SUPPLEMENTARY INFORMATION

Table of Contents

---

1. SUPPLEMENTARY RESULTS

***Detailed structural analysis of* Tannerella forsythia *mirolase* —** The KLIKK-peptidase mirolase is 791 amino acids in length, comprising an 18-residue signal peptide for secretion to the periplasm, a 164-residue N-terminal pro-domain involved in latency maintenance, a 348-residue catalytic domain (CD), and a 261-residue C-terminal extension (UniProt code (UP) A0A0A7KVG3). The latter features an 85-residue C-terminal domain (CTD) for export from the periplasm via a type-IX secretion system (T9SS) and a 176-residue pre-CTD fragment [1]. Based on its CD sequence, mirolase is assigned to the S8 family (subtilases) according to the MEROPS peptidase database (see https://www.ebi.ac.uk/merops; [1, 2]). The family prototype, subtilisin from *Bacillus subtilis*, is 381 amino acids in length in its pre-pro-form (UP P04189) and was the first bacterial serine endopeptidase (SP) to be characterised, 70 years ago [3]. Subtilisin Novo from *Bacillus amyloliquefaciens*, also termed BPN´ (UP P00782; 382 residues), and the biotechnologically relevant subtilisin Carlsberg from *Bacillus licheniformis* (UP P00780; 379 residues) are closely related. All have broad substrate specificity and a general preference for aromatic or aliphatic side chains in $\mathbf{P_1}$ (substrate and active-site cleft sub-site nomenclature, in bold, according to [4-6]). They are highly efficient and degrade protein substrates at turnover rates $>1 \times 10^6$ M^{-1} s^{-1} [7]. As previously described for subtilisin Novo [8] and other secreted bacterial peptidases [9-14], mirolase self-activated *in vitro* to yield the mature CD, which remained non-covalently associated with, and inhibited by, its pro-domain until the latter was fully degraded. The enzyme required calcium for self-processing, activity and stability, and displayed optimal activity at pH 9.5. It degraded fibrinogen, haemoglobin and the antimicrobial peptide LL-37, among other physiologically relevant human proteins [1].

Our analysis of the crystal structure revealed that the mirolase CD belongs to the parallel α/β category [15]. It comprises the central α/β/α-sandwich typical of subtilases, which was first described for subtilisin Novo

(Protein Data Bank code (PDB) 1SBT, [16]) and later for subtilisin Carlsberg (PDB 1SBC [17]). The mirolase CD contains 74 residues more than subtilisin Carlsberg's (25 % of the total) and pivots on a central eight-stranded β-sheet, twisted clockwise by ~90º (see Suppl. Fig. S9A,B). All strands are parallel except edge strand β9, which is antiparallel and interacts with the C-terminal half of preceding strand β8 while the N-terminal half of β8 interacts with downstream strand β10 in a parallel manner. This results in the overall topology β2↑-β3↑-β1↑-β6↑-β7↑-β8↑-β9↓+β10↑. The sheet is strongly arched between β3 and β10 but not significantly curled. This creates a concave surface, which accommodates six helices (α1–α3 and α6–α8), and a convex face, to which two more parallel helices are attached (α4 and α5). The axes of the latter helices, as well as those of α2, α3, α6 and α8, are roughly antiparallel to the central strands of the sheet.

Mirolase splits into an upper and a lower moiety, which span residues Q_{183}–S_{356} and S_{361}–P_{535}, respectively, and are strongly interdigitated. Central sheet strand β6 (I_{357}–N_{360}) acts as a hinge between the moieties, and the interface between them gives rise to the active-site cleft, which runs left (non-primed side) to right (primed side) on the front surface of the molecule from the perspective of Suppl. Fig. S9B. As in other α/β/α-sandwich enzymes, the active-site cleft is at the C-terminal end of the central sheet and is framed by loops: Lβ2α3 and Lβ3α4 delimit the top of the cleft; Lβ6α5 and Lβ7β8 shape the cleft bottom; and extended loop Lβ10α6, which includes β-ribbon β11β12 and calcium site 5 (see discussion of calcium sites below), delineates the cleft on its right side. The left side of the cleft is open. In particular, Lβ3α4 projects ~30 Å from the front surface above the cleft like a unicorn's horn and features the "protruding hairpin" β4β5 (W_{318}–Y_{335}), which is twisted clockwise by ~180º (Suppl. Fig. S9B). This hairpin is supported by "backing loop 1" (A_{250}–N_{260}) and "backing loop 2" (P_{268}–S_{279}; contains calcium site 1) provided by Lβ2α3. Overall, these three structural elements comprise 16, 4, and 15 residues more than the equivalent segments of subtilisin Carlsberg, respectively, and they create a unique structure for subtilases (Suppl. Fig. S9B, inset). Further remarkable deviations from subtilisin Carlsberg include an 11-residue insertion between α1 and α2, which accommodates calcium site 2; a five-residue insertion after α5; a four-residue insertion after β10; a seven-residue insertion preceding β12, which includes the aforementioned calcium site 5; and three extra residues within Lα7α8, which surround calcium site 4.

To gain more insight into the architecture of the cleft, which harbours the typical catalytic triad of subtilase SPs (D_{231}, H_{283} and S_{477}; Suppl. Fig. S9A,B), we obtained a model mimicking a bound substrate by superposing mirolase onto the complex of subtilisin Novo with its pro-domain, which is taken as a substrate complex for subtilisins (PDB 3CO0 [18]; Suppl. Fig. S9B,D). In these SPs, the principal contacts between substrate and enzyme in the cleft are made upstream of the scissile bond by segment $\mathbf{P_1}$–$\mathbf{P_4}$, which is pinched between an extended "upper-rim segment" above the cleft, equivalent to G_{204}–Y_{208} of subtilisin Carlsberg, and an extended "lower-rim segment" below the cleft, equivalent to L_{230}–G_{232}. Taken together, this yields a characteristic antiparallel three-stranded β-sheet [5, 18, 19]. In mirolase, a substrate would likewise bind in extended conformation, as generally found in the active-site clefts of peptidases [20]. However, although the enzyme possesses a lower-rim segment (S_{361}–Y_{364}; Suppl. Fig. S9D), which anchors the substrate model (except for the tyrosine side chain in $\mathbf{P_4}$ that clashes with Y_{364}), it unexpectedly lacks the upper rim, which is spatially replaced by the protruding hairpin and backing loop 2. This imposes a major difference in the way substrates are bound by mirolase and canonical subtilases.

As to the distinct sub-sites of the cleft, the $\mathbf{S_1}$ specificity pocket would easily accommodate middle-sized hydrophobic or polar — but probably not charged — amino acids in mirolase, which is in accordance with the generally broad substrate specificity of subtilisins. This sub-site would be framed by the backbone of the lower rim and the beginning of Lβ7β8, and contributing side chains would come from A_{396}, G_{398}, A_{410}, Q_{435} and T_{476}. In addition, cleft sub-sites $\mathbf{S_1}'$ and $\mathbf{S_2}'$ would be shaped by Y_{464}, F_{473}, and N_{474}; $\mathbf{S_2}$ by D_{232} and F_{315}; $\mathbf{S_3}$ by N_{277} and S_{279}; and $\mathbf{S_4}$ by I_{314}, I_{342}, W_{362}, Y_{364}, and I_{374}. The oxyanion hole characteristic of serine endopeptidases would be formed by S_{477}N and T_{476}N, as well as the side chain of N_{399}, which would stabilise the scissile carbonyl through its amide group like N_{155} in subtilisin Carlsberg. The N_{399} side chain is further fixed by a tight hydrogen bond between its Oδ1 atom and T_{476}Oγ1 at the cleft bottom (Suppl. Fig. S9D).

Another unique structural feature of mirolase is the presence of six calcium sites distributed around the structure (Suppl. Fig. S9B and Suppl. Table S6). In contrast, subtilisins, kexin, and furin have at most two such cations. Accordingly, mirolase has four unique sites for subtilases, shaped by segments within Lα1α2, Lβ2α3, Lβ11β12 and Lα7α8, respectively (Suppl. Fig. S9C). These calcium sites constitute an additional regulatory element for mirolase further to zymogenicity to prevent spatially and temporally undesirable activity because the enzyme is inactive in the absence of calcium [1]. This prevents the self-processing and activation of mirolase, which folds in the periplasm as other cargoes of the T9SS. In addition, the calcium sites also provide rigidity to and assist with the correct folding of mirolase [1].

2. SUPPLEMENTARY FIGURES

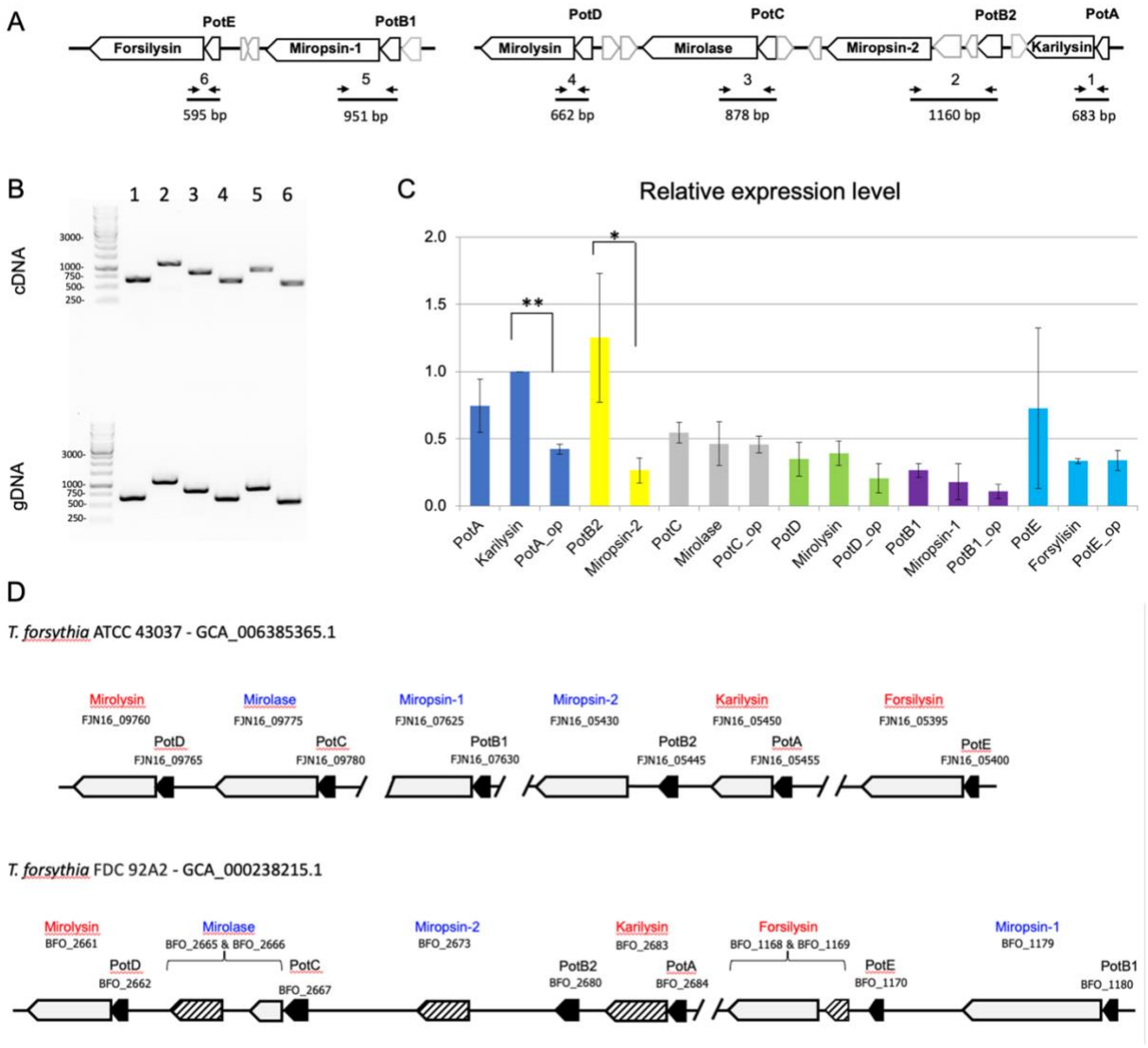
A

```
PotA/karilysin:      ...gaaaaacATGAAACAGCAAA...AATCAATtaatATGAAACGAT...AAAAAGtgatgcc...
PotB1/miropsin-1:    ...gaaaaacATGAAACAGAAAA...TACTCGAAACCATGAAAAAGA...AAAAAAtgagaaa...
PotB2/miropsin-2:    ...gaaaaacATGAAGACAAAAA...AGTCAtga...ATGAAAAAGA...AAAAAGtgagaag...
PotC/mirolase:       ...gaaaaacATGAAACAGAAAA...ATAAATCCATCATGAAAAAGA...AAAAAAtgaaaaa...
PotD/mirolysin:      ...gaaaaacATGAAACAGAAAA...AAGAAAtgatcATGAAAAAAA...AAGAAGtgagaaa...
PotE/forsilysin:     ...gaaaaacATGAAACAGCAAA...AAAtaaaaatcATGAAAAAAC...AAGAAGtgaaaaa...
```

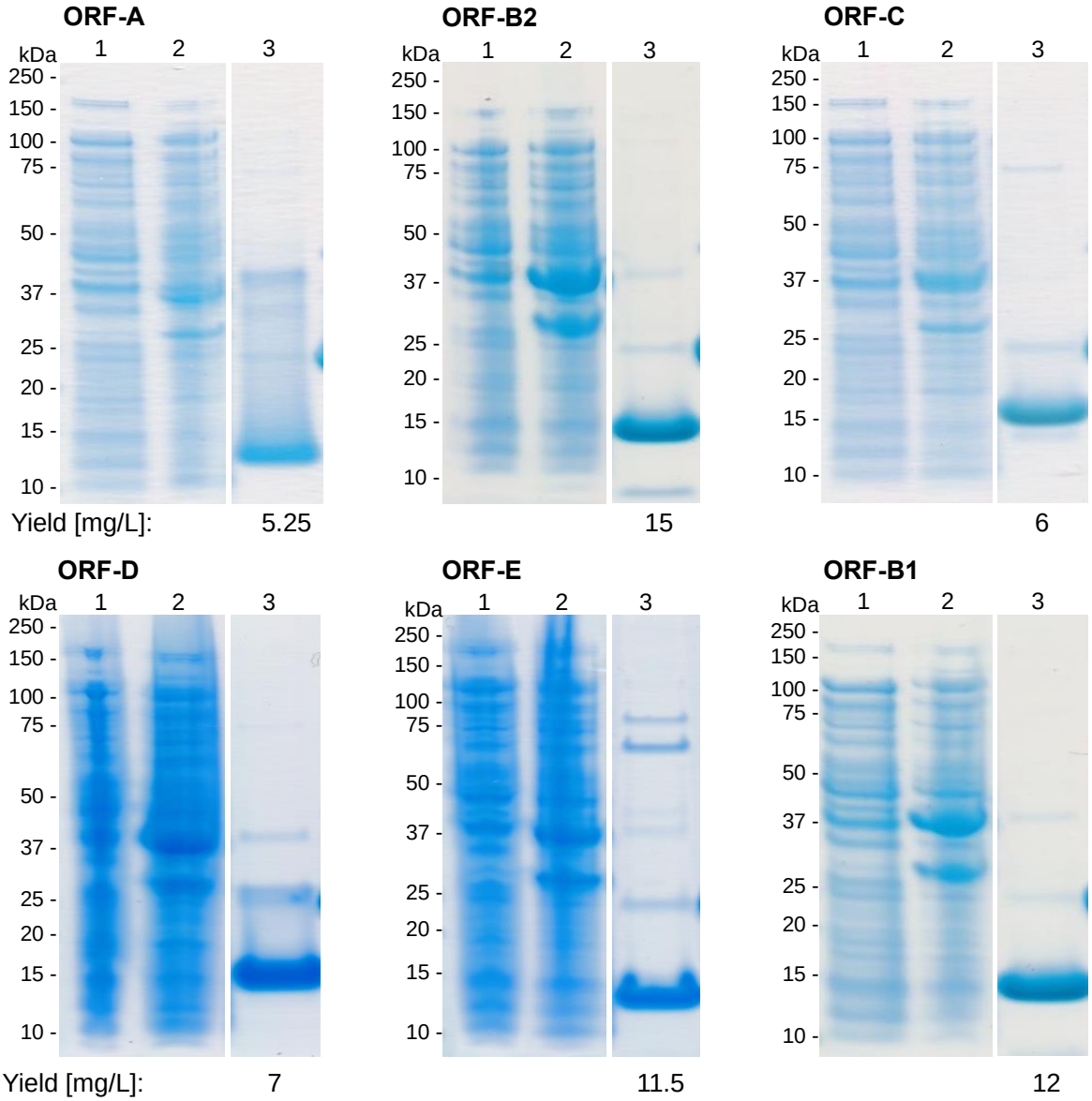
B

```
                          1
PotA  mkqqiilwigvllliggvgCKKDQSSCCDKE--------IIKDVSELTGIISYNT----   28
PotB1 mkqkiilwigalllltagtgCEKNKDYEVPTQ--------LTGTRWELAGIVDAKTG---   29
PotB2 mktkiilwisslllltagagCEKNKDYEVPTQ--------LTGTRWDLAGIVDTKTG---   29
PotC  mkqkiilwistlllltagagCKKETLPPNQAKGKVLGPTGPCQGYALYIEVENPKGIGLE   40
PotD  mkqkiilwistlllltagagCEKESHASCSCE--------CVEEKIPIVTLKNENA----   28
PotE  mkqqiilwigvllliggvgCENGQLHSPPAN--------PEQAILGKWELINSGGR---   29

PotA  -EVKR-------WYISVSDANSYDN-------VTLYFP--CNLDSKYMKEK--EKVIFSG   69
PotB1 -KITPLAPKGCYGFKFISETEAKGGTVLNQMTVHLTTPPFIGIVTMIGDEENGDAALFY-   87
PotB2 -KITPLAPKGSYWFKFISETKARGGSVLNIMAIDLTSPPFFNIETEIGDSHNGDAALFY-   87
PotC  GKGIPAGSGRTWNYRNAISVPLFNRIG--LPVELMEEGTWLHFEYREMTEEEKNRKLFQP   99
PotD  --HFRYIKRRNDFALEIENKELVRG--------LYLIPRGCDIPKKYKEDG--LPVIISG   76
PotE  -PIIP------TGYREFLPSGIVHK-------YDYTKEQYTSFQCEYSILN--DTVLLMC   73

PotA  QISKSTLKIT-LPAGTTSYCIN---LMSINKIN--    98
PotB1 RIIKTLESYT-WEKNELKFFYDNKQYYLLYKYSKP  121
PotB2 RIIRKLESYT-WEKNELKFFYDNKQYYLLYKYSKS  121
PotC  DEPVICLMNQ-IPPPANTYMITKIIAHKPLKINPS  132
PotD  EVFDCSEYIKPWIKRDPVYFIK---LSTIKKK---  105
PotE  NYRYKYLFYRDKMQLFPLDLIAIRDLTEIYQRKK-  107
```

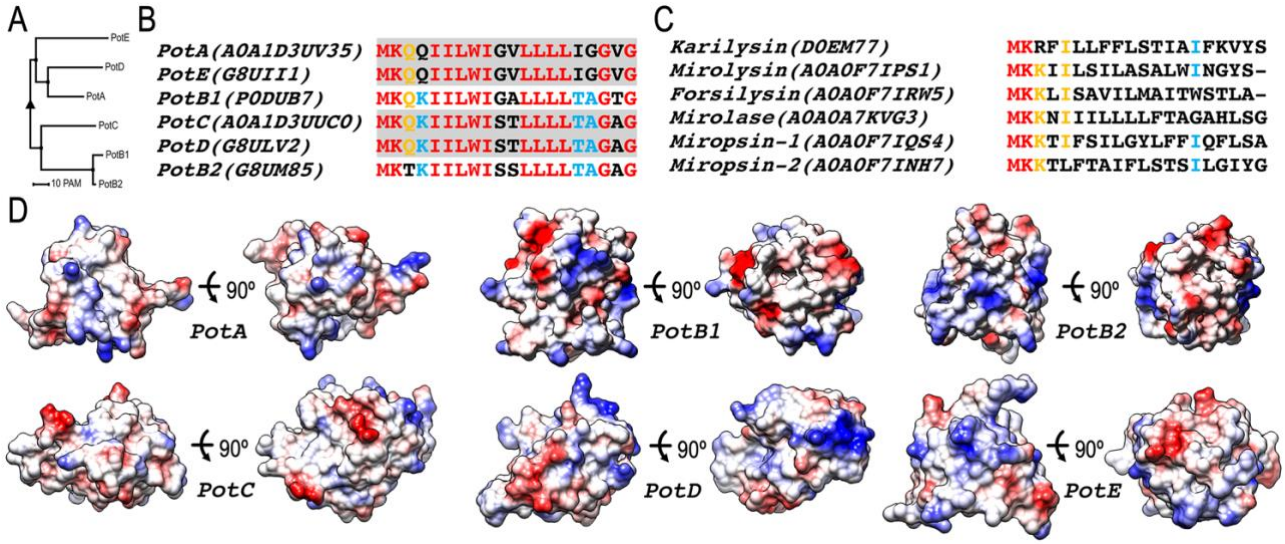
***Suppl. Fig. S1* — Comparison of potempin genes.** (**A**) Alignment of the 5´ and 3´ ends of the potempin coding DNA sequences (blue font) preceding the genes of the corresponding KLIKK-peptidases (in yellow) with start codons (ATG) in bold and stop codons framed to show of proximity of the genes (see also Fig. 1 and Suppl. Fig. S2). (**B**) Alignment of the six mature potempin sequences. For PotB1 and PotB2, identical residues are highlighted in green, signal peptides are in lowercase and residue numbering refers to the mature protein.



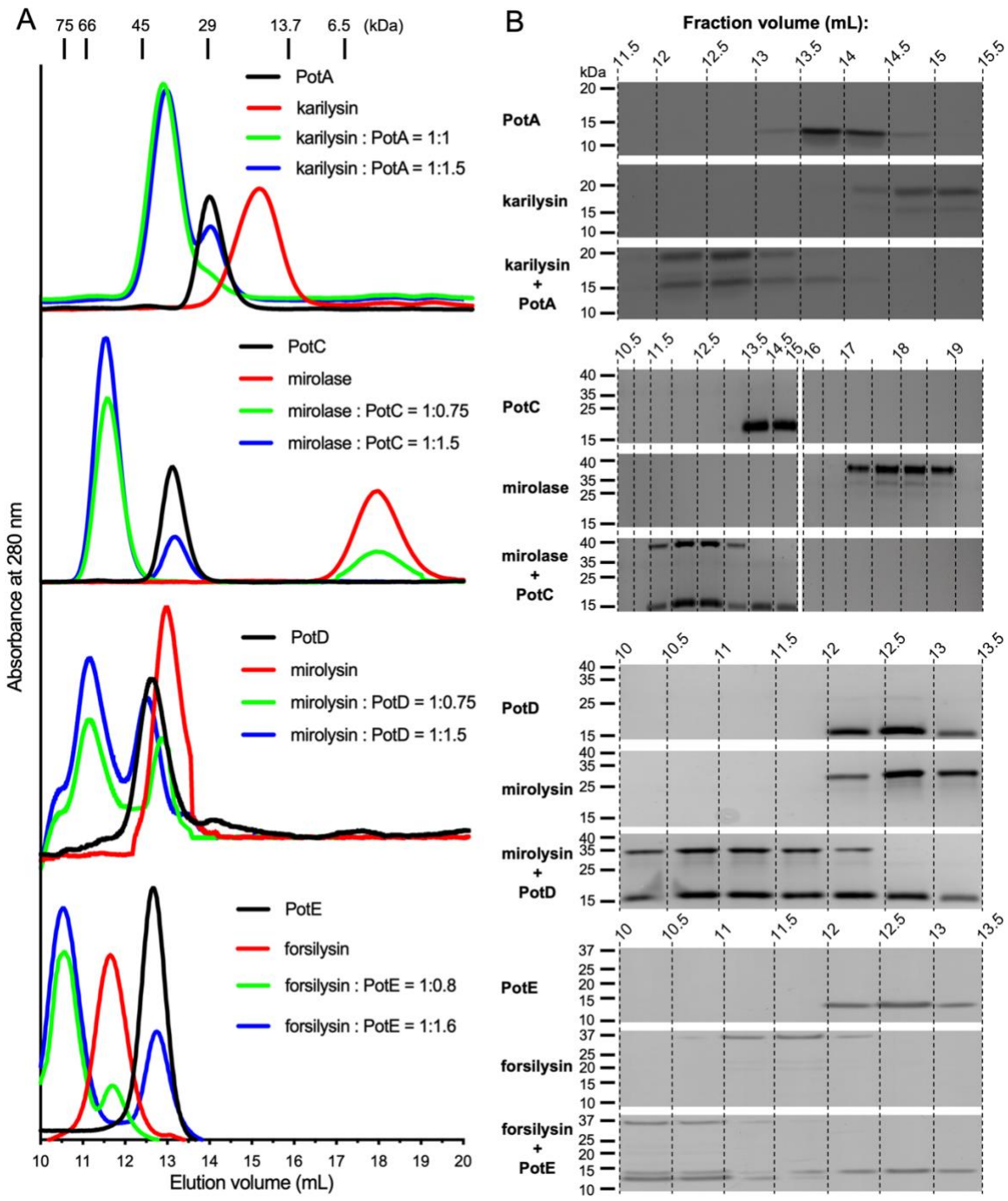
***Suppl. Fig. S2* — KLIKK-peptidases and potempins. (A)** Genomic arrangement of the two loci encoding the six KLIKK-peptidases and their co-transcribed potempins based on manually curated sequences from GenBank (access codes KP715369 and KP715368). For each peptidase:potempin intergenomic region (1–6), the position of the corresponding primers employed (see Suppl. Table S1) is indicated by arrows, as well as the amplicon size in base pairs (bp). Arrows outlined in grey denote other putative ORFs. **(B)** *Tannerella forsythia* mRNA was extracted in the middle of the exponential growth phase, residual genomic DNA was digested with DNase I and the pure mRNA was reverse transcribed into cDNA. Amplicons 1–6 from cDNA (upper panel) and genomic (g) DNA (lower panel) for each of the six intergenomic regions of (A) are shown next to the GeneRuler 1-kb DNA ladder (Thermo Fisher Scientific). **(C)** Comparative SYBR Green-based qRT-PCR analysis of transcript expression levels for each potempin, KLIKK-peptidase, and inhibitor:peptidase operon relative to karilysin (primers are listed in (Suppl. Table S1)). Data are means ± SD (n = 3 technical replicates) and are representative of three experiments. Statistical significance was determined by a one-way analysis of variance (ANOVA) using Tukey's honest significance difference test (* = $p < 0.05$; ** = $p < 0.01$). (**D**) Erroneous arrangement of potempin and KLIKK-peptidase *loci* in the genomic sequences of two *T. forsythia* strains deposited with GenBank under codes GCA_006385365.1 and GCA_000238215.1. These errors arose from the highly repetitive nature of the *T. forsythia* genome, which hampers unambiguous gene assembly based on the short reads generated by Illumina sequencing. For example, the KLIKK-peptidases share a ~700-bp identical segment, which precludes correct assembly. Moreover, the percentage of A/T bases in non-coding regions is very high due to the unique mechanisms of gene expression regulation within the Bacteroidetes phylum. Similar problems were reported during the genome assembly and annotation of *Porphyromonas gingivalis* [21]. Black arrows represent potempin genes, white arrows stand for KLIKK-peptidase genes and dashed arrows denote pseudogenes.



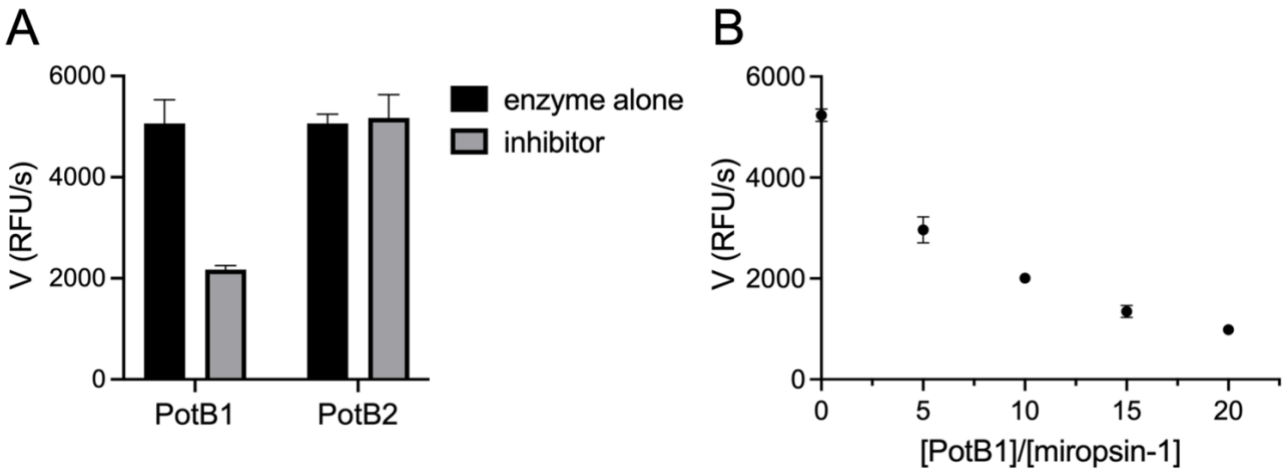
***Suppl. Fig. S3* — Expression and purification of potempins.** All six genes were amplified and cloned in expression vector pGEX-6P-1 without the predicted signal peptide for expression in *Escherichia coli* BL21 cells. The resulting GST-fusion proteins were immobilised on glutathione Sepharose 4 fast flow resin and digested with PreScission protease to elute the tag-free proteins. For each potempin, lanes 1 and 2 show cellular extracts before and 6 h after induction, and lane 3 shows the eluted tag-free proteins.



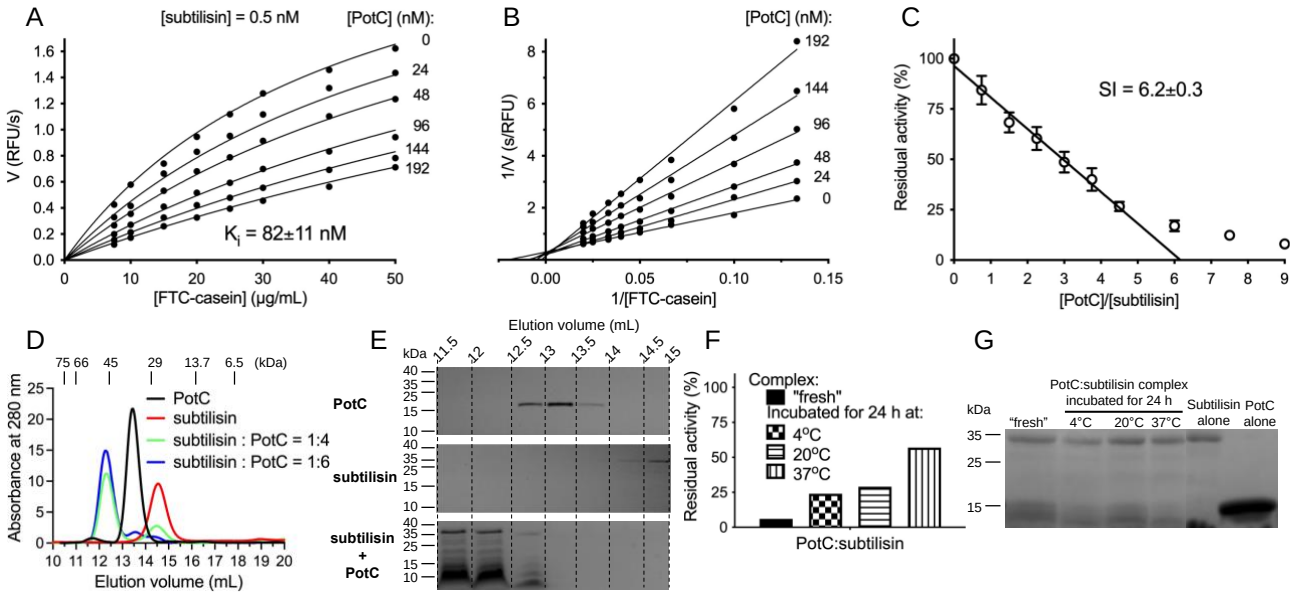
***Suppl. Fig. S4* — Phylogenetic analysis of the potempins and their signal peptides, and Coulombic surfaces. (A)** Sequence-based phenogram of the potempins. The horizontal bar represents 10 point accepted mutations (PAM). **(B)** Sequence alignment of the 20-residue signal peptides of the potempins. Identical residues in all six sequences are shown in red, whereas positions with five and four conserved residues are shown in orange and blue, respectively. Identical sequences are grouped and highlighted in grey. UniProt codes are shown in parentheses. **(C)** Sequence alignment of the 19/20-residue signal peptides of KLIKK-peptidases, with the same colour coding as in (B). **(D)** Coulombic surfaces of the potempins in the two orthogonal orientations of Fig. 6, respectively.



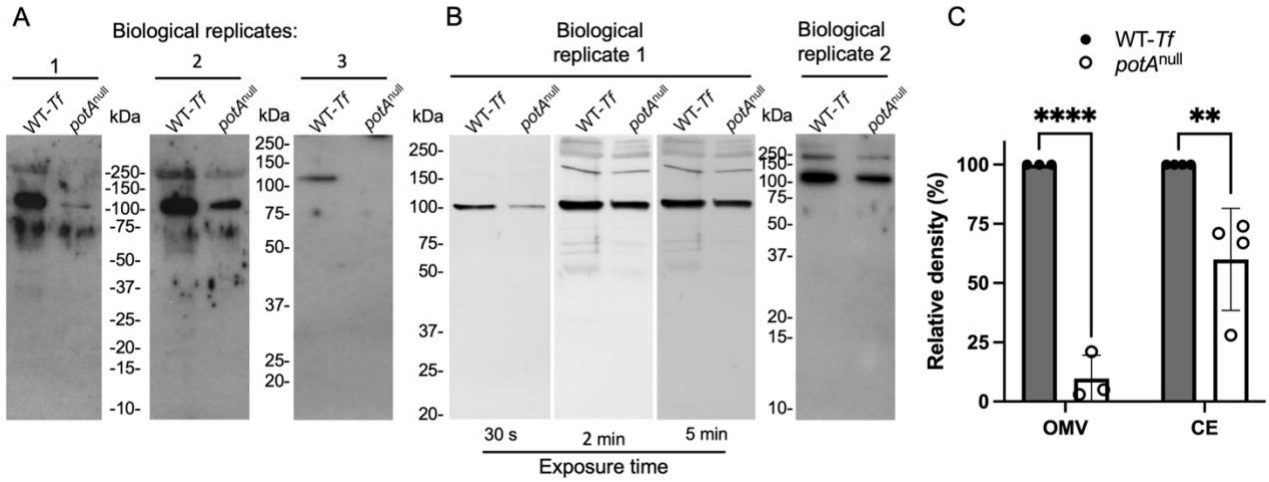
***Suppl. Fig. S5* — Stability of peptidase:inhibitor complexes. (A)** Assessment of 1:1 complex formation by size-exclusion chromatography (SEC) and **(B)** SDS-PAGE. Inhibitors (PotA, PotC, PotD and PotE) and peptidases (karilysin, mirolase, mirolysin and forsilysin) were incubated separately or together (karilysin+PotA, mirolase+PotC, mirolysin+PotD and forsilysin+PotE) at two different molar ratios for 15 min and loaded onto a calibrated Superdex 75 10/300 GL SEC column. Fractions at relevant retention volumes (lower *x*-axis in A) of the distinct runs (inhibitor only = black, peptidase only = red, mixtures = blue and green) were analysed by SDS-PAGE (B) to verify the presence of peptidases and inhibitors matching the anticipated molecular masses.



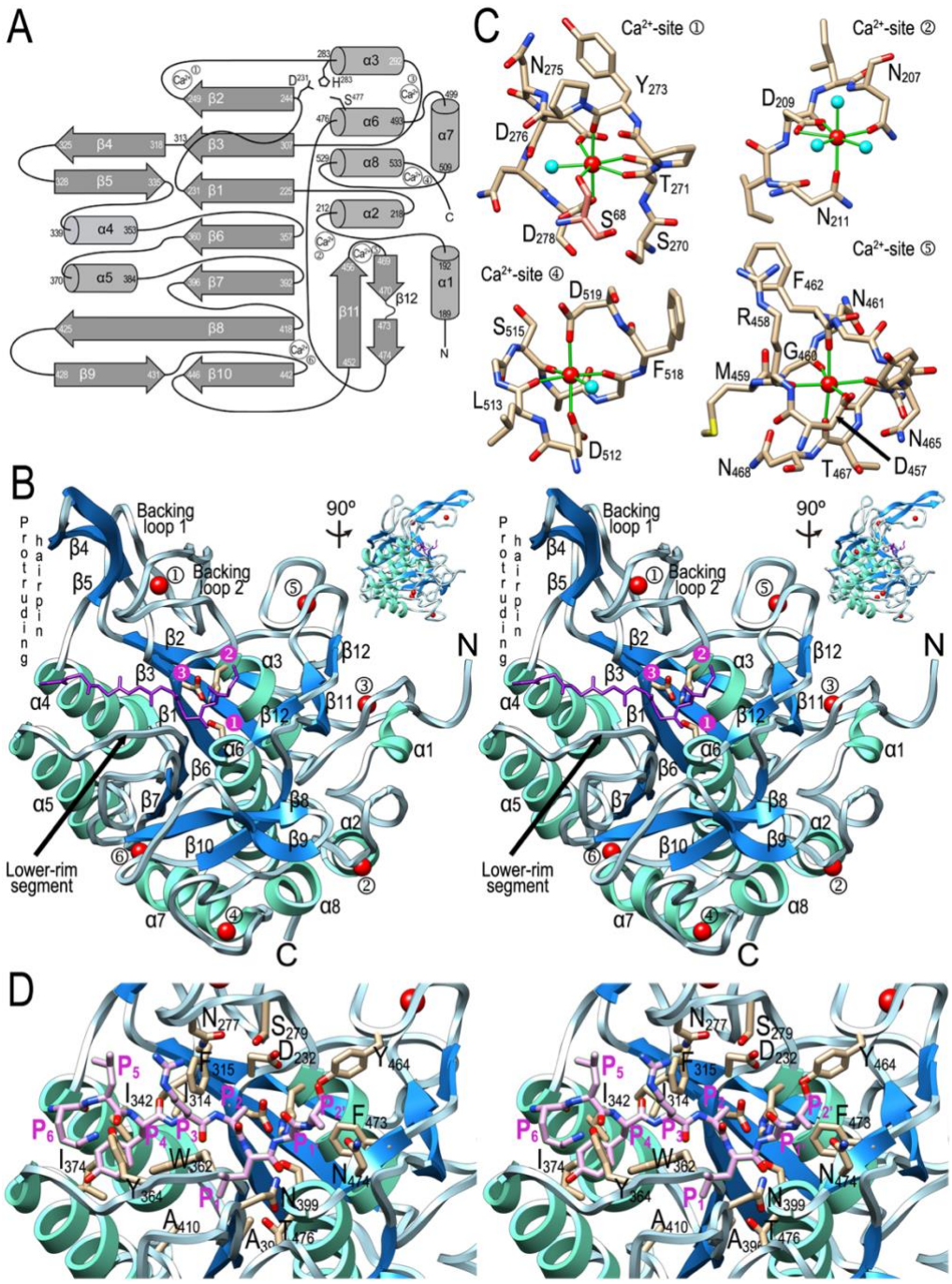
***Suppl. Fig. S6* — Miropsin-1 inhibition by PotB1.** The trypsin-like serine peptidase is inhibited by PotB1 but not the closely-related PotB2 (see also Suppl. Figs. S1B and S4A). **(A)** Miropsin-1 was pre-incubated with a 10-fold molar excess of PotB1 or PotB2 for 15 min, and residual enzyme activity on fluorescein-labelled casein (FTC-casein) was determined as the initial velocity (V) of fluorescence release. **(B)** Concentration-dependent inhibition of miropsin-1 by PotB1 at inhibitor:peptidase ratios of 5:1–20:1. Data are means ± SD (n = 3 technical replicates).



***Suppl. Fig. S7*— PotC inhibits subtilisin Carlsberg. (A)** Subtilisin was incubated with increasing amounts (0–192 nM) of PotC for 15 min, and the velocity (V) in relative fluorescence units (RFU) s^{-1} of the cleavage of FTC-casein at different concentrations was recorded. The readings are the average of two technical replicates, and the K_i values are means ± SD from three experiments. **(B)** The reciprocal of the reaction velocity (1/V) was plotted against the inverse of the substrate concentration, which revealed reversible competitive inhibition and an apparent K_i value of 82 ± 11 nM. **(C)** To determine the stoichiometry of inhibition (SI), subtilisin was pre-incubated with increasing amounts of PotC for 15 min. The residual enzyme activity was determined with Azocoll as the substrate and was plotted against the inhibitor:peptidase molar ratio as a percentage of the activity of the isolated peptidase. Data are means ± SD (n = 3). **(D)** Calibrated SEC in a Superdex 75 10/300 GL column of PotC (black), subtilisin (red), and pre-incubated mixtures of subtilisin:PotC at molar ratios of 1:4 (green) and 1:6 (blue), which reveal complex formation. **(E)** Relevant fractions of (D) were analysed by SDS-PAGE, which revealed the presence of the component proteins and their molecular mass. **(F)** The subtilisin:PotC complex was obtained at a 1:6 molar ratio and incubated at different temperatures for 48 h. The residual activity of the peptidase was determined with Azocoll as the substrate compared to the "fresh" complex (15 min incubation at room temperature), taking the activity of isolated subtilisin as 100%. **(G)** SDS-PAGE analysis of (F) showing proteolytic degradation of most PotC during the inhibitor reaction with subtilisin, which explains the high SI.



***Suppl. Fig. S8* – BspA as a marker for PotA activity.** Densitometry analysis of the intensity of the BspA immunoreactive band in the **(A)** outer-membrane vesicles (OMVs; three biological replicates) and **(B)** cell envelope (CE; two biological replicates) of wild-type *T. forsythia* (WT-*Tf*) and the isogenic PotA-deficient mutant (*potA*^{null}). **(C)** Bands were scanned and their intensity was analyzed using the ImageJ program (https://imagej.nih.gov/ij). The intensity of the band from WT-*Tf* was taken as 100%. Data are means ± SD [n = 3 (OMV) and 4 (CE)]. Statistical significance of difference was determined employing 2–way ANOVA with Šídák's multiple comparisons test (**p<0.01; ****p<0.0001).



***Suppl. Fig. S9* — Structure of mirolase.** **(A)** Topology scheme of mirolase depicting the 12 strands (β1–β12) and the eight helices (α1–α8) of the structure with the flanking residue numbers, as well as the six calcium-binding sites (①–⑥). Strand β12 is interrupted by a bulge. **(B)** Richardson-type plot in cross-eye stereo of mirolase viewed perpendicular to the active-site cleft, which runs horizontally from left (non-primed side) to right (primed side). The strands (blue arrows) and the helices (aquamarine ribbons) are shown and labelled, as well as the six calcium sites (red spheres, labelled ①–⑥). The residues of the catalytic triad (D_{231}, H_{238} and S_{477}) are shown as sticks and labelled in white over magenta (❶–❸). The "protruding hairpin", "backing loop 1", "backing loop 2", and "lower-rim segment" are labelled for reference. A model subtilase substrate was derived from PDB 3CO0 [18]. Residues K_{72}–A_{79} from the latter structure are displayed for their main chain (in purple) to map the active-site cleft of mirolase after optimal superposition of the catalytic domains. An orthogonal view of the catalytic domain is provided in the top-right inset. **(C)** From left to right and from top to bottom, close-up views of the four calcium-binding sites depicted in (B) that have not been found thus far in subtilisins or kexin (①, ②, ④ and ⑤). The chain segments of mirolase displayed are S_{270}–D_{278}, N_{207}–N_{211}, D_{512}–D_{519} and D_{457}–N_{468}, respectively. Calcium ions are shown in red, solvent molecules in cyan, and ion-coordinating atoms are linked with a green line (see also Suppl. Table S6). **(D)** Close up of (B) around the active site, highlighting selected residues that shape the cleft sub-sites as stick models with carbons in tan and black labels (except for the catalytic triad, for clarity). A full-atom substrate model is shown with carbons in pink (labelled **P_6**–**P_2´** in magenta).

3. SUPPLEMENTARY TABLES

***Suppl. Table S1*. List of primers (5´→3´), strains and plasmids used in this study.**	
Primers for cloning	
pGEX-6P-1-PotA	
pGEX_2684F	CACGGATCCGCAAAGAAAGATCAGTCTTCTTGTT
pGEX_2684R	ATTGCGGCCGCTTAATTGATTTTATTGATTGAC
pGEX-6P-1-PotB1	
pGEX_ 1181F	CACGGATCCGCGGAGAAGAACAAGGACTACGAG
pGEX_ 1181R	ATTGCGGCCGCTCATGGTTTCGAGTATTTATAAAG
pGEX-6P-1PotB2	
pGEX_2682F	CACGGATCCGCGGAGAAGAACAAGGACTACGA
pGEX_2682R	ATTGCGGCCGCTCATGACTTCGAGTATTTATAAAG
pGEX-6P-1-PotC	
pGEX_2667F	CACGGATCCGCAAAGAAAGAAACACTTCCTCC
pGEX_2667R	ATTGCGGCCGCTCATGATGGATTTATTTTTAATG
pGEX-6P-1-PotD first PCR	
pGEX_2662F	GACGAATTCGCTGAGAAAGAAAGCCATGCCTC
2662OpR_RTPCR	CCAATCCGTTATCCGCTGGT
pGEX-6P-1-PotD nested PCR	
pGEX_2662F	GACGAATTCGCTGAGAAAGAAAGCCATGCCTC
pGEX_2662R	ATTGCGGCCGCTCATTTCTTTTTTATAGTTGATAATTTAATA
pGEX-6P-1-PotE	
pGEX_1170F	CACGGATCCGCGGAGAATGGACAGTTGCATT
pGEX_1170R	ATTGCGGCCGCTATTTTTTACGTTGATAAATTTCTG
pKO-PotA	
KOPINA_pUC19R	CAGCGGATCCTCTAGAGTCGACCTG
KOPINA_UPF	CTAGAGGATCCGCTGGAAAGACGAAG
KOPINA_UPR	GAAGCTGCAGCGCATAACAATTAACAACTGAAT
KOPINA_ermF	ATGCGCTGCAGCTTCCGCTATTGCTTT
KOPINA_ermR	GTACCACGCATGCAATTTGCCAGCCGTTATG
KOPINA_DWF	AATTGCATGCGTGGTACACGCCGAAT
KOPINA_DWR	CCAGTGAATTCAAACTCTCGACAGTCAGTGTAAT
KOPINA_pUC19F	GAGTTTGAATTCACTGGCCGTCGTT
pdelKO-PotA	
DelKopin2F	ATGAAACGATTTATTCTTTTGTT
DelKopin2R	CAAAAGAATAAATCGTTTCATGTTTTTCTATTTTTATAA
pKO-PotA_C21A	
LipA2Rs	TCCTATCAGCAGAAGCAGGAC
LipA2Rt	TTTCTTAGATCCCACTCCTCCTATCAGCAGAAGCAGGAC
LipA2Fs	GATCAGTCTTCTTGTTGTGATAAAG
LipA2Ft	GGAGTGGGATCTAAGAAAGATCAGTCTTCTTGTTGTGATAAAG
PCR operon primers	
PINEopF/ PINEopR	CTTCCATCGGGTATTGTTCA / TGTCCGGTAATAGAAGGCTTG
PINB1opF / PINB1opR	TTACTAACAGCAGGAACGGGA / TGTATCCATAGTCGATGCGAG
PINDopF / PINDopR	ATACAAAGAAGATGGCCTCCC / TCGCCTAAATTACAGACCCAG
PINCopF / PINCopR	CACTTCGAATATCGGGAGATG / CATATCTTCGTGCGCCTCTAC
PINB2opF / PINB2opR	AGCTCGAATCGTACACATGG / TACATCGAGAGGAGGCATTG
PINAopF / PINAopR	ATGGTATATCTCGGTGTCCGA / TAGATGACCGATCTCATGTGC
qRT-PCR primers	
SSU	
16STFqPCR_F / 16STFqPCR_R	CCTGGTAGTCCACGCAGTAA / GTCAGGGTTGCGCTCGTTAT
PotA	
2684F_RTPCR / 2684R_RTPCR	TAGGTGTCCTGCTTCTGCTG / ACTATTCGCATCGGACACCG

Karilysin		
367f2 / 367rqRT	GGGAATCATGGTGACGGATATCC / GATGACCGATCTCATGTGCCG	
PotA_op		
2684OpR_RTPCR / 2684OpF_RTPCRn	GGCCATTATCGTATAGGCGCT / ACTTCCTGCGGGAACAACTT	
PotB2		
PINB2F_RTPCR / PINB2R_RTPCR	CGAGGTTCCTACGCAACTGA / AACCAGTAGCTGCCTTTCGG	
Miropsin-2		
364seqF2 / 364rqRT	ATCCAACGAAGAAGGAGATATGC / CCGAACTTCCTATCTGACCG	
PotC		
2667F_RTPCR / 2667R_RTPCR	AGCTTATGGAAGAGGGCACG / GGTATTGGCTGGAGGTGGAA	
Mirolase		
347fqRT / 347rqRT	GAATACATCCGTAAGAGTAGCGG / AGGGCTGTCATTGTCGTTCGG	
PotC_op		
PINCopR_RTPCR / PINCopF_RTPCR	GTCTGATGCGACGATACGGA / AGCTTATGGAAGAGGGCACG	
PotD		
2662F_RTPCR / 2662R_RTPCR	GAGCGGGGTGTGAGAAAGAA / CCCGTTTAATCCACGGCTTG	
Mirolysin		
341fqRT / 341rqRT	CTCGTAGTGTGCCTTCTTCCAC / GCCTGATCGGCATTCATTCGG	
PotD_op		
2662OpF_RTPCR / 2662OpR_RTPCR	CCCGCGGATGTGATATTCCT / CCAATCCGTTATCCGCTGGT	
PotB1		
PINB1F_RTPCR / PINB1R_RTPCR	GGAACGGGATGCGAGAAGAA / AATCCGTAGCAGCCTTTGGG	
Miropsin-1		
2174qRT2F / 2174qRT2R	ACGATAAGTTCTTGTTGCCT / GCACCGTTTGATTCGTTATT	
PotB1_op		
PINB1opF_RTPCR / PINB1opR_RTPCR	GCCGCATTGTTCTATCGCATT / AGACGGAGTGCTACCAAAGC	
PotE		
PINEF_RTPCR / PINER_RTPCR	GGTGTCCTGCTTCTGCTGAT / TCCCCAAGATTGCCTGTTCC	
Forsylisin		
Fors. qPCR2162F / Fors. qPCR2162R	GAGTCTGCGAATGACTGAACCAG/CTGTAATGCCCACTGCACGTCC	
PotE_op		
PINEopF_RTPCR / PINEopR_RTPCR	TGCCATACGCGATCTTACAG / CCCGCTTATCATCTGCCGAA	
Bacterial strains	***Description*** *(genotype, resistance)*	***Source / reference***
Escherichia coli DH5α	General cloning host	Thermo Fisher Scientific
Escherichia coli BL21 (DE3)	Expression of PorU protein	EMD-Millipore
Tannerella forsythia WT	Wild-type strain	ATCC 43037
Tannerella forsythia Δ*kly*	*Kar::ermF;* Em^r	[22]
Tannerella forsythia $potA^{null}$	*potA::ermF;* Em^r	This study
Plasmids		
pURgpB-E	Source of the *ermF* cassette	[23]
pKO-PINA-ermF	The master plasmid for *potA/kar* operon modification in the *T. forsythia* genome	This study
pdelKO-PINA	Plasmid for *potA* deletion in the *T. forsythia* genome	This study
pKO-PotA_C21A	Plasmid for PotA protein modification that results in its degradation	This study

***Suppl. Table S2*. Assays testing the inhibition of human and murine MMPs by PotA.**				
	Human catalytic domain	Human full-length enzyme	Mouse catalytic domain	Mouse full-length enzyme
MMP-1	–	No inhibition	–	–
MMP-2	–	No inhibition	–	–
MMP-3	–	No inhibition	–	–
MMP-7	–	No inhibition	–	–
MMP-8	–	Weak inhibition K_i > 100 nM	–	–
MMP-9	–	No inhibition	–	–
MMP-10	–	Weak inhibition K_i > 100 nM	–	–
MMP-12	Strong inhibition K_i ~ 10 nM	Strong inhibition K_i ~ 10 nM	Strong inhibition K_i ~ 5 nM	Strong inhibition K_i ~ 5 nM
MMP-13	–	Very weak inhibition K_i > 1 µM	–	–
MMP-14	–	No inhibition	–	–
MMP-20	Weak inhibition K_i > 100 nM	–	–	–

Suppl. Table S3. **Crystallographic data.**

Dataset	***PotA***	***PotA:karilysin***	***PotA:MMP-12***	***PotC:mirolase***	***PotD $I^{53}M$ (SeMet)***	***PotE (SeMet)***	***PotE***
Synchrotron / beamline	ESRF / ID-23-1	ALBA / XALOC	ALBA / XALOC	ALBA / XALOC	ALBA / XALOC	ALBA / XALOC	ALBA / XALOC
Space group / protomers or complexes per a.u. [a]	$P3_121$ / 1	$P2_12_12_1$ / 1	$P2_12_12_1$ / 2	$P2_1$ / 1	$P2_1$ / 6	$P2_1$ / 2	C2 / 1
Cell axes (a, b, c in Å)	76.21, 76.21, 30.52	38.24, 62.94, 108.68	54.41, 63.00, 135.37	45.66, 113.36, 50.20	70.96, 78.15, 71.46	37.13, 71.35, 39.15	86.92, 36.56, 37.16
Cell angles (α, β, γ in º)	90, 90, 120	90, 90, 90	90, 90, 90	90, 112.7, 90	90, 117.8, 90	90, 89.9, 90	90, 105.6, 90
Wavelength (Å)	0.97918	0.97926	0.97926	0.97950	0.97926	0.97919	0.97925
Measurements / unique reflections	216,001 / 11,454	715,341 / 57,705	424,940 / 40,486	1,153,014 / 184,353	181,967 / 27,047	45,003 / 25,564 [h]	63,917 / 10,402
Resolution range (Å) (outermost shell) [b]	65.9-1.70 (1.80-1.70)	62.9-1.35 (1.42-1.35)	67.7-1.85 (1.96-1.85)	42.9-1.10 (1.17-1.10)	63.2-2.40 (2.55-2.40)	71.3-2.00 (2.11-2.00)	41.9-1.80 (1.91-1.80)
Completeness (%)	100.0 (99.9)	98.6 (97.2)	99.9 (99.3)	96.9 (83.0)	99.2 (98.9)	94.2 (95.3)	98.1 (91.4)
R_{merge} [c]	0.114 (1.379)	0.060 (0.767)	0.097 (0.835)	0.052 (0.650)	0.147 (1.517)	0.060 (0.430)	0.052 (0.492)
$R_{r.i.m.}$ [= R_{meas}] [d] / $CC(^1/_2)$ in % [d]	0.117 (1.423) / 100 (83)	0.063 (0.800) / 100 (90)	0.102 (0.878) / 100 (90)	0.057 (0.740) / 100 (79)	0.159(1.639) / 100 (78)	0.083(0.589) / 100 (72)	0.057(0.560) / 100 (88)
Average intensity [e]	17.1 (2.6)	20.8 (3.5)	14.3 (3.8)	18.2 (2.2)	11.1 (2.4)	8.7 (2.1)	21.2 (3.2)
B-Factor (Wilson) ($Å^2$) / Aver. multiplicity	31.5 / 18.9	21.9 / 12.4	36.7 / 10.5	13.8 / 6.3	61.4 / 6.7	35.0 / 1.8	30.1 / 6.1
Number of heavy-atom sites used for phasing / f.o.m.	-	-	-	-	-	4 / 0.29 0.59	-
Resolution range used for refinement (Å)	66.0 – 1.70	36.1 – 1.35	67.7 – 1.85	42.9 – 1.10	63.2 – 2.40	37.3 – 2.00	35.8 – 1.80
Reflections used (test set)	10,873 (578)	56,942 (762)	39,686 (798)	183,591 (761)	26,453 (531)	13,070 (658) [h]	9852 (549)
Crystallographic R_{factor} (free R_{factor}) [c]	0.177 (0.214)	0.154 (0.171)	0.190 (0.228)	0.139 (0.148)	0.218 (0.236)	0.193 (0.251)	0.184 (0.201)
Non-H protein atoms / solvent molecules / ligands per a.u.	760 / 146 / 1 Ni^{2+}	2090 / 305 / 2 Zn^{2+}, 1 Ca^{2+}, 1 MES, 2 GOL	3947 / 276 / 4 Zn^{2+}, 6 Ca^{2+}	3618 / 752 / Na^+, 6 Ca^{2+}, 8 GOL, 2 EDO	4944 / 66 / 1 EDO	1605 / 107 / 2 GOL	850 / 88 / 1 EPE
Rmsd from target values							
bonds (Å)/angles (°)	0.010 / 1.12	0.010 / 1.07	0.009 / 0.99	0.008 / 1.02	0.009 / 1.20	0.010 / 1.07	0.010 / 0.95
Average B-factor ($Å^2$)	33.0	22.3	41.0	14.5	61.0	34.0	30.3
All-atom contacts and geometry analysis [f]							
Ramachandran favoured / outliers / all [g]	95 / 0 / 97	255 / 0 / 261	491 / 0 / 477	500 / 0 / 488	559 / 0 / 587	178 / 0 / 182	100 / 0 / 98
Protein bond-length / bond-angle / chiral. / planar. outliers	0 / 0 / 0 / 0	0 / 0 / 0 / 0	0 / 0 / 0 / 0	0 / 0 / 0 / 1	0 / 0 / 0 / 0	0 / 0 / 0 / 0	0 / 0 / 0 / 0
Side-chain outliers	0	2	15	0	16	2	2
All-atom clashscore	0.0	0.2	0.5	0.1	6.3	0.3	2.2
RSRZ outliers / F_{obs}:F_{calc} correlation	4 / 0.96	6 / 0.97	13 / 0.93	14 / 0.97	17 / 0.95	4 / 0.95	3 / 0.94
PDB access code	8B2M	8B2Q	8B2N	8EHE	8EHB	8EHC	8EHD
Crystallization conditions (reservoir buffer)	20% PEG 2000, 10 µM nickel chloride, 100 mM Tris·HCl, pH 8.5	25% PEG 6000, 100 mM MES, pH 6.0	30% PEG 3000, 200 mM sodium chloride, 100 mM Tris·HCl, pH 7.0	19% PEG MME 2000 in 100 mM succinic acid, sodium dihydrogen phosphate and glycine at 2:7:7 (pH 8.0)	20% PEG 3350, 0.2 M diammonium hydrogen citrate	20% PEG 1000, 100 mM Tris·HCl, pH 8.5	20% PEG 8000, 100 mM HEPES, pH 7.5

[a] Abbreviations: a.u., asymmetric unit; EDO, ethylene glycol; EPE, 2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid (HEPES); f.o.m., mean figure of merit before and after density modification plus averaging; GOL, glycerol; MES, 2-(N-morpholino)ethanesulfonic acid; MME, monomethyl ether; PEG, polyethylene glycol; RSRZ, real-space R-value Z-score; and Tris, 2-amino-2(hydroxymethyl)propane-1,3-diol. [b] Values in parentheses refer to the outermost resolution shell. [c] For definitions, see Table 1 in [24]. [d] For definitions, see [25, 26]. [e] Average intensity is $\langle I/\sigma(I)\rangle$ of unique reflections after merging according to the XDS program [27]. [f] According to the PDB Validation Service (https://validate-rcsb-1.wwpdb.org/validservice/). [g] Including residues with alternate main-chain conformations. [h] Friedel mates were kept separately for phasing but merged for refinement.

***Suppl. Table S4*. Homology model validation.**				
	PotB1	***PotB2***	***PotD:mirolysin***	***PotE:forsilysin***
MolProbity Score	1.15	1.18	1.69	2.96
Clash score	0.52	0.51	4.3	15.1
Ramachandran favoured	92%	92%	97%	91%
Ramachandran outliers	2 (2%)	4 (3%)	1 (0%)	10 (2%)
Rotamer outliers	0 (0%)	0 (0%)	8 (2%)	37 (10%)
Cβ deviations	0	0	0	0
Bad bonds (out of total)	0 (996)	0 (1016)	0 (3060)	0 (3549)
Bad angles (out of total)	2 (1349)	0 (1374)	6 (4155)	0 (4826)
Results obtained with *Molprobity* [28].				

***Suppl. Table S5*. Root-mean square deviations (*rmsd*) upon pairwise superposition of potempins.**

	PotC	***PotD***
PotA	2.05 Å (46)	1.64 Å (62)
PotC		2.54 Å (57)
	PotB2	***PotE***
PotB1	3.46 Å (111)	1.80 Å (67)
PotB2		1.99 Å (67)

Listed are the respective *rmsd* values and the number of superposed Cα atoms (in parenthesis) of OB-fold potempins (PotA, PotC and PotD) and β-hairpin-repeat barrels (PotB1, PotB2 and PotE). Results obtained with *Lsqkab* [29].

***Suppl. Table S6*. Calcium-binding sites in mirolase.**

Site	Coordination	In-plane ligands	Apical ligands	Distances to ion	Observations
1	Pentagonal bipyramidal	T_{271} O T_{271} Oγ1 D_{276} Oδ1 Solvent S^{68} Oγ1 [a]	Y_{273} O D_{278} Oδ2	2.29–2.45 Å	
2	Octahedral + 1	D_{209} Oδ1 and Oδ2 [c] N_{211} Oδ1 2 x Solvent	N_{207} Oδ1 Solvent	2.23–2.58 Å	
3	Octahedral + 1	D_{187} Oδ2 D_{239} Oδ1 and Oδ2 [c] N_{296} Oδ1 I_{300} O	N_{294} O M_{298} O	2.31–2.54 Å	Shared with subtilisins Carlsberg and Novo, and kexin [b]
4	Octahedral	L_{513} O S_{515} O G_{517} O Solvent	D_{512} Oδ1 D_{519} Oδ2	2.26–2.35 Å	
5	Octahedral	D_{457} Oδ2 R_{458} O G_{460} O N_{465} O	F_{462} O T_{467} O	2.31–2.34 Å	
6	Octahedral	T_{420} Oγ1 D_{441} Oδ2 2 x Solvent [d]	V_{418} O S_{439} O	2.32–2.43 Å	Shared with subtilisins Carlsberg and Novo, and kexin [b]

[a] Ligand provided by PotC, probably replaced by a solvent in unbound mirolase.
[b] See PDB entries 1R64 (three potassium and three calcium sites), 2SNI (two calcium sites), 1SBT (no site), 1SBC (one calcium site), 2SEC (three calcium sites) and 3CO0 (four zinc sites).
[c] Bidentate ligand, slightly larger distances than standard (2.36–2.39 Å, see [30]).
[d] One of the two solvent ligands is disordered and was modelled in two alternate positions.

4. SUPPLEMENTARY REFERENCES

1. M. Ksiazek, A. Y. Karim, D. Bryzek, J. J. Enghild, I. B. Thøgersen, J. Koziel and J. Potempa, *Biol. Chem.*, 2015, **396**, 261-275.
2. N. D. Rawlings, A. J. Barrett and R. Finn, *Nucleic Acids Res.*, 2016, **44**, D343-D350.
3. K. Linderstrøm-Lang and M. Ottesen, *Nature*, 1947, **159**, 807-808.
4. I. Schechter and A. Berger, *Biochem. Biophys. Res. Commun.*, 1967, **27**, 157-162.
5. C. A. McPhalen and M. N. G. James, *Biochemistry*, 1988, **27**, 6582-6598.
6. F. X. Gomis-Rüth, T. O. Botelho and W. Bode, *Biochim. Biophys. Acta*, 2012, **1824**, 157-163.
7. H. Grøn, M. Meldal and K. Breddam, *Biochemistry*, 1992, **31**, 6011-6018.
8. T. Gallagher, G. Gilliland, L. Wang and P. Bryan, *Structure*, 1995, **3**, 907-914.
9. C. Tallant, R. García-Castellanos, J. Seco, U. Baumann and F. X. Gomis-Rüth, *J. Biol. Chem.*, 2006, **281**, 17920-17928.
10. A. Y. Karim, M. Kulczycka, T. Kantyka, G. Dubin, A. Jabaiah, P. S. Daugherty, I. B. Thogersen, J. J. Enghild, K. A. Nguyen and J. Potempa, *Biol. Chem.*, 2010, **391**, 105-117.
11. T. Goulas, J. L. Arolas and F. X. Gomis-Rüth, *Proc. Natl. Acad. Sci. USA*, 2011, **108**, 1856-1861.
12. S. Trillo-Muyo, S. Martínez-Rodríguez, J. L. Arolas and F. X. Gomis-Rüth, *Chem. Sci.*, 2013, **4**, 791-797.
13. M. Ksiazek, D. Mizgalska, S. Eick, I. B. Thøgersen, J. J. Enghild and J. Potempa, *Front. Microbiol.*, 2015, **6**, 312.
14. J. L. Arolas, T. Goulas, A. P. Pomerantsev, S. H. Leppla and F. X. Gomis-Rüth, *Structure*, 2016, **24**, 25-36.
15. J. S. Richardson, *Adv. Prot. Chem.*, 1981, **34**, 167-339.
16. C. S. Wright, R. A. Alden and J. Kraut, *Nature*, 1969, **221**, 235-242.
17. D. J. Neidhart and G. A. Petsko, *Prot. Eng.*, 1988, **2**, 271-276.
18. T. Gallagher, B. Ruan, M. London, M. A. Bryan and P. N. Bryan, *Biochemistry*, 2009, **48**, 10389-10394.
19. W. Bode, E. Papamokos, D. Musil, U. Seemueller and H. Fritz, *EMBO J.*, 1986, **5**, 813-818.
20. P. K. Madala, J. D. A. Tyndall, T. Nall and D. P. Fairlie, *Chem. Rev.*, 2010, **110**, PR1-PR31.
21. L. Acuña-Amador, A. Primot, E. Cadieu, A. Roulet and F. Barloy-Hubler, *BMC genomics*, 2018, **19**, 54.
22. M. Jusko, J. Potempa, D. Mizgalska, E. Bielecka, M. Ksiazek, K. Riesbeck, P. Garred, S. Eick and A. M. Blom, *J. Immunol.*, 2015, **195**, 2231-2240.
23. D. Mizgalska, T. Goulas, A. Rodríguez-Banqueri, F. Veillard, M. Madej, E. Małecka, K. Szczesniak, M. Ksiazek, M. Widziołek, T. Guevara, U. Eckhard, M. Solà, J. Potempa and F. X. Gomis-Rüth, *Proc. Natl. Acad. Sci. USA*, 2021, **118**, e2103573118.
24. R. García-Castellanos, A. Marrero, G. Mallorquí-Fernández, J. Potempa, M. Coll and F. X. Gomis-Rüth, *J. Biol. Chem.*, 2003, **278**, 39897-39905.
25. M. S. Weiss, *J. Appl. Cryst.*, 2001, **34**, 130-135.
26. P. A. Karplus and K. Diederichs, *Science*, 2012, **336**, 1030-1033.
27. W. Kabsch, *Acta Crystallogr. sect. D*, 2010, **66**, 125-132.
28. C. J. Williams, J. J. Headd, N. W. Moriarty, M. G. Prisant, L. L. Videau, L. N. Deis, V. Verma, D. A. Keedy, B. J. Hintze, V. B. Chen, S. Jain, S. M. Lewis, W. B. Arendall 3rd, J. Snoeyink, P. D. Adams, S. C. Lovell, J. S. Richardson and D. C. Richardson, *Protein Sci.*, 2018, **27**, 293-315.
29. W. Kabsch, *Acta Crystallogr. sect. A*, 1976, **32**, 922-923.
30. M. M. Harding, *Acta Crystallogr. sect. D*, 2006, **62**, 678-682.