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

Oleg I. Klychnikov¹, Tatiana M. Shamorkina², Stephen D. Weeks¹, Hans C. van Leeuwen³, Jeroen Corver³, Jan W. Drijfhout⁴, Peter A. van Veelen², Nikolai N. Sluchanko^{5,6}, Sergei V. Strelkov¹, Paul J. Hensbergen²*

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: VWFA, Mucbp, Surface-layer homology Protein (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	VMSP (NCBI Reference	Contig (NCBI Reference	
•	/ Genbank ID)	/ Genbank ID)	/ Genbank ID)	
Paenibacillus alvei A6-6i	EPY14168	EPY11279/	ATMS01000012	
		EPY14167	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	Х	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
	195	171
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

	30	40	5 <u>0</u>	e 0	70	80	90
C.difficile_630_PPEP1	DSTTIQQNK <mark>DTlsq</mark>	IVVF PTGNY	KN EA N AM VN	RL ANIDGK <mark>YL</mark> NAL	QNN LKIKL L	S G K L T D E K E Y A	YLK
P.alvei_DSM29_PPEP2	QEQ <mark>SIL</mark> DK	LVVL <mark>P</mark> SGEYI	NHS <mark>EA</mark> A <mark>AM</mark> KQI	RLEKIPTS <mark>IL</mark> DA <mark>L</mark> M	S <mark>kgvkikltq</mark>	d G A I T N E P E L A	YLK
P.alvei_TS_15	QEQ <mark>TIL</mark> DK	LVVL <mark>P</mark> SGDY1	NRSEADAMKQI	RLEKIPTS <mark>VL</mark> NAL	SKGVKIKFT	d G A I T N E P E L A	YLK
P.alvei_A66ix	QEQ <mark>TIL</mark> DK	LVVL <mark>P</mark> NGDY1	NRSEAAAMKQI	RLEKIPTS <mark>VL</mark> NA <mark>L</mark> 3	S <mark>kgvkiklt(</mark>	dait d epel a	YLK
P.spec_KS1	QEQ <mark>TIL</mark> DK	LVVL <mark>P</mark> SGDY1	NRSEAAAMKQI	RLEKIPTS <mark>VL</mark> NA <mark>L</mark> 3	S <mark>kgvkiklt(</mark>	d G A I T N E P E L A	' <mark>Y T</mark> K
P.spec_UNC217MF	QEQ <mark>TIL</mark> DK	LVVL <mark>P</mark> NGDY1	NRSEAAAMKQI	RLEKFPTS <mark>VL</mark> NA <mark>L</mark> Y	(S <mark>kgvkiklt(</mark>	d G <mark>AIT</mark> D EPEL A	' <mark>Y L</mark> K
P.spec_NAIST15	QEQ <mark>TIL</mark> DK	LVVL <mark>P</mark> NGDY1	NRSEAAAMKQI	RLEKIPTS <mark>VL</mark> NAL	S <mark>kgvkiklt(</mark>	d G AIT D EPEL A	. YLK
P.spec_cl6col	QEQ <mark>TIL</mark> DK	LVVL <mark>P</mark> NGDY1	NRSEAAAMKQI	RLEKFPTS <mark>VL</mark> NAL	S <mark>kgvkiklt(</mark>	d G AIT D EPEL A	Y L K
P.alvei_E194	QEQ <mark>TIL</mark> DK	LVVL <mark>P</mark> NGDY1	NRSEAAAMKQI	RLEKIPTS <mark>VL</mark> NA <mark>L</mark> 3	S <mark>KGVKIKLT(</mark>	d G AIT D EPEL A	' <mark>Y T</mark> K
P.dendritiformis_C454	ED <mark>AII</mark> DK	LVVL <mark>P</mark> SSDY1	NVK EA E AM KEI	RLDKIPAR <mark>IL</mark> SML	O <mark>KGVKIKLTI</mark>	MD IIT N EPEL S	YLK
P.thiaminolyticus_B4156	ED <mark>AII</mark> NK	LVVL <mark>P</mark> SSEY1	NVK <mark>EA</mark> E <mark>AM</mark> KEI	RLAKIPAR <mark>IL</mark> NQ <mark>L</mark> M	(D <mark>kgvtiklti</mark>	MDIITNEPELS	YLK
P.popilliae_ATCC14706	KD <mark>aii</mark> ek	LVVL <mark>P</mark> SSDY1	NVK EA E AM KEI	rl<mark>akip</mark>ah<mark>il</mark>nk<mark>l</mark>h	ID <mark>kgvkiklti</mark>	MDIITNEPELS	YLK
P.terrigena_DSM21567	EG <mark>Al</mark> PDQ	LVVL <mark>P</mark> TGDY1	NVK EA K <mark>EM</mark> MEI	RLE <mark>KIP</mark> AP <mark>IL</mark> KTLA	AD <mark>KGVKVKLT1</mark>	M D I I T N E P E L N	I <mark>Y L</mark> K
P.apiarius_B23460	AD <mark>NTI</mark> DK	LVVL <mark>P</mark> SGDY1	NVN EA A <mark>AM</mark> KQI	RLEKIPAN <mark>IL</mark> NTL	'D <mark>kgvkiklti</mark>	M D M I T N E P E L N	I <mark>Y L</mark> K
P.spec_OSY_SE	ED <mark>STI</mark> DR	LVVLPSGDY	NVN EA T AM KQI	RLEKIPAN <mark>IL</mark> SALF	'N <mark>KGVKIKLTI</mark>	M D I I T N E P E L N	I Y L K
P.yonginensis_DCY84	QDAN <mark>SFI</mark> SS	MIQL <mark>P</mark> TGSY1	NTKEAQAMIGI	RIQRIPAP <mark>II</mark> KA <mark>I</mark> M	ID <mark>KGVKIILTI</mark>	M D I I T N V P E L N	I Y L K
P.spec_P1XP2	RE <mark>SVI</mark> QQ	LVVLPTGSY1	NAKEAKAMMGI	RLERIPAP <mark>VL</mark> EAL	ID <mark>kgvkvilt</mark> i	MDAITNVPELS	YLK
P.macquariensis_mac_def	QS <mark>SLL</mark> DQ	LVVLPTGSY1	NTKEAS <mark>AM</mark> IGI	RIE <mark>KIP</mark> AP <mark>IL</mark> KT <mark>L</mark> S	D KGVKIKLTI	MDIITNDPELS	YLK
P.spec_GM2	DT <mark>SLL</mark> KD	IVILPQGNY	OKL <mark>EA</mark> D <mark>KM</mark> MN	HLVLIPAP <mark>LL</mark> QGL	(D <mark>KÕAHIKT</mark> A)	r g k i t d e p e f s	QYK
P.antibioticophila_GD11	ETSAKLSSD <mark>Ell</mark> SQ	IVVL <mark>P</mark> KSDY1	NKA <mark>SA</mark> DKMVSI	RL SAIPAS <mark>LI</mark> QG <mark>I</mark> F	(E <mark>nsvkiklvi</mark>	NGKITDEPEFA	' <mark>Q X K</mark>
P.macerans	DDSQ <mark>SLV</mark> ND	IVVL <mark>P</mark> KD DF I	NKSDAG <mark>KM</mark> ISI	RL SEIPAS <mark>LL</mark> EG <mark>L</mark> F	ID <mark>NGVKIKL</mark> VI	M G K I T D E P E L A	. <mark>Q Y K</mark>
P.phoceensis_MT24	QEDNH <mark>NLV</mark> SQ	LVVLPQDHYI	NEAEADKMLG	RLSAIPANLIKGI	E HGVKIKLAN	JGKITDEPEFA	. <mark>QY</mark> K

		100	110	*	120	130	140	*	150		160
C.difficile_630_PPEP1	GVV	P K G W E G T G K 7	WDDVPGI	LGGS	TVALRIG <mark>F</mark> SI	NKGKGHDAI	ILE <mark>LHET</mark> AHAI	DHIV	LNDI	SKSAQE	KQIF
P.alvei_DSM29_PPEP2	GVV	P <mark>r</mark> gwe <mark>g</mark> tg <mark>l</mark> j	WD <mark>DV</mark> PG <mark>V</mark>	SER	VVAV <mark>RIG</mark> YSI	E <mark>KGKGHNS</mark> L	ILE <mark>I</mark> HET <mark>L</mark> HA <mark>V</mark>	DRL	/LNEV	SGTDEF	INIF
P.alvei_TS_15	GVV	P <mark>R</mark> GWE <mark>G</mark> TG <mark>L</mark>]	WD <mark>DV</mark> PG <mark>V</mark>	SER	VVAV <mark>RIG</mark> YSI	E <mark>KGKGHNS</mark> LI	ILE <mark>IHET<mark>L</mark>HA<mark>V</mark></mark>		FNEI	SGTEEF	NTIF
P.alvei A66ix	GVV	P <mark>r</mark> gwe <mark>g</mark> tg <mark>l</mark> j	WDDVPG	SER	VVAVRIG <mark>y</mark> si	E <mark>KGKGHNS</mark> L	ILE <mark>IHET</mark> LHA <mark>V</mark>		FNEI	SGTEEF	NTIF
P.spec_KS1	GVV	P <mark>R</mark> GWE <mark>G</mark> TG <mark>L</mark>]	WD <mark>DV</mark> PG <mark>V</mark>	/SER	<mark>VVAV</mark> RIG <mark>Y</mark> SI	E <mark>KGKGHNS</mark> LI	ILE <mark>I</mark> HET <mark>L</mark> HAV	DRL	FNEI	SGTEEF	NTIF
P.spec_UNC217MF	GVV	P <mark>R</mark> GWE <mark>G</mark> TG <mark>L</mark>]	WD <mark>DV</mark> PG <mark>V</mark>	SER	VVAV <mark>RIG</mark> YSI	E <mark>KGKGHNS</mark> LI	ILE <mark>I</mark> HET <mark>L</mark> HAV	DRL	FNEI	SGTEEF	NTIF
P.spec_NAIST15	GVV	P <mark>R</mark> GWE <mark>G</mark> TG <mark>L</mark>]	WD <mark>DV</mark> PG <mark>V</mark>	SER	VVAV <mark>RIG</mark> YS	E <mark>KGKGHNS</mark> FI	ILE <mark>IHETL</mark> HAV		FNEI	SGTEEE	NTIF
P.spec_cl6col	GVV	P <mark>r</mark> gwe <mark>g</mark> tg <mark>l</mark> j	WD <mark>DV</mark> PG <mark>V</mark>	SER	VVAV <mark>RIG</mark> YSI	E <mark>KGKGHNS</mark> L	ILE <mark>IHETL</mark> HA <mark>V</mark>		FNEI	SGTEEF	NTIF
P.alvei_E194	GVV	P <mark>R</mark> GWE <mark>G</mark> TG <mark>L</mark>]	WD <mark>DV</mark> PG <mark>V</mark>	SER	VVAV <mark>RIG</mark> YSI	E <mark>KGKGHNS</mark> F1	ILE <mark>IHETL</mark> HA <mark>V</mark>		FNEI	SGTEEF	NTIF
P.dendritiformis_C454	GVΤ	P <mark>r</mark> gwe <mark>g</mark> tg <mark>m</mark> 1	WDDVPG	SER	VVVARIG <mark>y</mark> si	K <mark>kgoghns</mark> f1	ILE <mark>IHET</mark> MHA <mark>V</mark>		FNEI	SSTEEF	KNIF
P.thiaminolyticus_B4156	GVΤ	P <mark>r</mark> gwe <mark>g</mark> tg <mark>m</mark> 1	WD <mark>DV</mark> PG <mark>V</mark>	JSER	IVAA <mark>RIG</mark> YSI	N <mark>KGHGHNS</mark> FI	ILE <mark>I</mark> HET <mark>M</mark> HAV		FNEI	SSTEEF	KNIF
P.popilliae_ATCC14706	GVΤ	P <mark>r</mark> gwe <mark>g</mark> tg <mark>k</mark> j	WD <mark>DV</mark> PG <mark>V</mark>	/SEH	<mark>VVVA</mark> RIG <mark>Y</mark> SI	K <mark>qgqghns</mark> f1	ILE <mark>L</mark> HET <mark>M</mark> HAV		FNEI	SSTKEF	KNIF
P.terrigena_DSM21567	GVΤ	P <mark>R</mark> GWE <mark>G</mark> TG <mark>L</mark>]	WD <mark>DV</mark> PG <mark>V</mark>	JSEK	<mark>VVVV</mark> RIG <mark>Y</mark> S]	K <mark>KGHGHNS</mark> Y	ILE <mark>IHETL</mark> HAV	DRF V	FNNA	SDSQDE	KDVF
P.apiarius_B23460	GVΤ	P <mark>R</mark> GWE <mark>G</mark> TG <mark>K</mark> 7	WD <mark>DV</mark> PG <mark>V</mark>	SER	<mark>VVVA</mark> RIG <mark>Y</mark> SI	NTGKGHNSL	ILE <mark>I</mark> HET <mark>M</mark> HAV	DRF V	FNEV	SSTEEE	KNIF
P.spec_OSY_SE	GVΤ	P <mark>R</mark> GWE <mark>G</mark> TG <mark>L</mark> 7	WD <mark>DV</mark> PG <mark>V</mark>	SER	<mark>VVVA</mark> RIG <mark>Y</mark> SI	N <mark>KGKGHNS</mark> L	ILE <mark>I</mark> HET <mark>M</mark> HAV	DRF V	FNEV	SSTEEE	KNIF
P.yonginensis DCY84	GVΤ	P <mark>r</mark> gwe <mark>k</mark> tg <mark>l</mark> j	WD <mark>DV</mark> PG <mark>V</mark>	JSEK	D V V V R I G Y S	O <mark>KGKGHNS</mark> FN	ILE <mark>IHET</mark> MHA <mark>V</mark>	DRF	FNTV	SSSQEE	KDIF
P.spec_P1XP2	GKT	P <mark>r</mark> gwe <mark>g</mark> tg <mark>l</mark> j	WDDVPG	SEK	NVVVRIG <mark>Y</mark> SI	K <mark>kgeghns</mark> f1	ILE <mark>IHET</mark> MHAV	DRF V	FGNA	SQSAEE	QRIF
P.macquariensis_mac_def	GVΤ	P <mark>r</mark> gwe <mark>g</mark> tg <mark>l</mark> j	WD <mark>DV</mark> PG <mark>V</mark>	JSEK	N I V V R I G <mark>Y</mark> S I	K <mark>kgkghss</mark> y1	ILE <mark>I</mark> HET <mark>M</mark> HAV	DRF V	/FNNV	SDSGDE	KDVF
P.spec_GM2	GVΤ	P <mark>R</mark> GWE <mark>N</mark> TG <mark>L</mark>]	WD <mark>DI</mark> PG <mark>V</mark>	JS SN	IVIA <mark>RIG</mark> YSI	D <mark>KGKGHN</mark> GQ1	ILE <mark>L</mark> HET <mark>F</mark> HAI	DRL	/LGN I	SSSASE	TDIW
P.antibioticophila_GD11	GVΤ	P <mark>R</mark> GWE <mark>N</mark> TG <mark>L</mark>]	WD <mark>DV</mark> PG <mark>V</mark>	JS SN	VVIA <mark>RIG</mark> YSI	N <mark>KGKGHN</mark> GQ1	ILE <mark>L</mark> HET <mark>F</mark> HA <mark>I</mark>	DRL	/LNNV	SS <mark>S</mark> AAF	LNVW
P.macerans	GΙΤ	P <mark>R</mark> GWE <mark>K</mark> TG <mark>L</mark>]	WD <mark>DV</mark> PG <mark>V</mark>	/S MN	<mark>VVIV</mark> RIG <mark>Y</mark> SI	N <mark>KGKGHN</mark> GQI	ILE <mark>L</mark> HET <mark>F</mark> HA <mark>I</mark>		/LNNI	SSSLEE	TEIW
P.phoceensis_MT24	GVΤ	P <mark>R</mark> GWE <mark>N</mark> TG <mark>L</mark>]	WD <mark>NV</mark> PG <mark>V</mark>	/S MD	IVIV <mark>RIG</mark> YSI	N <mark>KGKGHN</mark> GQI	ILE <mark>L</mark> HET <mark>F</mark> HA <mark>I</mark>	DRL	/LNNI	SSSQEI	VDLW

	170	180	190 2	00 210
C.difficile_630_PPEP1	AKEGRS.LGNVN	Y L G V Y P E E F F A E	SFAYYYLNODTNSK	LKSACPOTYSFLQNLAK
P.alvei_DSM29_PPEP2	NKEASVKYKGDG	Y <mark>VSAYP</mark> TE <mark>YF</mark> AE	A <mark>ASLYLYSD</mark> A <mark>TR</mark> SD	LKDSMPLTYEFMAKLFAN
P.alvei_TS_15	NKEASIKYKGDG	Y <mark>VSTYP</mark> TE <mark>YFA</mark> E	AASLYLFSD T <mark>TR</mark> DD	LKSSMPLTYEFMAKLFAS
P.alvei_A66ix	NKEASVKYKGDG	Y <mark>VSMYP</mark> TE <mark>YF</mark> AE	A <mark>ASLYLFSD</mark> T <mark>TR</mark> D <mark>D</mark>	LKSS <mark>mplTyefm</mark> a <mark>klf</mark> as
P.spec_KS1	NKEASVKYKGDG	Y <mark>VSTYP</mark> T <mark>EYFA</mark> E	A <mark>ASLYLFSD</mark> T <mark>TR</mark> DD	LKSS <mark>mplTyefm</mark> a <mark>klf</mark> as
P.spec_UNC217MF	NKEASVKYKGDG	Y <mark>VSTYP</mark> T <mark>EYFA</mark> E	AASLYLFSD T <mark>TR</mark> DD	LKSS <mark>MPLTYEFM</mark> A <mark>KLF</mark> A <mark>S</mark>
P.spec_NAIST15	NKEASVKYKGDG	Y <mark>VSTYP</mark> T <mark>EYF</mark> AE	AASLYLFSD T <mark>TR</mark> DD	LKSS <mark>MPLTYEFM</mark> A <mark>KLF</mark> A <mark>S</mark>
P.spec_cl6col	NKEASVKYKGDG	Y <mark>VSTYP</mark> T <mark>EYFA</mark> E	AASLYLFSDT <mark>TR</mark> DD	LKSSMPLTYEFMAKLFAS
P.alvei_E194	NKEASVKYKGDG	Y <mark>VSTYP</mark> T <mark>EYFA</mark> E	AASLYLFSD T <mark>TR</mark> DD	LKSSMPLTYEFMAKLFAS
P.dendritiformis_C454	NQEANVNYNGDG	Y <mark>VSVHP</mark> T <mark>EYFA</mark> E	T ASLYVYSD A <mark>TR</mark> EE	LKETTPLTYEFMDKLFAS
P.thiaminolyticus_B4156	NQEANIKYNGDG	Y <mark>VSVHP</mark> TE YFA E	T <mark>ASLYVYSD</mark> A <mark>TR</mark> EE	LKELTPLTYEFMDKLFANP.
P.popilliae_ATCC14706	NQEANINYYEDG	Y <mark>VLAYP</mark> V <mark>EYF</mark> TE	T ASLYVYSD A <mark>TR</mark> EE	LKKTTPLTYEFMDKLFTNPS
P.terrigena_DSM21567	TKEAASNYH.DG	Y <mark>VSAYP</mark> A <mark>eyfa</mark> e	S ASLYLYND T TR K E	LKKDAPLTYDFLDKLFNQ
P.apiarius_B23460	NAEANVKYSGDG	Y <mark>VSVQP</mark> T <mark>EYFA</mark> E	T ATLYVYSE N <mark>TR</mark> ND	LKNSMPLTYEFLDKLFAS
P.spec_OSY_SE	NAEASVKYSGDG	Y <mark>VSVQP</mark> T <mark>EYFA</mark> E	T ASLYVYSE N <mark>TR</mark> ND	LKNS <mark>MPLTYEFL</mark> DKLFTS
P.yonginensis_DCY84	NKEASVNYKGDG	Y <mark>F S A Y P</mark> E <mark>E Y F A E</mark>	A ASMYVYND N <mark>TR</mark> KQ	LKDSTPLTYAFFDKLFNK
P.spec_P1XP2	NQEAAAEFGNDG	Y <mark>F S A Y T</mark> E <mark>E Y F A E</mark>	A <mark>A S L Y V Y N E</mark> N <mark>T R</mark> S E	LKANAPLTYQYLDKLFNQ
P.macquariensis_mac_def	NKEASVNYNKDG	Y I SVYP A E YF <mark>A</mark> E	TASLYLYNNTTREE	LKNSTPLTYDFMDKLFNK
P.spec_GM2	KKEADIDYKGDG	Y <mark>VSAYA</mark> TE YFA E	T ATLYLYSE Q <mark>TQ</mark> AR	LKQD <mark>MPLTYQFM</mark> KDLFSKYQ
P.antibioticophila_GD11	K K <mark>E A D V</mark> N <mark>Y</mark> K N D G	Y <mark>VSVY</mark> STE YFA E	TATLYFYSEETRKE	LKEDMPLTYDFLDKLFSNL.
P.macerans	KK <mark>EAN</mark> ND Y S GDG	Y LSAYSNEYFAE	T <mark>S T L Y F Y S E</mark> E <mark>T K</mark> K H	LKEHMPLTYEFL D KLY A N WK
P.phoceensis_MT24	K S EANN DYE GDG	Y V S A Y A E <mark>E Y F A E</mark>	TATLYFYSDKTRER	LK QD MPLTFEFL D KLF A N AG

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).

	Δ	L
1	_	۱

Species	PPEP cleavage sites	Amino acids
P.alvei_TS-15	SSSYYPSSK <mark>PLPPVP</mark> PLPPVP-KLETQYHYNYINGYPDG	1226 - 1263
P.alvei_A66ix	STSYYPSSK <mark>PLPPVP</mark> PLPPVP-KLETQYHYNYINGYPDG	614 - 651
P.spec_UNC217MF	STSYYPSSK <mark>PLPPVP</mark> PLPPVP-KLETQYHYNYINGYPDG	1287 - 1324
P.spec_NAIST15_1	STSYYPSSK <mark>PLPPVP<mark>PLPPVP</mark>-KLETQYHYNYINGYPDG</mark>	1218 - 1255
P.spec_cl6col	STSYYPSSK <mark>PLPPVP<mark>PLPPVP</mark>-KLETQYHYNYINGYPDG</mark>	1220 - 1257
P.alvei_E194	SSK <mark>PLPPVPPLPPVP<mark>PLPPVP</mark>-KLETQYHYNYINGYPDG</mark>	1159 - 1196
P.alvei_DSM29	SSK <mark>PLPPVP</mark> PVQ <mark>PLPPVP</mark> -KLETSYHYNYINGYPDG	1101 - 1135
P.dendritiformis_C454	SYTITPSPK <mark>PLPPVP</mark> PKLDMENHYNYINGYPDG	461 - 499
P.thiaminolyticus_B4156	TTTITPSPK <mark>PLPPVP</mark> PKLDMENHYNYINGYPDG	1447 - 1486
P.popilliae_ATCC14706	SSTITPSPKPLL <mark>PLPPVP</mark> PKLDMENHSNYINGYPDG	855 - 890
P.terrigena_DSM21567	STPT <mark>PLPPV</mark> PLPPVP <mark>PLPPVP</mark> PQLEKDKHFNYINGYPDG	407 - 446
P.apiarius NRRL B-23460	TTNPSNPAN <mark>PLPPVP<mark>PLPPVP</mark>PKLETDNHYNYINGYPDG</mark>	1437 - 1476
P.spec_OSY_SE	TTNPTNPVN <mark>PLPPVP<mark>PLPPVP</mark>PKLEMDNHYNYINGYPDG</mark>	342 - 380
P.yonginensis_DCY84	DNNVT PLPPAP	1096 - 1133
P.spec_P1XP2	SGSVT <mark>PLPPAP</mark> PVVP <mark>PLPPAP</mark> PKLDTANHFNYINGYPDG	1096 - 1133
P.macquariensis_mac/def	TSNPT <mark>PLPPAP</mark> PVVA <mark>PLPPPP</mark> PKLEAENHYNYINGYPDG	1297 - 1335
P.spec_GM2	SSYTTPSPP <mark>PAPPVP<mark>PLPPVP</mark>PKLDTENHFDYIQGYPDG</mark>	1073 - 1111
P.antibioticophila_GD11	PSSSVTPLPAVP <mark>PLPPLP</mark> AVPPVLDRENHYDYINGYPDG	290 - 328
P.macerans	PSPSTPEVD <mark>PLP<mark>PLPPLP</mark>AVPPKLETENHYDYINGYPDG</mark>	1100 - 1138
P.phoceensis MT24	GNPTPPSTPEVTPLPPLEAVPPTLDTENHFOYIHGYPDG	1500 - 1538







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 (VWFA, Mucbp, Surfacelayer 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	5 1319.7336	5 2.94	0	108	3e-10	1	U	K.IPTSILDALYSK.G
3437	716.8299	1431.6452	2 1431.6414	2.69	0	(42)	0.0016	1	U	K.DSMPLTYEFMAK.L
3520	724.8257	1447.6368	3 1447.6363	0.34	0	58	3.2e-05	1	U	K.DSMPLTYEFMAK.L
4499	564.3273	1689.9600	0 1689.9552	2.80	1	48	0.0002	1	U	R.LEKIPTSILDALYSK.G
4724	880.9757	1759.9369	9 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	0 1906.8692	4.05	1	(39)	0.0041	1	U	R.SDLKDSMPLTYEFMAK.L
5114	639.3262	1914.9567	7 1914.9509	3.01	0	67	7.9e-06	1	U	K.LVVLPSGEYNHSEAAAMK.Q
5171	644.6575	1930.9508	3 1930.9458	2.55	0	(47)	0.00099	1	U	K.LVVLPSGEYNHSEAAAMK.Q
5247	980.9595	1959.9044	1959.8963	3 4.13	0	57	5.3e-05	1	U	R.GWEGTGLTWDDVPGVSER.V
5365	668.0445	2001.1116	5 2001.1146	6 -1.48	1	32	0.009	1	U	K.IKLTQGAITNEPELAYLK.G
5496	514.5168	2054.0380	0 2054.0293	4.23	0	74	1.6e-06	1	U	K.GHNSLNLEIHETLHAVDR.L
5970	757.0913	2268.2521	1 2268.2478	3 1.92	1	53	7.9e-05	1	U	K.LTQGAITNEPELAYLKGVVPR.G
6231	829.0903	2484.2492	2 2484.2398	3.78	1	76	9.6e-07	1	U	R.GWEGTGLTWDDVPGVSERVVAVR.I
6496	947.4776	2839.4110	2839.4062	2 1.69	1	95	1.1e-08	1	U	A.QEQSILDKLVVLPSGEYNHSEAAAMK.Q
6505	952.8104	2855.4093	3 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	3 2856.4327	3.89	1	(49)	0.00036	1	U	A.QEQSILDKLVVLPSGEYNHSEAAAMK.Q
6512	719.1173	2872.4399	2872.4276	6 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 ($25 \mu M$) 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.



В

MKRRVKLAVTWLCVFAVFFGFIPHTLOSADAAAODPIWPEPGAVKLSKTAKPTGTPGEW QITLTAEGKNVEAGDTDVVLVLDKSDSMRSYKRLDNAKKAAEKFVNTLLFSNVKTRIAL VTFSTSSEDTSSGFLGQNDKQQLLNNIKGISAS GGTNIQSALYTAHRLLATKSSAKNKF IVLLSDGAPTFSSPAKKAASYSWPNNKYNFVLSDFDYTKTIGSGNSYQLS IGNYTVKDNGIPTISQAKASKDAGINIYSIGLEVGNDANAVYVLKNSASANNYYPSDSA ELEKVFSELAGKISYAAENAVVKDPMGEMFNLIFKGNAPVENQDYKVSQGTVKWDAASE TFVWDVGNIVEGSPATLTYTIKMDSSKNPSADVNYPTNGRTTMEYTDVNQQRTSKEFEV PQVSFGKGSILVKGYKVNDQGEPINADGKVVERVDLAEQLYSGSFTVNGKDALDIGPKY KVPAPKLDNYVLKKGTDPTEETLTIQEPNKTIWFGYALAPHKLTIKYVDRETGKSIAND DSSLSGTIGQTITLKALDIPGYTPEKTEVQYTFKAKDNEYTFFYKKMDLPITIKFVDKD TSKSIADDDTSHSGKVGEAVTLKALDIPGYTPEKAEITYTVKATDNVVTFFYTANPOKI TIHYIDQDTKLPIAKDSSKDGKTGESVTLTALEITGYTSVNPEVKYTFTAEANQEHTFF YTKNAPVERTVHVHYKEADTGAVLKESTSTSGKVGETVTLKAEDITINGVVYVPNAYEH EYTITDKEEQSYTFIYTKKDVIPOERTVHVHYKDAATGTVLKESTSTSGKVGATITLRA EDITVDGVIYVPASYQHEYAVTEQTEQSYTFFYSKKELPTYFVTVHHLEEGTDAVLHEP TTVSGKSGEVVHVTAEPISVADAVYQPIKFNHDITIEGTPEQLYTIYYKKGEPVEPLLL TVLHLDKETLAPLAEPTSLKGKAGEHITLKPTPITVTDAVYNPEHAAYDYVFTGEATOT FSILYIKNPITHPDOHVTIKYVEOGTGKOLVNPTTKNGKAGEKVELTALSISGYTPVKS TDTYTFTDKEGQEYIFYYTRNSSGGGSSSSGGSSSYYPSSK**PLPPVP**PVQ**PLPPVP**KLET SYHYNYINGYPDGKIKPENLISREEVAVIFYRLMENSTRANFLKGTNSYKDIAEKRWSN RHISTMENAGI IKGYPDGSFLPERPITRAEFAAIASRFDKLNEOPNTMFSDISGHWAEK YIVSAANKGWIKGYKNGTFKPNQYITRAEAMAFINSVLNRKVHNDEIHKDAKKWPDITP TNWYYSDVMEATNYHDYHRIKDNFESWDKVNADVVYH

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 $\beta 3/\beta 4$ 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.