Discovery of a new Pro-Pro endopeptidase, PPEP-2, provides mechanistic insights into the differences in substrate specificity within the PPEP family

Oleg I. Klychnikov[1], Tatiana M. Shamorkina[2], Stephen D. Weeks[1], Hans C. van Leeuwen[3], Jeroen Corver[3], Jan W. Drijfhout[4], Peter A. van Veelen[2], Nikolai N. Sluchanko[5,6], Sergei V. Strelkov[1], Paul J. Hensbergen[2]*

Two supporting tables: Table S1-2

Six supporting figures: Fig. S1-7

Table S1: List of PPEP-1 homologs from the Paenibacillus genus used for alignment (Fig. S1). All sequences were retrieved from the NCBI database and NCBI Reference / Genbank IDs are indicated. VMSP: **V**WFA, **M**ucbp, **S**urface-layer homology **P**rotein (see text for explanation). The sequences of OAB34809 and OAB31907 are identical and therefore combined in Fig. S1. From *Paenibacillus sp*. KS1 no full VMSP sequence could be retrieved. For *Paenibacillus terrigena* DSM 21567 genes were not automatically annotated. * PCR analysis followed by Sanger sequencing demonstrated that *P. alvei* DSM29 genes PAV_1c07810 (encoding EJW19794) and PAV_1c07820 (encoding EJW19795) form one single open reading frame encoding 1335 amino acids.

Species	**PPEP** (NCBI Reference / Genbank ID)	**VMSP** (NCBI Reference / Genbank ID)	**Contig** (NCBI Reference / Genbank ID)
***Paenibacillus alvei* A6-6i**	EPY14168	EPY11279/ EPY14167	ATMS01000012 ATMS01000059
***Paenibacillus alvei* TS-15**	EPY03961	EPY03960	NZ_ATMT01000102
***Paenibacillus alvei* DSM 29**	EJW19796	EJW19794/EJW19795*	AMBZ01000001
***Paenibacillus alvei* E194**	WP_044353226	WP_052648330	NZ_JMDX01000004
***Paenibacillus sp*. NAIST15-1**	GAV13913	GAV13912	BBYF01000009
***Paenibacillus sp*. cl6col**	SDF26896	SDF26925	FNAZ01000004
***Paenibacillus sp*. KS1**	OBY76659	x	MAIS01000128
***Paenibacillus sp*. UNC217MF**	WP_028533504	WP_028533505	NZ_KK366029
***Paenibacillus popilliae* ATCC 14706**	GAC42002	GAC42003	BALG01000064
***Paenibacillus dendritiformis* C454**	EHQ62738	EHQ62737	AHKH01000016
***Paenibacillus sp*. OSY-SE**	WP_026011046	WP_036704090	NZ_ALKF01000177
***Paenibacillus yonginensis* DCY84**	ANS76224	ANS76225	CP014167
***Paenibacillus macquariensis* subsp. defensor**	OAB34809	OAB34808	LVJG01000033
***Paenibacillus macquariensis* Macquariensis**	OAB31907	OAB31908	LVJF01000029
***Paenibacillus sp*. P1XP2**	KHF34792	KHF34791	JRNV01000019
***Paenibacillus phoceensis* (MT24)**	WP_068785363	WP_068784776	NZ_FCOQ01000018
***Paenibacillus sp*. GM2 (numidis)**	WP_068778763	WP_068778762	NZ_CTEK01000019
***Paenibacillus macerans* strain 8244**	KFN10106	KFN10123	JMQA01000020
***Paenibacillus antibioticophila* sp. nov. GD11**	WP_044478267	WP_044478266	NZ_HG005139
***Paenibacillus thiaminolyticus* strain NRRL B-4156**	WP_087440397	WP_087440396	NDGK01000005
***Paenibacillus apiarius* strain NRRL B-23460**	WP_087434869	WP_087434735	NDGJ01000039.1
***Paenibacillus terrigena* DSM 21567**	-	-	ARGP01000031

Table S2: Overall SAXS-derived hydrodynamic parameters of PPEP-1 and PPEP-2

	PPEP-1	**PPEP-2**
Monomeric Mw (calculated from amino acid sequence)	21.6 kDa	21.1 kDa
Number of residues	195	191
Rg (reciprocal space)	17.33±0.03 nm	17.61±0.05 nm
Rg (real space)	17.32±0.03 nm	17.63±0.04 nm
Porod Volume	30.7 nm^3	32.8 nm^3
Mw Porod	19.4 kDa	20.5 kDa
Dmax	5.95 nm	5.90 nm

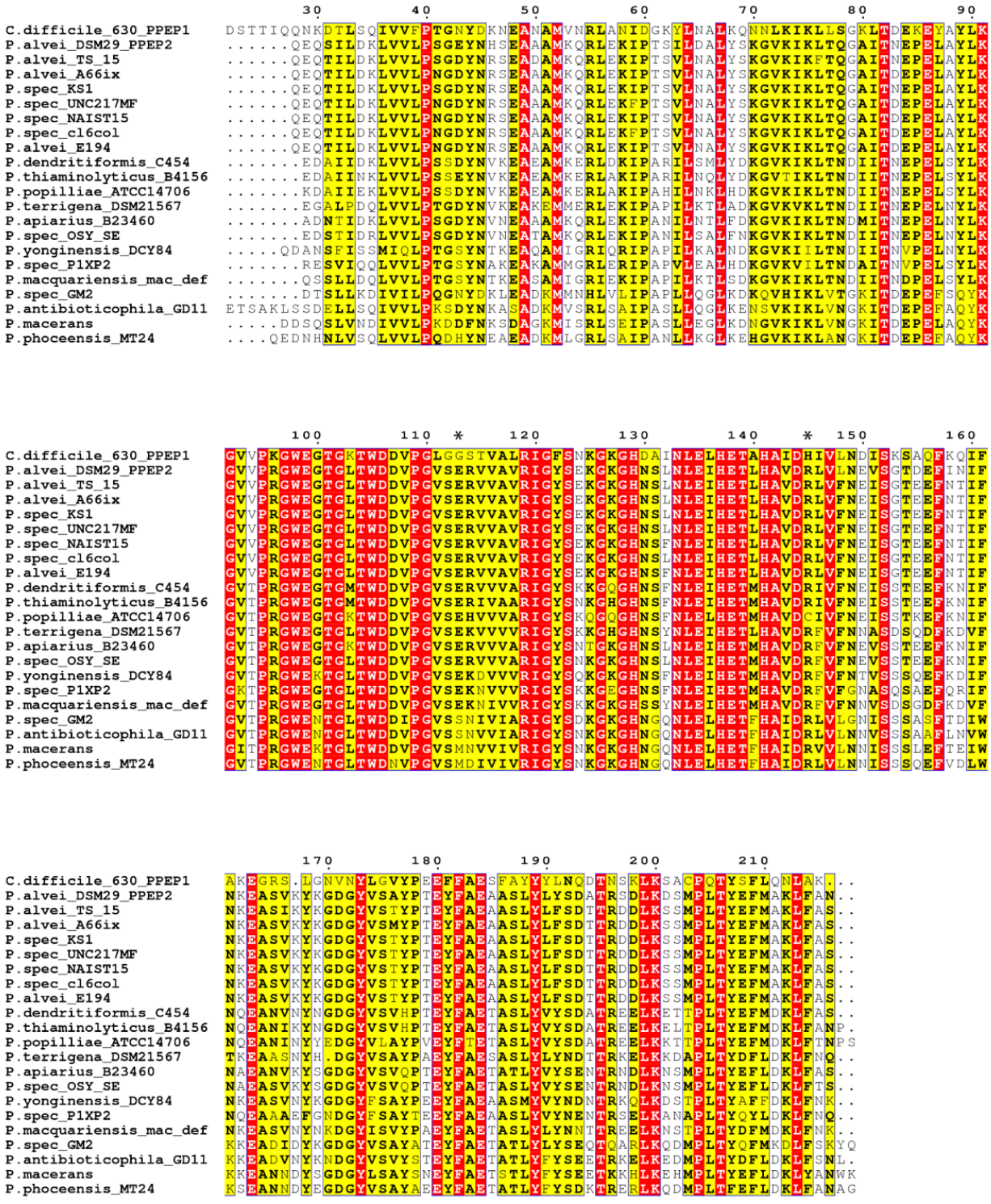


Figure S1: Sequence alignment of *C. difficile* PPEP-1 with its homologs from Paenibacillus.

The multiple alignment was performed using ClustalW Omega, and the final output was processed using the program ESPript 3.0. Signal sequences as predicted by SignalP4 were removed from all proteins. The numbering refers to PPEP-2 from *P. alvei* DSM29. Identical residues are in white letters with red background, similar residues are in black letters with yellow background, varied residues are in black letters, and dots represent gaps. Designations: P, Paenibacillus; C, Clostridium. *: residues which form a salt bridge in PPEP-2 (see text regarding Figure 5b for further details).

A

Species	PPEP cleavage sites	Amino acids
P.alvei_TS-15	SSSYYPSSKPLPPVPPLPPVP-KLETQYHYNYINGYPDG	1226 - 1263
P.alvei_A66ix	STSYYPSSKPLPPVPPLPPVP-KLETQYHYNYINGYPDG	614 - 651
P.spec_UNC217MF	STSYYPSSKPLPPVPPLPPVP-KLETQYHYNYINGYPDG	1287 - 1324
P.spec_NAIST15_1	STSYYPSSKPLPPVPPLPPVP-KLETQYHYNYINGYPDG	1218 - 1255
P.spec_cl6col	STSYYPSSKPLPPVPPLPPVP-KLETQYHYNYINGYPDG	1220 - 1257
P.alvei_E194	SSKPLPPVPPLPPVPPLPPVP-KLETQYHYNYINGYPDG	1159 - 1196
P.alvei_DSM29	SSK---PLPPVPPVQPLPPVP-KLETSYHYNYINGYPDG	1101 - 1135
P.dendritiformis_C454	SYTITPSPKPLPPVPPLPPVPPKLDMENHYNYINGYPDG	461 - 499
P.thiaminolyticus_B4156	TTTITPSPKPLPPVPPLPPVPPKLDMENHYNYINGYPDG	1447 - 1486
P.popilliae_ATCC14706	SSTITPSPKPLL---PLPPVPPKLDMENHSNYINGYPDG	855 - 890
P.terrigena_DSM21567	STPTPLPPVPLPPVPPLPPVPPQLEKDKHFNYINGYPDG	407 - 446
P.apiarius NRRL B-23460	TTNPSNPANPLPPVPPLPPVPPKLETDNHYNYINGYPDG	1437 - 1476
P.spec_OSY_SE	TTNPTNPVNPLPPVPPLPPVPPKLEMDNHYNYINGYPDG	342 - 380
P.yonginensis_DCY84	DNNVTPLPPAPPVT-ALPPAPPTLDKENHYNYIFGYPDG	1096 - 1133
P.spec_P1XP2	SGSVTPLPPAPPVVPPLPPAPPKLDTANHFNYINGYPDG	1096 - 1133
P.macquariensis_mac/def	TSNPTPLPPAPPVVAPLPPPPPKLEAENHYNYINGYPDG	1297 - 1335
P.spec_GM2	SSYTTPSPPPAPPVPPLPPVPPKLDTENHFDYIQGYPDG	1073 - 1111
P.antibioticophila_GD11	PSSSVTPLPAVPPLPPLPAVPPVLDRENHYDYINGYPDG	290 - 328
P.macerans	PSPSTPEVDPLPPLPPLPAVPPKLETENHYDYINGYPDG	1100 - 1138
P.phoceensis_MT24	GNPTPPSTPEVTPLPPLPAVPPTLDTENHFQYIHGYPDG	1500 - 1538

B

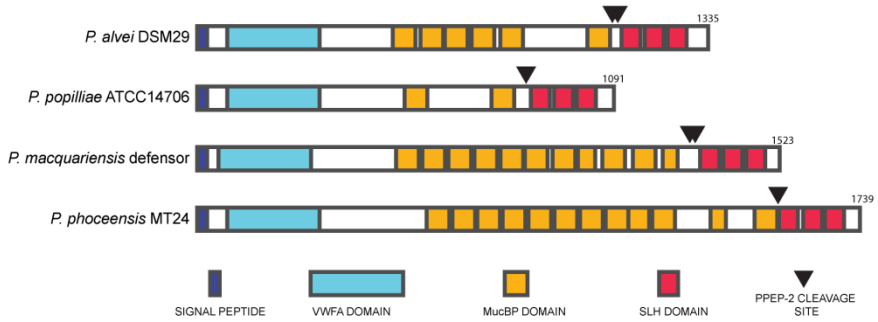
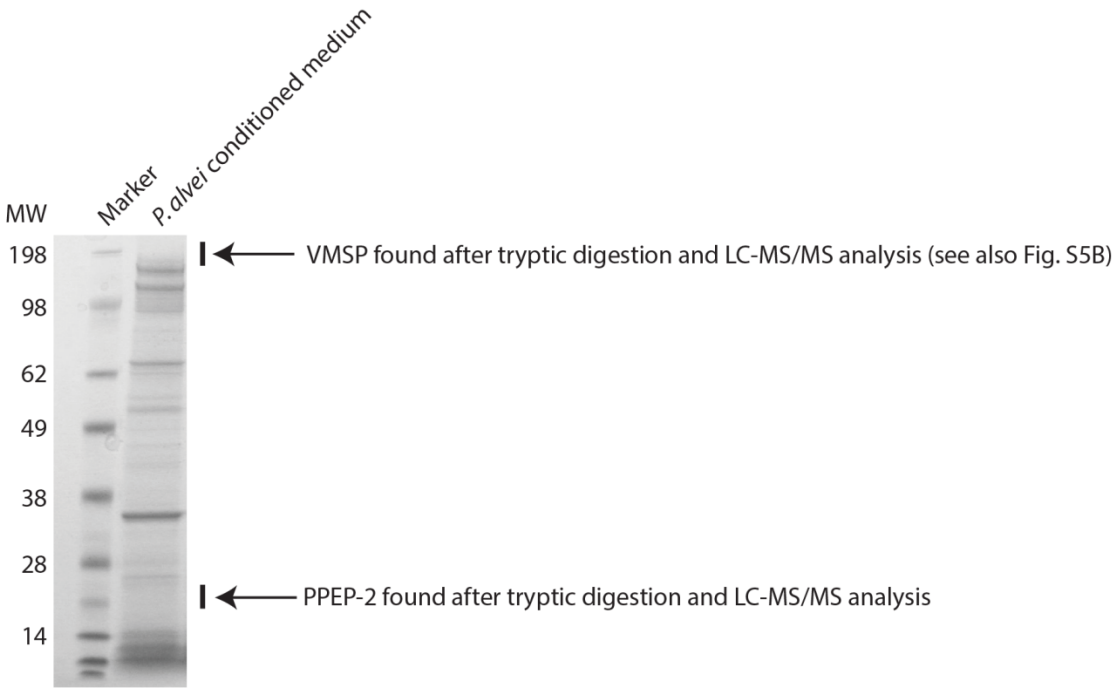


Figure S2: Bioinformatics identification of putative PPEP-2 cleavage sites in proteins encoded by adjacent open reading frames in Paenibacillus genomes

a) Paenibacillus species are indicated with the (putative) PPEP-2 cleavage sites (PLPPXP) highlighted in red, yellow and green. Position of protein region shown is indicated on the right

b) Schematic representation of the modular organisation of the PPEP-2 substrates VMSP (**V**WFA, **M**ucbp, **S**urface-layer homology **P**rotein), encoded by the gene adjacent to *ppep-2*. Black triangles indicate the (putative) PPEP-2 cleavage site (see also A). Domains are illustrated in different colours: Blue, VWFA = Von Willebrand factor type A domain (IPR002035); Orange, MucBP = Mucin-Binding Protein domain (IPR009459); Red, SLH = S-layer homology domain (IPR001119). Functional domain analysis was performed using InterPro [http://www.ebi.ac.uk/interpro/].



PPEP-2 identification by LC-MS/MS analysis of tryptic digests and Mascot data base searching

WP_005543643.1 Mass: 24056 Score: 579 Matches: 21(20) Sequences: 14(14) emPAI: 21.41
hypothetical protein [Paenibacillus alvei]

	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
184	348.6822	695.3498	695.3490	1.21	0	38	0.0046	1	U	R.IGYSEK.G
1126	480.7530	959.4914	959.4924	-1.03	0	40	0.0053	1	U	A.QEQSILDK.L
2971	660.8760	1319.7375	1319.7336	2.94	0	108	3e-10	1	U	K.IPTSILDALYSK.G
3437	716.8299	1431.6452	1431.6414	2.69	0	(42)	0.0016	1	U	K.DSMPLTYEFMAK.L
3520	724.8257	1447.6368	1447.6363	0.34	0	58	3.2e-05	1	U	K.DSMPLTYEFMAK.L
4499	564.3273	1689.9600	1689.9552	2.80	1	48	0.0002	1	U	R.LEKIPTSILDALYSK.G
4724	880.9757	1759.9369	1759.9356	0.73	0	71	2.3e-06	1	U	K.LTQGAITNEPELAYLK.G
5012	938.4523	1874.8900	1874.8794	5.65	1	66	9.1e-06	1	U	R.SDLKDSMPLTYEFMAK.L
5093	636.6329	1906.8770	1906.8692	4.05	1	(39)	0.0041	1	U	R.SDLKDSMPLTYEFMAK.L
5114	639.3262	1914.9567	1914.9509	3.01	0	67	7.9e-06	1	U	K.LVVLPSGEYNHSEAAAMK.Q
5171	644.6575	1930.9508	1930.9458	2.55	0	(47)	0.00099	1	U	K.LVVLPSGEYNHSEAAAMK.Q
5247	980.9595	1959.9044	1959.8963	4.13	0	57	5.3e-05	1	U	R.GWEGTGLTWDDVPGVSER.V
5365	668.0445	2001.1116	2001.1146	-1.48	1	32	0.009	1	U	K.IKLTQGAITNEPELAYLK.G
5496	514.5168	2054.0380	2054.0293	4.23	0	74	1.6e-06	1	U	K.GHNSLNLEIHETLHAVDR.L
5970	757.0913	2268.2521	2268.2478	1.92	1	53	7.9e-05	1	U	K.LTQGAITNEPELAYLKGVVPR.G
6231	829.0903	2484.2492	2484.2398	3.78	1	76	9.6e-07	1	U	R.GWEGTGLTWDDVPGVSERVVAVR.I
6496	947.4776	2839.4110	2839.4062	1.69	1	95	1.1e-08	1	U	A.QEQSILDKLVVLPSGEYNHSEAAAMK.Q
6505	952.8104	2855.4093	2855.4011	2.86	1	(56)	8.3e-05	1	U	A.QEQSILDKLVVLPSGEYNHSEAAAMK.Q
6507	953.1530	2856.4370	2856.4327	1.51	1	(32)	0.02	1	U	A.QEQSILDKLVVLPSGEYNHSEAAAMK.Q
6508	715.1182	2856.4438	2856.4327	3.89	1	(49)	0.00036	1	U	A.QEQSILDKLVVLPSGEYNHSEAAAMK.Q
6512	719.1173	2872.4399	2872.4276	4.27	1	(43)	0.0013	1	U	A.QEQSILDKLVVLPSGEYNHSEAAAMK.Q

Figure S3: Identification of endogenous PPEP-2 and its substrate VMSP in conditioned medium of *P. alvei* cell cultures.

Conditioned medium from *P. alvei* cell cultures was concentrated and analyzed by SDS-PAGE (upper panel). In-gel tryptic digestion followed by LC-MS/MS analyses and database searching identified PPEP-2 (lower panel) and VMSP (Fig. S5).

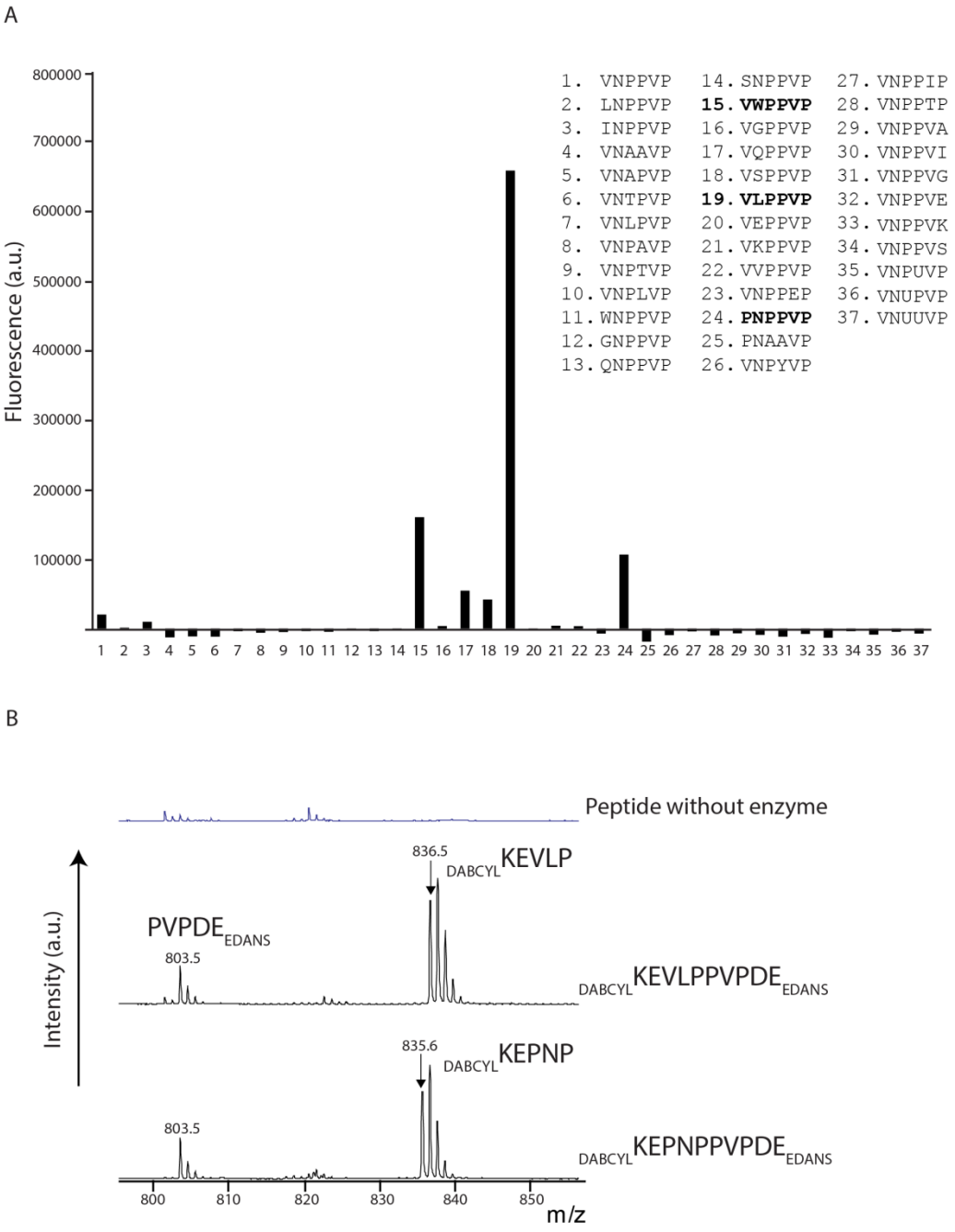
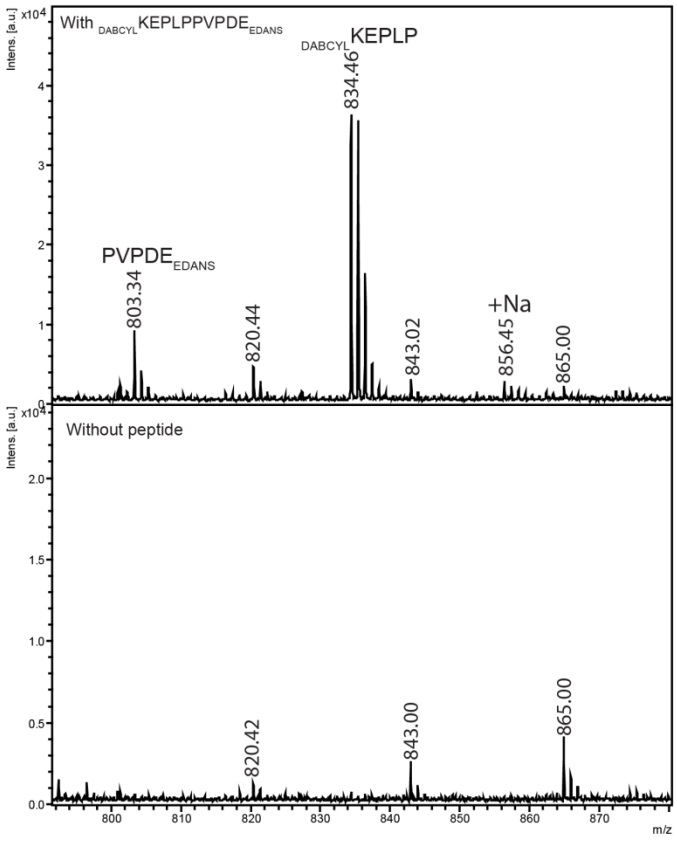


Figure S4: Screening of a FRET substrate peptide library with PPEP-2 confirms the preference for a proline at the P3 and leucine at the P2 position of the cleavage site.

a) A set of FRET-peptides which had previously been used to identify the optimal substrate peptide for PPEP-1 was tested with recombinant PPEP-2. The fluorescence increase after 1 h of incubation of a substrate peptide (25 µM) with 1 µg of PPEP-2 is shown. The core sequence of the peptides (P3-P3') is indicated. In bold, the three best cleaved peptides are indicated. U: hydroxyproline.

b) MALDI-ToF MS analysis of the cleavage products of the FRET-peptides containing the VLPPVP and PNPPVP motif after incubation with recombinant PPEP-2.

A



B

*MKRRVKLAVTWLCVFAVFFGFIPHTLQSADA*AAQDPIWPEPGAVKLSKTAKPTGTPGEW
QITLTAEGKNVEAGDTDVVLVLDKSDSMRSYKRLDNAKKAAEKFVNTLLFSNVKTRIAL
VTFSTSSEDTSSGFLGQNDKQQLLNNIKGISASGGTNIQSALYTAHRLLATKSSAKNKF
IVLLSDGAPTFSSPAKKAASYSWPNNKYNFVLSDFDYTKTIGSGNSYQLSSGLFGDRYT
IGNYTVKDNGIPTISQAKASKDAGINIYSIGLEVGNDANAVYVLKNSASANNYYPSDSA
ELEKVFSELAGKISYAAENAVVKDPMGEMFNLIFKGNAPVENQDYKVSQGTVKWDAASE
TFVWDVGNIVEGSPATLTYTIKMDSSKNPSADVNYPTNGRTTMEYTDVNQQRTSKEFEV
PQVSFGKGSILVKGYKVNDQGEPINADGKVVERVDLAEQLYSGSFTVNGKDALDIGPKY
KVPAPKLDNYVLKKGTDPTEETLTIQEPNKTIWFGYALAPHKLTIKYVDRETGKSIAND
DSSLSGTIGQTITLKALDIPGYTPEKTEVQYTFKAKDNEYTFFYKKMDLPITIKFVDKD
TSKSIADDDTSHSGKVGEAVTLKALDIPGYTPEKAEITYTVKATDNVVTFFYTANPQKI
TIHYIDQDTKLPIAKDSSKDGKTGESVTLTALEITGYTSVNPEVKYTFTAEANQEHTFF
YTKNAPVERTVHVHYKEADTGAVLKESTSTSGKVGETVTLKAEDITINGVVYVPNAYEH
EYTITDKEEQSYTFIYTKKDVIPQERTVHVHYKDAATGTVLKESTSTSGKVGATITLRA
EDITVDGVIYVPASYQHEYAVTEQTEQSYTFFYSKKELPTYFVTVHHLEEGTDAVLHEP
TTVSGKSGEVVHVTAEPISVADAVYQPIKFNHDITIEGTPEQLYTIYYKKGEPVEPLLL
TVLHLDKETLAPLAEPTSLKGKAGEHITLKPTPITVTDAVYNPEHAAYDYVFTGEATQT
FSILYIKNPITHPDQHVTIKYVEQGTGKQLVNPTTKNGKAGEKVELTALSISGYTPVKS
TDTYTFTDKEGQEYIFYYTRNSSGGGSSSGGSSSYYPSSK**PLPPVP**PVQ**PLPPVP**KLET
SYHYNYINGYPDGKIKPENLISREEVAVIFYRLMENSTRANFLKGTNSYKDIAEKRWSN
RHISTMENAGIIKGYPDGSFLPERPITRAEFAAIASRFDKLNEQPNTMFSDISGHWAEK
YIVSAANKGWIKGYKNGTFKPNQYITRAEAMAFINSVLNRKVHNDEIHKDAKKWPDITP
TNWYYSDVMEATNYHDYHRIKDNFESWDKVNADVVYP

Figure S5: Identification of endogenous PPEP-2 activity in conditioned medium of *P. alvei* cell cultures

a) Conditioned medium from a *P. alvei* cell culture was concentrated and incubated with a FRET substrate peptide containing the sequence PLPPVP. Following a 1 h incubation, MALDI-ToF MS was used to identify the specific product peptides as a result of PPEP-2 cleavage.

b) Conditioned medium from *P. alvei* cell cultures was concentrated and analyzed by SDS-PAGE (see Fig. S2). In-gel tryptic digestion followed by LC-MS/MS analyses and database searching identified three peptides (indicated in red) from *P. alvei* DSM29 VMSP. The PPEP-2 cleavage sites are indicated in bold and the putative signal peptide in italic.

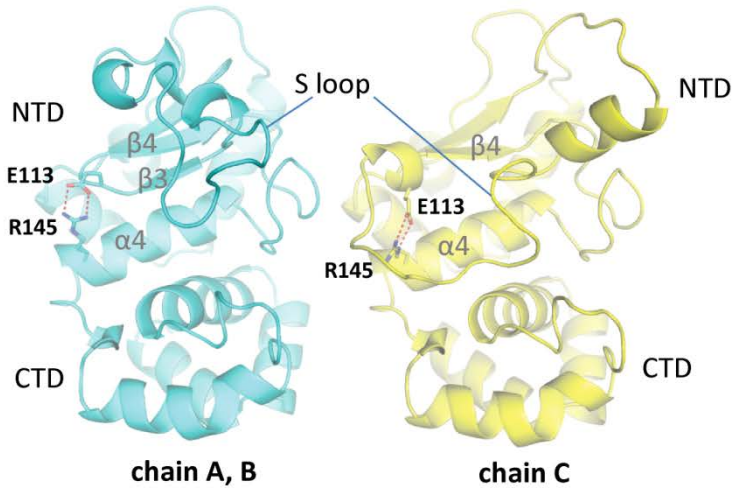


Figure S6: Positions of the S-loop in different chains of the PPEP-2 crystal

Different chains of the PPEP-2 crystal structure show different positions of the S-loop. The chains A through C are presented whereas chain D is not shown because of the uncertain position of the S-loop. Note that in all chains the salt bridge between Glu113 and Arg145 (shown in red dashed lines) is preserved.

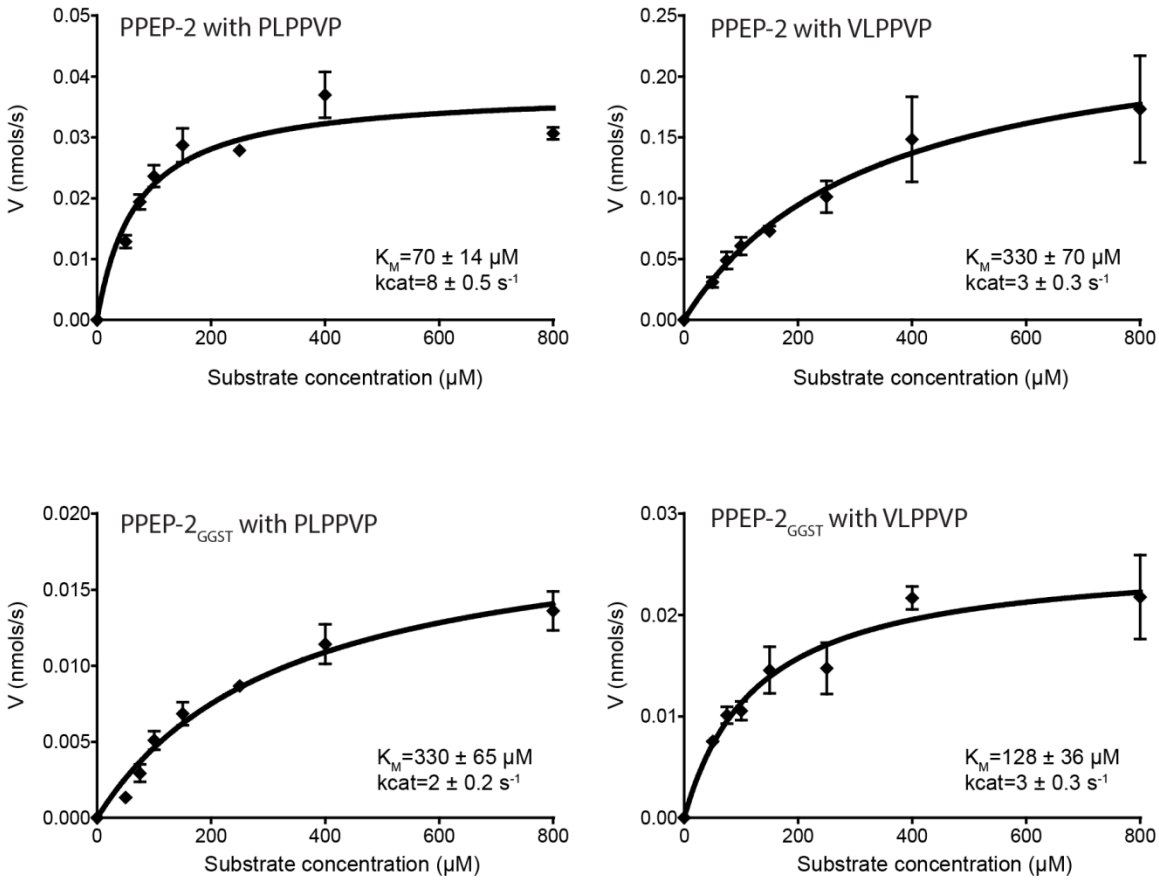


Figure S7: The specificity of PPEP-2 can be modified towards that of PPEP-1

A recombinant modified PPEP-2 was produced where the β3/β4 loop from PPEP-2 (SERV) was replaced by that from PPEP-1 (GGST). Shown are the Michaelis-Menten plots for PPEP-2 and PPEP-2_{GGST} with FRET substrate peptides containing either PLPPVP or VLPPVP. PPEP-2 (100 ng for the experiments with the PLPPVP peptide and 2 μg for the experiments with the VLPPVP peptide) and PPEP-2_{GGST} (200 ng for both peptides) were incubated with different concentrations of the substrate peptides. See Material and Methods for further details. A minimum of three experiments were used for the determination of the Michaelis-Menten constants.