A structure-derived snap-trap mechanism of a multispecific serpin from the dysbiotic human oral microbiome

Theodoros Goulas 1,4, Miroslaw Ksiazek 2,3,9, Irene Garcia-Ferrer 1,7, Alicja M. Sochaj-Gregorczyk 2,4, Irena Waligorska 2, Marcin Wasylewski 2, Jan Potempa 2,3 & F. Xavier Gomis-Rüth 1,5

1 Proteolysis Lab; Structural Biology Unit (“María de Maeztu” Unit of Excellence); Molecular Biology Institute of Barcelona, CSIC; Barcelona Science Park, Helix Building; c/ Baldirí Reixac, 15-21; 08028 Barcelona (Spain).
2 Department of Microbiology; Faculty of Biochemistry, Biophysics and Biotechnology; Jagiellonian University; ul. Gontosatowa 7; 30-387 Kraków (Poland).
3 Department of Oral Immunology and Infectious Diseases; University of Louisville School of Dentistry; Louisville, KY 40202 (USA).
4 Malopolska Centre of Biotechnology, Jagiellonian University; ul. Gontosatowa 7A; 30-387 Kraków (Poland).
5 Correspondence to F. Xavier Gomis-Rüth (xgcr@ibmb.csic.es).
6 These authors contributed equally and share first authorship.
7 Present address: EMBL Outstation Grenoble, 71 Avenue des Martyrs, CS 90181, 38042 Grenoble Cedex 9 (France).

Keywords: endopeptidase inhibition; host-microbiome interaction; Tannerella forsythia; red complex; miropin; X-ray crystal structure

ABSTRACT

Enduring host-microbiome relationships are based on adaptive strategies within a particular ecological niche. Tannerella forsythia is a dysbiotic member of the human oral microbiome that inhabits periodontal pockets and contributes to chronic periodontitis. To counteract endopeptidases from the host or microbial competitors, T. forsythia possesses a serpin-type proteinase inhibitor called miropin. While serpins from animals, plants, and viruses have been widely studied, those from prokaryotes have received only limited attention. Here we show that miropin uses the serpin-type suicidal mechanism. We found that similarly to a snap trap, the protein transits from a metastable native form to a relaxed triggered or induced form after cleavage of a reactive-site target bond in an exposed reactive-center loop. The prey peptidase becomes covalently attached to the inhibitor, is dragged 75Å apart, and is irreversibly inhibited. This coincides with a large conformational rearrangement of miropin, which inserts the segment upstream of the cleavage site as an extra β-strand in a central β-sheet. Standard serpins possess a single target bond and inhibit selected endopeptidases of particular specificity and class. In contrast, miropin uniquely blocked many serine and cysteine endopeptidases of disparate architecture and substrate specificity owing to several potential target bonds within the reactive-center loop and to plasticity in accommodating extra β-strands of variable length. Phylogenetic studies revealed a patchy distribution of bacterial serpins, incompatible with a vertical descent model. This finding suggests that miropin was acquired from the host through horizontal gene transfer, perhaps facilitated by T. forsythia’s long and intimate association with the human gingiva.

The human oral microbiome was first described in 1683 by Antoni van Leeuwenhoek (1). It is the second largest microbial community after the gut microbiome (2) and comprises over 600 species or phylotypes (3; http://www.homd.org). Under conditions of dysbiosis, it is dominated by opportunistic pathogens causing periodontal disease (PD), and likely contributing to the development of systemic diseases (4,5). PD is a chronic inflammatory disease driven by bacteria in the subgingival dental plaque, which occurs in 5–20% of the adult population worldwide (6). It is mainly caused by the "red complex" (7), a bacterial consortium including Tannerella forsythia, Porphyromonas gingivalis, and Treponema denticola, which exclusively reside in periodontal pockets (8). Analysis of contemporary and ancient oral microbiomes revealed that these pathogenic bacteria have inhabited our oral cavity over several thousand years (5). Indeed, paleomicrobiological studies showed that all three species were abundant in ancient dental calculus samples from medieval individuals with PD dated to ~800-1000 years ago (5). P. gingivalis and members of the Tannerella and Treponema genera were also identified in Polish mesolithic/paraneolithic, German neolithic and English Bronze-Age calculus samples dated to ~7,550-5,450, ~7,400-6,725 and ~4,200-3000 years ago, respectively (9). Moreover, P. gingivalis and T. denticola were detected in the mummy of the Tyrolean iceman "Ötzi" from ~5,300 years ago, who also suffered from PD (10). In addition, P. gingivalis was found in human dental calculus samples from Chile and Argentina dated to ~3,700-4,500 years ago (11). Finally, the three red-complex members were recently identified in ~49,000-year-old dental calculus samples from Homo neanderthalensis (12), which suffered from dental caries and PD like modern humans (13) and likely interbred with them across Eurasia (14). The persistence of these bacterial species within the human oral microbiome can only be explained by the development of adaptive
mechanisms to thrive in a harsh environment, which is characterized by competing microorganisms and host defenses.

*T. forsythia* is an anaerobic, Gram-negative bacterium from the Bacteroidetes phylum that is associated with grievous PD (15). It inhibits dental plaques in periodontal pockets (16) that contain an inflammatory exudate: gingival crevicular fluid. This fluid is rich in defensive cysteine endopeptidases (CEPs) and serine endopeptidases (SEPs) such as neutrophil elastase and cathepsin G, from the host immune system (17). It also contains extracellular peptidases engaged in virulence and colonization from *T. forsythia* (18) and other "red complex" partners, which compete for resources at the site of infection. Indeed, *P. gingivalis* secretes the CEPs calpain-like peptidase Tpr (19), periodontain (20), and gingipains K and R (21), as well as SEP PepK (22); and *T. denticola* contributes with CEP dentipain (23) and SEP dentilyisin (24). To keep these peptidases in check, *T. forsythia* possesses miropin, a serine proteinase inhibitor from the serpin family (25-27).

Serpins are comparatively large proteins of ~350-400 residues that are grouped into peptidase inhibitor family 14 in the MEROPS database (http://merops.sanger.ac.uk; (28)). They span over 3,000 members and also include non-inhibitory variants with other functions (27,29,30). Serpins have been extensively studied in humans and other mammals, where they participate in inflammation, coagulation, fibrinolysis, intracellular signaling, and complement activation (29,31). They are also widespread in other animals, plants, and some viruses (32). In contrast, they are only sporadically found in prokaryotes (33,34), where they have been mainly studied from environmental microbiota (34). Here, their precise biological roles remain obscure (27).

Generally, serpins target in a highly specific manner certain chymotrypsin- and/or subtilisin-like SEPs, as well as CEPs from the papain, cathepsin, and caspase families (31,32,35,36). These inhibitors are one of three covalent suicide inhibitor families (37), which also include the $\alpha_\text{I}$-macroglobulins ($\alpha_\text{I}$Ms; family I39; (38)) and the relatives of baculovirus p35 protein (family I50). Serpins behave like pseudo-substrates and inhibition is initiated by formation of the Michaelis complex, which occurs without significant structural changes in either enzyme or serpin (39). Subsequently, nucleophilic attack of the catalytic hydroxyl or sulphydryl group from the target SEP or CEP, respectively, on the scissile carbonyl of a specific reactive-site bond (RSB; bond $P_{1}$-$P_{1}'$; nomenclature of substrate sub-sites in the active-site cleft according to (40)) cleaves the bond and forms a covalent (thio)acyl-enzyme intermediate. The RSB occurs in an exposed, flexible 20-to-24-residue reactive-center loop (RCL; (41)) or serpin binding loop (42). At this stage, large conformational rearrangement of the serpin (29,32) causes the loose segment upstream of the cleavage site to insert as a new strand into a central $\beta$-sheet of the inhibitor moiety (sheet sA). This occurs under translocation of the covalently attached peptidase (27,32,43). The residues in $P_{1}$ and $P_{1}'$ are then far apart and thus unavailable for enzymatic resynthesis of the RSB, which may occur in some standard-mechanism protein inhibitors (see e.g. (44)), so the cleavage reaction is irreversible. Furthermore, the steric collisions caused by translocation cause deformation of the peptidase (41). The peptidase becomes unable to catalyze hydrolysis of the (thio)acyl-enzyme intermediate—which is chemically stable and resistant to SDS-PAGE—and strongly susceptible to proteolysis (41).

This inhibitory mechanism also entails that the serpin undergoes a "stressed-to-relaxed" transition (45,46) between an intact, high-energy, $"S$" or native conformation and a cleaved, triggered, low-energy, $"R$" or induced conformation, which is reminiscent of a snap trap and thermodynamically driven by the energy derived from strand insertion. The process results in an induced serpin that is substantially more stable than the native molecule (47). Some intact serpins have been further described in two additional conformers: the "latent form" and the "$\delta$ form," but these forms are rare (29,32).

Since the discovery of the first inhibitory serpin structures of induced $\alpha_\text{I}$-trypsin inhibitor alias $\alpha_\text{I}$-proteinase inhibitor ($\alpha_\text{I}$PI) (48) and native human antithrombin (49), a number of structural studies have contributed to shedding light on the working mechanism of animal, plant and viral serpins (see (29,31,32,37,46,50) for reviews). However, the only prokaryotic serpins that have been structurally investigated to date are *Thermobifida fusca* thermopin (51,52) and *Caldanaerobacter subterraneus* tengpin (53,54), which both originate in environmental thermophiles. To expand these data, we assessed the inhibitory capacity of miropin against physiologically relevant SEPs and CEPs and found it uniquely blocks many serine and cysteine endopeptidases of disparate architecture and substrate specificity owing to several potential target bonds within the RCL. We further investigated miropin's structure-based molecular mechanism, which is the first of a bacterial serpin from a human microbiome member of biomedical relevance, and found that broad specificity is due to plasticity in accommodating extra $\beta$-strands of variable length. To this aim, we analyzed its native and trypsin- plus subtilisin-induced wild-type forms, as well as a trypsin-induced mutant affecting the RCL and a mutant ablating a disulfide bond between vicinal residues. We further performed phylogenetic studies to hypothesize about the evolutionary origin of miropin and other bacterial serpins and suggest that it was acquired from the host through horizontal gene transfer.

**RESULTS AND DISCUSSION**

Miropin broadly inhibits SEPs and CEPs through various reactive-site bonds — *T. forsythia* miropin has a signal peptide (M$^{1.43}$; residue numbers in superscripts), which is cleaved off upon secretion. This leaves a
cysteine at the N-terminus (C), which suggests post-translational modification for insertion into the membrane through a "lipobox"-like mechanism (55) and location to the outer layer of the outer membrane (27). For functional and structural studies, we produced a soluble fragment spanning residues E-E (hereafter "wild-type miropin") and found it had higher activity (data not shown) than the previously described N-terminally extended variants (27). This suggests that the first ~20 residues of the secreted protein are dispensable for function and likely arranged as a flexible spacer from the bacterial membrane surface across the periplasm. The periplasmic location of miropin contrasts with that of most animal serpins, which are secreted and soluble (32). This paradox is reminiscent of αL families inhibitors, whose animal members are also generally secreted to the circulation (38) while bacterial members undergo post-translational modifications for periplasmic location similar to miropin (56).

Miropin inhibited physiologically relevant SEPs such as trypsin, neutrophil elastase, pancreatic elastase, subtilisin and cathepsin G, which have disparate substrate specificities and architectures, but not chymotrypsin and thrombin (see Figure 1A,B and (27)). Moreover, to assess whether the broad inhibitory spectrum of miropin could be extended to CEPs, we tested papain (Figure 1A,B) and two physiologically-relevant peptidases secreted by red-blood cell partner P. gingivalis at the site of infection, viz. calpain-like peptidase Tpr and ginpapain K (A.M. Sochaj-Gregorczyk, I. Waligorska, M. Wasylewska, M. Ksiazek and J. Potempa, unpublished results). When covalent enzyme-inhibitor complexes were incubated, trapped peptidases underwent processing over time owing to destabilization, as is usually observed in serpin complexes (41). Only fragments spanning up to ~3 kDa remained covalently attached through the catalytic residue to the C-terminal residue of the miropin cleaved bond. For inhibition, miropin employed not only the theoretic RSB, P-P (T-S; assigned based on (27)), but also upstream bonds P-P (K-T) and P-P (V-K), i.e. up to two positions upstream within the RCL. In stark contrast, the overwhelming majority of inhibitory serpins described are specific for one or few more endopeptidases of equivalent specificity and one class, and employ a single RSB as bait (57-60). In these cases, cleavage outside the RSB within the RCL results in inactivated serpins without endopeptidase inhibition, similarly to snap traps that are triggered but do not catch the prey (29,61,62).

To characterize inhibition in more detail and determine whether there was any preferred RSB in miropin, we produced two mutants in which the lysine in P, of the wild type was moved to P, (mutant K-K) or P, (mutant V-K), while the pl of the molecule was kept intact. We then compared their inhibitory capacity against trypsin with that of the wild type. Similar mutation studies that moved the target bond had revealed cleavage, but no peptidase inhibition, in other serpins (32). Given the mechanism of serpins as suicide substrate-like inhibitors ("branched-pathway mechanism," see Fig. 6 in (32)), their inhibitory activity is usually assessed by the "stoichiometry of inhibition" (SI). This is defined as the number of molecules of serpin needed to inhibit one molecule of proteinase and is ~1 for most serpins. A second parameter, which is more relevant for efficacy, is the second-order association rate constant, k, which is ~10^-10 M^-1 s^-1 for most inhibitory serpins (63). Wild-type miropin and mutant K-K inhibited trypsin with SI ~ 1.5, while V-K did so with SI ~ 5. The k_s values, in contrast, were very similar (~10^-5 M^-1 s^-1; Figure 2A-C), thus indicating that the RBSs are equivalent with respect to inhibitory power in miropin.

Accordingly, miropin is unique in its inhibitory capacity against endopeptidases of variable specificity, from two enzyme classes. It employs various bonds of the RCL, which is reminiscent of the "bait region" from the likewise suicidal, but otherwise structurally and mechanistically unrelated αMs (38,64).

**Structure of native miropin** — To shed light on the mechanistic basis of this broad specificity, we determined crystal structures of native and induced miropin. The structure of native wild-type miropin was solved using diffraction data to 3.0 Å-resolution (see Table 1). Miropin is an elongated ellipsoid with two similar equatorial semi-axes of ~35Å and a long vertical semi-axis of ~80Å (Figure 3A; molecule orientation of reference). The structure contains a central sandwich consisting of two perpendicular, partially overlapping, twisted β-sheets, sA and sB (consensus nomenclature of serpins, see (32)), of five vertical strands (from left to right, s6A, s5A, s3A, s2A, s1A) and seven horizontal strands (top to bottom, s0B-s6B; see Figure 3A,D) respectively. Sheet sB is antiparallel and includes an uppermost extra strand, s0B, that is absent from other serpins, while sA is mixed parallel-antiparallel. A third four-stranded antiparallel β-sheet (sC; left to right, s1C-s4C) is placed on top of sA, forming the roof of the molecule and making a second β-sandwich with the left half of sB. In sC, strands s1C, s2C and the C-terminal segment of s3C are antiparallel. The latter strand is bent leftwards at half-length, so that its N-terminal half does not interact with neighboring strand s2C. Instead, it forms a β-ribbon with strand s4C, which protrudes from the front surface of the molecule. Overall, the three β-sheets constitute the central core of the protein, which is decorated with nine helices (hA-hI). All are located behind the sheets, except for frontal helix hF (Figure 3A).

N-terminal helix hA runs horizontally along the back surface of miropin right to left, approximately at half height of sA, and leads to the lowermost β-strand of sB, s6B (Figure 3A,D). This strand leads to hB, which nestles in the back surface of sheet sA and roughly parallels its strands. Helices hA-hE and hI are arranged as a helical cluster behind sA and shape the back surface of the lower half of miropin. Among them, hB, hC, and hD are
contiguous and linked by short loops, and they connect s6B with s2A. Helix hE, in turn, is inserted between the latter strand and s1A and helix hI is placed in an extended loop connecting strands s5A and s6A. In the upper part of the molecule, behind sB, two more tandem helices (hG and hH) shape the left back surface of the molecule and connect s3B with s0B. Finally, on the front surface of sA, helix hF and the downstream extended loop leading to s3A (loop-hF-s3A) cover sA like a "front flap."

Interestingly, a highly strained disulfide bond links the side chains of vicinal residues C245 and C246 (Figure 4A). To investigate the role of this unique feature for serpins, we constructed a mutant, in which both residues were replaced with alanine (mutant C245A/C246A). We found that this mutant had inhibitory properties undistinguishable from wild-type miropin (data not shown). Next, as disulfide bridges generally contribute to protein stability, we further compared mutant and wild-type miropin by differential scanning calorimetry. These studies revealed a single peak for both species in the analyzed range, but the temperature at which the excess capacity was maximal was 4.7°C lower in the mutant (56.4°C vs. 61.1°C; Figure 4B). This reveals that the intra-molecular disulfide bond plays a role in the stability, but not in the activity, of miropin.

Native miropin exhibits an exposed RCL, which connects s5A with s1C and protrudes upwards from the molecular body (Figure 3A,D,F). It spans 24 residues (E353-P376, corresponding to positions P17-P24) and does not interact with the protein moiety between A360 and T375. This suggests that it is likely to be flexible and disordered in solution, so it can easily adapt to the active-site clefts of disparate prey peptidases. In the crystal structure, the RCL is defined in the final Fourier map (Figure 3F) owing to non-crystallographic-symmetry contacts. It contains a short 1.5-turn helix (hRCL; T362-V367; P2-P7) shortly before the theoretic RSB (Figure 3A), which is reminiscent of the native structures of non-inhibitory serpin ovalbumin from chicken (65) and of a variant of human α1-antichymotrypsin (66). The RCL is subdivided into a "hinge region" (27) spanning E353-V361 (P17-P9; Figure 3A), which is essential for the conformational rearrangement of the RCL upon induction (see the next section), and an "exposed loop" region from P9 to P14 (T362-P376; Figure 3F), which contains the theoretic RSB (P2-P7). Another region of functional importance is the "breach": the point of initial strand insertion at the top of sA (see the next section). Following the consensus serpin architecture (67), the breach includes residues F201, K202, G203, W205 and F209 from s3A and loop-s3A-s4C; M202 from s3C; Y253 from s2B; and E253 plus G255 from the RCL hinge region. Also relevant is the "shutter", engaged in sheet opening and other changes prior to strand insertion (see the next section), which in miropin encompasses F17 from hA; S67, P68, S70 and L75 from loop-s6B-hB and hB; L94 from hC; I168, N169, C172, T176, D178 and I180 from the front flap; L195 and N197 from s3A; N345 from s5A; A358 from the hinge region; and L398, F399 and G401 from s5B and the C-terminal tail. Collectively, the residues of shutter and breach contribute to a set of 51 positions, which are required for core structure integrity and are conserved among >70% of inhibitory serpins (see Table 2 in (67)). In miropin, 45 positions are strictly conserved (Figure 5A) and the exceptions are conservative replacements that do not interfere with the general core architecture: C172 instead of valine, D178 instead of glycine, A201 instead of threonine, R203 instead of lysine, F38 instead of leucine, and V406 instead of proline. Moreover, inhibitory serpins also show a consensus sequence pattern in the hinge region, which is required for efficient and rapid strand insertion upon induction (see also the next section): E-(E/K/R)-(G)-(T/S)-X-(A/G/S)4 (42,68). Mutation of these residues in inhibitory serpins often resulted in inhibitor cleavage but not peptidase inhibition (67,69). Inspection of the miropin structure reveals that it matches the consensus sequence (E353-E-G-T-E-A-A-A-V361). Accordingly, the structure of native miropin fulfills all the structural requirements described for functional inhibitory serpins and contains a long and flexible RCL, which is potentially targetable by endopeptidases.

Structures of induced miropin and reaction mechanism — We obtained covalent complexes of induced wild-type miropin with two SEPs of disparate specificity and structure, viz. trypsin (to 1.6 Å resolution) and subtilisin (to 1.7 Å resolution) (see Table 1), which cleaved the RCL at bonds P2-P3 and P14-P15, respectively. These sites are consistent with the respective endopeptidase specificities (70,71). We also solved the structure of trypsin-induced mutant V360/K361A (to 1.5 Å resolution), which was cleaved at P1-P2. The high resolutions of these structures contrast with the low resolution of the native structure, which is consistent with the difference between the relaxed, low-energy induced conformation and the stressed, metastable native conformation normally found in serpins. As mentioned above, the enzyme:inhibitor complexes underwent processing, so that the crystallized samples contained miropin covalently linked only to small fragments of trypsin and subtilisin via the Oγ atoms of the respective catalytic serines. These fragments could not be resolved in the crystal structures, with the exception of dipeptide D194-S195 (trypsin numbering in subscripts) in the complex between wild-type miropin and trypsin (Figure 3B,C,E,G).

Superposition of the three induced miropin structures shows they are practically indistinguishable (rmsd values of ~0.5 Å). As expected from an inhibitory serpin, induction-cleavage causes the downstream segment of the RCL to loosely protrude from the top molecular surface and be flexible, following a similar chain trace to native miropin only from P370 onwards (Figure 3A,B). In contrast, the upstream segment of the RCL undergoes large rearrangement, which drags the covalently bound peptidase fragment from one pole of the serpin to the opposite (Figure 3A,B,C). This results from a ~180°-
rotation of bond Cα-C of P16 residue E354 (Figure 3A,B), next to "fulcurn" (29) residue E353 in P17, which causes the downstream polypeptide chain to be rotated ~90° downwards and inserted between sA strands s5A and s3A as new strand s4A (G355-K368). In the three induced structures, strand insertion is equivalent until V360 (K360 in mutant V360K/K360A), which is the last residue buried in the molecular moeity and visible in the Fourier map of the mutant complex. The trypsin complex of wild-type miropin includes extra residue K368, which is linked to the catalytic serine of trypsin (Figure 3G), and the subtilisin complex even further includes T369 as the last defined residue. Thus, miropin uniquely selects segments spanning between 14 and 16 residues, apparently without destabilization of the complex, which provides a structural explanation for its flexible inhibitory capacity. This is exceptional, because while the length on the free C-terminal side of the reactive site bond generally varies between five and nine residues (P1-P5/P5'; 32) among serpins, the segments for s4A strand insertion strictly span 16 or 17 residues (P1-P16/P17; 32)). This has been hailed as critical for enzyme:inhibitor complex stabilization upon cleavage and rearrangement (29). Serpin variants that were two residues longer or shorter were cleaved, but no inhibitory complexes were formed, as the peptidase was still able to hydrolyze the acyl-enzyme intermediate (72).

Strand insertion transforms sA from a mixed five-stranded β-sheet to a more stable antiparallel six-stranded β-sheet (Figure 3B,C,E,G). This movement causes the cleaved residue upstream of the scissile bond—and thus bound pre-Y peptidase fragments—to be pulled downwards by ~75 Å. The insertion is accounted for by a horizontal sliding motion, a rigid-body left-shift of strands s6A and s5A of up to ~2 Å and a right-shift of strands s3A, s2A, and s1A of maximally ~3.5 Å (Figure 3C). This "stage curtain opening" mechanism was supported by analysis of the theoretical molecular flexibility of native and trypsin-induced wild-type miropin based on the elastic network model, which identified three hinge points centered at positions Y200-F201, T304-C306 and T317-F318 (scores of 0.85-0.92; see (73)). These are central residues of s3A, s6A, and s5A, which highlight the importance of the horizontal section at half-height of sA in conformational rearrangement upon induction. This rearrangement cascades down to elements from the shutter region: the C-terminal segment of the molecule moves ~1.5 Å to the left, and this causes C β-ribbon s3C-s4C to be rotated backwards ~7°. Moreover, loop-hA-s6B from the shutter is shifted leftwards by ~2 Å. On the right hand side of sA, displacement of s2A causes slight rearrangement of the C-terminal turn of hD, which leads loop-s4B-s5B to be shifted backwards ~2 Å. However, the most relevant of all the changes affects the front flap (Figure 3C). This structural element may play a role in stabilizing the five-strand conformation in native serpins, thus preventing untimely triggering of the rearrangement, and in controlling opening of the sheet (32,74). The front flap has to be pulled away to allow for strand insertion and then positioned back to protect the refurbished β-sheet sA in what is known as a "coupling mechanism" (see Fig. 10 in (32)). In miropin, displacement of s1A causes downstream helix hF and loop-hF-s3A from the flap to be rearranged under maximal displacement of ~3.5 Å (at N177), in particular due to a χ1-rotation of the side chain of Y200. In contrast to all these changes, the helical cluster behind sA and the rest of the molecule remain essentially unchanged upon induction.

The induced structures also explain the aforementioned conservation of the hinge-region in miropin: generally small residues are required to prevent clashes after strand insertion with side chains behind sheet sA, which are unaltered upon induction. Indeed, a glycine (G355) is the only residue allowed in P14 to avoid steric hindrance with W205 and Y353, maximally a threonine (T356) is allowed in P14 owing to the aforementioned aromats plus M260, V349, and V551, and small or middle-sized side chains are also required at the positions of downstream s4A residues A358, A360, T362 and V364. Notably, a methionine (M366) is found at P14, whose bulky hydrophobic side chain contributes to a hydrophobic cluster with F192, M193 and L195 from neighboring strand s3A; upstream s4A residue V364; I338 and I400 from the segment preceding s5A; T78 from hB; and F330 plus I333 from loop-hl-s5A. Taken together, these methionine-mediated interactions provide a strong anchor for s4A to the subjacent moeity in miropin, as found in plasminogen activator inhibitor 1 and peptidase inhibitor 6 among human serpins, which have an equivalent methionine (32).

To sum up, structures of induced miropin provide the basis for its potentially targetability by endopeptidases of different architecture and specificity.

Structural comparison and evolutionary considerations — Structural similarity searches with miropin identified an entire cohort of very similar serpins (Z-scores > 40; (75)). Interestingly, the closest relative was antithrombin, followed by other mammalian serpins such as human intracellular serpin SCA1 (34.67), rather than tengpin (Z-scores ~45) or thermopin (Z-scores 36-41), which are the only bacterial serpins structurally analyzed and did not appear within the 50 top hits. Antithrombin had the closest similarity to both native miropin (antithrombin PDB code 3KCG, (76); Z-score = 48.2, rmsd = 2.1 Å, length of alignment = 370 residues, 419 residues in total, 27% sequence identity) and induced miropin (antithrombin PDB code 1ATT, (77); 50.2, 1.4 Å, 356, 420, 29%).

In turn, sequence identity searches revealed close matches within bacteria (see Figure 5A for a selection). Miroprin is the only serpin of T. forsythia, and several potential orthologs were found in other Tannerella species that displayed pairwise sequence identities of 57-76%. In contrast, red-complex partners P. gingivalis and T. denticola lacked serpins despite sharing the habitat...
with *T. forsythia*, possibly because they themselves secrete several SEPs and CEPs that would be cancelled out by endogenous serpins (78). However, other *Porphyromonas* strains, such as human intestinal *Porphyromonas asaccharolytica* and canine gingival *Porphyromonas gingivalis* and *Porphyromonas crevioricanis*, encoded potential serpins sharing 26-49% sequence identity but they clustered separately from miropin (Figure 5B). Clustering with miropin were sequences from *Bacteroides* species residing in the human gastrointestinal tract, with identities spanning 33-50%. However, as within the genus *Porphyromonas*, other *Bacteroides* species such as *Bacteroides fragilis* and *Bacteroides thetaiotaomicron* lacked serpins.

Serpin sequences were also found in other human symbionts such as *Bifidobacterium* and *Prevotella* (27,33,79) and in human pathogenic *Mycobacterium ulcerans* but not in *Mycobacterium tuberculosis*. Their presence was also detected in bacteria of distinct environmental origin, such as the plant symbiont *Rhizobium leguminosarum*, the free-living bacteria *Streptomyces albus* and *Bacillus subtilis*, and the cyanobacteria *Anabaena variabilis* and *Arthrospira platensis*.

The high structural similarity and sequence identity of miropin to the human intracellular serpin SCCA1 (37%) is not an isolated case. Other bacterial serpin sequences shared even higher identities with eukaryotic homologs, such as *Chondromyces croatus*, a myxobacterium isolated from dead herbarium specimens and animal faeces (80). This potential serpin shares 46% identity with the potential homolog from American sparrow *Zonotrichia albicollis*, and it clusters together with the eukaryotic sequences (Figure 5B).

To sum up, the presence of serpins and serpin-like sequences in bacteria is widespread and includes free-living, pathogenically invasive, symbiotic colonizing and saprophytic species, which indicates that bacterial serpins can be housekeeping proteins but may also participate in colonization or virulence by providing protection against attacking peptidases of human or bacterial origin (27,29). However, taken together, the occurrence of bacterial serpin sequences is patchy.

**Concluding remarks** – Our resident microbes play a pivotal role in health and disease and they can be envisaged as an additional organ contributing to our human condition (81). Unfortunately, we know remarkably little about their diversity, variation and evolution, so unravelling the molecular mechanisms they have derived to adapt to us and microbial competitors is key to our understanding of both healthy symbiosis and pathogenic dysbiosis.

We hereby describe for the first time the mechanism of action of the bacterial serpin that enables *T. forsythia* to better thrive in the harsh and crowded human oral cavity. Miropin is capable of broadly inhibiting SEPs and CEPs of disparate class, architecture and substrate specificity. This is achieved by offering several target bonds of the RCL for cleavage within a bait region, instead of a single RSB as found in canonical serpins. In addition, promiscuous inhibitory capacity is facilitated by the capacity to insert strands deviating from the canonical length into the central sheet SA, while keeping the prey peptidase bound and inactivated. Despite the apparent flexibility of RCLs, so far only three serpins were shown to inhibit target proteases using two separate—though overlapping—sites (57-60). Hence, the structural adaptation of miropin to provide a relaxed inhibitory specificity, which allows for formation of inhibitory complexes using different sites, is unique among serpins. It can be hypothesized that this adaptation evolved in response to the highly proteolytic environment of the dysbiotic bacterial biofilm, which is crowded with microbial species secreting a broad array of SEPs and CEPs and saturated with host phagocyte-derived proteases. For *T. forsythia*, a bacterium whose integrity depends on the semi-crystalline surface layer of proteins (S-layer) (82), protection from proteolytic degradation is a matter of life or death. To this end, outer-membrane anchored miropin, with its ability to inhibit a variety of both host and bacterial proteases, seems to be a perfect solution. It would be interesting to see if other miropin-like serpins from highly proteolytic environments have the same propensity. A long RCL, which in many cases evinces one or two additional residues, is found in putative serpins from gut- and dental-plaque dwelling *Bacteroides* sp. and *Prevotella* sp., but not in environmental specimens, which seems to argue that these serpins may also possess a relaxed inhibitory specificity. Also in this context, it is also interesting that *Tannerella* species from the oral microbiome carry a cluster of 4 genes encoding serpins that are very closely related to miropin and comprise a RCL of variable length. Analysis on the inhibitory properties of further miropin-like serpins from species inhabiting different environments will verify our hypothesis.

Because of the wide distribution of serpins in animals and plants (for example, 36 paralogs are found in humans (36)), they should also be broadly and homogeneously present among ancestral organisms if they had an ancient origin (34). However, proteinase inhibitors are generally rare within prokaryotes (33), and this also holds for serpins: their distribution is patchy in bacteria, where they have only been identified recently (34). This contradicts the vertical descent model and would only be partially explained by massive loss-of-function events during evolution, so bacterial serpins may not share a common ancestor. Instead, xenologous horizontal gene transfer from our interfacial epithelial cells to the bacterial cells of our microbiomes, partially supported by phylogenetic data, may explain the possible origin and evolution of at least a fraction of bacterial serpins, including miropin (27,34,67,83). The intimate host-microbiome interaction and persistence, which can be traced back several millennia for red-complex partners, would support such transfer. It is further backed by the
proposal that an exchange of DNA in plaque biofilms by a transformation-like process may provide an important ecological advantage for the survival and persistence of oral microbiome bacteria (84). Finally, this hypothesis is in line with those postulated for the origin of other human microbiome effectors, such as bacterial α-Ms (56) and bacterial metalloproteases such as T. forsythia karilysin (85) and Bacteroides fragilis fragilysin (86).

**MATERIALS AND METHODS**

**Protein production and purification** — A fragment of T. forsythia strain ATCC43037 miropin (GenBank code WP_041590947; UniProt code G8UQY8) spanning residues E<sup>59</sup>-E<sup>308</sup> (mutation R<sup>174</sup>Q attributed to natural variability within T. forsythia strains, hereafter "wild-type miropin") was produced from a construct derived from plasmid pGEX-6P-1_Tfs46 (27), which attaches an N-terminal glutathione-S-transferase (GST) tag and a PreScission protease cleavage site to the protein of interest, by employing the Phusion Site-Directed Mutagenesis Kit (Thermo Fisher Scientific) and phosphorylated forward primer GAAAGATAGAAAA AGACAAATGCCTTTGCTC and reverse primer GATCCGGGGGCCCCCTGGAC. The resulting plasmid was transformed into Escherichia coli Rosetta (DE3) cells, which were grown in Luria Bertani medium supplemented with ampicillin (100 μg/ml) and chloramphenicol (33 μg/ml) at 37°C to an OD<sub>600</sub> of 0.75−1 and then incubated for 30 min at 4°C. Recombinant protein expression was induced with 0.1 mM isopropyl-β-D-1-thio-galactopronanoside. After 6 h at 20°C, cells were harvested by centrifugation (15 min, 6,000 × g, 4°C), resuspended in PBS plus 0.02% sodium azide (15 ml per pellet from 1 l of culture), and subsequently lyzed by sonication (cycle of 30 × 0.5 s pulses at 70% amplitude per pellet from 1 l of culture) using a Branson Digital 450 Sonifier (Branson Ultrasonics). Cell lysates were clarified by centrifugation (40 min, 40,000 × g, 4°C) and loaded onto a glutathione-Sepharose 4 Fast Flow column (GE Healthcare Life Sciences; bed volume 10 ml), previously equilibrated with PBS plus 0.02% sodium azide at 4°C. Tag-free recombinant miropin was obtained by in-column cleavage of the GST moiety with PreScission protease (GE Healthcare Life Sciences), which left five residues (G<sup>4</sup>-P<sup>2</sup>-L<sup>3</sup>-G<sup>2</sup>-S<sup>1</sup>) attached to the N-terminus of the protein. Protein-containing fractions were pooled, concentrated and further purified by size-exclusion chromatography in a HiLoad 16/600 Superdex 75 pg column (GE Healthcare) previously equilibrated with 5 mM Tris-HCl, 50 mM sodium chloride, 0.02% sodium azide, pH 8.0 using an AKTA Purifier 900 FPLC system (GE Healthcare) at a flow rate of 1.5 ml/min. Protein identity and purity were assessed by 15% Tricine SDS-PAGE stained with Coomassie blue, peptide-mass fingerprinting of tryptic protein digests, N-terminal sequencing through Edman degradation, and mass spectrometry. The latter three approaches were carried out at the Protein Chemistry Service and the Proteomics Facilities of the Centro de Investigaciones Biológicas (Madrid, Spain). Ultrafiltration steps were performed with Vivaspin 15 and Vivaspin 500 filter devices of 5-kDa cutoff (Sartorius Stedim Biotech). Protein concentrations were estimated with the help of the respective theoretical extinction coefficients by measuring A<sub>280</sub> in a spectrophotometer (NanoDrop). When required, concentrations were also determined more precisely with the BCA Protein Assay Kit (Thermo Scientific) with bovine serum albumin as a standard.

Miropin double mutants V<sup>367</sup>deletions K<sup>368</sup>A, K<sup>368</sup>A/T<sup>369</sup>K, and C<sup>245</sup>A/C<sup>246</sup>A were obtained by site-directed mutagenesis of the plasmid encoding wild-type miropin employing the QuikChange II Site-Directed Mutagenesis Kit (Agilent Technologies) according to the manufacturer’s instructions. To this aim, the following forward (F) and reverse (R) primers were used: K<sup>368</sup>A-F, CCGTAGAAATGTGTAAGGTACCATCCCCCTC; K<sup>368</sup>A-R, GAGGGGATAGCTGCTACCATTTCTCA; V<sup>367</sup>K-F, GAAACACGCCCCAATAGGAAGAAGAAGCA CTGCTACCCCCCTC; V<sup>367</sup>K-R, GAGGGGATAGCAGGCTTCTATT TC; K<sup>368</sup>−<sup>369</sup>K-F, GAAATGGCTACCATTTCTCA; K<sup>368</sup>A/R, AGAGGGATAGCTGCTACCATTTCTC; C<sup>245</sup>A/C<sup>246</sup>A-F, CGCGTTATAGCCAGAGGGGCGCGGCAATACCT GAGATGGAC; and C<sup>245</sup>A/C<sup>246</sup>A-R, GTCCATCTCA AGATTATGGCGGCTC GTGGTCTGATGACGC. Protein production and purification proceeded as described with the wild type.

**Peptidase cleavage assays** — Native wild-type miropin and double mutants V<sup>367</sup>K/K<sup>368</sup>A and K<sup>368</sup>A/T<sup>369</sup>K were used to assay inhibition of SEPs and CEPs through incubation for different time spans (30 s to 1 h) at room temperature. The SEPs that were tested included trypsin from bovine pancreas, elastase from porcine pancreas (both from Sigma-Aldrich), subtilisin Carlsberg from B. subtilis (CalBiochem) and human neutrophil elastase (Elastin Products Company, Inc.). The CEPs that were tested included papain from papaya latex (Sigma-Aldrich). Reactions were carried out in 50 mM Tris-HCl, 150 mM sodium chloride, pH 7.5 at peptidase:miropin ratios between 1:1 and 1:2. Reactions were stopped through inhibition with 4 mM Pefabloc SC (Roche Life Sciences) or precipitation with 2,2,2-trichloroacetic acid (Sigma-Aldrich). Reaction products were visualized directly on (SDS-)PAGE or after purification by size-exclusion chromatography.

**Differential scanning calorimetry** — The thermostability of wild-type miropin and mutant C<sup>245</sup>A/C<sup>246</sup>A was studied using NANO DSC III apparatus (model 6300) with capillary cells. Samples in 5 mM Tris-HCl, 50 mM sodium chloride, 0.02% sodium azide, pH 8.0 were degassed for 10 minutes, and centrifuged at 20,000 × g for 10 minutes. Buffer was used for controls. Measurements were performed within the temperature range 25-75°C, with a scanning rate of 1°C/min. The
thermograms of the protein solutions and the buffer were used to calculate excess molar heat capacity (MHC) curves using NanoAnalyze software (TA Instruments).

**Stoichiometry of inhibition** — The number of molecules of miropin needed to inhibit one molecule of target protease (stoichiometry of inhibition, SI) was determined as previously reported (27) by incubating constant amounts of active-site titrated bovine trypsin (50 nM) in 100 mM Tris-HCl, 150 mM sodium chloride, 5 mM calcium chloride, 0.02% Tween-20, pH 7.6 with increasing concentrations of miropin (wild type, V\(^{567}K/K^{308}A\) or K\(^{368}A/T^{509}K\)) to yield molar ratios of enzyme:inhibitor ranging from 0 to 5. After 15 min of incubation at 37°C, an equal volume of a solution of chromogenic trypsin substrate N\(_2\)-benzoyl-L-Arg-4-nitroanilide hydrochloride (3 mM) was added and enzymatic hydrolysis of the substrate was monitored for 30 min at 37°C at \(\lambda = 410\) nm in a SpectraMAX microplate reader. Residual activity was plotted as a function of the molar ratio of miropin:proteinase. The SI was considered to be the value at which the fitted line intersected the x-axis.

**Determination of the association rate constant** — Kinetic parameters of inhibition of trypsin by miropins were determined by the progress curve method (87). Briefly, mixtures containing constant concentrations of substrate (Boc-Gln-Ala-Arg-MCA; at 40 \(\mu\)M), and increasing concentrations of miropin (wild type, V\(^{567}K/K^{308}A\) or K\(^{368}A/T^{509}K\)) in a total volume of 100 \(\mu\)l were prepared in microtiter plates. Next, 100 \(\mu\)l of trypsin (0.2 nM) was added and the rate of substrate hydrolysis was recorded (\(\lambda_{exc} = 360\) nm, \(\lambda_{em}= 460\) nm) employing a SpectraMax Gmini XS microplate reader (Molecular Devices). The pseudo-first-order association rate constant, \(k_{on}\), was determined as previously described (27).

**Crystallization and diffraction data collection** — For crystallization, native wild-type or V\(^{567}K/K^{308}A\) mutant miropin was reacted with either trypsin or subtilisin, and the residual activity was inhibited by adding Pefabloc\textsuperscript{®} SC or phenylmethylsulfonyl fluoride (Roche). Proteins were subsequently buffer-exchanged to 20 mM Tris-HCl, 2 mM sodium chloride, pH 7.5 and further purified by ion-exchange chromatography in a TSKgel DEAE-2SW column (TOSOH Bioscience) equilibrated with the same buffer. A gradient of 4–60% of 20 mM Tris-HCl, 500 mM sodium chloride, pH 7.5 was applied over 20 ml and samples were collected and pooled. Finally, each pool was concentrated by ultrafiltration and subjected to size-exclusion chromatography in a Superdex 75, 10/300 column (GE Healthcare) equilibrated with 20 mM Tris-HCl, 150mM sodium chloride, pH 7.5. Under these conditions, only small peptidase fragments of 370-3,580 Da—as measured by MALDI-TOF analysis—remained covalently attached through the catalytic serine to induced miropin forms due to autoproteolysis. These specimens were crystallized. In parallel, a certain fraction of intact enzyme:inhibitor complexes with trypsin or subtilisin could be isolated by rapid inhibition of the SEPs and subsequent size-exclusion chromatography, but these samples did not crystallize.

Crystallization assays were performed by the sitting-drop vapor diffusion method using 96 × 2-well MRC plates (Innovadyne) at the joint IBMB/IRB Automated Crystallography Platform at Barcelona Science Park. To this end, reservoir solutions were prepared by a Tecan robot and 100 nL-crystallization drops were dispensed by a Phoenix nanodrop robot (Art Robbins) or a Cartesian Microsys 4000 XL robot (Genomic Solutions). Plates were stored in Bruker steady-temperature crystal farms at 4°C or 20°C. Successful conditions were scaled up to the microfilter range in 24-well Cryschem crystallization dishes (Hampton Research). The best crystals of native wild-type miropin were obtained at 20°C from 1 \(\mu\)l : 1 \(\mu\)l drops of protein solution (at 6-15 mg/ml concentration in 20 mM Tris-HCl pH 7.4, 100 mM sodium chloride) and 2.4 M disodium malonate, pH 7.0 as reservoir solution. In turn, all induced miropin variants were crystallized similarly but the reservoir solution contained 200 mM sodium iodide, 100 mM Bis-Tris, 20% [w/v] polyethylene glycol 3,550, pH 6.5 instead. Carefully washed and dissolved crystals were analyzed by N-terminal Edman degradation and mass spectrometry, which revealed that native crystals contained intact miropin spanning residues E\(^{39}\)–E\(^{368}\). In contrast, induced miropin variants were cleaved in the N-terminal segment (after T\(^{41}\) or K\(^{60}\)) and within the RCL (trypsin/wild type, after K\(^{368}\); subtilisin/wild type, after T\(^{369}\); and trypsin/mutant V\(^{567}K/K^{308}A\), after K\(^{463}\)).

Crystals were cryo-protected by rapid passage through drops containing increasing concentrations of either 2.4-3.5 M disodium malonate, pH 7.0 (native crystals) or the crystallization buffer plus glycerol up to 15% [v/v] (induced crystals). Complete diffraction datasets were collected at 100 K from liquid-N\(_2\) flash cryo-cooled crystals (Oxford Cryosystems 700 series cryostream) on a Pilatus 6M pixel detector (from Dectris) at beam line XALOC (88) of the ALBA synchrotron (Barcelona, Spain). Further data were collected on the same detector type at beam line ID30A of the ESRF synchrotron (Grenoble, France) within the Block Allocation Group “BAG Barcelona.” Diffraction data were integrated, scaled, merged and reduced with programs XDS and XSCALA (89). Native miropin crystals belonged to space group P4\(_1\)2\(_1\)2\(_1\), contained one dimer per asymmetric unit, and diffracted to maximally 3.0 Å. Trypsin-induced and subtilisin-induced wild-type miropin crystals, as well as trypsin-induced V\(^{567}K/K^{308}A\) mutant miropin crystals, belonged to space group P2\(_1\)2\(_1\)2\(_1\), contained a monomer per asymmetric unit, and diffracted to maximally 1.6 Å, 1.7 Å and 1.5 Å resolution, respectively. Table I provides a summary of data collection and processing.

**Structure solution and refinement** — An initial sequence similarity search identified horse leukocyte elastase inhibitor as the closest relative of miropin among the proteinase-activated serpin structures reported (PDB...
code 1HLE; (90)). Its coordinates were used to solve the structure of trypsin-induced wild-type miropin by likelihood-scored molecular replacement with the PHASER program (91). A clear solution was found at 293.6, 116.5, 332.1 (α, β, γ in Eulerian angles) and 0.293, 0.960, 0.209 (x, y, z, as fractional unit-cell coordinates) after rigid-body refinement. This solution gave an initial Z-score of 12.4 for the rotation function and 19.6 for the translation function, as well as a final log-likelihood gain of 508. A subsequent density improvement step with ARP/wARP (92), which employed refinement program REFMAC5 (93), yielded a partial model and a Fourier map that enabled straightforward manual model building with the COOT program (94). The latter alternated with crystallographic refinement with PHENIX (95) and BUSTER/TNT (96) under inclusion of TLS refinement, until the final refined model was obtained. The latter comprised residues D<sup>34</sup>-K<sup>368</sup> and S<sup>373</sup>-E<sup>408</sup> from miropin and D<sup>34</sup>S<sub>194</sub> from trypsin (residue numbers in subscripts) covalently linked between atoms S<sub>194</sub> Oγ and K<sup>368</sup> C. Two tentative glycerol molecules and 409 solvent molecules completed the structure.

Among the reported structures of native serpins, human squamous cell carcinoma antigen 1 (SCCA1, PDB code 2ZZV; (97)) displayed the closest sequence similarity with miropin and was employed—with all side chains mutated to alanine—to solve the native structure of the latter, for which a dataset to 3.3 Å resolution was initially available (data not shown). Two solutions were found with PHASER at 15.1, 84.3, 247.1, 0.660, 0.337, 0.219 and 251.4, 87.5, 79.3, 0.841, 0.213, 0.834 after rigid-body refinement. These solutions yielded initial Z-scores of 4.7/15.1 and 3.3/11.9 for the respective rotation/translation functions, as well as a final log-likelihood gain of 422. This calculation was followed by a density modification and model extension step with the AUTOBUILD protocol of PHENIX (98), which produced an improved Fourier map and a partial model. Model building and refinement proceeded as for trypsin-induced wild-type miropin. At the final stages of model completion, an isomorphous dataset to higher resolution (3.0 Å) became available, which was used to complete the structure (see Table 1). The final model of native miropin contained residues E<sup>39</sup>-E<sup>408</sup> plus residues P<sup>4</sup>L<sup>-1</sup>G<sup>-1</sup>S<sup>-1</sup> from the N-terminal tag for each of the two molecules (A and B) present in the asymmetric unit. A further 16 solvent molecules completed the structure. Molecule A was significantly more rigid and better defined than molecule B (overall thermal displacement parameter of 66.0 Å<sup>2</sup> vs. 90.8 Å<sup>2</sup>) and superposition of the two molecules revealed essentially identical chain traces (rmsd value of 0.9 Å), so molecule A was used for the presentation of the results and discussion.

The structure of subtilisin-induced wild-type miropin was solved by molecular replacement with the protein coordinates of trypsin-induced miropin after omitting segments E<sup>353</sup>-K<sup>368</sup> and S<sup>373</sup>-I<sup>381</sup> and the trypsin dipeptide. These calculations yielded a solution at 255.0, 174.0, 77.1, -0.472, 0.546, -0.055 after rigid-body refinement, which had initial Z-scores of 4.1 and 6.1 for the rotation and translation functions, respectively, as well as a final log-likelihood gain of 9,201. Subsequent density modification, model building and refinement proceeded as with trypsin-induced miropin. The final model comprised miropin residues K<sup>368</sup>-V<sup>369</sup> and the following (partially occupied) tentatively-assigned ligands: seven iodide ions, one potassium ion, four glycerol molecules, and 402 solvent molecules.

Finally, the structure of trypsin-induced miropin mutant V<sup>360</sup>K/K<sup>368</sup>A was solved as that of subtilisin-induced wild-type miropin. The molecular replacement calculations yielded a top peak at 192.3, 178.3, 12.2, -0.494, 0.497, -0.001 after rigid-body refinement, which had initial Z-scores of 4.1 and 6.6 for the rotation and translation functions, respectively, as well as a final log-likelihood gain of 12,722. Subsequent density modification, model building and refinement proceeded as with trypsin-induced miropin. The final model comprised miropin mutant residues E<sup>39</sup>-E<sup>408</sup> plus residues G<sup>-1</sup>S<sup>-1</sup> from the fusion construct, as well as the following (partially occupied) tentatively-assigned ligands: two zinc ions, four iodide ions, one potassium ion, six chloride ions, two glycerol molecules, one tris(hydroxymethyl)aminomethane molecule, and 448 solvent molecules.

**Phylogenetic Analysis** – The miropin sequence was used for BLAST searches within the NCBI database (National Centre for Biotechnology Information) (99), which were initially performed within phyl Fibrobacteres, Chlorobi and Bacteroidetes—the FCB group of bacteria—and subsequently expanded to other bacterial families and eukaryotes. Only selected sequences were used and further analyzed due to the very high number of sequences retrieved. These were aligned with MAFFT using the G-INS-i algorithm (100). A phylogenetic tree was constructed with PHYML (101) using LG as substitution model, with 100 bootstrapping replicates, 4 substitution rates, optimization in topology/length/rate and topology search best of NNI and SPR search within the Genious platform.

**Miscellaneous** – Structural similarity searches were performed with DALI (75), and structure figures were prepared with the CHIMERA program (102). Structures were validated with MOLPROBITY (103). Inter-domain flexibility was ascertained with HINGEPROT using standard settings ((73); http://bioinfo3d.cs.tau.ac.il/HingeProt). The final coordinates of native, trypsin-induced, and subtilisin-induced wild-type miropin, as well as of trypsin-induced miropin mutant V<sup>360</sup>K/K<sup>368</sup>A, were deposited with the PDB at www.pdb.org (respective access codes 5NCS, 5NCT, 5NCU and 5NCW).
FUNDING INFORMATION

This study was funded in part by grants from European, US, Polish, Spanish, and Catalan agencies (UMO-2015/17/B/NZ1/00666; “Mobility Plus” I306/MOB/IV/2015/0; NIH/NIDCR R21DE026280; 2012/04/A/NZ1/00051; 2016/21/B/NZ1/00292; FP7-HEALTH-2012-306029-2 “TRIGGER”; BU2015-64487-R; MNiSW 2975/7.PR/13/2014/2; BIO2013-49320-EXP; MDM-2014-0435 and 2014SGR9). TG acknowledges a “Juan de la Cierva” research contract (ICI-2012-13573) from the Spanish Ministry for Economy and Competitiveness. Funding for data collection was provided in part by ESRF. The Faculty of Biochemistry, Biophysics and Biotechnology of the Jagiellonian University is a partner of the Leading National Research Center (KNOW) supported by the Ministry of Science and Higher Education. The funders had no role in the study design, data collection and interpretation, or the decision to submit the work for publication. In addition, the authors declare no competing financial interests.

ACKNOWLEDGMENTS

We are grateful to Joan Pous, Laia Vives and Xandra Kreplin from the joint IBMB/IRB Automated Crystallography Platform for assistance during crystallization experiments. We further acknowledge the help provided by local contacts at the ESRF and ALBA synchrotrons.

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest with the contents of this article.

AUTHOR CONTRIBUTIONS

T.G., F.X.G.-R., M.K., and J.P. designed the research. T.G., M.K., A.M.S-G., I.W., M.W. and F.X.G.-R. performed the experiments and/or data analysis. F.X.G.-R. wrote the paper with input from all authors.

REFERENCES

1. Leevvenhoeck, A. (1683) An abstract of a letter from Mr. Anthony Leevvenhoeck at Delft, dated Sep. 17. 1683, containing some microscopical observations, about animals in the scurf of the teeth, the substance call’d worms in the nose, the cuticula consisting of scales. Phil. Trans. 14, 568-574


human proteinase inhibitor 8.


Endopeptidase inhibition mechanism of Tannereilla forsythia miropin


Endopeptidase inhibition mechanism of Tanneraella forsythia miropin


FOOTNOTES

This study was funded in part by grants from European, US, Polish, Spanish, and Catalan agencies (UMO-2015/17/B/ZN1/00666; “Mobility Plus” 1306/MOB/IV/2015/0; NIH/NIDCR R21DE026280; 2012/04/A/ZN1/00051; 2016/21/B/ZN1/00292; FP7-HEALTH-2012-306029-2 “TRIGGER”; BFU2015-64487-R; MNiSW 2975/7.PR13/2014/2; BIO2013-49320-EXP; MDM-2014-0435 and 2014SGR9). TG acknowledges a “Juan de la Cierva” research contract (JCI-2012-13573) from the Spanish Ministry for Economy and Competitiveness. The Structural Biology Unit of IBMB is a “María de Maeztu” Unit of Excellence of the Spanish Ministry of Economy, Innovation and Competitiveness. Funding for data collection was provided in part by ESRF. The Faculty of Biochemistry, Biophysics and Biotechnology of the Jagiellonian University is a partner of the Leading National Research Center (KNOW) supported by the Ministry of Science and Higher Education. The funders had no role in the study design, data collection and interpretation, or the decision to submit the work for publication. In addition, the authors declare no competing financial interests.

FIGURE LEGENDS

Figure 1. Miropin complexes with serine or cysteine peptidases. (A) SDS-PAGE analysis of complexes. Wild-type or V367K/K368A-mutant miropin were incubated for various time points with peptidases in an equimolar ratio or with twofold molar excess of the inhibitor. Incubation periods varied between 30 seconds and one hour. Complexes
with subtilisin, papain and trypsin with the mutant were further purified by gel filtration. *Stars*, high molecular mass complexes; black arrows, miropin in complex with peptidase-derived peptides after autolysis; open arrows, unbound peptidases; dotted frames, miropin fragments from the respective RSB to the C-terminus. (B) Peptidase cleavage sites within miropin at the N-terminus (i) and within the RCL of the wild type (ii) and double-point mutants V^{367}_A/K^{368}_A and K^{368}_A/T^{369}_K (iv). Mutated residues are shown in red. Vertical arrows pinpoint the cleavage sites of the distinct peptidases. Legend: *Try*, trypsin; *Sub*, subtilisin; *Pap*, papain; *HNE*, human neutrophil elastase; and *Etp*, porcine pancreatic elastase.

**Figure 2.** Effect of reactive-site residue scrambling on miropin inhibition. Inhibition of trypsin was assayed by (A) wild-type inhibitor, (B) double mutant V^{367}_A/K^{368}_A, and (C) double mutant K^{368}_A/T^{369}_K. Inhibition was characterized by determining the stoichiometry of inhibition (SI) ([upper panels]) and the association rate constant (k_{on}) ([lower panels]). The data shown are means and standard deviation of the means from three replicates.

**Figure 3.** *Structures of native and induced miropin.* (A) Native miropin in the reference "front view" (according to (32)) as a Richardson-type plot, with strands as arrows and helices as purple spirals (labeled hA-hH plus hRCL). The strands are arranged in three β-sheets (sheet sA, yellow strands s6A, s5A, s3A-s1A; sheet sB, orange strands s0B-s6B; and sC, red strands s1C-s4C). The RCL (E^{355-370}) connects strand s5A with s1C and is subdivided into the "hinge region" (E^{355-356}; brown ribbon) and the "exposed loop" (T^{362-376}; blue ribbon). The residues flanking the theoretical RSB (P_{10}(E^{355}) and P_{17}(E^{354}) residues is indicated by lines. (B) Same as (A) but showing trypsin-induced wild-type miropin as representative of the induced miropin structures. β-Strand s4A from sheet sA, absent in (A), is shown in the colors of the corresponding segment of (A) and labeled. Dipeptide D_{194-195} of trypsin is covalently attached through atom S_{194}Oγ to the carbonyl of K^{368} after cleavage of bond P_{2}-P_{1} (K^{368}-T^{369}). On the primed side, the chain is only defined from S^{373}(P_{2}) onwards. (C) Superposition of native and subtilisin-induced wild-type miropin as Cα-traces in cross-eye stereo depicting sheet sA and the RCL (respectively, in orchid and turquoise), and the flap consisting of helix hF and downstream loop hF-s3A (pink and blue). Upon induction, rotation around the curved orange arrow leads to strand insertion, which entails a left shift for strands s6A and s5A and a right shift for s3A, s2A and s1A (small orange arrows). The visible ends of the cleaved region of induced miropin (T^{369}; P_{1} and P^{373}; P_{2}) are pinpointed by black arrows. Topology scheme of native (D) and induced (E) miropin in the coloring of (A) and (B). Relevant positions for the mechanism are depicted in the notation of Schecter and Berger (40) (P_{10}-P_{17}; miropin residues D^{352-368}; see also Fig. 7 in (27)). The "exposed loop" of the RCL (blue coil) spans P_{2}-P_{1} (T^{369}-S^{376}), includes helix hRCL, and contains the theoretical RSB (P_{1}-P_{2}). The "hinge region" of the RCL (brown coil) spans P_{2}-P_{6} (E^{355}-V^{361}). Upon productive cleavage at the RCL and induction (red scissors, (D)), P_{2}-P_{2} (G^{355-358}; trypsin cleavage at P_{2}-P_{1} or P_{6}-P_{1} (G^{355-370}; subtilisin cleavage at P_{1}-P_{1}) becomes inserted into sheet sA as s4A (in brown/blue) between s3A and s5A. The residues of each regular secondary structural element are indicated in *italics*, those differing in native and induced miropin are in red. The covalently attached endopeptidase is symbolized by a green and yellow ellipse. (F) Fragment of refined native miropin depicting the five strands of sheet A (from left to right, s6A, s5A, s3A, s2A and s1A) and strand s1C in magenta and the RCL in orange superposed with the final refined (2mF_{o-b}-DF_{o-b})-type Fourier map (turquoise mesh) shown with a zone radius of 2 Å and a contour level of 1 σ. Residues E^{355} ("fulcrum") and E^{354} (Cα-C rotation leads to insertion) are labeled, and the rotation occurring in the latter upon induction is pinpointed by a magenta arrow. (G) Same as (F) but showing trypsin-induced miropin around sheet A only (from left to right, strands s6A-s1A) as magenta sticks, except for strand s4A and the preceding "hinge region," in orange as in (F). Trypsin dipeptide D_{194}-S_{195} is shown as blue sticks at the sheet bottom. (H) Same as (G) but showing the structure of subtilisin-induced miropin. (I) Same as (G) but depicting the structure of trypsin-induced miropin V^{367}_A/K^{368}_A.

**Figure 4.** *Effect of disulfide bond C^{245-246} on protein stability.* (A) Detail in cross-eye stereo of the structure of trypsin-induced miropin centered on the disulfide bond C^{245-246} superimposed with the final (2mF_{o-b}-DF_{o-b})-type Fourier map. (B) Wild-type (WT) and C^{245A/C^{246}A-mutant miropin were subjected to differential scanning calorimetry. The thermograms of temperature vs. molar heat capacity (MHC) are representative of triplicate experiments.

**Figure 5.** *Phylogenetic studies.* (A) Multiple alignment of fragments of selected serpin sequences including *Tannarella forsythia* miropin (GenBank code WP_041590947). The amino-acid numbering and the secondary structure elements—yellow arrows for β-strands and green rods for α-helices—correspond to miropin. Red squares identify the 51 conserved residues in serpins (67). Sequence stretches with low similarity are replaced by three black dots over a salmon background. Gram-negative bacteria (in parenthesis, the UniProt codes, except for those preceded by GB, GenBank code): *Bacteroides ovatus* (A7LRL73), *Bacteroides uniformis* (R9HW7), *Bacteroides vulgatus* (18Z335), *Arthospira platensis* (D4ZTF4), *Anabaena variabilis* (Q3M416), *Porphyronomas cangiingivalis* (A0A0A2VE67), *Porphyromonas crevioricinis* (A0A0A2FPU3), *Porphyromonas uenosin* (GB WP_025883851), *Rhizobium leguminosarum* (19N4F0), *Tannerella* sp. oral taxon_1 (W2C1Q2), *Tannerella* sp. oral taxon_2 (W2CNM8) and *Chondromyces crocatus* (A0A0K1ECH9); Gram-positive bacteria: *Legionella feelei* (A0A0W0TMA7), *Mycobacterium ulcerans* (X8FN27), *Streptomyces albus* (M95ZH9), *Bifidobacterium bifidum* (E1AWB1), *Bacillus subtilis* (A0A0K6L0C3) and *Clostridium butyricum* (A0A0S3L311); eukaryotes: *Drosophila*
erecta (B3NAL5), Homo sapiens_1 (α,PI; P01009), Homo sapiens_2 (SCCA-1; P29508), Homo sapiens_3 (neuroserpin; Q99574), Rattus norvegicus (D3ZJK2) and Zonotrichia albicollis (GB XP_005481738). (B) Circular phylogenetic tree reflecting evolutionary distances among the sequences shown in (A). Potential bacterial proteins from the Fibrobacteres, Chlorobi and Bacteroides (FCB) group are depicted with a blue and yellow background. Other bacterial and eukaryotic sequences are shown over grey and red backgrounds, respectively. The bar represents 0.5 substitutions per site.
### Table 1. Crystallographic data.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>native miropin</th>
<th>trypsin-induced miropin</th>
<th>subtilisin-induced miropin</th>
<th>trypsin-induced miropin V^{367}K/K^{368}A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>P4,2,2</td>
<td>P2,2,2</td>
<td>P2,2,2</td>
<td>P2,2,2</td>
</tr>
<tr>
<td>Cell constants (a, b, c, in Å)</td>
<td>78.5, 78.5, 351.7</td>
<td>62.53, 73.75, 84.45</td>
<td>64.00, 70.04, 90.86</td>
<td>62.60, 74.28, 84.38</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.9677</td>
<td>0.9795</td>
<td>0.9793</td>
<td>0.9789</td>
</tr>
<tr>
<td>No. of measurements / unique reflections</td>
<td>497,174 / 23.150</td>
<td>589,220 / 51.750</td>
<td>547,809 / 45.392</td>
<td>811,419 / 63.557</td>
</tr>
<tr>
<td>Resolution range (Å) (outermost shell)</td>
<td>47.0 – 3.00 (3.18 – 3.00)</td>
<td>47.7 – 1.60 (1.70 – 1.60)</td>
<td>55.5 – 1.70 (1.80 – 1.70)</td>
<td>84.4 – 1.50 (1.59 – 1.50)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.9 (99.9)</td>
<td>98.9 (93.0)</td>
<td>99.5 (96.7)</td>
<td>99.8 (99.1)</td>
</tr>
<tr>
<td>R&lt;sub&gt;merge&lt;/sub&gt;</td>
<td>0.172 (1.295)</td>
<td>0.036 (0.169)</td>
<td>0.051 (0.479)</td>
<td>0.074 (0.843)</td>
</tr>
<tr>
<td>R&lt;sub&gt;i.m.&lt;/sub&gt; [= R&lt;sub&gt;meas&lt;/sub&gt; / CC(1/2)]</td>
<td>0.176 (1.325) / 0.999 (0.895)</td>
<td>0.037 (0.182) / 1.000 (0.984);</td>
<td>0.054 (0.513) / 1.000 (0.931);</td>
<td>0.077 (0.886) / 1.000 (0.842);</td>
</tr>
<tr>
<td>Average intensity</td>
<td>20.5 (3.3)</td>
<td>42.5 (9.4)</td>
<td>29.9 (4.4)</td>
<td>21.8 (2.8)</td>
</tr>
<tr>
<td>B-Factor (Wilson) / Aver. multiplicity</td>
<td>58.7 / 21.5 (22.4)</td>
<td>26.1 / 11.4 (7.0)</td>
<td>29.2 / 12.1 (7.6)</td>
<td>24.0 / 12.8 (10.7)</td>
</tr>
<tr>
<td>Resolution range used for refinement (Å)</td>
<td>47.0 – 3.00</td>
<td>23.4 – 1.60</td>
<td>24.9 – 1.70</td>
<td>55.8 – 1.50</td>
</tr>
<tr>
<td>No. of reflections used (test set)</td>
<td>22,400 (735)</td>
<td>50,954 (752)</td>
<td>44,619 (756)</td>
<td>62,749 (807)</td>
</tr>
<tr>
<td>Crystallographic R&lt;sub&gt;factor&lt;/sub&gt; (free R&lt;sub&gt;factor&lt;/sub&gt;)</td>
<td>0.176 (0.216)</td>
<td>0.145 (0.160)</td>
<td>0.162 (0.184)</td>
<td>0.163 (0.183)</td>
</tr>
<tr>
<td>No. of protein atoms / solvent molecules / ligands</td>
<td>5,886 / 16 / –</td>
<td>2,960 / 409 / 2 glycerols</td>
<td>4 glycerols, 7 iodides, 1 potassium</td>
<td>2 glycerols, 1 tris, 4 iodides, 2 zincs, 6 chlorides, 1 potassium</td>
</tr>
<tr>
<td>R&lt;sub&gt;rmsd&lt;/sub&gt; from target values</td>
<td>0.009 / 1.19</td>
<td>0.010 / 1.05</td>
<td>0.010 / 1.03</td>
<td>0.010 / 1.03</td>
</tr>
<tr>
<td>bonds (Å) / angles (°)</td>
<td>78.3</td>
<td>23.2</td>
<td>29.1</td>
<td>22.5</td>
</tr>
<tr>
<td>All-atom contacts and geometry analysis</td>
<td>723 (97.2%) / 0 / 744</td>
<td>353 (98.9%) / 0 / 357</td>
<td>353 (98.6%) / 0 / 358</td>
<td>357 (98.3%) / 0 / 363</td>
</tr>
<tr>
<td>with poor rotamers / bad bonds / bad angles</td>
<td>31 (4.8%) / 0 / 1 (0.13%)</td>
<td>2 (0.6%) / 0 / 0</td>
<td>3 (1.0%) / 0 / 0</td>
<td>0 (0.13%) / 0 / 0</td>
</tr>
<tr>
<td>with Cβ deviations&gt;0.25Å / clashscore</td>
<td>1 / 10.55 (979th percentile)</td>
<td>0 / 2.45 (999th percentile)</td>
<td>0 / 2.61 (999th percentile)</td>
<td>0 / 3.09 (989th percentile)</td>
</tr>
<tr>
<td>MolProbity score</td>
<td>2.21 (98th percentile)</td>
<td>1.03 (100th percentile)</td>
<td>1.05 (100th percentile)</td>
<td>1.18 (98th percentile)</td>
</tr>
</tbody>
</table>

---

* Data processing values in parenthesis refer to the outermost resolution shell.
* For definitions, see Table 1 in (104).
* For definitions, see (105,106).
* Average intensity is <I/o(I)> of unique reflections after merging according to the XDS program (89).
* According to Engh and Huber (107).
* According to MOLPROBITY (103,108).
Fig. 1