Supporting Information

Exploring the specificity of extracellular wastewater peptidases to inform the design of sustainable peptide-based antibiotics

Michael T. Zumstein[1,2*], Jeffrey J. Werner[3], and Damian E. Helbling[1]

[1]School of Civil and Environmental Engineering, Cornell University, Ithaca, NY, 14853, USA
[2]Swiss Federal Institute of Aquatic Science and Technology (Eawag), 8600 Dübendorf, Switzerland
[3]Chemistry Department, SUNY-Cortland, Cortland, NY, 13045, USA

[*]To whom correspondence should be addressed:
E-mail: mtz27@cornell.edu, mtzumstein@gmail.com

This PDF file (15 pages) includes:

Supplementary Materials and Methods

Chemicals. Bovine serum albumin (BSA, article number: A2153), sodium phosphate monobasic (S8282), sodium phosphate dibasic (S7907), ammonium bicarbonate (09830), and calcium chloride (C4901) were obtained from Sigma-Aldrich. Trypsin (90058), formic acid (28905), dithiothreitol (DTT, R0861), and sodium dodecyl sulfate (SDS, 064382) were obtained from Thermo Fisher Scientific. Iodoacetamide (IAA, 122270050) was obtained from Acros Organics. Trifluoroacetic acid (TFA, 20420) was obtained from Honeywell. Acetic acid (044721) was obtained from Oakwood Chemical. LC-MS grade water (58201) and methanol (58215) were obtained from OmniSolv. Milli-Q water was prepared with an Advantage A10 (Millipore).

Total suspended solids quantification. We determined the total suspended solid content of wastewater sludge by transferring 30 mL of suspended sludge onto a pre-weighted glass microfiber filter (Whatman GF/F, product number: 1825-047) that had been mounted into a filter holder (Nalgene, 300-4000). After applying a vacuum and drying the filter paper (120 °C, 2 h), we re-reweighted the filter paper and calculated the solid content from the mass difference.

Protein concentration measurements. We used the Pierce bicinchoninic acid (BCA) Protein Assay kit (Thermo Fisher, 23225) and a microplate reader (Tecan, Infinite M200-pro) to determine the protein concentrations of the enzyme extracts (*14*).

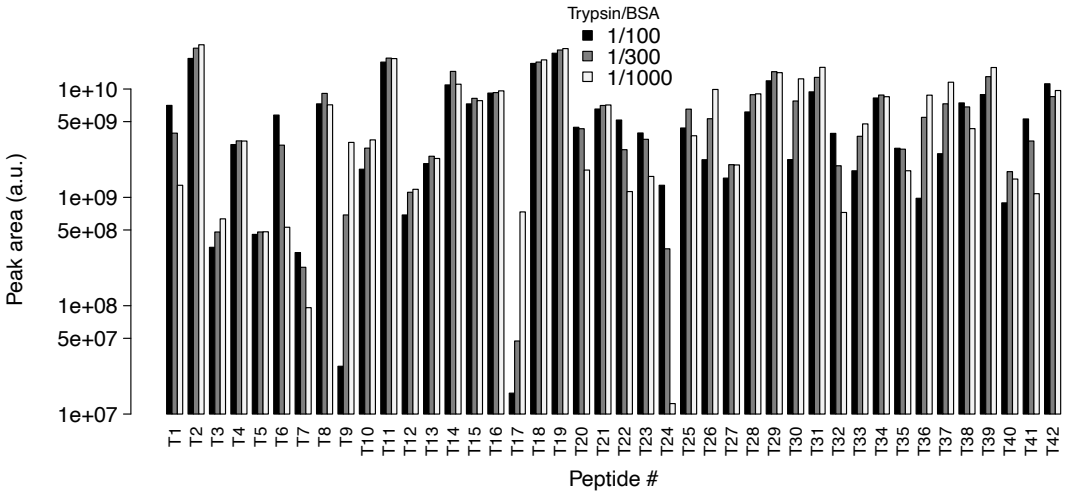

Figure S1. Peak areas of parent peptides resulting from the digestion of bovine serum albumin (BSA) with trypsin at different trypsin/BSA ratios (w/w).

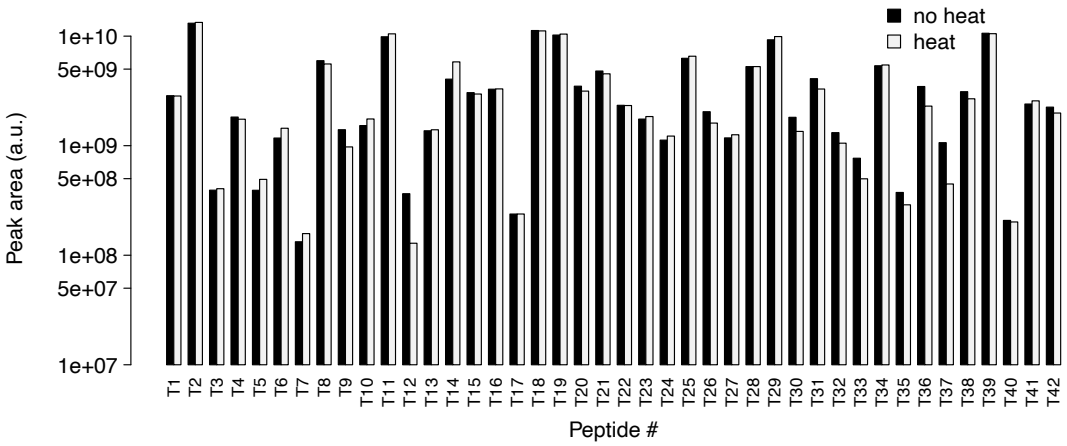

Figure S2. Peak areas of parent peptides before and after heat treatment.

Table S1. Amino-acid sequences (single-letter abbreviations) and high-performance liquid chromatography high-resolution mass spectrometry (HPLC-HRMS) acquisition parameters of parent peptides.

Peptide	Sequence	Position within BSA	Exact m/z	Mass deviation	Charge State	Retention time (min)	Fragment ions
T1	DLGEEHFK	D36-K43	487.7325	-0.7	2	34	y6, y4, y3
T2	LVNELTEFAK	L65-K74	582.3190	1.1	2	57	y8, y6, y5
T3	TCVADESHAGCEK	T75-K87	488.5345	0.0	3	22	y11, y10, y5
T4	VASLR	V100-R104	273.1739	-0.4	2	19	y3
T5	ETYGDMADCCEK	E105-K116	739.7653	0.7	2	35	y9, y7, y6
T6	NECFLSHK	N122-K129	345.4956	-0.4	3	30	y6, y5, y3
T7	DDSPDLPK	D130-K137	443.7113	0.0	2	34	y6, y5, y3
T8	LKPDPNTLCDEFK	L138-K150	526.2607	0.2	3	52	y7, y6, y5
T9	LKPDPNTLCDEFKADEK	L138-K154	673.9946	-0.2	3	55	y15, y13
T10	KFWGK	K155-K159	333.1921	-0.2	2	35	y3
T11	YLYEIAR	Y160-R166	464.2504	0.6	2	46	y5, y4, y3
T12	HPYFYAPELLYYANK	H168-K182	630.3138	-0.3	3	73	y9, y6, y5
T13	YNGVFQECCQAEDK	Y183-K196	874.3562	1.5	2	42	y10, y9, y8
T14	GACLLPK	G197-K203	379.7151	0.5	2	35	y5, y4, y3
T15	IETMR	I204-R208	325.1705	-0.7	2	22	y3
T16	AWSVAR	A235-R240	345.1901	-1.0	2	31	y4, y3
T17	LSQKFPK	L241-K247	424.2554	0.5	2	29	y5, y4, y3
T18	AEFVEVTK	A248-K255	461.7477	0.0	2	37	y6, y5, y4
T19	LVTDLTK	L256-K262	395.2395	-0.1	2	36	y5, y4, y3
T20	ECCHGDLLECADDR	E266-R279	583.8924	0.8	3	41	y6, y5
T21	YICDNQDTISSK	Y285-K296	722.3247	-0.3	2	34	y10, y9
T22	ECCDKPLLEK	E299-K308	431.2055	-0.2	3	29	y8, y5
T23	SHCIAEVEK	S309-K317	358.1746	-0.5	3	28	y5, y4, y3
T24	DAIPENLPPLTADFAEDK	D318-K335	978.4835	0.5	2	74	y15, y11
T25	DAIPENLPPLTADFAEDKDVCK	D318-K229	820.0651	0.8	3	69	y19, y15
T26	DAFLGSFLYEYSR	D346-R358	784.3750	-0.1	2	81	y10, y9, y8
T27	EYEATLEECCAK	E374-K385	751.8105	0.3	2	39	y8, y7, y6
T28	DDPHACYSTVFDK	D386-K398	518.8892	0.1	3	46	y11, y5, y3
T29	HLVDEPQNLIK	H401-K411	653.3617	0.0	2	51	y9, y8, y6
T30	LGEYGFQNALIVR	L420-R432	740.4014	-0.2	2	66	y10, y9, y7
T31	KVPQVSTPTLVEVSR	K436-R450	547.3174	0.3	3	52	y6, y5, y4
T32	VPQVSTPTLVEVSR	V437-R450	756.4250	0.6	2	55	y13, y10, y8
T33	MPCTEDYLSLILNR	M468-R481	862.9209	-0.4	2	80	y13, y11
T34	LCVLHEK	L482-K488	300.1654	-0.3	3	31	y5, y4, y3
T35	CCTESLVNR	C498-R506	569.7526	-0.2	2	31	y7, y6, y5
T36	RPCFSALTPDETYVPK	R507-K522	627.6452	0.9	3	59	y8
T37	LFTFHADICTLPDTEK	L528-K543	636.6451	0.1	3	67	y14, y5
T38	QTALVELLK	Q548-K556	507.8133	1.2	2	60	y7, y5, y4
T39	TVMENFVAFVDK	T568-K579	700.3499	0.3	2	74	y10, y9, y8
T40	CCAADDKEACFAVEGPK	C580-K596	643.2709	-0.1	3	41	y15, y14
T41	EACFAVEGPK	E587-K596	554.2606	1.1	2	39	y7, y6, y3
T42	LVVSTQTALA	L597-A606	501.7951	0.4	2	55	y8, y3

MKWVTFISLLLLFSSAYSRGVFRRDTHKSEIAHRFKDLGEEHFKGLVLIAFSQYLQQCPFDEHVKLVNEL
TEFAKTCVADESHAGCEKSLHTLFGDELCKVASLRETYGDMADCCEKQEPERNECFLSHKDDSPDLPK
LKPDPNTLCDEFKADEKKFWGKYLYEIARRHPYFYAPELLYYANKYNGVFQECCQAEDKGACLLPKIET
MREKVLASSARQRLRCASIQKFGERALKAWSVARLSQKFPKAEFVEVTKLVTDLTKVHKECCHGDLLE
CADDRADLAKYICDNQDTISSKLKECCDKPLLEKSHCIAEVEKDAIPENLPPLTADFAEDKDVCKNYQEA
KDAFLGSFLYEYSRRHPEYAVSVLLRLAKEYEATLEECCAKDDPHACYSTVFDKLKHLVDEPQNLIKQN
CDQFEKLGEYGFQNALIVRYTRKVPQVSTPTLVEVSRSLGKVGTRCCTKPESERMPCTEDYLSLILNRL
CVLHEKTPVSEKVTKCCTESLVNRRPCFSALTPDETYVPKAFDEKLFTFHADICTLPDTEKQIKKQTALV
ELLKHKPKATEEQLKTVMENFVAFVDKCCAADDKEACFAVEGPKLVVSTQTALA

Figure S3. Amino acid sequence (single-letter abbreviations) of bovine serum albumin (BSA, UniProt-ID: P02769). Sequences of parent peptides that were detected by high-performance liquid chromatography coupled to high-resolution mass spectrometry (HPLC-HRMS) are highlighted in grey.

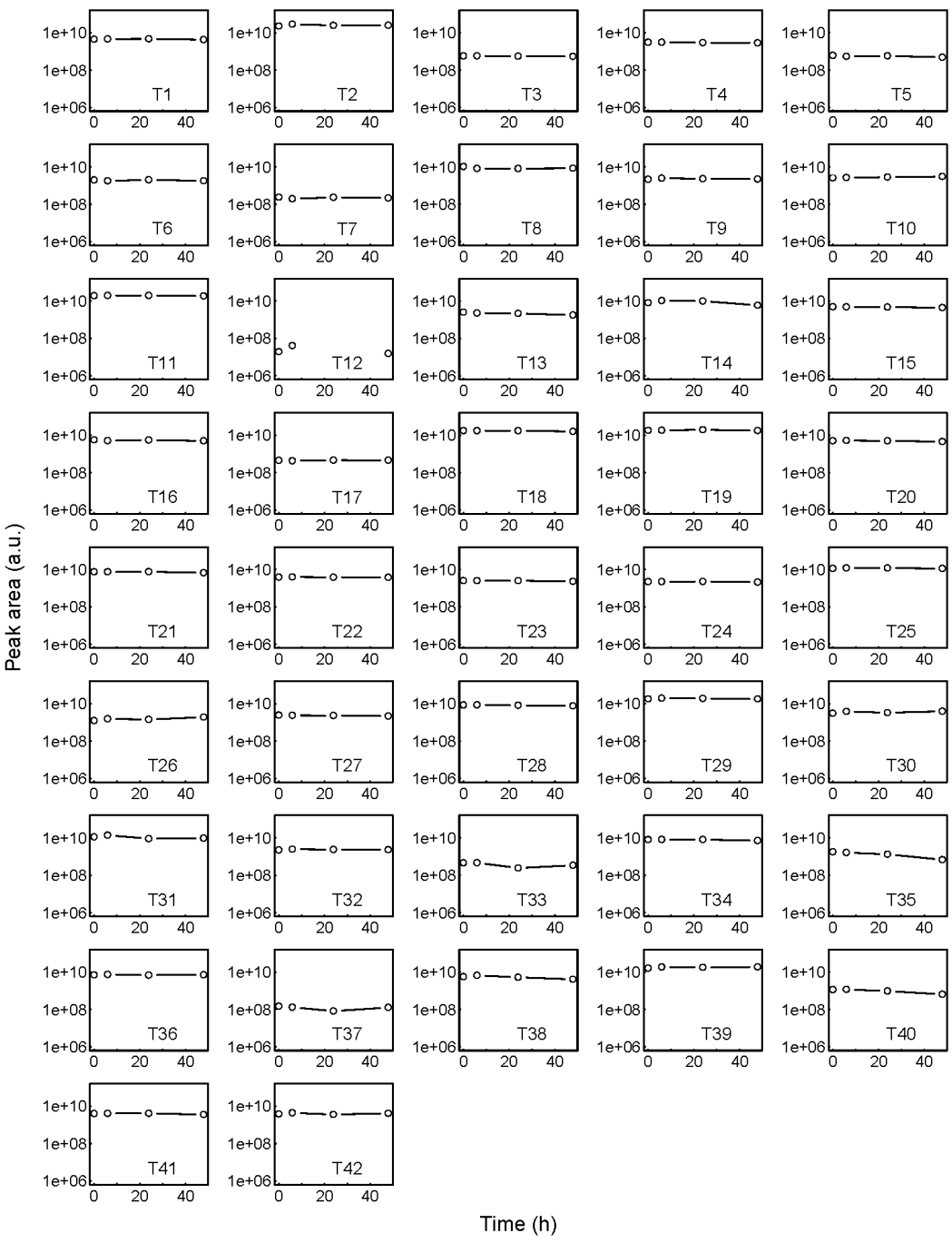

Figure S4. Peak area progress curves of parent peptides during their incubation in pH-buffered Milli-Q water. Points represent data from single measurements. Peptide T12 was not detected at 24 h.

Table S2. Characteristics of the activated sludge process of the wastewater treatment facilities from which microbial communities (MCs) were derived.

	Location	Process[1)]	Hydraulic retention time (h)	Solids retention time (d)	TSS (g/L) [2)]	Protein concentration (mg/L) [3)]	Protease activity (a.u.) [3)]
MC1	Ithaca	CAS	2.7	5.4	1.90 ± 0.02	14.43 ± 0.07	1.00 ± 0.06
MC2	Dryden	SBR	25	25	1.80 ± 0.02	9.90 ± 0.06	1.57 ± 0.03
MC3	Trumansburg	CAS	20	22	3.49 ± 0.03	10.35 ± 0.03	1.75 ± 0.04

[1)]CAS: conventional activated sludge, SBR: sequencing batch reactor, note: none of the three treatment facilities was designed for biological phosphorus removal. [2)]TSS: total suspended solids. [3)]Protein concentration and protease activity were measured on dissolved extracellular enzyme extracts. Protease activities were normalized to the activity measured for MC1.

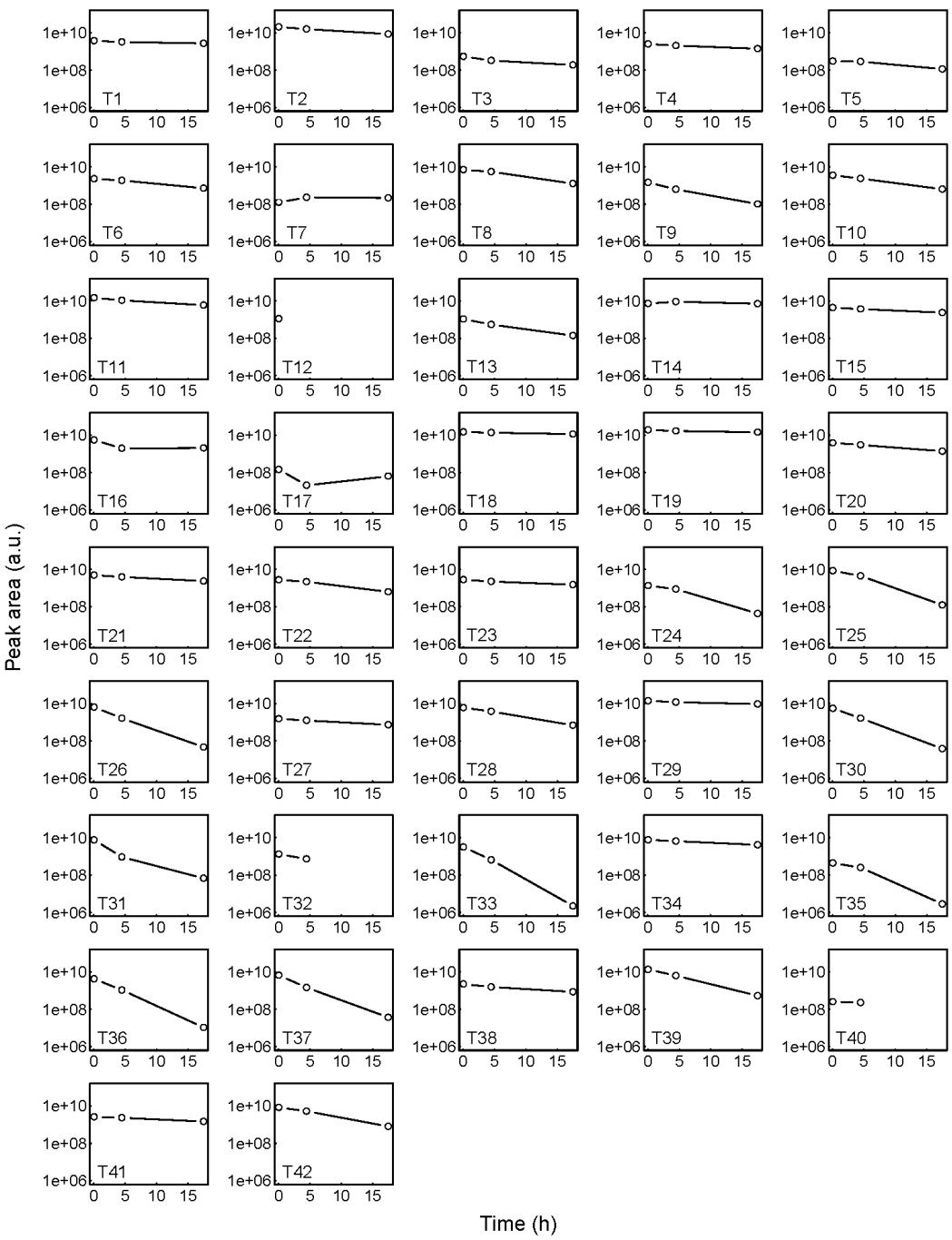

Figure S5. Peak area progress curves of parent peptides during their incubation with enzyme extract from MC1. Points represent data from single measurements. The missing data points for peptides T12, T32, and T40 indicate that these peptides were not detected at later sampling points.

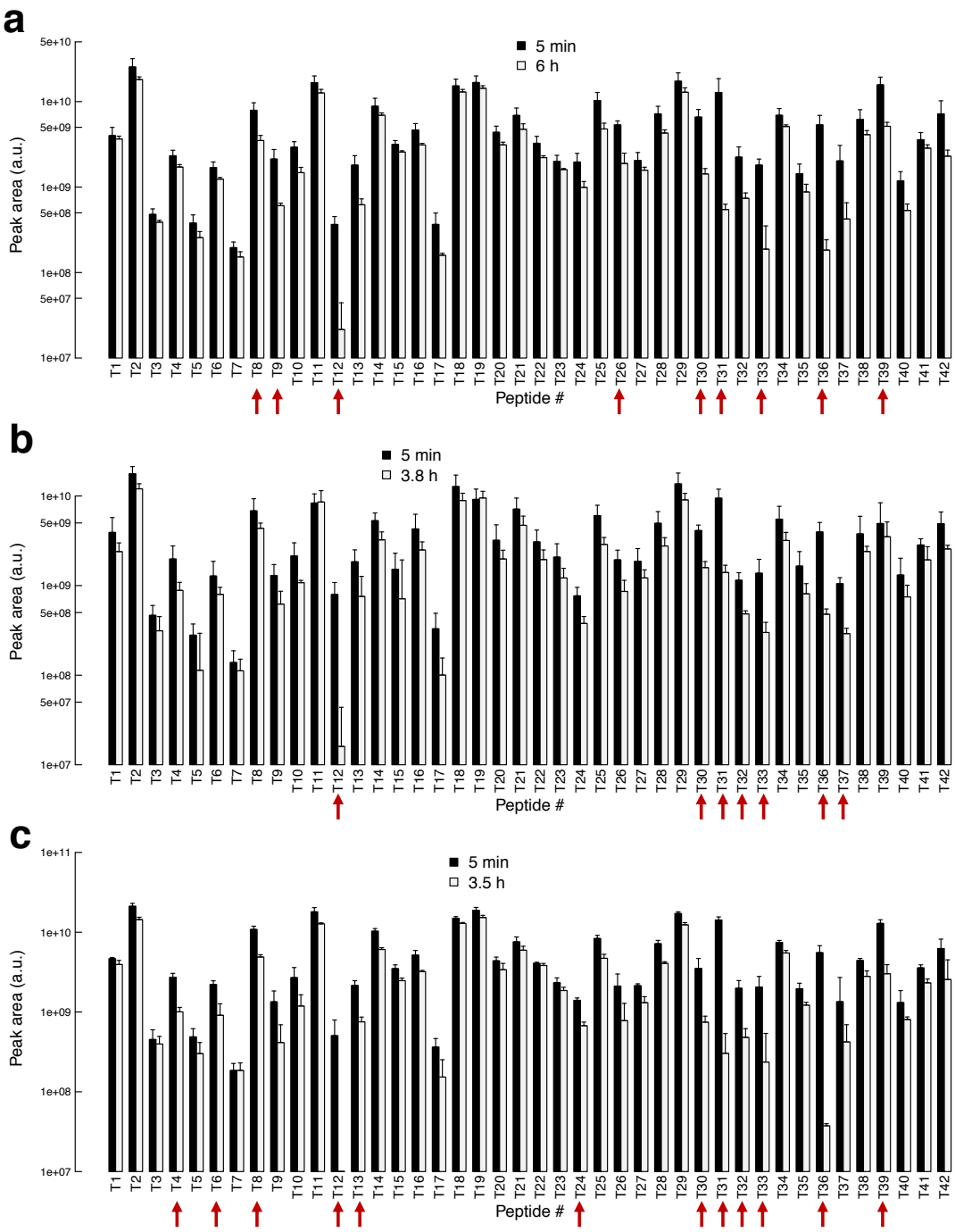

Figure S6. Peak areas of parent peptides at the onset and at the end of their incubation with enzyme extracts from MC1 (**a**), MC2 (**b**), and MC3 (**c**). Red arrows indicate parent peptides that were selected for product identification. Error bars represent standard deviations of triplicates.

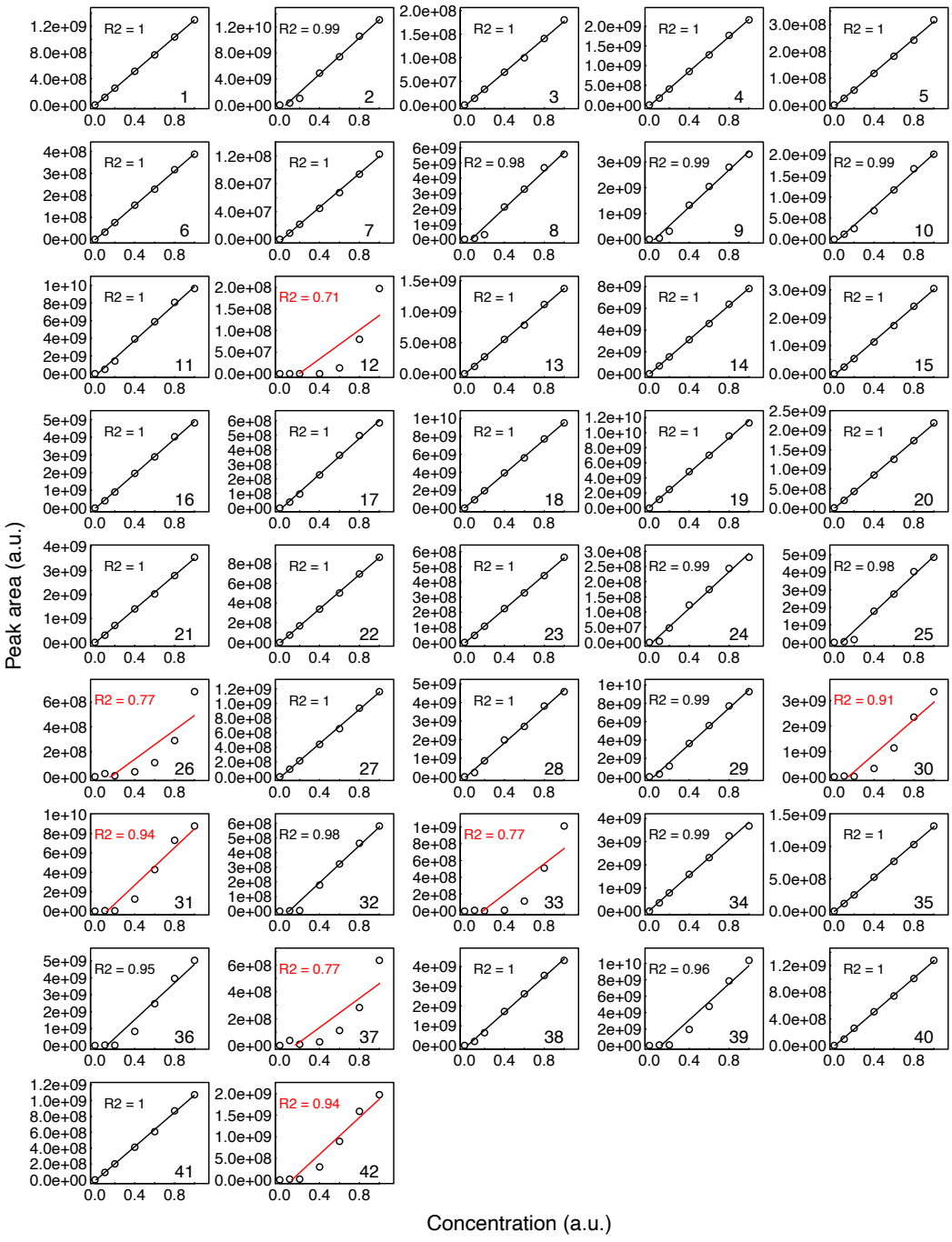

Figure S7. Peak areas of parent peptides at different concentrations. The concentration (in arbitrary units) is provided relative to the starting concentration of the incubation experiments. Lines represent curves of a linear fit. R^2 values (rounded to two decimals) of the linear fit are depicted next to the curves. For peptides that did not show linear relationship between peak area and concentration (based on our criteria: $R^2 > 0.95$), we depicted curves and letters in red. We note that this control experiment was conducted in a different laboratory and with a different peptide solution (i.e., a new digestion with the same reagents was conducted) than the remaining experiments of the study.

Table S3. Amino acid sequences (single-letter abbreviations) and high-performance liquid chromatography high-resolution mass spectrometry (HPLC-HRMS) acquisition parameters of product peptides that were detected for MC1.

Product sequence	Exact m/z	Mass deviation	Charge State	Retention time (min)	Fragment ions
ELLYANK	508.2687	-1.3	2	43	y6, y5, y4
LLYYANK	442.7475	0.6	2	37	y5, y3, y2
HPYFYAPEL	568.7742	0.8	2	63	y4, y3, y2
GSFLYEYSR	561.2667	-1.4	2	50	y6, y5, y4
SFLYEYSR	532.7560	-0.2	2	49	y6, y5, y4
FQNALIVR	480.7849	-1.3	2	46	y6, y5, y4
KVPQVS	329.2001	-1.3	2	22	y4, y3, y2
KVPQVSTPT	478.7742	-1.4	2	33	y7, y4, y3
VPQVS	265.1527	0.4	2	11	y4
VPQVSTPT	414.7267	1.4	2	38	y3, y2
VPQVSTPTL	471.2687	0.5	2	57	y3, y2
SLILNR	358.2267	-0.2	2	37	y4, y3
MPCTEDYLSLI	671.3069	0.3	2	82	y10, y3, y2
LTPDETYVPK	581.8032	-0.6	2	43	y8, y2
TPDETYVPK	525.2611	-0.7	2	35	y8, y7, y2
TYVPK	304.1761	-1.1	2	22	y4, y3, y2
YVPK	253.6523	0.8	2	16	y3, y2
AFVDK	290.1605	-0.2	2	21	y4, y3, y2
FVDK	254.6419	0.7	2	15	y3, y2
LKPDPNT	392.7136	0.3	2	21	y6, y5, y3
DAFL	465.2344	-0.8	1	55	y3, y2
DAFLG	522.2558	0.7	1	50	y3, y2
DAFLGS	609.2879	-1.1	1	51	y4, y3, y2
LGEY	481.2293	0.2	1	34	y3, y2
LVEVSR	702.4145	-0.6	1	26	y5, y3, y2
TVMENF	740.3284	-1.4	1	51	y4, y3, y2

Table S4. Amino acid sequences (single-letter abbreviations) and high-performance liquid chromatography high-resolution mass spectrometry (HPLC-HRMS) acquisition parameters of product peptides that were detected for MC2.

Product sequence	Exact m/z	Mass deviation	Charge State	Retention time (min)	Fragment ions
ELLYYANK	507.2687	1.6	2	44	y6, y5, y4
HPYFYAPELL	625.3162	1.3	2	74	y5, y4, y2
YYANK	329.6634	-0.8	2	16	y4, y3, y2
LIVR	250.6814	1.3	2	26	y3, y2
KVPQVS	329.2001	1.5	2	23	y4, y2
TPTLVEVSR	501.2849	-0.6	2	41	Y7, y6, y5
TLVEVSR	402.2347	1.6	2	34	y5, y4, y3
KVPQVSTPT	478.7742	1.2	2	35	y7, y3, y2
LVEVSR	351.7109	-0.4	2	27	y5, y4, y3
KVPQVSTPTL	535.3162	-0.8	2	52	y8, y4, y3
KVPQVSTPTLVE	649.3717	0.5	2	54	y10, y5, y2
LSLILNR	414.7687	0.9	2	54	y6, y5, y4
LILNR	314.7107	1.4	2	34	y4, y3, y2
MPCTEDYLSLI	671.3069	1.7	2	83	y10, y3, y2
ALTPDETYVPK	617.3217	1.8	2	47	y8, y2
RPCFSA	369.1736	0.0	2	31	y2
LTPDETYVPK	581.8032	1.9	2	43	y8, y2
RPCFSAL	425.7156	-0.8	2	48	y2
YVPK	253.6523	0.7	2	17	y3, y2
LPDTEK	351.6871	0.1	2	20	y5, y4, y3
LGEY	481.2293	-0.8	1	34	Y3, y2
LGEYG	538.2508	0.9	1	31	y4, y3, y2
LGEYGFQ	813.3777	-1.6	1	53	y4, y3, y2

Table S5. Amino acid sequences (single-letter abbreviations) and high-performance liquid chromatography high-resolution mass spectrometry (HPLC-HRMS) acquisition parameters of product peptides that were detected for MC3.

Product sequence	Exact m/z	Mass deviation	Charge State	Retention time (min)	Fragment ions
LKPDPNT	391.7134	-1.6	2	23	y6, y5, y3
LCDEFK	406.1864	-1.0	2	35	y5, y4, y3
LKPDPNTL	449.2556	1.3	2	39	y6, y4
CDEFK	349.6443	-1.1	2	22	y4, y3, y2
LKPDPNTLC	529.2710	1.1	2	38	y7, y4
LKPDPNTLCDE	651.3057	-0.6	2	42	y9, y3, y2
LLYYANK	442.7475	-1.5	2	38	y5, y4, y3
LYYANK	386.2054	-0.9	2	27	y5, y4, y3
VFQECCQAEDK	707.2923	-1.4	2	30	y9, y8, y7
LTADFAEDK	505.2455	-0.1	2	41	y8, y7, y6
LGEYGFQ	407.1925	-1.4	2	35	y6, y2
KVPQVS	329.2001	0.4	2	25	y5, y4, y3
VEVSR	295.1688	-0.5	2	18	y4, y3, y2
KVPQVSTPTLVE	649.3717	-1.4	2	55	y10, y5, y2
VPQVSTPT	414.7267	-0.5	2	38	y4, y3, y2
MPCTEDY	458.1648	1.3	2	41	y5, y2
LILNR	314.7107	0.8	2	35	y4, y3, y2
RPCFSA	369.1736	-1.2	2	33	y3, y2
LTPDETYVPK	581.8032	-0.1	2	43	y8, y2
YVPK	253.6523	0.4	2	19	y3, y2
FVDK	254.6419	0.1	2	18	y3, y2
HPY	416.1928	-1.3	1	21	y2

Table S6. Information on parent and product peptides for peptide bonds that were hydrolyzed by enzymes derived from two or more microbial communities.

P1P1'	Parent #	Parent sequence[1)]	Product sequence	Peak area (x 10^7 a.u.) of product peptide[2)]		
				MC1	MC2	MC3
AF	T39	TVMENFV**AF**VDK	FVDK	13.5 ± 1.5		12.9 ± 0.9
AL	T36	RPCFS**AL**TPDETYVPK	LTPDETYVPK	67 ± 4	80 ± 16	270 ± 30
			RPCFSA		13 ± 3	28 ± 2
	T30	LGEYGFQN**AL**IVR	LIVR		26 ± 4	
EL	T12	HPYFYAP**EL**LYYANK	LLYYANK	66 ± 6		51.0 ± 1.1
EV	T31	KVPQVSTPTLV**EV**SR	KVPQVSTPTLVE		6.8 ± 1.9	11 ± 2
GF	T30	LGEY**GF**QNALIVR	FQNALIVR	33 ± 5		
			LGEYG		18 ± 3	
IL	T33	MPCTEDYLSL**IL**NR	MPCTEDYLSLI	33 ± 5	12.8 ± 1.8	
LL	T12	HPYFYAPE**LL**YYANK	HPYFYAPEL	14.2 ± 1.2		
			LYYANK			29 ± 10
LT	T36	RPCFSA**LT**PDETYVPK	TPDETYVPK	173 ± 9		
			RPCFSAL	19.2 ± 1.7	11 ± 4	
LV	T31*	KVPQVSTPT**LV**EVSR*	KVPQVSTPTL	20 ± 4	27 ± 15	7 ± 2
			VEVSR			13.3 ± 1.5
			VPQVSTPTL	5.2 ± 0.8		
PE	T12	HPYFYA**PE**LLYYANK	ELLYYANK	9.35 ± 0.16	5.5 ± 0.8	
QN	T30	LGEYGF**QN**ALIVR	LGEYGFQ		6.9 ± 0.5	5.02 ± 0.19
SL	T33	MPCTEDYL**SL**ILNR	LILNR		27 ± 9	117 ± 5
ST	T31*	KVPQV**ST**PTLVEVSR*	KVPQVS	75 ± 2	19 ± 4	64 ± 9
			VPQVS	16.0 ± 1.4		
			TPTLVEVSR		70 ± 40	
TL	T31*	KVPQVSTP**TL**VEVSR*	KVPQVSTPT	180 ± 30	164 ± 2	
			VPQVSTPT	16.4 ± 1.4		28 ± 10
			LVEVSR		130 ± 30	
	T8**	LKPDPN**TL**CDEFK**	LKPDPNT	22.1 ± 1.5		51 ± 3
			LCDEFK			58 ± 7
	T37	LFTFHADIC**TL**PDTEK	LPDTEK		16.5 ± 1.0	
TY	T36	RPCFSALTPDE**TY**VPK	YVPK	2.8 ± 0.4	1.27 ± 0.16	4.3 ± 1.1
YG	T30	LGE**YG**FQNALIVR	GFQNALIVR	6.9 ± 0.3		
			LGEY	21.8 ± 1.0	4.2 ± 0.9	
YL	T33	MPCTED**YL**SLILNR	LSLILNR		5.9 ± 1.3	
			MPCTEDY			5.1 ± 1.0

[1)]Amino acid residues flanking the hydrolyzed peptide bond (i.e., P1P1') are shown in bold. [2)]Peak areas of product peptides at the end of the incubations (i.e., at 6 h, 3.8h, and 3.5 h for experiments with enzymes from MC1, MC2, and MC3, respectively) are provided as mean ± standard deviation of triplicates. *Some products might alternatively result from the hydrolysis of parent peptide T32. **Some products might alternatively result from the hydrolysis of parent peptide T9.

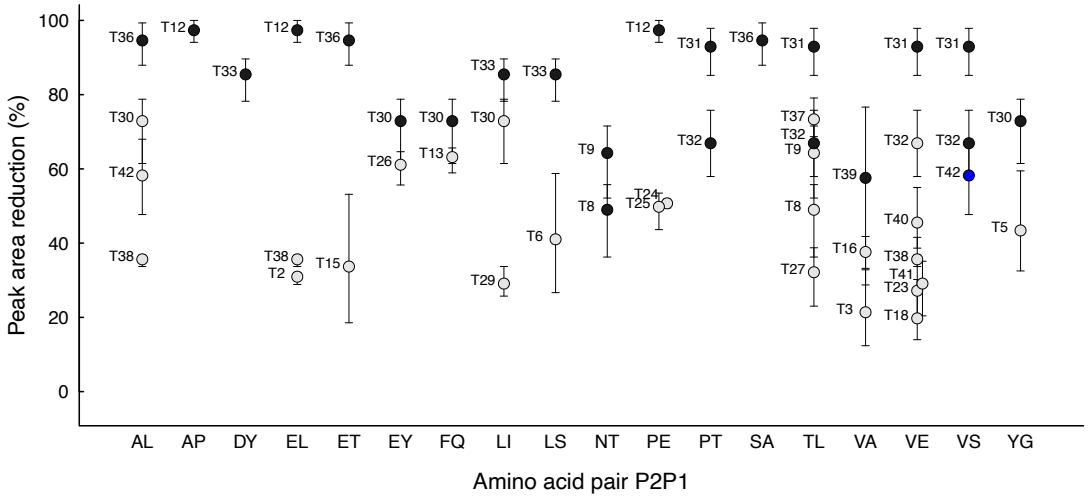

Figure S8. Peak area reduction of parent peptides that contain the amino-acid residue pairs that occurred at the P2P1 site of peptide bonds that were hydrolyzed by enzymes extracted from at least two wastewater microbial communities. Points and error bars represent means and ranges, respectively, of the peak area reduction measured for the three microbial communities. Parent peptide numbers are provided next to the respective points. Black points represent parent peptides for which we detected the hydrolysis of the respective peptide bond by enzymes extracted from at least two microbial communities (**Table S6**). An additional search for products of the hydrolysis of peptide bonds with EY, FQ, and VS at the P2P1 site (note: these amino-acid residue pairs were selected because they occurred in at least two parent peptides and all of the respective parent peptides showed a mean peak area reduction across the three microbial communities of >50%) resulted in the identification of the following products: (i) amino-acid sequence: TQTALA; exact mass: 604.3301; mass deviation: -1.1 ppm; charge state: +1; retention time: 26 min; detected fragments: y4, y2; peak area (x10E7) 7.2 ± 1.5 (MC1), (ii): amino-acid sequence: LVVS; exact mass: 417.2708; mass deviation: -1 ppm; charge state: +1; retention time: 31 min; detected fragments: y3, y2; peak area (x10E7) 9 ± 5 (MC2). These are products of the hydrolysis of the peptide bond in T42 that contains the amino-acid residue pair VS at the P2P1 site (indicated in blue).

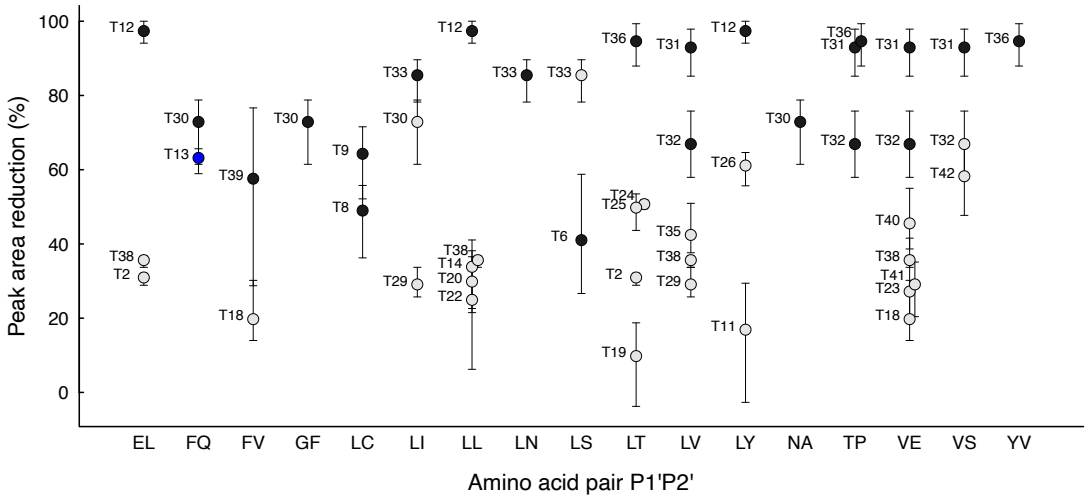

Figure S9. Peak area reduction of parent peptides that contain the amino-acid residue pairs that occurred at the P1'P2' site of peptide bonds that were hydrolyzed by enzymes extracted from at least two wastewater microbial communities. Points and error bars represent means and ranges, respectively, of the peak area reduction measured for the three microbial communities. Parent peptide numbers are provided next to the respective points. Black points represent parent peptides for which we detected the hydrolysis of the respective peptide bond by enzymes extracted from at least two microbial communities (**Table S6**). An additional search for products of the hydrolysis of peptide bonds with FQ and VS at the P1'P2' site (note: these amino-acid residue pairs were selected because they occurred in at least two parent peptides and all of the respective parent peptides showed a mean peak area reduction across the three microbial communities of >50%) resulted in the identification of the following product: amino-acid sequence: YNGV; exact mass: 452.2140; mass deviation: 1.1 ppm; charge state: +1; retention time: 25 min; detected fragments: y3, y2; peak area (x10E7) 0.90 ± 0.14 (MC1), 1.8 ± 0.4 (MC3). This is the product of the hydrolysis of the peptide bond in T13 that contains the amino-acid residue pair FQ at the P1'P2' site (indicated in blue).

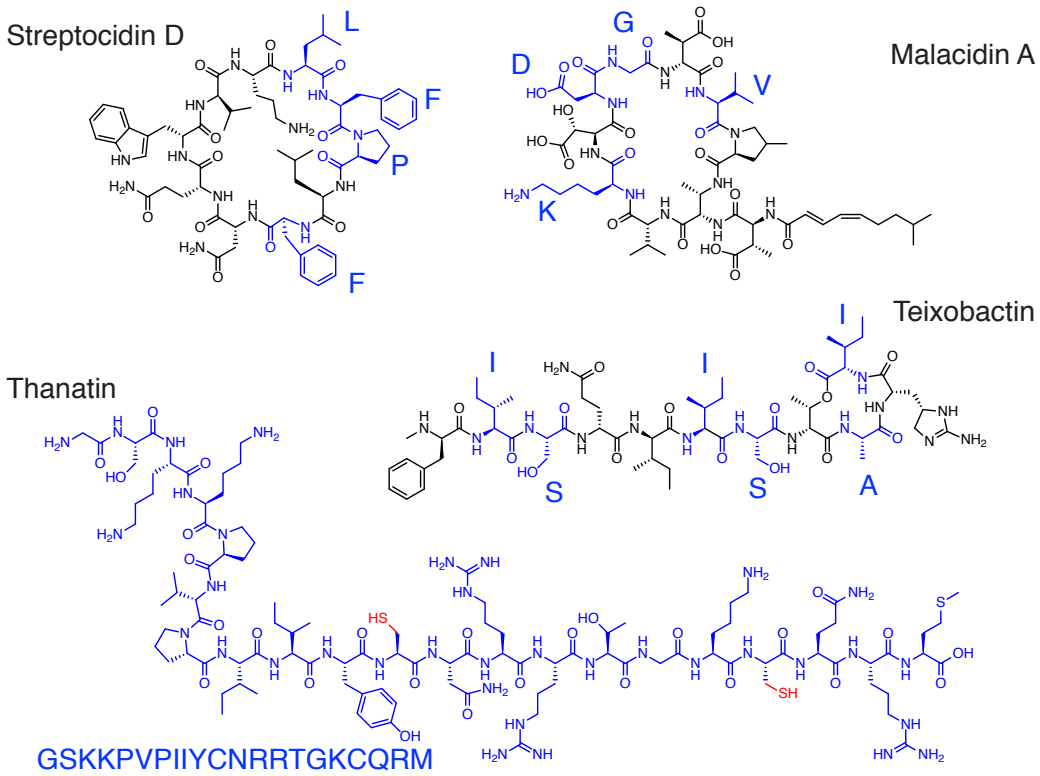

Figure S10. Names and structures of selected antimicrobial peptides Amino acid residues with L-stereochemistry are shown in blue and their single-letter abbreviation is provided next to the respective side chain. Thanatin entirely consists of amino acid residues with L-stereochemistry and its sequence is presented below the structure. Cysteine side chains that form a disulfide bond are shown in red.