirect horizontal gene transfer from eukaryotic donors (9,19,20). This is reminiscent of the evolutionary origin postulated for fragilysin, which is the only molecular virulence factor described for enterotoxigenic Bacteroides fragilis and for which no similar proteins have been reported, not even from other B. fragilis strains (21). Structural studies supported the view that the catalytic domain of this MMP is the result of horizontal gene transfer of a member of the ADAM/adamalysin family, which has 38 orthologs in humans (8,22-25), from a mammalian host to this bacterium, which thrives in the intestinal tract (26,27).

Returning to MMPs, karylsin from the human periodontopathogen Tannerella forsythia is the only bacterial family member to have been analyzed biochemically to date (9,28-33). In addition to karylsin, only MmpZ from Bacillus anthracis has been functionally assessed at the genetic level through knock-out studies in B. anthracis cells, but it has not been isolated or characterized (34). Similarly to vertebrate MMPs, karylsin showed preference for medium-sized to bulky hydrophobic residues (leucine, tyrosine and methionine) in the specificity pocket, S,’ (30); for active-site clef homology sub-site nomenclature, see (35)). It inactivates antimicrobial peptide LL-37 and integrants of the complement system, including ficolin-2, ficolin-3, C4 and C5, by proteolysis and may thus contribute to evasion of the innate host immune response (29,31). Karylsin is sequentially and evolutionarily closer to MMPs from winged insects that are transmission vectors of human diseases (47% sequence identity with Dm1 from Aedes aegypti and Anopheles gambiae; (9)) and mammals (44% identity with human MMP-11, -13, and -20; (9)) than to the few other bacterial sequences found in genomic sequences. Accordingly it was likewise suggested that it may be the result of horizontal gene transfer of an MMP gene from an animal to an intimate bacterial pathogen, which inhabits a biofilm on the tooth surface in humans (9).

The metzincins are characterized by a consensus sequence responsible for binding of the catalytic zinc ion (CSBZ), H-E-X-X-H-X-X-[G/N]-X-X-[H/D] (amino acid one-letter code; X stands for any residue), and a conserved methionine-containing turn, the “Met-turn” (1-5,36). In MMPs, the CSBZ encompasses three histidine zinc ligands, the general base/acid glutamate for catalysis, and a structurally-relevant glycine (3). In addition, the distinct MMP paralogs are multi-domain proteins that display a disparate domain organization that is the result of successive polyplication, gene fusion and exon shuffling (11). The only domains common to all animal and plant MMPs are a signal peptide, which is removed after secretion, a pro-domain and a catalytic domain, as found e.g. in human MMP-7 and MMP-26, and in plant MMPs (12,16,18).

Most peptidases are biosynthesized as zymogens containing pro-segments, which are required for latency maintenance to prevent unbridled activity but also sometimes to assist in proper folding of the usually downstream catalytic moieties (37-40). Metzincin exceptions lacking pro-segments include the arachematzincins, for which no hydrolytic activity has so far been reported, i.e. they might not need to be kept latent (41,42); the toyxilysin ExxA from Escherichia coli, whose soluble expression requires co-expression with its cognate ExxB subunit, thus pointing to a chaperone-like function for this ancillary subunit (43-45); the cholerilysin SteE from E. coli, for which an N-terminal immunoglobulin-like domain may assist the downstream catalytic moiety in proper folding (46); and igalsins, where an all-ß-domain of similar topology to immunoglobulin-like domains is likewise found at the N-terminus of the catalytic moiety (see Protein Data Bank [PDB] access codes 4DF9 and 3P1V, and (5)).

MMP pro-domains (see Table 1 in (47)) span 60-90 residues and include a conserved sequence motif, P-R-C-G-(V/N)-D, engaged in a “cysteine-switch” or “Velero” mechanism of latency (10,16,48-51). It has been suggested that this mechanism may be shared by variants within other metzincin families, for which conserved cysteines were described upstream of the catalytic domain, such as the ADAMs/adamains (motif P-K-M-C-G-V, (8,52-54)), leishmanolysins (motif H-R-I-H-D; (2)), and pappalysins (motif C-G; (55)). In contrast, the 472 residues encoded by the karylsin gene (see UniProt sequence database [UP] access code D0E7M7) only comprise a short 14-residue potential pro-peptide—which lacks cysteines—between the 20-residue signal peptide and the 161-residue mature catalytic moiety (Fig. 1A). A C-terminal domain of 277 residues of unknown function and sequence unrelated to any domain found in eukaryotic MMPs completes the protein. This strongly suggests a potentially different mechanism of latency maintenance, hitherto unseen not only in MMPs but also in metzincins in general, as the shortest pro-segments described to date are those of members of the astacin family, which span >34 residues (7,56-58).

We had previously determined the structure of the catalytic domain of karylsin (termed Kly18; (9)). To shed light on the molecular determinants of the first mechanism of latency maintenance of a bacterial MMP, in this work we assayed the possible function of the pro-peptide in folding, stability, and activity inhibition of Kly18. We further solved the X-ray crystal structure of an active-site mutant of a construct spanning the pro-peptide and Kly18 affecting the catalytic glutamate, pKly18-E150A, to circumvent autolysis. The mechanism derived was supported by site-directed mutagenesis and it is discussed in the context of general MMP latency maintenance.

EXPERIMENTAL PROCEDURES

Protein production and purification — The gene coding for full-length wild-type T. forsythia pro-karylsin without the 20-residue signal peptide (hereafter pKly;
52KDa; residues Q21-K372 according to UP D0EM77, see also Fig. 1A) was cloned at BamHI and XhoI restriction sites into vector pGEX-6P-1 (GE Healthcare) as described elsewhere (30). The resulting vector, pKAR1 (see Table 1 for an overview of vectors and constructs used), confers resistance towards ampicillin and attaches an N-terminal glutathione-S-transferase (GST) moiety followed by a human rhinovirus 3C proteinase (HR3CP) recognition site (L-E-V-L-F-Q-L-G-P; HR3CP cleavage leaves two extra residues—underlined—at the N-terminus of the recombinant protein after digestion; three extra residues, L-G-S, are further present due to the cloning strategy). Single-residue point mutants pKly-Y35A and pKly-E156A (pKAR2 and pKAR3, respectively) were generated using the QuikChange Site-Directed Mutagenesis Kit (Stratagene) according to the manufacturer's instructions as described (30). Double mutant pKly-D35A+Y35A (pKAR4) was similarly generated using pKAR2 as a template. Genes coding for the E156A-mutated catalytic domain of kalirin, with and without the pro-peptide (hereafter pKly18-E156A and Kly18-E156A; 20KDa and 18KDa; residues Q21-S201 and residues Y35-S201, respectively), were also cloned into vector pGEX-6P-1 (pKAR5 and pKAR6, respectively). Genes coding for pKly18 and its mutant proteins pKly18-Y35A and pKly18-D35A+Y35A were cloned into the same vector (pKAR7, pKAR8, and pKAR9, respectively) following a strategy previously described (59). Genes coding for pKly18, pKly18-E156A, Kly18 and Kly18-E156A were furthermore cloned at NcoI and XhoI restriction sites into vector pCRI-7a (59), which confers resistance towards kanamycin and does not attach fusion proteins (pKAR10-pKAR13, respectively). In these cases, the cloning strategy entailed that residues M-G were attached at the N-terminus. All constructs were verified by DNA sequencing.

Proteins encoded by vectors pKAR1-pKAR9 were produced by heterologous overexpression in E. coli BL21 (DE3) cells, which were grown at 37°C in Luria Bertani medium supplemented with 100μg/ml ampicillin. Cultures were induced at an A600 of 0.8 with 0.2-mM IPTG and incubated either for 5h at 37°C or overnight at 18°C. Cells were harvested by centrifugation at 7,000xg for 30min at 4°C, washed in buffer A, resuspended in the same buffer, and further lysed in an ice-bath using a digital sonifier (Branson). After centrifugation at 15,000xg for 30min at 4°C, both cell debris and supernatant were analyzed by 15% Tricine-SDS-PAGE stained with Coomassie blue.

Protein identity and purity were assessed by mass spectrometry using an Autoflex Bruker apparatus and by N-terminal sequencing through Edman degradation at the Proteomics Facility of Centro de Investigaciones Biológicas (Madrid, Spain). Ultrafiltration steps were performed with Vivaspin 15 and Vivaspin 4 filter devices of 5-KDa cut-off (Sartorius Stedim Biotech). Approximate protein concentration was determined by measuring A280 in a spectrophotometer (NanoDrop) using the calculated absorption coefficients E1% = 2.32 and 2.42 for pKly18-E156A and Kly18-E156A, respectively.

**Autolytic activation and pro-peptide inhibitory activity assays** —Mutants pKly-Y35A (from pKAR2), pKly-D35A+Y35A (pKAR4), and pKly18-Y35A (pKAR8) were incubated in buffer B at 37°C and at 0.4mg/ml final protein concentration for up to 120h to assay autolysis. Reactions were stopped at specific time points by boiling aliquots in reducing/denaturing buffer, and samples were further analyzed by 10% or 15% Tricine-SDS-PAGE stained with Coomassie blue. Kly18, obtained by autolysis from pKAR1-encoded protein, was incubated at 0.025μg/ml final protein concentration for 30min with 0.1-10mM of peptide Q-R-L-Y-D-N-G-P-L-T (purchased from GL Biochem Ltd), which mimics the pro-peptide sequence. Proteolytic activity was subsequently measured at 37°C in buffer C on substrate Mca-R-P-K-P-V-E-Nva-W-R-K(dnp)-NH2 (Bachem; at 10μM) in a microplate fluorimeter (Infinite M200, Tecan).

**Thermal shift assays** — Aliquots were prepared by mixing 7.5μl of 300x Sypro Orange dye (Molecular Probes) and 42.5μl of either pKly18-E156A (from pKAR5) or Kly18-E156A (pKAR6) at 1-2mg/ml in buffer C in the absence and in the presence of 1-5mM CaCl2. Four replicates of each aliquot were analyzed in an iQ5 Multi Color Real Time PCR Detection System (Bio-Rad) in 96-well PCR plates sealed with optical tape. Samples were heated from 30°C to 95°C at 0.5°C/min, and the change in absorbance (Δm = 490nm; λem = 575nm) was monitored over time. The temperature of midtransition (Tm; 60) was determined for both proteins from the inflection point of each curve using iQ5 software.

**Crystallization and data collection** — Crystallization assays of pKAR5-encoded pKly18-E156A protein were carried out at the IBMB/IRB Crystallography Platform by the sitting-drop vapor diffusion method using 96x2-well MRC plates (Innovadyne). A TECAN Freedom EVO robot was used to prepare reservoir solutions, and a Phoenix/RE (Art Robbins) robot or a Cartesian Microsys
RESULTS AND DISCUSSION

Roles of the pro-peptide in vitro and in cellula

— Wild-type karylysins is secreted as a zymogen with a 14-residue N-terminal pro-peptide (Q1-R-E-L-Y-D-N-G-P-L-T-G-D-N-N35), which is cleaved off at position N33-Y35 during maturation (Fig. 1A). This is the primary activation cleavage and it releases an active 48-KDa form (Kly48; (30)). In recombinant protein production, subsequent cleavages within the C-terminal domain give rise to Kly38 and, finally, to a stable form of 18KDa (Kly18), which corresponds to the isolated mature catalytic domain (CD; (28,30,33,78)). These cleavages were shown to be autolytic as activation was repressed by general chelating MP inhibitors and in the inactive active-site variant, E156A, which ablated the catalytic glutamate of the CSBZ (1,5,30,79,80). In addition, cleavage-site mutant V35A, which does not match the enzyme’s specificity, was activated only slowly when compared with the wild type (30,33).

To assess whether the pro-peptide had a chaperone-like function on the downstream catalytic moiety, we cloned the genes encoding pKly18-E156A and Kly18-E156A in a vector that does not attach a fusion protein at the N-terminus that would assist in proper folding (pKAR11 and pKAR13, respectively; see “Experimental procedures”). We found that the active-site mutant pKly18-E156A was successfully overexpressed in soluble form (Fig. 1B). In contrast to thezymogen, Kly18-E156A was produced only in insoluble form (Fig. 1B). Moreover, when expressed from the pKAR6 vector, which attaches an N-terminal glutathione-S-transferase fusion protein (see Table 1), Kly18-E156A was obtained with ~10-times lower yield than the pro-protein (vector pKAR5). We conclude that the pro-peptide plays a major role in proper folding of Kly18 as previously described for other MPs such as fragilysin (26,27), funnellin, metallocarboxypeptidases (79,81,82), and ADAMs/adamalysins (54) but not for mammalian MMPs (83).

We further examined the effect of the pro-peptide in response to denaturation by a thermal shift assay following the thermofluor approach (60). Purified pKly18-E156A (pKAR5) showed two unfolding transitions compatible with unfolding of pro-peptide and CD, with Tm values of 60°C±0.5°C and 74°C±2.0°C (Fig. 1C). In contrast, the unfolding of purified Kly18-E156A (pKAR6) showed a single transition, with a Tm of 49°C±2.2°C. The addition of a physiological concentration of calcium resulted in a substantial increase in stability of both Kly18-E156A and Kly18-E156A. Accordingly, the former showed Tm values of 67.5°C±1.7°C and 76.5°C±1.2°C, and 69.5°C±1.7°C and 79°C±2.2°C, in the presence of 1mM and 5mM CaCl2, respectively, whereas those of Kly18-E156A were 52.5°C±1.2°C and 54.5°C±1.4°C. This result is in agreement with the important role of calcium in Kly18 activity, as addition of 2-5mM CaCl2 is reported to enhance activity about three times (30). Thus, regardless of calcium, the 14-residue pro-peptide redounded to a dramatic increase in Tm, underpinning that it plays a major role in the thermal stability of the zymogen. Finally, we assayed the effect of a decapetide spanning pro-peptide sequence Q21-T30 on the activity of purified mature Kly18 (pKAR1) in the presence of a fluorogenic peptide (Fig. 1D).
observed a weak but consistently concentration-dependent inhibitory effect as previously shown for other MPs when their pro-peptides or pro-domains were added in trans, among others funnels (79,81), ADAMs/adamalysins (84), and mammalian MMPs (85-87).

Summarizing, the pro-peptide of karilysin is the shortest currently described to date for an MP, and it exerts all roles, which collectively or selectively had been previously described for peptidase pro-peptides or pro-domains: latency maintenance, folding assistance during biosynthesis, stability to thermal denaturation, and inhibition of peptidolytic activity (38,39,81).

**Overall structure of pKly18** — Due to rapid autolytic processing of recombinant wild-type pro-karilysin (30), crystals of pKly18 could only be obtained for an inactive variant affecting the catalytic glutamate (pKly18-E158A), as already reported for other MP zymogens (88-92). This protein crystallized as monoclinic crystals diffracting to 2Å resolution with two molecules per asymmetric unit. These were essentially identical (Ca-atom rmsd = 0.53Å) except for segments N95-H57 and N34-G39. The latter flanks the primary activation cleavage plane and is flexible. It is stabilized through an interaction with segment N87-N90 of the second molecule present in the asymmetric unit of the crystal though in different conformations in molecules A and B, so the discussion hereafter is centered on molecule A if not otherwise stated. When two values are indicated, they refer to both molecules.

The protein reveals a compact, almost spherical shape of ~40Å in diameter and is subdivided into three moieties (Fig. 2A): the N-terminal pro-peptide (Q51-N153), and a CD split into a larger N-terminal upper sub-domain moiety (Y35-G162; NTS) and a smaller C-terminal lower sub-domain moiety (115-P186; CTS, see also (9,28)) if viewed in the standard orientation for MPs (35). NTS and CTS conform to the overall fold of vertebrate MMPs (47,93) and are separated by a shallow active-site cleft. The NTS is an α/β-sandwich consisting of a twisted five-stranded pleated β-sheet (strands β1-β7; see Fig. 2A), which is parallel for its first four strands and antiparallel for its lowest one, β4. The sheet accommodates on its concave side two α-helices (the “backing helix” αA and the “active-site helix” αB; for numbering and extension of repeating secondary structure elements, see Fig. 2C of (9)). The right-handed twist of the helices coincides with the right-handed twist of the sheet and both helices’ axes intersect the strands of the sheet at an angle Ω ≈ 35° (94). The two helices pack against each other interacting through A46-A70 of αA and L155-A158 of αB at a crossing angle Ω ≈ 50°, which corresponds to a class-II helix interaction (94). The loop connecting strands β31 and β36 (L31β36β36β37) contains the “S-loop” (G106-L115), which encompasses first a binding site for a structural calcium (Zn998) and, downstream, a binding site for a structural calcium (Ca997; see Fig. 2B). The zinc is tetrahedrally coordinated by H102 Ne2, D104 O82, H117 Ne2, and H133 N81, while the calcium is octahedrally coordinated by six oxygens: D109 O61, G110 O, T112 O, L114 O, D132 O82, and E138 O2 (see Fig. 2B and the legend for details). The presence of calcium is consistent with its crucial role in catalysis (30) and in protein stability (see Fig. 1C). Such calcium is found in mammalian MMP structures (47,93), but it was not found in previous mature Kly18 structures (see below and (9,28)). At G102 of the CSBZ, the polypeptide chains takes a sharp turn and enters the CTS (Fig. 2A), which mainly contains the “C-terminal helix” αC and the Met-turn—centered on M173—which forms a hydrophobic base for the catalytic metal-binding site and is required for its integrity in MMPs and metzincins in general (47).

The active-site cleft contains the catalytic zinc ion (Zn999) at half width coordinated by the three histidines of the CSBZ (H135, H139, and H989) through their Ne2 atoms at distances 2.00-2.05Å (Fig. 2A,C). The cleft is top-framed on its non-primed side (see (35,95)) by the “upper-ribbon strand” βIV of the NTS β-sheet, which in MMPs binds substrates in extended conformation from left to right through antiparallel β-ribbon-like interactions. On its primed side, the cleft is top-framed by the final stretch of the S-loop, termed the “bulge-edge segment” (T112-L115), and bottom-framed by the segment bridging the Met-turn and helix αC. This segment includes the “S1’-wall forming segment” (P175-Y177) at the front and the “specificity loop” (G179-Q183) at the back. Together with the first turn of the active-site helix αB, the latter structural elements contribute to the size and chemical nature of the S1’ pocket, which confers specificity to Kly18 and also MMPs in general (47,93), here for medium-sized to bulky hydrophobic residues (30). Side chains participating in pocket shaping include L115, A116, T151, V152, H155, L172, Y177 and K181.

**Inhibition by the pro-peptide** — The 14-residue pro-peptide starts at the front right and runs in extended conformation across the active-site cleft, thus blocking access to the cleft, though in the opposite direction to a substrate, i.e. right to left (Fig. 2A,C). This reverse orientation of the pro-peptide in the cleft may contribute to attenuate autolysis, as previously suggested for zymogens of cysteine peptidases and mammalian MMPs (39). The interaction with the CD buries a surface of 2,100±35Å², which is much larger than the average of monomeric protein-protein domain intra-chain interfaces (1,193Å²; (96)) but is slightly lower than the range of typical MMP-protein inhibitor interaction surfaces (2,400-2,700Å²; see (97)). The interaction includes 13 hydrogen bonds, a double salt bridge, one metalorganic bond, and hydrophobic carbon-carbon contacts between eight residues from the pro-peptide and eleven from the CD (see Table 3). Segments involved include almost the entire pro-peptide (R22-G31) and, from the CD, mainly N111-Y120 from the bulge-edge segment and the upper-ribbon strand, and P175, Y177 from the S1’-wall forming segment. Further involved are Y106, A124, E138 and the zinc-liganding histidine side chains. Four inter-main chain hydrogen bonds form on the primed side of the cleft (two with the S1’-wall forming segment and two with the bulge-edge segment and strand βIV) and three more on the upstream non-primed side (with βIV and LβIVβV; Fig. 2C). In particular, R22 contacts the base of the S-loop: it doubly salt-bridges E138, which is also one of the calcium ligands (see above, Table 3 and Fig. 2B), and hydrogen bonds three carboxyl oxygens of the S-loop, N111, G113, and T112, which, again, is also a calcium ligand. In addition, the R22 carboxyl oxygen binds the S1’-wall forming segment and its side chain performs a
hydrophobic interaction with L115. Accordingly, this residue plays a major role in the stabilization of the Ca997 site and, thus, thezymogen in general, which explains its enhanced stability in response to thermal denaturation (see above). In addition, superposition of pKly18-E156A onto mature Kly18 in complex with a tetrapeptidic cleavage product in the primed side (see below) and human MMP-8 with a modeled substrate traversing its cleft based on inhibitor structures (98) indicates that R22 occupies the S1' position of the cleft.

However, the most important interaction of the pro-peptide with the CD is exerted by D25, which approaches the catalytic zinc from the top and monodentately occupies through its Oδ1 atom the fourth position of the tetrahedral coordination sphere of the metal (2.00/2.04Å apart; Fig. 2C) further to H155, H159, and H165 Ne2 atoms. The preceding carbonyl group of Y24 binds strand βIV, and its aromatic side chain penetrates the deep hydrophobic S1' pocket, mainly interacting with the H155 ring face-to-face. The π-rings are ~3.5Å apart and parallel but slightly displaced along the ring planes to form a half-overlapping sandwich, which gives rise to an optimal π-stacked structure (99). Downstream in the chain, P8 is in a pocket, probably S2, framed by H159, E164, and Y120—the latter two interact through a tight hydrogen bond (Y120 Oη-E164 Oε2, 2.61Å). Residue L29 is surrounded by the side-chains of Y106, H117, and F119, which may feature S2 (Fig. 2C). After G31, the polypeptide abandons the active-site cleft moving outward to reach the primary activation cleavage point, N34-Y35 (Fig. 2A), after which the chain folds back towards the molecular moiety and enters strand βII of the NTS β-sheet.

A novel activation mechanism in MMPs —

Previous work had yielded three structures of mature wild-type Kly18 in complexes with tri- and tetrapeptidic cleavage products, as well as an inhibitory tetrapeptidic in the non-primed side of the cleft (PDB 2XS3, 2XS4, and 4IN9; (9,28)). These were obtained both in the presence and in the absence of magnesium and showed deviating chain traces for segment N53-H27 (LβIαA) in the two molecules found in the asymmetric unit of the magnesium-unbound structure (PDB 2XS3; (9)) and in the single molecules found in magnesium-bound (PDB 2X4; (9)) and inhibitor-bound crystals (PDB 4IN9; (28)). In addition, significant differences were also found in the second half of the S-loop including the bulge-edge segment, which was metal-free in all structures, as the aforementioned magnesium—which coincides with a potassium site in the inhibitor-bound form—was found on the opposite surface of the CD (see Fig. 1A,C in (9) and Fig. 1A in (28)), in a place that suggests little if any functional or structural relevance. In these structures, either an outward- or an inward-folded flap was found for the S-loop (Fig. 1E in (9) and Fig. 1D in (28)), which suggests intrinsic flexibility of this protein segment to adapt to different substrates. Among the distinct Kly18 coordinates, molecule A of the magnesium-unbound structure (PDB 2XS3) was chosen here for comparison with pKly18-E156A as it showed the lowest divergence in the overall chain trace (Fig. 2D-F).

Superposition revealed that the mature CD is preformed in the zymogen and, with some notable local exceptions (see below), is uncovered by removal of the pro-peptide, as found in mammalian MMPs (47) and other MPs such as funnelins (79,82). Removal occurs through cleavage at N34-Y35, which is solvent exposed on the molecular surface and thus readily accessible for processing (Fig. 2A). This explains why the wild-type zymogen undergoes rapid autolysis, so it cannot be isolated intact (see (30) and first section of “Results and discussion”). This was the first cleavage observed in vitro, thus termed primary activation cleavage site, and no further cleavage was detected either within the pro-peptide or in the CD. The site is consistent with most vertebrate MMPs being activated at X-F/Y' bonds, which are found at similar regions in all structures (10). Pro-peptide removal occurs under loss of a number of protein-protein interactions (see Table 3 and preceding section), which explains why the mature enzyme is less stable to thermal denaturation (see first section of “Results and discussion”). In particular, R22 plays a key role in stabilizing the Ca997 site (see above), and its removal may contribute to cation-site and S-loop flexibility, leading to metal loss. This site is easily created from the unbound form by two glycine-mediated main-chain rotations (peptide flip of bond T112-G113 and ~70° rotation of peptide bond G110-N111), so as to orient the carboxyl oxygens towards the interior, and cation binding should largely compensate for the energetic cost of such minor rearrangement. However, the finding that none of the mature Kly18 structures—which were partially obtained in the presence of calcium, see (9)—contained an intact calcium-site supports the requirement of R22 as an additional stabilizing factor for site integrity.

Activation further entails that the position occupied by D25 Oδ1 in the ligand sphere of the catalytic zinc (see preceding section) is taken over by a catalytic solvent molecule, which renders a competent active site following an “aspartate-switch” mechanism. Such a competent zinc environment has also been reported for several mature MPs (see e.g. (64,80,100)). To date, aspartate-switch zymogenic mechanisms have been described only for astacins (7,88) and fragilysins (26), which are only distantly related MPs grouped with MMPs within the metzincins. To verify the function of D25 in latency in pKly18, we used mutant pKly18-Y35A (pKAR8), as the wild-type form (pKAR7) was insoluble. While this mutant was produced with a yield similar to that of pKly18-E156A and was stable for several days, mutant pKly18-D33A+Y35A (pKAR9) was insoluble. We further assessed the function of D25 in full-length karylsin using the slowly-autoytic mutant pKly-Y35A (pKAR2), as the reaction in the wild type is too rapid (30). While pKly-Y35A was essentially intact after 5 days at 37°C, pKly-D33A+Y35A (pKAR4) had been entirely transformed into the 38-KDa and 18-KDa forms after this time (Fig. 1E,F). Taken together, these results support the essential role of D25 in latency maintenance.

As to further changes upon maturation, segment P122-A129 from LβIβVβV is slightly shifted downwards by ~2Å and the side chains of Y120 and E164 rotate towards the zinc site (Fig. 2E). Activation only entails major rearrangement of the new N-terminal segment Y35-S40, on the left surface (Fig. 2D,E), which is rotated downward around bonds C-Cu and Cu-N of S40. In this way, this segment nests in a surface cavity framed by helix αC and
the first segment of the CTS between G162, and the “family specific residue”, which is a serine in MMPs (1,101; here S166). This entails that the new α-amino group of Y35, which is translated 25Å, establishes an intra-molecular salt bridge with D187 of αC, which is vaguely reminiscent of the activation of trypsin-like serine peptidases (102). D187, in turn, is itself further bound to S166 and is adjacent to a second aspartate, D188, which binds two main-chain amides of the Met-turn. This electrostatic network is characteristic of physiologically-relevant mature MMPs, also referred to as “superactive forms” (47,103). With the exception of the mature N-terminal fragment, the rest of this electrostatic network is already present in the zymogen (Fig. 2E).

Intensive studies of the activation of mammalian MMPs have produced the structures of pro-MMP-1 (PDB 1SU3; (90)), pro-MMP-3 (PDB 1SLM; (104)), pro-MMP-9 (PDB 1L6J; (105)), and pro-MMP-2 (PDB 1EAK; (89)). These studies revealed that the mammalian MMP zymogens contain a pre-formed competent protease moiety and true pro-domains, which span between 66 and 91 residues, as shown for pro-MMP-2 (Fig. 3A) (47). The pro-domains include elongated N-terminal extensions that may interact with ancillary domains, such as the fibronectin type-II insertions found in MMP-2 and MMP-9, followed by globular cores of ~55 residues. These are made up of three α-helices that are arranged around a three-fold axis with a left-handed twist.

The pro-domain globular core serves as a scaffold to place a downstream peptide, which runs in extended conformation in the opposite direction to a bound substrate and thus blocks the active-site cleft (Fig. 3A,B). This peptide encompasses the conserved motif involved in cysteine-switch or Velcro latency characteristic of animal and plant MMPs (48-50), P109-R-C-G-N-P-D106 (MMP-2 residues in italics; see PDB 1EAK and UP P08253), which is equivalent to pKly18 segment L23-Y-D-N-G-P-L29 (Fig. 3C,D). Both the cysteine- and aspartate-switch motif show an intricate electrostatic network producing a unique scaffold to interact with the mature catalytic domain moiety. In contrast to pKly18, where the first cleavage occurs in the primary activation cleavage site, however, classical mammalian pro-MMPs are activated by conformational changes in the pro-domain induced by cleavage in a so-called “bait region” by several peptidases such as trypsin, plasmin, and other MMPs. Activation follows a “stepwise activation” process to eventually yield the final cleavage site X-F/Y accessible for processing and dissociation of cysteine and zinc to generate a functional active site (48,49,51,106-108). As in Kly18, after cleavage at N109-Y110, the new N-terminus is rearranged and participates in the electrostatic network centered on the conserved aspartate of helix αC, D246 in MMP-2.

**Conclusions** — This examination of the structure and function of the zymogen of the first bacterial MMP to be studied biochemically has uncovered several features of the activation mechanism of pKly18, which are shared with animal and plant MMPs: (i) the relevant cleavage site is X-F/Y; (ii) the scissile bond is located in similar regions of the structure; (iii) activation entails rearrangement of the segment equivalent to Y35-S160 to yield a salt-bridge between the new α-amino group and the first of two conserved aspartates in helix αC; (iv) this aspartate is bound to the family-specific serine; (v) the aspartate immediately downstream binds two main-chain amides of the Met-turn; (vi) the inhibitory segments runs across the cleft in the opposite direction to a genuine substrate and metal blockage occurs through the side chain of an intervening residue, not through a chain terminus; and (vii) the catalytic moiety is largely preformed in the zymogen. All these features are related to the highly conserved CD itself. In contrast, all the features of the mechanism related to the segment preceding this conserved CD diverge: (i) in pKly the pro-peptide spans just 14 residues and does not contain repetitive secondary structure elements, while vertebrate and plant MMPs feature a true protein pro-domain that folds into a pseudosymmetric three-helix bundle followed by a segment in extended conformation; (ii) no relevant sequence identity is found between the pro-regions; (iii) in eukaryotic MMPs activation occurs through a cysteine-switch mechanism exerted by residues from a conserved sequence motif, while in pKly18 this motif is absent and activation follows an aspartate-switch mechanism; (iv) multiple cleavages are apparently required in eukaryotic MMPs to liberate the CD while a single cleavage suffices in pKly; and (v) the pro-domain is not required for (re)folding of the catalytic moieties in eukaryotic MMPs while it is in karilysin. In addition, pKly shares parts of its mechanism of latency with otherwise unrelated MPs from the astacin and fragilysin families. Accordingly, this overall novel mechanism unveiled for MMPs supports previous hypotheses, according to which Kly18 originated from an animal MMP CD co-opted through horizontal gene transfer by T. forsythia. This transfer was fostered by the intimate coexistence of the latter with the human blood-irrigated gingival crevice. Subsequently, Kly18 would have evolved in a bacterial environment, where it was furnished with unique flanking domains that contribute to a mechanism of zymogenicity similar to distantly related MPs only (9).
REFERENCES


22. Klein, T., and Bischoff, R. (2011) Active metalloproteinases of the A Disintegrin and Metalloproteinase (ADAM) family: biological function and structure. J. Proteome Res. 10, 17-33


ACKNOWLEDGMENTS

We are grateful to the joint IBMB/IRB Automated Crystallography Platform for assistance during crystallization experiments and to Robin Rycroft for very helpful contributions to the manuscript. We acknowledge the help provided by local contacts at the ESRF synchrotron. The clone for production of human rhinovirus 3C proteinase was a generous gift from Arie Gerlof (EMBL, Hamburg).

FOOTNOTES

*This study was financially supported in part by grants from European, US American, Polish, Spanish, and Catalan agencies (UMO-2012/04/A/NZ1/00051; 2137/7.PR-EU/2011/2; 2975/7.PR/13/2014/2; DE09761; DE022597; FP7-HEALTH-2010-261460 “Gums&Joints”; FP7-PEOPLE-2011-ITN-290246 “RAPID”; FP7-HEALTH-2012-306029-2 “TRIGGER”; BU2012-32862; BIO2013-49320-EXP; CSD2006-00015; and 2014SGR9). Funding for data collection was provided in part by ESRF. The Faculty of Biochemistry, Biophysics and Biotechnology of the Jagiellonian University is a beneficiary of structural funds from the European Union (POIG.02.01.00-12-064/08).

FIGURE LEGENDS

**Figure 1. Effect of the Kly18 pro-peptide in vitro and in vivo.** (A) Scheme depicting the domain structure of T. forsythia karilysin. Numbering according to UP D0EM77. (B) SDS-PAGE of cultures of wild-type and E156A variants of pKly18 and Kly18. Lanes 1 and 2, insoluble and soluble fractions of wild-type pKly18 (from pKAR10), respectively. Lanes 3 and 4, insoluble and soluble fractions of wild-type Kly18 (pKAR12), respectively. Lanes 5 and 6, insoluble and soluble fractions of pKly18-E156A (pKAR11), respectively. Lanes 7 and 8, insoluble and soluble fractions of Kly18-E156A (pKAR13), respectively. Overexpressed proteins are labeled with an asterisk. (C) Unfolding transition curves showing temperature-dependent change in fluorescence of pKly18-E156A (pKAR5; solid line) and Kly18-E156A (pKAR6; dashed line) in the absence (black curve) and in the presence of CaCl2 (1mM, red curve; 5mM, green curve). The blank curve is indicated with a dotted line. (D) Proteolytic activity of Kly18 (pKAR1) at 37°C using substrate Mca-R-P-K-P-V-E-Nva-W-R-K(dnp)-NH3 at 10µM in the absence (0) and in the presence of 0.1, 0.5, 1, 2, and 5mM pro-peptide mimic. (E,F) Stability of mutant pKly-Y35A (E) and mutant pKly-Y35A+D25A (F) over time. Kly48, Kly38, and Kly18 are indicated by arrows.

**Figure 2. Overall structure of pKly18-E156A.** (A) Ribbon-type plot of pKly18-E156A in standard orientation (35). Depicted are the pro-peptide (ribbon and carbon atoms in turquoise) and the mature enzyme moiety (β-strands in yellow labeled βL-βV; α-helices in salmon labeled αA-αC; and coils and carbon atoms in tan). Further shown are the catalytic zinc ion (Zn2+; bottom magenta sphere), the structural zinc ion (Znκ2; top magenta sphere), and the structural calcium ion (Caκ2; red sphere), as well as the side chains of the three catalytic zinc ligands (H155, H159, and H165), the Met-turn methionine (M172), the alamino replacing the catalytic glutamate (A156), and the residues flanking the primary activation cleavage point, N34-Y35. (B) Close-up of the window of (A) as stick model highlighting the structural zinc and calcium sites. Protein segments depicted are N101-G105 from the first part of the S-loop (carbons in gold), D109-F113 from the second part of the S-loop (carbons in sandy brown), H131-E138 from βV-LβVeB (carbons in tan), and the side chain of R32 from the pro-peptide (carbons in turquoise). The zinc is bound by H155-Nε2, D109-Oδ2, H159-Nε2, and H165-N61 at distances 1.99-2.06Å, and the calcium is bound by D109-O61, G110-O, T112-O, I114-O, D135-O62, and E138-O2 at distances 2.34-2.39Å. These distances agree with standard zinc- (1.99-2.09Å; (109)) and calcium-binding (2.56-2.39Å; (109)) distance values for oxygens and nitrogens. (C) Close-up of (A) in wall-eye stereo centered on the catalytic zinc after a horizontal ~30°-rotation upwards. Selected hydrogen and ionic bonds (see also Table 3) are depicted as green lines. Residues and ions labeled in (A) are not labeled here for clarity. The pro-peptide is shown in cyan to distinguish it from the mature catalytic moiety (in tan/yellow/orange) and its chain direction is pinpointed by a cyan arrow and labels of the N- and C-terminal parts depicted. (D) Superposition in wall-eye stereo of pKly18-E156A (ribbon in tan for the mature enzyme moiety and in brown for the pro-peptide, zinc ions in magenta, and calcium ion in red; stick model for the side chains of S29,Y35 with carbons in brown) and Kly18 (ribbon and zinc ions in pink, see PDB 2XS3, molecule A; (9)), which was obtained in a product complex with peptide A-F-T-S bound to the primed side of the cleft (stick model with carbons in red). Y35 is shown for both structures. (E) Detail of (D) in wall-eye stereo depicting the large rearrangement of the N-terminus at Y35 after maturation cleavage at N34-Y35. The α-amino group of Y35 makes a salt bridge with the side chain of D187 in the mature enzyme. Aside from Y120 and E164 (significantly) and P222, A129, (slightly; see black arrows), maturation does not entail major conformational rearrangement of the rest of the structure.

**Figure 3. Structural comparison with mammalian pro-MMPs.** (A) Cartoon depicting the structure of pro-MMP-2 (PDB 1EAK; (89); MMP-2 residues in italics with superscripted numbering), shown only for its CD (Y140-D145 in cyan; the fibronectin type-II domains spanning Q19-D392 have been omitted, the black arrows pinpoint the insertion points) and pro-domain (P45-A66 in pink, without the first eleven residues in extended conformation). The orientation displayed corresponds to that of Fig. 2A after applying a horizontal rotation of 15°. Residues of the conserved motif (P100-D106) key for structural integrity of the inhibitory segment are depicted for their side chains. (B) Close-up of (A) after removal of pro-domain segment P45-A66 to provide insight into the interactions of the conserved motif. Key electrostatic interactions are shown as green lines. The catalytic glutamate, E404, is replaced by a glutamine, the histidines from the CSBZ are H465, H469, and H471.
(C,D) Scheme depicting the interaction modi of the pro-peptides of pro-MMP-2 through a cysteine-switch mechanism (C) and pKly18 through an aspartate-switch mechanism (D). The catalytic zinc ions are shown as magenta spheres and relevant interactions are shown as yellow dashed lines.
<table>
<thead>
<tr>
<th>Name</th>
<th>Original vector</th>
<th>Antibiotic resistance</th>
<th>Restriction sites</th>
<th>Insert</th>
<th>Fusion construct (N-term.)</th>
<th>Additional N-terminal residues</th>
</tr>
</thead>
<tbody>
<tr>
<td>pKAR1</td>
<td>pGEX-GP-1 (GE Healthcare)</td>
<td>amp</td>
<td>BamHI / Xho</td>
<td>wt Q&lt;sup&gt;21&lt;/sup&gt;-K&lt;sup&gt;472&lt;/sup&gt; (pKly)</td>
<td>GST + HR3CPr</td>
<td>G-P-L-G-S *</td>
</tr>
<tr>
<td>pKAR2</td>
<td>pGEX-GP-1 (GE Healthcare)</td>
<td>amp</td>
<td>BamHI / Xho</td>
<td>Y&lt;sup&gt;35&lt;/sup&gt;A mut. Q&lt;sup&gt;21&lt;/sup&gt;-K&lt;sup&gt;472&lt;/sup&gt; (pKly-Y&lt;sup&gt;35&lt;/sup&gt;A)</td>
<td>GST + HR3CPr</td>
<td>G-P-L-G-S *</td>
</tr>
<tr>
<td>pKAR3</td>
<td>pGEX-GP-1 (GE Healthcare)</td>
<td>amp</td>
<td>BamHI / Xho</td>
<td>E&lt;sup&gt;156&lt;/sup&gt;A mut. Q&lt;sup&gt;21&lt;/sup&gt;-K&lt;sup&gt;472&lt;/sup&gt; (pKly-E&lt;sup&gt;156&lt;/sup&gt;A)</td>
<td>GST + HR3CPr</td>
<td>G-P-L-G-S *</td>
</tr>
<tr>
<td>pKAR4</td>
<td>pGEX-GP-1 (GE Healthcare)</td>
<td>amp</td>
<td>BamHI / Xho</td>
<td>D&lt;sup&gt;25&lt;/sup&gt;A+Y&lt;sup&gt;35&lt;/sup&gt;A mut. Q&lt;sup&gt;21&lt;/sup&gt;-K&lt;sup&gt;472&lt;/sup&gt; (pKly-D&lt;sup&gt;25&lt;/sup&gt;A+Y&lt;sup&gt;35&lt;/sup&gt;A)</td>
<td>GST + HR3CPr</td>
<td>G-P-L-G-S *</td>
</tr>
<tr>
<td>pKAR5</td>
<td>pGEX-GP-1 (GE Healthcare)</td>
<td>amp</td>
<td>BamHI / Xho</td>
<td>E&lt;sup&gt;156&lt;/sup&gt;A mut. Q&lt;sup&gt;21&lt;/sup&gt;-S&lt;sup&gt;201&lt;/sup&gt; (pKly18-E&lt;sup&gt;156&lt;/sup&gt;A)</td>
<td>GST + HR3CPr</td>
<td>G-P- *</td>
</tr>
<tr>
<td>pKAR6</td>
<td>pGEX-GP-1 (GE Healthcare)</td>
<td>amp</td>
<td>BamHI / Xho</td>
<td>E&lt;sup&gt;156&lt;/sup&gt;A mut. Y&lt;sup&gt;35&lt;/sup&gt;-S&lt;sup&gt;201&lt;/sup&gt; (Kly18-E&lt;sup&gt;156&lt;/sup&gt;A)</td>
<td>GST + HR3CPr</td>
<td>G-P- *</td>
</tr>
<tr>
<td>pKAR7</td>
<td>pGEX-GP-1 (GE Healthcare)</td>
<td>amp</td>
<td>BamHI / Xho</td>
<td>wt Q&lt;sup&gt;21&lt;/sup&gt;-S&lt;sup&gt;201&lt;/sup&gt; (pKly18)</td>
<td>GST + HR3CPr</td>
<td>G-P- *</td>
</tr>
<tr>
<td>pKAR8</td>
<td>pGEX-GP-1 (GE Healthcare)</td>
<td>amp</td>
<td>BamHI / Xho</td>
<td>Y&lt;sup&gt;35&lt;/sup&gt;A mut. Q&lt;sup&gt;21&lt;/sup&gt;-S&lt;sup&gt;201&lt;/sup&gt; (pKly18-Y&lt;sup&gt;35&lt;/sup&gt;A)</td>
<td>GST + HR3CPr</td>
<td>G-P- *</td>
</tr>
<tr>
<td>pKAR9</td>
<td>pGEX-GP-1 (GE Healthcare)</td>
<td>amp</td>
<td>BamHI / Xho</td>
<td>D&lt;sup&gt;25&lt;/sup&gt;A+Y&lt;sup&gt;35&lt;/sup&gt;A mut. Q&lt;sup&gt;21&lt;/sup&gt;-S&lt;sup&gt;201&lt;/sup&gt; (pKly18-D&lt;sup&gt;25&lt;/sup&gt;A+Y&lt;sup&gt;35&lt;/sup&gt;A)</td>
<td>GST + HR3CPr</td>
<td>G-P- *</td>
</tr>
<tr>
<td>pKAR10</td>
<td>pCRI-7a (59)</td>
<td>kan</td>
<td>NcoI / Xho</td>
<td>wt Q&lt;sup&gt;21&lt;/sup&gt;-S&lt;sup&gt;201&lt;/sup&gt; (pKly18)</td>
<td>none</td>
<td>G-P- *</td>
</tr>
<tr>
<td>pKAR11</td>
<td>pCRI-7a (59)</td>
<td>kan</td>
<td>NcoI / Xho</td>
<td>E&lt;sup&gt;156&lt;/sup&gt;A mut. Q&lt;sup&gt;21&lt;/sup&gt;-S&lt;sup&gt;201&lt;/sup&gt; (pKly18-E&lt;sup&gt;156&lt;/sup&gt;A)</td>
<td>none</td>
<td>M-G-</td>
</tr>
<tr>
<td>pKAR12</td>
<td>pCRI-7a (59)</td>
<td>kan</td>
<td>NcoI / Xho</td>
<td>wt Y&lt;sup&gt;35&lt;/sup&gt;-S&lt;sup&gt;201&lt;/sup&gt; (Kly18)</td>
<td>none</td>
<td>M-G-</td>
</tr>
<tr>
<td>pKAR13</td>
<td>pCRI-7a (59)</td>
<td>kan</td>
<td>NcoI / Xho</td>
<td>E&lt;sup&gt;156&lt;/sup&gt;A mut. Y&lt;sup&gt;35&lt;/sup&gt;-S&lt;sup&gt;201&lt;/sup&gt; (Kly18-E&lt;sup&gt;156&lt;/sup&gt;A)</td>
<td>none</td>
<td>M-G-</td>
</tr>
</tbody>
</table>

*amp*, ampicillin; GST, glutathione-S-transferase; HR3CP, human rhinovirus 3C peptidase; HR3CPr, recognition sequence for human rhinovirus 3C peptidase; *kan*, kanamycin; mut., mutant; wt, wild-type.

* After cleavage with HR3CP.
Table 2. Crystallographic data.

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group / cell constants (a, b, c in Å and β in °)</td>
<td>P2₁ / 36.18, 121.69, 41.88, 105.26</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.97242</td>
</tr>
<tr>
<td>No. of measurements / unique reflections</td>
<td>104,406 / 22,975</td>
</tr>
<tr>
<td>Resolution range (Å) (outermost shell)</td>
<td>60.8 – 2.01 (2.06 – 2.01)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.8 (96.2)</td>
</tr>
<tr>
<td>R&lt;sub&gt;merge&lt;/sub&gt;&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.105 (0.534)</td>
</tr>
<tr>
<td>R&lt;sub&gt;r.i.m.&lt;/sub&gt; (= R&lt;sub&gt;meas&lt;/sub&gt;)&lt;sup&gt;a&lt;/sup&gt; / CC(1/2)</td>
<td>0.116 (0.616) / 99.6 (84.3)</td>
</tr>
<tr>
<td>Average intensity (&lt;[&lt;I&gt; / s(&lt;I&gt;)]&gt;</td>
<td>10.3 (2.5)</td>
</tr>
<tr>
<td>B-Factor (Wilson) (Å&lt;sup&gt;2&lt;/sup&gt;) / Average multiplicity</td>
<td>29.0 / 4.5 (4.0)</td>
</tr>
<tr>
<td>Resolution range used for refinement (Å)</td>
<td>∞ – 2.01</td>
</tr>
<tr>
<td>No. of reflections used (test set)</td>
<td>22,047 (741)</td>
</tr>
<tr>
<td>Crystallographic R&lt;sub&gt;factor&lt;/sub&gt; (free R&lt;sub&gt;factor&lt;/sub&gt;)&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.165 (0.194)</td>
</tr>
<tr>
<td>No. of protein atoms&lt;sup&gt;d&lt;/sup&gt; / solvent molecules / ligands / ions</td>
<td>2,843 / 226 / 3 glycerols / 4 zinc, 2 calcium</td>
</tr>
<tr>
<td>R&lt;sub&gt;r.m.s.d&lt;/sub&gt; from target values</td>
<td></td>
</tr>
<tr>
<td>bonds (Å) / angles (°)</td>
<td>0.010 / 0.98</td>
</tr>
<tr>
<td>Overall average B-factor (Å&lt;sup&gt;2&lt;/sup&gt;)</td>
<td>30.8</td>
</tr>
<tr>
<td>Model validation&lt;sup&gt;e&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Residues in favored regions / outliers / all residues</td>
<td>345 (97.5%) / 0 / 355</td>
</tr>
<tr>
<td>Residues with bad bonds / bad angles / poor rotamers / Cβ dev&gt;0.25Å</td>
<td>0 / 0 / 7 (2.4%) / 0</td>
</tr>
</tbody>
</table>

Values in parentheses refer to the outermost resolution shell.

<sup>a</sup> R<sub>merge</sub> = Σ<sub>hkl</sub> Σ<sub>i</sub> |I<sub>i(hkl)</sub>| - <I<sub><I(hkl)</sub>|> / Σ<sub>hkl</sub> Σ<sub>i</sub> I<sub>i(hkl)</sub>; R<sub>r.i.m.</sub> = Σ<sub>hkl</sub> |I<sub>i(hkl)</sub>| / Σ<sub>hkl</sub> Σ<sub>i</sub> I<sub>i(hkl)</sub>; R<sub>p.i.m.</sub> = Σ<sub>hkl</sub> |I<sub>i(hkl)</sub>| / Σ<sub>hkl</sub> Σ<sub>i</sub> I<sub>i(hkl)</sub>, where |I<sub>i(hkl)</sub>| is the i-th intensity measurement and n<sub>hkl</sub> the redundancy of reflection hkl—including symmetry-related reflections—and <I<sub><I(hkl)</sub>|> its average intensity. R<sub>r.i.m.</sub> (alias R<sub>meas</sub>) and R<sub>p.i.m.</sub> are improved multiplicity-weighted indicators of the quality of the data, the redundancy-independent merging R factor and the precision-indicating merging R factor. The latter is computed after averaging over multiple measurements (for details, see (110,111)). According to Karplus & Diederichs (112).<sup>b</sup> According to MOLPROBITY (76).
Table 3. Direct interactions between the pro-peptide (PP) and the catalytic domain (CD).

<table>
<thead>
<tr>
<th>PP</th>
<th>CD</th>
<th>Distance (Å) molecule A / B</th>
<th>PP</th>
<th>CD</th>
<th>Distance (Å) molecule A / B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hydrogen bonds</td>
<td></td>
<td>Hydrophobic carbon-carbon interactions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R^{22} O</td>
<td>Y^{177} N</td>
<td>2.73 / 2.73</td>
<td>R^{22} L</td>
<td>L^{155}</td>
<td></td>
</tr>
<tr>
<td>R^{22} Nη2</td>
<td>N^{111} O</td>
<td>3.19 / 3.19</td>
<td>L^{23} Y</td>
<td>Y^{176}</td>
<td></td>
</tr>
<tr>
<td>R^{22} Nη2</td>
<td>T^{112} O</td>
<td>3.20 / 3.21</td>
<td>Y^{24} H</td>
<td>H^{155}</td>
<td></td>
</tr>
<tr>
<td>R^{22} Nη2</td>
<td>G^{113} O</td>
<td>3.24 / 3.22</td>
<td>Y^{24} Y</td>
<td>Y^{177}</td>
<td></td>
</tr>
<tr>
<td>R^{22} Ne</td>
<td>G^{113} O</td>
<td>3.18 / 3.19</td>
<td>D^{25} T</td>
<td>T^{114}</td>
<td></td>
</tr>
<tr>
<td>Y^{24} N</td>
<td>P^{175} O</td>
<td>3.41 / 3.37</td>
<td>G^{27} H</td>
<td>H^{159}</td>
<td></td>
</tr>
<tr>
<td>Y^{24} O</td>
<td>L^{115} N</td>
<td>2.82 / 2.79</td>
<td>G^{27} H</td>
<td>H^{165}</td>
<td></td>
</tr>
<tr>
<td>Y^{24} O</td>
<td>A^{116} N</td>
<td>3.15 / 3.11</td>
<td>P^{28} Y</td>
<td>Y^{120}</td>
<td></td>
</tr>
<tr>
<td>L^{29} N</td>
<td>A^{118} O</td>
<td>2.81 / 2.81</td>
<td>P^{28} H</td>
<td>H^{159}</td>
<td></td>
</tr>
<tr>
<td>L^{29} O</td>
<td>Y^{120} N</td>
<td>2.89 / 2.90</td>
<td>L^{29} Y</td>
<td>Y^{106}</td>
<td></td>
</tr>
<tr>
<td>T^{30} O</td>
<td>Y^{106} Oη</td>
<td>3.16 / 3.15</td>
<td>L^{29} H</td>
<td>H^{117}</td>
<td></td>
</tr>
<tr>
<td>T^{30} Oη1</td>
<td>Y^{106} Oη</td>
<td>3.75 / 4.18</td>
<td>L^{29} F</td>
<td>F^{119}</td>
<td></td>
</tr>
<tr>
<td>G^{31} N</td>
<td>A^{124} O</td>
<td>2.93 / 2.98</td>
<td>T^{30} Y</td>
<td>Y^{120}</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Salt bridges</th>
<th></th>
<th>Metallorganic bonds</th>
</tr>
</thead>
<tbody>
<tr>
<td>PP</td>
<td>CD</td>
<td>Distance (Å) molecule A / B</td>
</tr>
<tr>
<td>R^{22} Nη2</td>
<td>E^{138} Oε2</td>
<td>2.85 / 2.85</td>
</tr>
<tr>
<td>R^{22} Nη1</td>
<td>E^{138} Oε1</td>
<td>3.08 / 3.01</td>
</tr>
<tr>
<td>PP</td>
<td>CD</td>
<td>Distance (Å) molecule A / B</td>
</tr>
<tr>
<td>D^{25} Oδ1</td>
<td>Zn^{999}</td>
<td>2.00 / 2.04</td>
</tr>
</tbody>
</table>
Fig. 3