# Letter to the Editor

# Prediction of a caspase-like fold in *Tannerella forsythia* virulence factor PrtH

Jimin Pei<sup>1,\*</sup> and Nick V. Grishin<sup>1,2</sup>

<sup>1</sup>Howard Hughes Medical Institute; and <sup>2</sup>Department of Biochemistry; University of Texas Southwestern Medical Center; Dallas, TX USA

Key words: *Tannerella forsythia* PrtH, cysteine protease, periodontal disease, forsythia detaching factor, caspase-like fold, structure prediction

Tannerella forsythia is a bacterial pathogen involved in periodontal disease. A cysteine protease PrtH has been characterized in this bacterium as a virulence factor. PrtH has the activity of detaching adherent cells from substratum, and the level of PrtH is associated with periodontal attachment loss. No reports exist on the structure, active site, and catalytic mechanism of PrtH. Using comparative sequence and structural analyses, we have identified homologs of PrtH in a number of bacterial and archaeal species. PrtH was found to be remotely related to caspases and other proteases with a caspase-like fold, such as gingipains from another periodontal pathogen Porphyromonas gingivalis. Our results offer structural and mechanistic insights into PrtH and its homologs, and help classification of this protease family.

Tannerella forsythia is an anaerobic, Gram-negative bacterial pathogen found in human oral cavity. A number of clinical studies have linked T. forsythia to periodontitis, an infectious disease that causes inflammatory destruction of periodontal tissues.<sup>1</sup> In an effort to isolate putative proteolytic enzymes, a DNA fragment from T. forsythia was identified to cause hydrolysis of milk proteins when transformed to Escherichia coli.<sup>2</sup> A putative gene prtH in this fragment was predicted to encode a protein of 423 amino acid (aa) residues. The product of the prtH gene, PrtH, was suggested to be a cysteine protease based on protease inhibitor analysis, although no significant sequence similarity to other proteases was found.<sup>2</sup> Clinical studies suggested that PrtH is a virulence factor, as the level of PrtH is associated with periodontal attachment loss.<sup>3</sup> In an independent study to explore the cytopathic effect of T. forsythia, a protein named forsythia detaching factor (FDF) was identified that can detach human adherent cells from substratum.<sup>4</sup> Cloning of the *fdf* gene revealed that the original *prtH* 

\*Correspondence to: Jimin Pei; University of Texas Southwestern Medical Center; Howard Hughes Medical Institute; 6001 Forest Park Rd.; Room ND10.120; Dallas, TX 75390 US; Tel.: 001.214.645.5951; Fax: 001.214.645.5948; Email: jpei@ chop.swmed.edu

Submitted: 02/17/09; Revised: 02/18/09; Accepted: 02/19/09

Previously published online as a *Cell Cycle* E-publication: http://www.landesbioscience.com/journals/cc/article/8243 gene was part of the *fdf* gene, which encodes a protein of 536 residues. The original *prtH* turns out to be a partial gene that resulted from an incorrect prediction of the translation start site. Consistent with the initial studies of *prtH*, recombinant FDF protein was shown to possess proteolytic activity.<sup>4</sup> In this work, we use the name PrtH to refer to FDF, the authentic 536aa product of the *fdf* gene.

The structure, active site and catalytic mechanism have not been reported for PrtH. Submissions of the PrtH sequence to public protein domain databases such as Pfam<sup>5</sup> and CDD<sup>6</sup> did not yield significant predictions of known domains in this protein. A PSI-BLAST<sup>7</sup> search for the full length PrtH protein (gene identification (gi) number: 38707242) converged to about 20 proteins with an e-value inclusion threshold of 0.001. More PSI-BLAST iterations starting from multiple representatives of found PrtH homologs revealed limited, yet significant sequence similarities to protein domains in the peptidase C14 family, which includes eukaryotic caspases. For example, a PSI-BLAST search from gil 166367720 of Microcystis aeruginosa, a close homolog of PrtH, found a Desulfovibrio salexigens protein (gi 218151631) annotated as 'peptidase C14 caspase catalytic subunit p20' with significant e-value (0.0009) in the third iteration. A Pfam database search confirmed that this protein contains the 'Peptidase C14, Caspase domain' (Pfam accession number: PF00656). PSI-BLAST alignments revealed that the catalytic residues (a histidine and a cysteine) in the peptidase C14 family are also conserved in PrtH and its homologs. HHpred,<sup>8</sup> a profile-profile based sequence similarity search method, also consistently found structures with a caspase-like fold as top hits using PrtH and its homologs as queries. Structure predictions for PrtH were made by the 3D-Jury Meta server,<sup>9</sup> which assembles the results of various fold recognition methods and computes consensus scores for the predictions. Several fold recognition methods found structures with a caspase-like fold as top hits (e.g., PDB id: 2j31, 1jxq and 3bij). The top hits of the 3D-Jury consensus results have significance scores above 60. These results indicate that PrtH indeed has a caspase-like fold.

The MEROPS peptidase database<sup>10</sup> classifies cysteine peptidases with a caspase-like fold in clan CD, currently including several remotely-related peptidase families: clostripain (family C11), legumain (family C13), caspase (family C14), gingipain (family C15), separase (family C50) and RTX self-cleaving toxin (family C80). Extensive sequence similarity searches have detected new members of clan CD peptidases, as well as a new family HetF.<sup>11</sup> CPDadh, a new family of caspase-like cysteine peptidases homologous to the cysteine protease domains of multifunctional autoprocessing RTX toxins, has been recently identified in a group of bacterial cell adhesion molecules.<sup>12</sup> MEROPS classifies PrtH and its close homologs in peptidase family C84, which has not been assigned to a known clan. The homologous relationship of the PrtH protease to caspases suggests that peptidase family C84 belongs to clan CD. Interestingly, PrtH and gingipains from Porphyromonas gingivalis have related functions as they are virulence factors linked to periodontitis.

The structures of caspases and gingipains have a Rossmann-fold like core characterized by a mainly parallel central  $\beta$ -sheet and surrounding  $\alpha$ -helices on both sides.<sup>11</sup> The structure of the cysteine protease domain from *V. cholerae* RTX toxin<sup>13</sup> (peptidase family C80) exhibits a similar core as well as noticeable differences in peripheral regions as compared to the structures of caspases and gingipains. Multiple sequence alignment and secondary structure predictions reveal that PrtH and its close homologs have the core structural elements of the caspase-like fold (Fig. 1). The most conserved parts of caspase-like domains in PrtH homologs surround the two catalytic residues. For example, the catalytic histidine is sandwiched by two conserved small residues, usually glycines. The N-terminal regions before the  $\beta$ -strand preceding the catalytic histidine show high sequence divergence. Compared to other structures with a Rossmannlike fold, a distinct feature of caspases and gingipains lies in their

#### A caspase-like fold in Tannerella forsythia PrtH

			eeeeee	e		hhhhhhh	hhh		eee	eee		eee	eeee			hhh	hhhh		eeee	eeeee
38707242 (PrtH	Tf	30	IYGACSIE	DFL	12	AEGFLGY	PTS		FTPANF	RKDG	20	DAVKV	FYHSG	GNML	18	ALSSK	MAFAN	EK	LRYLF	STCLS
108743329	Sk	59	WWGAFSIQ	KFV	11	ATGWLAY	LQQ		FYDRNF	FADG	20	DAVCA	VYHSG	GGMD	18	AISNR	MALGN	EK	VNYVF	STCDS
166367720	Ma	48	LYGGCSIE	VGC	13	AQGWLDY	VKK		FTPLNF	NYQDC	20	DAVMA	FYHSG	GGMD	21	SSSDK	LVWAN	IER	VRYIE	LSTCLS
119386630	Pd	71	IYGACSVE	TYR	11	AGGFLDA	VDR		FATPDE	RRDG	20	DSVRV	FYHAG	GRMD	19	LTSDR	MRLGS	GA	LRYLF	STSQS
222528938	At	41	PNGSCTYG	DRA		GAHYVYN	ILAD	5	AWVTAY	LIKLN	14	SOADL	LVYSG	GLCF	29	ARTNE	VREGH	ISN	LKWVI	MYTCNW
146297187	Cs	41	PNGSCTYG	DRA		GAHYVYN	ILAD	5	AWVTAY	LIKLN	14	SOADL	LVYSG	GLCL	29	ARTNE	VRFGR	SN	LKWII	MYTCNW
78044552	Ch	39	LYPDCTYG	DRA		GASYVNN	ILYN	4	KFTLGF	XYLDN	13	NNVDF	FSFAG	GKNY	30	ARTNE	TREGH	INK	LKWVN	MYCCWW
162452479	Sc	428	VVGAEWVG	LCG	8	VDGLLKS	FRE	1	GIDSRE	WGDO	20	DSVDL	TFYTG	ANGV	12	LHFSD	ARWGN	AN	LEWMV	AACGP
118048019	Cau	475	SEGAEGNW	DYO	15	VTGERSG	MLG	1	GYTORE	WTNA	21	DRAAF	VYYAG	GGPG	10	TWFSG	ANARY	ON	LRWVG	TASCOT
163848691	Cag	474	TEGIEGNW	DYP	15	VNGLREG	MLS	1	GYTORE	WSNS	21	DRAAF	VYYAG	GGPG	10	TWFDG	TNARY	OS	LRWVG	TASCOT
84497076	Js	4	YVGSWWTR	DTV	6	AEGEAST	LGG		LSEFDW	TONHG	19	DOVDA	MOMSS	GNPN	9	NASEA	DEGK	ND	LETFA	THACDL
194337677	Pph	33	EEGGEVCR	EES	2	VDETWNE	TKH	1	SYEONY	AVPE	10	DNMDT	AFESC	GONE	13	FASCA	SLGD	VD	LEFLT	TDACSV
194337435	Pph	37	EGGYVGSK	TEG	2	TTPVWNF	TNN	1	SYDOYY	AFPE	10	DKMDF	SYFVC	GSPW	14	GSTTH	KGWGD	YN	AFFUT	HACNV
194334999	Pa	37	FCCVVCSK	TEG	2	TTOWNE	TPH	1	DYNOYY	ADETE	10	DAMDE	SVEVC	CSPW	14	CSFAAL	FCWCD	TN	SEEVT	HACKV
149924090	Pna	19	FCTIVUN	DED	16	VCVLEDK	TDD	1	CEDQUV	CND	25	DCADT	CMMVC	CCKM	26	ACND	SWWCD	TD	INAMT	DIRECCO
154150067	rpa	500	TYCMTMIO	MDC	10	HCUNONN	TET	2	CHENTE	UTED	11	DNATE	FVUEC	CUTT	20	DCDUVI	KKNCC	0-	NKHUF	DACGT
20080304	Mag	199	FUCUEWUN	DVT	12	ANCI VND	LCN	1	CHTREE	INCING	17	DAUDT	ATVCC	COTD	10	UVUDE	FWCD	ND ND	LEWIC	
106170074	Mac	21	LVGVEWVN	NVD	11	ANGLINE	LGN	1	GWINSTI	INGNO	22	DAVUI	ALISG	GSID	10	TCVCD	ALWGD	UTO I	MONTO	DOCULA
1201/80/4	Pin	25	ICCAEWVN	NIP	11	AQGEIIQ	2DGG	4	SWWGDE	LIGDG A	10	DUTHE	AFFSG	GAPE	10	LSISD	ALWGN	Q II	MDWIT	TDACTV
11499161	AI	33	DVGAAVAF	SED	0	VSQEVEL	155	4	TWVQHFI	MDIP	19	DITEL	SLVLG	GVVV	10	ALPDD	IRLGI	AS 0	SVWIP	TTOODAN
18977023	PI	30	DVGGVMTF	TDD	10	VENEVEI	ISS	4	HWIQGEI	KDIP .	20	DITEL	SLVLG	SCVY	13	ALPEK	VRLGY	KS 0	AIWIT	LIQCSV
73669560	MD	28	SISVIIIE	NYS	12	IGIVENW	LHD	10	GWNEYF	DSET	1/	DDADE	HYHLG	GVDD	20	AQUVYI	KKWDN	IN-	NEWVL	LHSCHI
169846526	Ce	435	RVGRIVVR	NDI	3	TASARSE	LSS	10	DNTHEY	ISRPP.	10	NNVDI	ALTEA	GSFH	10	DIPAG	STUCI	GA 9	LSYWV	THSCEV
119390377 2j3	A Hs	10	EMGLCIII	NNK	14	DVDAANI	RET	2	NLKYEV	RNKND 2	20	RSSFV	CVLLS	GEEG	9	DLKKI	INFFR	GD 8	PKLFI	IQACRG
42543031 1m7	A Sf	31	HRGMAIIF	NHE	13	NVDSDNL	SKV	2	TLGFKV	CVFPN 2	20	ADCLL	VAVLT	GELG	9	KPDNL	NY <mark>Y</mark> FT	AD 8	PKLFF	IQACQG
7245522 1cv:	A Pg	141	WLGQALCI.	ASA	12	IQHENVI	ANL	2	QYGYTK	LIKCY	14	GGISL	VNYTG	GSET	7	GTTHV	KQLTN	ISN 1	LPFIF	DVACVN
162330276 3bi	jA Gs	2	PKGIALAL	GLN	17	EADAEDM	IAAI	2	ERGFAV	TLMT	21	GDIFM	LSYSC	GGQV	22	IDDEL	YALLG	KF 4	RVLVF	SDSCHS
209870506 3eel	A Vc	30	RFDGQIIV	QME	5	AKAAANI	AGK		-HAESS	VVVQL	16	DGKLR	WQLVG	GRDH	12	ADELA	VKLAK	FQ 12	PDHIS	IVGCSL
			hhhhhhh		6	eeeeee			hhhl	hhhhhh	h		hhhi	hhhhhł	hhh		e	eeeeee		
38707242 (PrtH	Tf	10	PTWPLANK	1	GLE	MIEGYET	VSYD	S	ARYG	SEFWKQW	KKG	;	KSFSD	AFIEAS	SWSL	FRNQT	P	VVCACG	N 224	[536]
			LAT MILTINIA			TITT OT DA				VV DINE VNI									0 0 5 1	[572]
108743329	Sk	10	RTWAGPNI		GFI	MIFGFET	TSID	S	GDYG	UUL MEUM	RAG	;	QTYCD	AWLNAS	SWDI	HGQA	P	SVAAVG	1 231	10.01
108743329 166367720	Sk Ma	10 10	RTWAGPNI		GFI	RMIFGFET RMLFGYET	TSID VSWD	SN	GDYGH PNYGH	KFFWEEW	RAG	;	QTYCD KPLST	AWLNAS	SWDI SWRI	AHDQA	P	SVAAVG SVVACG	A 245	[561]
108743329 166367720 119386630	Sk Ma Pd	10 10 10	RTWAGPNI RTWHPSNR HSWVRANQ		GFH GWH GLH	RMIFGFET RMLFGYET RMLFGFDS	TSID VSWD ICWD	S N S	GDYGI PNYGI GRYGI	KFFWEEWI NFWHHW	RAG NKN QMG	; ; ;	QTYCDI KPLSTI KPFAQI	AWLNAS AWLDAS AWLDGF	WDI WRI WDI	(HGQA AHDQA APDQS	PP	SVAAVG SVVACG VACACA	A 245 P 264	[561] [584]
108743329 166367720 119386630 222528938	Sk Ma Pd At	10 10 10 10	RTWAGPNI RTWHPSNR HSWVRANQ ENIYKTFE		GFF GWF GLF GAT	RMIFGFET RMLFGYET RMLFGFDS FLVMGFAS	TSID VSWD ICWD TMYL	S N S D	GDYGH PNYGH GRYGJ SREAN	KFFWEEW ANFWHHW 7DFVKFL	RAG NKN QMG TGY	1	QTYCD KPLST KPFAQ DSFVK	AWLNAS AWLDAS AWLDGF AFVKAF	WDI WRI WDI ARIY(	THGQA AHDQA APDQS OPORO	P P P 4 S	SVAAVG SVVACG VACACA	A 245 P 264 Y 237	[561] [584] [281]
108743329 166367720 119386630 222528938 146297187	Sk Ma Pd At Cs	10 10 10 10	RTWAGPNI RTWHPSNR HSWVRANQ ENIYKTFE ENIYKTFE		GFH GWH GLH GAN	RMIFGFET RMLFGYET RMLFGFDS FLVMGFAS	TSID VSWD ICWD TMYL TMYL	S NI S D D	GDYGH PNYGH GRYGH SREAN SREAN	KFFWEEW ANFWHHW VDFVKFL' VDFVKFL'	RAG NKN QMG TGY TGE	1	QTYCD KPLST KPFAQ DSFVK LSFKE	AWLNAS AWLDAS AWLDGF AFVKAF AFIKAF	SWDI SWRI AWDI ARIY( ASIY(	YHGQA AHDQA APDQS QPQRQ QPQRK	P P 4 S 4 S	SVAAVG SVVACG VACACA IVRIMG LVRIMG	A 245 P 264 Y 237 Y 237	[561] [584] [281] [282]
108743329 166367720 119386630 222528938 146297187 78044552	Sk Ma Pd At Cs Ch	10 10 10 10 10	RTWAGPNI RTWHPSNR HSWVRANQ ENIYKTFE ENIYKTFE ONFYKMFE		GFH GWH GLH GAT GAT	RMIFGFET RMLFGYET RMLFGFDS FLVMGFAS FLVMGFAS FLMLGFSS	TSID VSWD ICWD TMYL TMYL STMYL	SNSDDD	GDYGH PNYGH GRYGH SREAN SREAN SREAN	KFFWEEW ANFWHHW VDFVKFL VDFVKFL IEFGOKL	RAG NKN QMG TGY TGE VNG	1	QTYCD KPLST KPFAQ DSFVK LSFKE YTIKT	AWLNAS AWLDAS AWLDGF AFVKAF AFIKAF AFIKAF	SWDI SWRIA AWDIA ARIY( ASIY( AKKY(	YHGQA AHDQA APDQS QPQRQ QPQRK QPQRK	P P 4 S 4 S 3 S	SVAAVG SVVACG VACACA IVRIMG LVRIMG IATVIG	A 245 P 264 Y 237 Y 237 Y 232	[561] [584] [281] [282] [275]
108743329 166367720 119386630 222528938 146297187 78044552 162452479	Sk Ma Pd At Cs Ch Sc	10 10 10 10 10 10	RTWAGPNI RTWHPSNR HSWVRANQ ENIYKTFE ENIYKTFE QNFYKMFE YRWAGAFD		GFF GWF GLF GAT GAT GAT	RMIFGFET RMLFGYET RMLFGFDS FLVMGFAS FLVMGFAS FLMLGFSS HLLLGYAT	TSID VSWD ICWD TMYL TMYL STMYL SVMYL ESFD	SNSDDDD	GDYGH PNYGH GRYGH SREAN SREAN SREGT TTEGS	KFFWEEW ANFWHHW VDFVKFL' VDFVKFL' TEFGQKL' SMFARTL	RAG NKN QMG TGY TGE VNG LDD		QTYCDA KPLSTA KPFAQA DSFVKA LSFKEA YTIKTA APVROA	AWLNAS AWLDAS AWLDG# AFVKA# AFVKA# AFLEA# AWVTT#	SWDI SWRI AWDI ARIY( ASIY( AKKY( AIEV(	YHGQA AHDQA APDQS QPQRQ QPQRK QPQRA QPQDE	P P 4 S 4 S 3 S V	SVAAVG SVVACG VACACA IVRIMG LVRIMG IATVIG IYAVMG	A 245 P 264 Y 237 Y 237 Y 232 A 614	[561] [584] [281] [282] [275] [653]
108743329 166367720 119386630 222528938 146297187 78044552 162452479 118048019	Sk Ma Pd At Cs Ch Sc Cau	10 10 10 10 10 10 9 13	RTWAGPNI RTWHPSNR HSWVRANQ ENIYKTFE ENIYKTFE QNFYKMFE YRWAGAFD REWFNAFO		GFI GWI GLI GAT GAT GLI GAT	RMIFGFET RMLFGYET RMLFGFDS FLVMGFAS FLVMGFAS FLMLGFSS HLLLGYAT	TSID VSWD ICWD TMYL TMYL VMYL ESFD	S N S D D D D V	GDYGH PNYGH GRYGJ SREAN SREAN SREGT TTEGS -AFGO	KFFWEEW ANFWHHW VDFVKFL VDFVKFL FEFGQKL SMFARTL SNLAANM	RAG NKN QMG TGY TGE VNG LDD		QTYCDI KPLSTI KPFAQI DSFVKI LSFKEI YTIKTI APVRQI LTIAQI	AWLNAS AWLDAS AWLDG# AFVKA# AFIKA# AFLEA# AWVTT# AWVKT#	SWDI SWRI AWDI ARIY( ASIY( AKKY( AIEV( AFDMI	YHGQA AHDQA APDQS QPQRQ QPQRK QPQRA QPQDE HAGK-	P P 4 S 4 S 3 S V P	SVAAVG SVVACG VACACA IVRIMG LVRIMG IATVIG IYAVMG AYIYAR	A 245 P 264 Y 237 Y 237 Y 232 A 614 S 677	[561] [584] [281] [282] [275] [653] [711]
108743329 166367720 119386630 222528938 146297187 78044552 162452479 118048019 163848691	Sk Ma Pd At Cs Ch Sc Cau Cau	10 10 10 10 10 10 10 10 13 13	RTWAGPNI RTWHPSNR HSWVRANQ ENIYKTFE ENIYKTFE QNFYKMFE YRWAGAFD REWFNAFQ RRWFNAFQ		GFI GWI GAN GAN GAN GAN GAN	RMIFGFET RMLFGFDS RLVMGFAS FLVMGFAS FLMLGFSS HLLLGYAT HMLLGFNS	TSID VSWD ICWD TMYL TMYL VMYL ESFD SNMAD	S N S D D D D V V	GDYGH PNYGH GRYGJ SREAN SREAN SREGT TTEGS -AFGO -AFGO	KFFWEEW ANFWHHW VDFVKFL VDFVKFL TEFGQKL SMFARTL 3NLAANM 30LMNNM	RAG NKN QMG TGY TGE VNG LDD KMP RMP	1 1 3 11 11	QTYCDI KPLSTI KPFAQJ DSFVKJ LSFKEJ YTIKTJ APVRQJ LTIAQJ LTIAQJ	AWLNAS AWLDG AFVKA AFVKA AFIKA AFLEA AWVTT AWVKT AWVKT	SWDI SWRIA AWDIA ARIYO ASIYO AKKYO AIEVO AFDMI AFDMI	YHGQA AHDQA APDQS QPQRQ QPQRK QPQRA QPDDE HAGK- NAGK-	P P 4 3 3 V P P	SVAAVG SVVACG VACACA IVRIMG LVRIMG IATVIG IYAVMG AYIYAR	A 245 P 264 Y 237 Y 237 Y 232 A 614 S 677 S 676	[561] [584] [281] [282] [275] [653] [711] [710]
108743329 166367720 119386630 222528938 146297187 78044552 162452479 118048019 163848691 84497076	Sk Ma Pd Cs Ch Sc Cau Cau Cag Js	10 10 10 10 10 10 10 10 13 13 9	RTWAGPNI RTWAPSNR HSWVRANQ ENIYKTFE ENIYKTFE QNFYKMFE YRWAGAFD REWFNAFQ RRWFNAFQ GRWIPAFQ		GFF GWF GLF GAT GAT GAT GAT GAT GAT	AMIFGFET AMLFGFET AMLFGFDS FLVMGFAS FLVMGFAS FLMLGFSS HLLLGYAT HMLLGFNS HMLLGFNS HMLLGFNS	TSID VSWD TCWD TMYL TMYL TMYL STMYL STMYL STMYL STMAD SNMAD SNMAD	S N S D D D D V V G	GDYGI PNYGI GRYGJ SREAN SREAN SREG TTEGS -AFGC -AFGC 9 FAMYS	KFFWEEWI ANFWHHW VDFVKFL' VDFVKFL' SMFARTLI SNLAANMI SQLMNNMI SAWLYYW	RAG NKN QMG TGY TGE VNG LDD KMP RMP SGS	1 1 3 11 11 11	QTYCDI KPLSTI KPFAQI DSFVKJ LSFKEJ YTIKTI APVRQJ LTIAQJ IPVREJ	AWLNAS AWLDG AFVKA AFIKA AFIEA AWVTT AWVKT AWVKT AWVKT AWAEAN	SWDI SWRI AWDI ARIY ASIY ASIY ASIY ASIY AFDMI AFDMI AFDMI NEIVI	YHGQA AHDQA APDQS QPQRQ QPQRK QPQRA QPDDE HAGK- NAGK- EGSNV	P P P 4 S 3 S V P P T	SVAAVG SVVACG VACACA IVRIMG LVRIMG IATVIG IYAVMG AYIYAR AYIYAR WAYLRA	A 245 P 264 Y 237 Y 237 Y 232 A 614 S 677 S 676 E 195	[561] [584] [281] [282] [275] [653] [711] [710] [227]
108743329 166367720 119386630 222528938 146297187 78044552 162452479 118048019 163848691 84497076 194337677	Sk Ma Pd Cs Ch Sc Cau Cag Js Pph	10 10 10 10 10 10 13 13 9 11	RTWAGPNI RTWHPSNR HSWVRANQ ENIYKTFE ENIYKTFE QNFYKMFE YRWAGAFD REWFNAFQ GRWIPAFQ GRWIPAFQ SGWWGVFH		GEFF GWFF GLFF GAT GAT GAT GAT GAT GAT BLFF RLFF	AMIFGFET AMIFGFET AMIFGFDS FLVMGFAS FLWGGFAS FLMLGFS HLLLGYAT HMLLGFNS HMLLGFNS HMLLGFNS HMLLGFNS	TSID VSWD ICWD TMYL TMYL TMYL STMYL STMYL STMYL STMYL STMYL STAN STAN STAN STAN STAN STAN STAN STAN	S N S D D D D V V G D	GDYGI PNYGI GRYGJ SREAV SREAV SREG TTEGS -AFGC -AFGC 9 FAMYS GRVEI	KFFWEEWI ANFWHHW VDFVKFL VDFVKFL SMFARTLI SNLAANMI SQLMNNMI SAWLYYW SOYAINL	RAG NKN QMG TGY TGE VNG LDD KMP RMP SGS	1 1 3 11 11 6	QTYCDI KPLSTI KPFAQI DSFVKI LSFKEI YTIKTI APVRQI LTIAQI LTIAQI IPVREI OKYIDI	AWLNAS AWLDGA AFUCAA AFIKAA AFIEAA AVTTA AWVTTA AWVKTA AWVKTA	SWDI SWRI AWDI ARIY ASIY ASIY ASIY AFDMI AFDMI AFDMI AFDMI ASVE	YHGQA AHDQA APDQS QPQRQ QPQRK QPQRA QPDDE HAGK- NAGK- EGSNV RSGVY	P P P 4 S 4 S V P P T T 3 C	SVAAVG SVVACG. VACACA IVRIMG LVRIMG IATVIG IYAVMG. AYIYAR AYIYAR WAYLRA AVIWAY	A 245 P 264 Y 237 Y 237 Y 232 A 614 S 677 S 676 E 195 P 206	[561] [584] [281] [282] [275] [653] [711] [710] [227] [238]
108743329 166367720 119386630 222528938 146297187 78044552 162452479 118048019 163848691 84497076 194337637 194337435	Sk Ma Pd At Cs Ch Sc Cau Cag Js Pph	10 10 10 10 10 10 10 10 9 13 13 9 11	RTWAGPNI RTWHPSNR HSWVRANQ ENLYKTFE ENLYKTFE QNFYKMFE YRWAGAFD REWFNAFQ GRWIPAFQ GRWIPAFQ GGWIPAFQ SGWWGVFH ADWYSAWT		GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT	RMIFGFET RMLFGFDS FLVMGFAS FLVMGFAS FLMLGFSS HLLLGYAT HMLLGFNS HMLLGFNS HYMLGFNN HQLLSFRT HOVCGFRT	TSID VSWD ICWD TMYL TMYL TMYL STMYL STMYL STMYL STMAD SNMAD SNMAD SNMAD SNMAD TGWYL	SNSDDDDVVGDS	GDYGI PNYGI GRYGJ SREAT SREAT SREG TTEGS -AFGC 9 FAMYS GRVEI 2 ODISJ	KFFWEEWI ANFWHHW VDFVKFL VDFVKFL SMFARTLI SNLAANMI SQLMNNMI SAWLYYW SQYAINLI XYFGSRII	RAG NKN QMG TGY TGE VNG LDD KMP SGS LSG KSG		QTYCD KPLST KPFAQ DSFVK LSFKE YTIKT APVRQ LTIAQ LTIAQ IPVRE QKYID AAVWO	AWLNAS AWLDAS AWLDG# AFVKA# AFIKA# AFIEA# AWVTT# AWVKT# AWVKT# AWAEAN AWFNA#	SWDI SWRI AWDI ARIYO ASIYO ASIYO ALEVO AFDMI AFDMI AFDMI ASIYO SLYO	YHGQA AHDQA APDQS 2PQRQ 2PQRK 2PQRA 2PDDE HAGK- NAGK- EGSNV 2SGVY 3DGSE	PP PP 4 S 3 S V PP T 3 C R	SVAAVG SVVACG. VACACA IVRIMG LVRIMG IATVIG IYAVMG. AYIYAR AYIYAR WAYLRA AVIWAY GAAVMY	A 245 P 264 Y 237 Y 237 Y 232 A 614 S 677 S 676 E 195 P 206 P 213	[561] [584] [281] [282] [275] [653] [711] [710] [227] [238] [240]
108743329 166367720 119386630 222528938 146297187 78044552 162452479 118048019 163848691 84497076 194337677 194337435 19433499	Sk Ma Pd At Cs Ch Sc Cau Cag Js Pph Ppa	10 10 10 10 10 10 10 13 13 9 11 7 7	RTWAGPNI RTWHPSNR HSWVRANQ ENLYKTFE ENLYKTFE QNFYKMFE YRWAGAFD REWFNAFQ GRWIPAFQ GRWIPAFQ GGWYSAWT		GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT	RMIFGFET RMLFGFDS FLVMGFAS FLVMGFAS FLMLGFSS HLLLGFNS HMLLGFNS HMLLGFNS HMLLGFNS HMLLGFNS HQLLSFRT HQVCGFRT	TSID VSWD ICWD TMYL TMYL STMYL STMYL STMYL STMYL STMAD SNMAD SNMAD SNMAD TGWYL TASM	SNSDDDDVVGDSS	GDYGI PNYGI GRYGJ SREAN SREAN SREG TTEGS -AFGC -AFGC 9 FAMYS GRVEI 2 QDISI	KFFWEEWI ANFWHHW VDFVKFL' VDFVKFL' SMFARTLI SNLAANMI SQLMNNMI SAWLYYW SQYAINLI YFGSRII YFGIRMI	RAG NKN QMG TGY TGE VNG LDD KMP SGS LSG KSG	1 1 3 11 11 11 6	QTYCDI KPLSTI KPFAQJ DSFVKI LSFKEI YTIKTI APVRQJ LTIAQJ IPVREJ QKYIDJ AAVWQJ GAVWQJ	AWLNAS AWLDAS AWLDGA AFVKAA AFIKAA AFIEAA AWVTTA AWVKTA AWVKTA AWVKTA AWFDAJ SWFDAJ	SWDI SWRIA AWDIA ARIY ASIY AKKY AIEV AFDMI AFDMI NEIVI ANSVI ISLY UMY	YHGQA AHDQA APDQS 2PQRQ 2PQRK 2PQRK 2PQRK 2PQRK 2PQDE HAGK- WAGK- EGSNV 3DGSE 3DGSE	PPP4S3VPPT3CR 3CR 3CR	SVAAVG SVVACG VACACA IVRIMG LVRIMG IATVIG IYAVMG AYIYAR AYIYAR AVIYAR AVIWAY GAAVMY GAAVMY	A 245 P 264 Y 237 Y 232 A 614 S 677 S 676 E 195 P 206 P 213 P 213	[561] [584] [281] [282] [275] [653] [711] [710] [227] [238] [240] [240]
108743329 166367720 119386630 222528938 146297187 78044552 162452479 118048019 163848691 84497076 194337677 194337435 194337435 19433499	Sk Ma Pd At Cs Ch Sc Cau Cag Js Pph Pph Pa	10 10 10 10 10 10 10 10 10 10 10 10 10 1	RTWAGPNI RTWHPSNR HSWVRANQ ENIYKTFE QNFYKMFE YRWAGAFD REWFNAFQ RRWFNAFQ GRWIPAFQ SGWWGVFH ADWYSAWT GDWYSAWT	7	GFF GWF GLF GAT GAT GAT GAT GAT GAT GAT GAT SCH SCH SCH SCH SCH SCH SCH SCH SCH SCH	RMIFGFET RMLFGFDS FLVMGFAS FLVMGFAS FLMLGFAS HLLLGYAT HMLLGFNS HMLLGFNS HMLLGFNS HMLLGFNS HQUCGFRT QVCGFRT ATFLGFHG	TSID VSWD ICWD TMYL TMYL STMYL STMYL STMYL STMAD SNMAD SNMAD SNMAD THSYS TGWYL TASM	SNSDDDDVVGDSSS	GDYGI PNYGI GRYGJ SREAN SREAN SREAN SREAN TTEGS -AFGC -AFGC 9 FAMYS GRVEI 2 QDISJ 2 QDISJ 2 QDISJ 8 KHT	KFFWEEWI ANFWHHW VDFVKFL' VDFVKFL' SMFARTLI SNLAANMI SQLMNNMI SAWLYYW OQYAINLI VYFGIRMI INFEDFVI	RAG NKN QMG TGY TGE VNG LDD KMP SGS SGS KSG KAG		QTYCDI KPLSTI KPFAQI DSFVKI LSFKEJ YTIKTI APVRQI LTIAQI IPVREJ QKYIDI AAVWQI GAVWQI GAVWQI	AWLNAS AWLDAS AWLDGA AFVKAA AFIKAA AFIEAA AWVTTA AWVKTA AWVKTA AWVKTA AWFDAJ SWFDAJ SWFDAJ	WDI WRI ARIY ASIY ASIY ASIY AFDMI AFDMI AFDMI AFDMI ASIY USIY USIY UMMY ATBBI	YHGQA AHDQA APDQS 2PQRQ 2PQRK 2PQRK 2PQRA 2PDDE HAGK- VAGK- EGSNV 3DGSE 2NGSE 2NGSE	PPP4SS 4SSVPPT 3CRR 3CRR 3	SVAAVG SVVACG. VACACA IVRIMG IATVIG IATVIG AYIYAR AYIYAR WAYLRA AVIYAR GAAVMY GAAVMY GAAVMY	1   231     1   237     2   237     1   237     237	[561] [584] [282] [275] [653] [711] [710] [227] [238] [240] [240] [240]
108743329 166367720 119386630 222528938 146297187 78044552 162452479 118048019 163848691 84497076 194337677 194337435 194334999 149924090 154150967	Sk Ma Pd At Cs Ch Sc Cau Cag Js Pph Pph Pa Ppa	10 10 10 10 10 10 10 13 13 9 11 7 7 8	RTWAGPNI RTWHPSNR HSWVRANQ ENIYKTFE ENIYKTFE YRWAGAFD REWFNAFQ GRWIPAFQ GRWIPAFQ GRWIPAFQ GRWIPAFQ GRWIPAFQ SAWFNAFQ GDWYSAWT SAYFHADG KRWGAIN	7 7	GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT	RMIFGFET RMLFGFUET RMLFGFDS FLVMGFAS FLVMGFAS FLMLGFSS HLLLGYAT HMLLGFNS HYMLGFHN HQLLSFRT HQVCGFRT HQVCGFRT HQVCGFRT HQVCGFRT	TSID VSWD TCWD TMYL TMYL TMYL TSFD TMAD TSST TGWY TASM TASM TASM	S X S D D D D V V G D S S S	GDYGI PNYGI GRYGJ SREAN SREAN SREGJ TTEGG -AFGG -AFGG 9 FAMY1 GRVEI 2 QDISI RKHTI TDI-T	KKFWEEWI ANFWHHW VDFVKFL VDFVKFL SMFARTLI SNLAANMI SAWLYW SQLMNNM SAWLYW SQYAINLI AYFGSRI IVFGIRMI INFEDFVI	RAG NKN QMG TGY TGE VNG LDD KMP RMP SGS LSG KSG KAG DNS		QTYCDJ KPLSTJ KPFAQJ DSFVKJ LSFKEJ YTIKTJ APVRQJ LTIAQJ IPVREJ QKYIDJ AAVWQJ GAVWQJ GAVWQS NGLGD	AWLNAS AWLDGA AVLDGA AFVKAA AFIKAA AFIKAA AWVTTA AWVKTA AWVKTA AWVKTA AWFDA1 AWFDA1 SWFDA1 SWFDA1 SWFDA1 SWFDA1	WDI SWRIA AWDIA ARIY( ASIY( ASIY( AFDMI AFDMI AFDMI AFDMI AFDMI SLY( UMMY( ATRRI	YHGQA AHDQA APDQS DPQRQ DPQRQ DPQRK DPQRQ	PP 4 SS 4 SS VP TC R C R C R C A	SVAAVG SVVACG. VACACA IVRIMG IVRIMG IATVIG IATVIG AVIYAR AVIYAR AVIYAR AVIYAR GAAVMY GAAVMY AVAIIF AVAIIF	1 231   1 231   245 245   1 237   2 237   2 237   2 237   2 237   2 237   2 237   2 237   2 237   2 232   2 195   2 195   2 195   2 213   2 232   2 232   2 807	[561] [584] [282] [275] [653] [711] [710] [227] [238] [240] [240] [275] [840]
108743329 166367720 119386630 222528938 146297187 78044552 162452479 118048019 163848691 84497076 194337435 194337435 194334999 149924090 154150967 20089394	Sk Ma Pd At Cs Cau Cag Js Pph Pa Ppa Ppa	10 10 10 10 10 13 13 9 11 7 7 8 3 2	RTWAGPNI RTWHPSNR HSWVRANQ ENIYKTFE ENIYKTFE VRWAGAFD REWFNAFQ GRWIPAFQ GRWIPAFQ GRWIPAFQ GRWIPAFQ GRWYSAWT GDWYSAWT SAYFHADG CDFSSCIM	77	GAN GAN GAN GAN GAN GAN GAN RLH RLH GVH SVH SVH SVH SVH SVH SVH SVH SVH SVH S	MIFGFET MLFGFDS TLVMGFAS TLVMGFAS TLUMGFSS HLLLGYAT MLLGFNS HYMLGFNS HYMLGFNS HYMLGFNS HYMLGFNS HYMLGFRT HQVCGFRT TTFLGFEG HGLLGFQS	TSID VSWD TCWD TMYL TMYL TMYL TSFD NMAD SN	S X S D D D D D V V G D S S S S S	GDYGI PNYGI GRYGI SREAN SREAT SREAT TTEGI -AFGG 9 FANG 9 FANG 9 QUISI 2 QDISI 2 QDISI RKHTT TDLPI -DLGI	KEFWELM KEFWELM VDFVKFL/ VDFVKFL/ VDFVKFL/ IEFGQKL IEFGQKL SALAANM SAL	RAG NKN QMG TGY TGE VNG LDD KMP SGS SGS KSG KSG KSG KSG FVR		QTYCD KPLST KPFAQ DSFVK LSFKEJ YTIKTJ APVRQ LTIAQJ LTIAQJ IPVREJ QKYIDJ AAVWQ GAVWQ: NGLGDJ ESIVQ: YTUKM	AWLNAS AWLDGA AVLDGA AFVKAA AFIKAA AFLEAA AWVTTA AWVKTA AWVKTA AWVKTA AWFDA1 SWFDA1 SWFDA1 SWFDA1 SVFDA1 SVF20	WDI SWRIA AWDIA ARIY ASIY ASIY ASIY AFDMI AFDMI AFDMI AFDMI SLY UMMY CIRRI CIAS ADVD	YHGQA AHDQA APDQS DPQRQ DPQRQ DPQRK DPQRK DPQRK DPQRK DPQRK DPQRK DPQRK ARGK- EGSNV RSGVY SDGSE SNGSE PIGSN DAGY- DPSGV	PPP4SSVPPTCRRCA 3CRRCA 3CRRCA 3CA	SVAAVG SVVACG. VACACA IVRIMG IVRIMG IATVIG AYIYAR AYIYAR WAYLRA AVIWAY GAAVMY GAAVMY GAAVMY AVIIF PAVIFA	1 231   2 237   2 245   Y 237   Y 237   Y 237   Y 237   Y 237   Y 237   Y 232   A 614   S 676   S 676   S 195   P 206   P 213	[561] [584] [282] [275] [653] [711] [710] [227] [238] [240] [240] [240] [240] [255] [840]
108743329 166367720 119386630 222528938 146297187 78044552 162452479 118048019 163848691 84497076 194337677 194337435 194334999 149924090 154150967 20089394 126178074	Sk Ma Pd At Cs Ch Sc Cau Cau Js Pph Pa Ppa Ppa CMD	10 10 10 10 10 13 13 9 11 7 7 8 3 2 6	RTWAGPNI RTWAGPNI RTWHPSNR HSWVRANQ ENIYKTFE ENIYKTFE QNFYKMFE YRWAGAFD REWFNAFQ GRWIPAFQ GRWIPAFQ GRWIPAFQ GRWIPAFQ SGWWGVFH ADWYSAWT SAYFHADG KRWGGALN GDFSSSLN	7	GFF GWF GLF GAT GAT GAT GAT GAT GAT STA STA STA STA STA STA STA STA STA S	MIFGFET MLFGFES FLVMGFAS FLMGFSS FLMLGFSS HLLGYAT MLLGFNS HMLGFNS HMLGFNS HMLGFNS HQVGFRT 40VCGFRT GUCGFRT GUCGFST HLLGFST	TSID VSWD TCWD TMYL TMYL TMYL TSFD TMAD TSFD TMAD TASM TASM TASM TASM TASM TSYD SKYV	SNSDDDDVVGDSSSS	GDYGI PNYGI GRYGJ SREAN SREAT SREAT SREAT SREAT GRYGI 2 QDISJ 2 QDISJ 2 QDISJ 2 QDISJ 2 QDISJ 2 QDISJ 5 RKHTT TDLP -DLGI SDPCT	NREWLAN KFFWEENW NDFVKFL VDFVKFL VDFVKFL SMFARTL SMFARTL SMFARTL SMFARTL SALANNMI SAWLYYW SAWLYYW SAWLYYW NFEDFVI NFEDFVI NFEDFVI NFEDFVI NFEDFVI NFEDFVI SAWLYN SMFUHDW SAWLYN SMFUHDW SAWLYN SMFUHDW SAWLYN SMFUHDW SMFUH	RAG NKN QMG TGY TGE VNG LDD KMP SGS SGS KAG KAG DNS FVR VDD		QTYCDJ KPLSTJ KPFAQJ DSFVKJ LSFKEJ YTIKTJ APVRQJ LTIAQJ LTIAQJ LTIAQJ GAVWQJ GAVWQJ GAVWQJ MGLGDJ ESIVQJ YTVKNS	AWLNAS AWLDAS AWLDGA AFUKAA AFUKAA AFIKAA AFIEAA AWVTTA AWVTTA AWVKTA AWVKTA AWFDAI SWFDAI SWFDAI SWFDAI SWFDAI SWFJGA	SWDIS SWRIA AWDIA ARIYO ASIYO AKKYO AFDMI AFDMI AFDMI AFDMI ISLYO UMMYO JTRRI CIASO ADVDO	YHGQA AHDQA APDQS 2PQRQ 2PQRA 2PDDE HAGK- 2PDDE HAGK- CGSNV AGK- EGSNV 3DGSE 2IGSN 2AGY- 2PSGV	PPP 4 SSV 3 SV PPT 3 RR 3 CA 3 CA	SVAAVG SVVACG VACACA IVRIMG IVRIMG ILVRIMG IATVIG AYIYAR AYIYAR WAYLRA AVIWAY GAAVMY GAAVMY AVIAIF PAVIFA AKVFGE AAXIP	1   231     2   245     P   264     Y   237     Y   206     P   213     Y   224     Y   224	[561] [584] [282] [275] [653] [711] [710] [227] [238] [240] [240] [240] [275] [840] [256] [256]
108743329 166367720 119386630 222528938 146297187 78044552 162452479 118048019 163848691 84497076 194337677 194337435 19433499 149924090 154150967 20089394 126178074	Sk Ma Pd At Cs Ch Sc Cau Cag Js Pph Pa Ppa Ppa Mac Mac	10 10 10 10 13 13 9 11 7 7 7 8 3 2 6	RTWAGPNI RTWAGPNI RTWHPSNR HSWVRANQ ENIYKTFE QNFYKMFE YRWAGAFD REWFNAFQ RRWFNAFQ GRWIPAFQ SGWWGVFH ADWYSAWT GDWYSAWT GDWYSAWT GDFSSSLN TNWYNAFG DRWIOTTC	7	GFF GWH GLE GAN GAN GAN GAN GAN GAN GAN GAN GAN GVH NF7 TSI GAN GLE GLE GLE GLE GLE GLE GLE GLE GLE GVH	MIFGFET RMLFGYET RUNGFAS FLWGFAS FLMLGFSS FLMLGFSS HLLGYAT MLLGFNS HMLLGFNS HQVCGFRT QVCGFRT QVCGFRT GILGYQS HLLGFST HSMTGFHT	TSID VSWD TCWD TMYL TMYL TMYL TSFD TMAD TASM TASM TASM TASM	SNSDDDDVVGDSSSSV	GDYGI PNYGI GRYGJ SREAN SREAN SREG TTEG -AFG -AFG GRVEI 2 QDISI 2 QDISI 2 QDISI RKHT TDLPI -DLGI SDRG	ALE WELL AN FWHEW VDFVKFL VDFVKFL SMFARTL SMLAANM SALAANM SALAANM VC SQLMNNM SALAANM VC SALAY	RAG NKN QMG TGY TGE VNG KNP RMP SGS SGS KSG KSG KSG KSG VDD VGT		QTYCD KPLST KPFAQ DSFVK LSFKE LSFKE LSFKE QKYID AAVWQ GAVWQ GAVWQ GAVWQ GAVWQ GAVWQ GAVWQ GAVWQ GAVWQ GAVWQ GAVWQ GAVWQ C I I I I I I I I I I I I I I I I I I	AWLDAS AWLDAS AWLDGA AFVKAA AFVKAA AFIKAA AVVTTA AWVKTA AWVKTA AWVKTA AWVKTA AWVDAI SWFDAI SWFDAI SWFDAI SWFDAI SWFJAA SWFYGA AWFIAA AWFIAA	SWDIS SWRIA AWDIA ARIYO ASIYO AKKYO AFDMI AFDMI AFDMI AFDMI ISLYO ISLYO ISLYO ISLYO ANSVI ISLYO ISLYO ANDO ISLYO ANDO ISLYO ISLYO	YHGQA AHDQA APDQS 2PQRQ 2PQRX 2PQRA 2PDDE HAGK- VAGK- CGSNV RSGVY RSGVY 2GSNV 2GSNV 2GSNV 2GSNV 2GSNV 2GSV 2DGSE 2DGSE 2PSGV 2PSGV	4 SS V P P T C R R C A T Y P	SVAAVG SVVACG. VACACA VACACA IVRIMG IVRIMG IIVRIMG AVIYAR AVIYAR AVIYAR AVIYAR GAAVMY GAAVMY GAAVMY AVIIFA AVAIIF AVAIFA AVAIAR AVAIAR	1   231     1   231     1   245     1   245     1   237     1   236     1   236     1   236     1   236     1   236     1   237     1   236     1   237     1   238     1   237     1 <t< th=""><th>[561] [584] [281] [282] [275] [653] [711] [227] [238] [240] [240] [240] [240] [240] [275] [840] [256] [256]</th></t<>	[561] [584] [281] [282] [275] [653] [711] [227] [238] [240] [240] [240] [240] [240] [275] [840] [256] [256]
108743329 166367720 119386630 22528938 146297187 78044552 162452479 118048019 163848691 84497076 194337677 194337435 194337435 194334999 149924090 154150967 20089394 126178074 11499161 198977023	Sk Ma Pd At Cs Ch Sc Cau Cag Js Pph Pph Ppa Ppa CMb Mac Mm Af	10 10 10 10 13 13 9 11 7 7 8 3 2 6 6	RTWAGPNI RTWHPSNR HSWVRANQ ENIYKTFE QNFYKMFE YRWAGAFD REWFNAFQ GRWIPAFQ GRWIPAFQ GRWIPAFQ GRWIPAFQ GDWYSAWT GDWYSAWT GDWYSAWT GDFSSSLN TNWYNAFG DDWLQTLQ	77	GFFF GWFF GLE GAT GAT GAT GAT GAT GVFF GVFF GVFF GVFF GVFF GAT GAT GAT GAT GAT	MIFGFET RMLFGYET RMLFGFDS FLVMGFAS FLMGFAS FLLGYAT HILLGFNS HILLGFNS HILLGFNS HULLGFNS HULGFAT HOVCGFRT HOVCGFRT HOVCGFRT HOVCGFRT HILGFAS HILLGFAS	TSID VSWD TCWD TCWD TMYL TMYL TSFD SMMAD S	SNSDDDDVVGDSSSSVSA	GDYGI PNYGI GRYGI SREAN SREAN SREGI TTEGI -AFGG -AFGG 2 QDISI RKHTI TDLPR -DLGI SDRGI NADFI	ALE WEAK AN FWHHW VDFVKFL VDFVKFL PFGQKL SMFARTL SNLAANMI SOLMNNM SAWLYYS SAWLYYS SAWLYYS SAWLYYS SAWLYYS SAWLYYS SAWLYYS SAWLYYS SAWLYYS SAWLYYS SAWLYYS SAWLYYS SAWLYYS SAWLYYS SAWLYYS SAWLY	RAG NKN QMG TGY TGE VNG LDD KMP SGS SGS KSG KSG KSG KSG VDD VGT		QTYCD KPLST KPFAQJ DSFVKL LSFKEJ YTIKTJ APVRQJ LTIAQJ LTIAQJ LTIAQJ GAVWQJ GAVWQJ GAVWQJ GAVWQJ YTVKN: QPIITJ EKVESJ EKVESJ	AWLDAS AWLDGA AFIKAA AFIKAA AFIKAA AFIKAA AWVTTF AWVKTF AWVKTF AWVKTF AWVKTF AWVKTF AWFDAI SWFDAI SWFDAI SWFYGF AWFIAF AFFSTF	WDI SWRIA AWDIA ARIYO ASIYO AFDMI AFDMI AFDMI AFDMI SLYO UWMYO TRRI CIASO ADVDO AKDKSI	YHGQA AHDQA APDQS 2PQRQ 2PQRK 2PQRA 2PDDE HAGK- SAGK- SAGK- SAGK- SAGK- SAGK- SAGK- SAGK- SAGK- 2PGSN 2PGSN 2PSGV 2PSGV 2PSGV 2PSGV 2PSGV	PPPPS 4 SSVPPTCRRCATYR 3 RCATYR 2 2	SVAAVG SVVACG. VACACA IVRIMG IVRIMG ILVRIMG AYIYAR AYIYAR AYIYAR AYIYAR AYIYAR AYIYAR AYIYAR AYIYAR AYIYAR AYIYAR AYIYAR AYIYAR AYIYA AXIF AXIF AAALAR ARIIAE	1   231     1   237     2   245     P   264     Y   237     Y   2313     P   2132     P   2132     P   2132     P   2132     P   2234     Y   234	[561] [584] [282] [275] [653] [711] [710] [227] [238] [240] [240] [240] [240] [256] [247] [256] [247] [263]
108743329 166367720 119386630 222528938 146297187 78044552 162452479 118048019 163848691 84497076 194337677 194337435 194334999 149924090 154150967 20089394 126178074 11499161 18977023 73669560	Sk Ma Pd Cs Cs Cau Cau Cau Cau Pph Pa Pph Pa Ppa Mac Mm Af	10 10 10 10 13 13 9 11 7 7 8 3 2 6 6 6 6	RTWAGPNI RTWAGPNI RTWHPSNR HSWVRANQ ENIYKTFE ENIYKTFE VRWAGAFD REWFNAFQ GRWIPAFQ GRWIPAFQ GRWIPAFQ GRWIPAFQ GRWIPAFQ GDWYSAWT GDWYSAWT GDFSSSLN TNWYNAFG DDWLQTLQ NNWLEALA	77	GFF GWH GLL GAT GAT GAT GAT GAT GAT GAT GVH NFF TSE GAT GLL GVH GLL GVH GAT	MIFGFET MLFGFET MLFGFDS FLVMGFAS FLVMGFAS FLMLGFSAS HLLLGYAT MLLGFNS HYMLGFNS HYMLGFET HOVCGFRT HOVCGFRT HOVCGFRT HOUCGFRT HSMTGFHT HMULGFAS HMLLGFAS HMLLGFAS HMLLGFAS HMLGF	TSID VSWD TCWD TCWD TMYL TMYL TMYL TSFD SMAD TSSMAD TASM TASM TASM TASM TASM TASM TASM TASM	S N S D D D D V V G D S S S S V S A N	GDYGi PNYGi GRYGJ SREAN SREG TTEG -AFGG 9 FAMYI 2 QDISJ 2 QDISJ RKHTI TDLPI -DLGJ SDRG NADFI LGDA	NE WEAK AN FWHEW AN FWHEW VDFVKFL' VDFVKFL' FE FGQKL' SMEARTL' SNLAANMI GOLMNNM GOLMNNM GOLMNNM GOLMNNM SAULYY SAULYY VFGSRI MYFGSRI MYFGSRI MYFGSRI MYFGSRI SNFVHRM PELAYRL' VERTYW	RAG NKN QMG TGY TGE VNG LDD KMP SGS KAG DNS FVR VDD VGT TGT		QTYCDZ KPLSTZ KPFAQJ DSFVKL LSFKEJ YTIKTJ APVRQZ LTIAQJ LTIAQJ LTIAQJ GAVWQZ GAVWQZ GAVWQZ GAVWQZ GAVWQZ SZ SZ YTVKN QPIITZ EKVESZ ESVHDZ	AWLDAS AWLDAS AWLDAS AFUKAA AFUKAA AFUKAA AFUKAA AWVKTA AWVKTA AWVKTA AWVKTA AWVFDAI SWFDAI SWFDAI SWFDAI SWFDAI SWFJAA AFFSTE AFFSTE AFFDTY	SWDIY SWRIA AWDIA ARIY ASIY ASIY ASIY ASIY ASIY ASIY ASIY AS	YHGQA AHDQA APDQS 2PQRQ 2PQRK 2PQRK 2PQRK 2PQRK 2PQRK 2PQRK 2PQRK 2PQRK 2PQRK 2PQRK 2PQRK 2PQRK 2SQV 2DQSQV 2PSGV 2PSGV 2PSGVHD 2GTHN 2GTHN	PPPPS 4 SSVPPTCRRCATYRI 3 CATYRI 2 2	SVAAVG SVVACG. VACACA IVRIMG LIVRIMG LIVRIMG IIATVIG IIATVIG AYIYAR AYIYAR AYIYAR AYIYAR AYIYAR AYIYAR AYIYA GAAVMY GAAVMY AVAIIF AKVFGE AAALAR ARIIAE ARIIAE	1   231     1   231     2   245     P   264     Y   237     Y   238     Y   234     Y   235     Y   236     Y   237	[561] [584] [282] [275] [653] [711] [710] [227] [238] [240] [240] [275] [840] [256] [256] [256] [256] [256] [256]
108743329 166367720 119386630 222528938 146297187 78044552 162452479 118048019 163848691 84497076 194337677 194337435 194334999 149924090 154150967 20089394 126178074 11499161 18977023 73669560	Sk Ma Pd Cs Ch Sc Cag Js Pph Pa Pph Pa CMO Mac Ma C Mo Mac	100 100 100 100 100 100 100 100 100 100	RTWAGPNI RTWAGPNI RTWHPSNR HSWVRANQ ENIYKTFE QNFYKMFE YRWAGAFD REWFNAFQ GRWIPAFQ GRWIPAFQ GRWIPAFQ GRWIPAFQ GRWIPAFQ SAYFHADG KRWGGALN GDFSSSLN TNWYNAFG DDWLQTLQ NNWLEALA NDWAGALE	77	GFF GWH GLL GAT GAT GAT GAT GAT GAT GAT GVH NFF TSE GAT GUH GUH GT	MIFGFET MLFGYET MLFGFDS FLVMGFAS FLMLGFSS HLLGYAT MLLGFNS MULGFNS MULGFNS GVCGFRT GVCGFRT GVCGFRT HSMTGFHT MVLGFAS MILGFAS MILGFAS	TSID VSWD TVSWD TTYL TMYL TMYL TSTMYL TSFD SMAD SMAD TSSFD SMAD TSSFD SMAD TSSFD SMAD TSSFD SSFT SSFT SSFT TST TST TST TST TST TST	S N S D D D D V V G D S S S S V S A N O	GDYGI PNYGI GRYGI SREAT SREAT SREAT SREAT SREAT GRVEI 2 QDISJ 2 QDISD 2 QDISJ 2 QDISJ 2 QDISJ	NREWENN ANFWHHW VDFVKFL VDFVKFL VDFVKFL SMFARTL SMFARTL SMFARTL SMFARTL SAMLYW SQLMNMM SQLMNMM SQLMNMM SQLMNMM VFGIRM VFGIRM VFGIRM VFGIRM VFGIRM VFGIRM VFGIRM VFGIRM SNFVHRW SELAYRL VKFFYT	RAG NKN QMG TGY TGE VNG LDD KMP SGS SGS KAG DNS FVR VDD VGT TGT TGT		QTYCD QTYCD KPFAQJ DSFVKL LSFKEJ YTIKT APVRQJ LTIAQJ IPVREJ QKYIDJ AAVWQJ GAVWQJ NGLGDI ESIVQJ YTVKNS QPIITI ESVHDJ DEICDJ UBVUSJ	AWLDAS AWLDAS AWLDAS AFUKAA AFUKAA AFUKAA AFUKAA AWUTTA AWUTTA AWUTTA AWUTTA AWUTAA AWFDA1 SWFDA1 SWFDA1 SWFDA1 SWFYGA AFFSTFA AFFSTFA AFFSTFA	SWDIY SWRIA ANDIARIY ASIY ASIY ANDIA AFDMH AFDMH AFDMH AFDMH CISLY	YHGQA AHDQA APDQS 2PQRQ 2PQRK 2PQR 2PDDE 2PQRK 2PQR 2PQRK 2PQR 2PQRK 2PQR 2PQR 2PQR 2PQR 2PQR 2PQR 2PQR 2PQR	4 SSVPPTCRRCATYR 3 CRRCATYR 3 CATYR 2 LR	SVAAVG SVACACA VACACA IVRIMG IVRIMG IIVRIMG AYIYAR WAYLRA AVIYAR WAYLRA AVIYAR WAYLRA AVIYAR WAYLRA AVIYAR AVIYA AVIYA AVIYA AVIYA AVIYA AVIYA AVIYA AVIYA AVIYA	1   231     1   231     1   231     2   245     P   264     2   237     Y   237     S   676     S   676     S   213     S   232     P   213     S   224     S   2200     N   234     N   225     Y   236     Y   236	[561] [584] [282] [275] [653] [711] [710] [227] [240] [240] [240] [240] [256] [247] [263] [255] [247] [263]
108743329 166367720 119386630 222528938 146297187 78044552 162452479 118048019 163848691 84497076 194337677 194337435 19433499 149924090 154150967 20089394 126178074 11499161 18977023 73669560 169846526	Sk Ma Pd Cs Ch Sc Cau Cau Cau Js Pph Pa Ppa Ppa CMb Mac Mb Cc	100 100 100 100 100 100 99 133 133 99 111 77 78 8 32 26 66 66 66 66 64 421	RTWAGPNI RTWAGPNI RTWHPSNR HSWVRANQ ENIYKTFE QNFYKMFE YRWAGAFD REWFNAFQ GRWIPAFQ SGWWGVFH ADWYSAWT GDWYSAWT GDWYSAWT GDFSSSLN TNWYNAFG DDWLQTLQ DNWLEALA NDWAGALE APWWDVFN	77	GFI GWI GAU GAU GAU GAU GAU GAU GVI GVI GVI GVI GVI GVI GUI SU GUI GUI SU GUI SU SU SU SU SU SU SU SU SU SU SU SU SU	MIFGFET RMLFGYET RMLFGFDS FLWGFAS FLMLGFSS FLMLGFSS HLLGYAT HMLGFNS HMLGFNS HMLGFNS HOVCGFRT HOVCGFRT HOVCGFRT HOVCGFRT HMLGFAS HMLGFAS HMLGFAS HMLGFAS HMLGFAS HMLGFTT HAVLSYRT	TSID VSWD TVSWD TSTMYLL TMYLL TSTMYLL TSFD SMAD SMAD TSST TASM TASM TASM TASM TASM TASM TASM	S N S D D D D V V G D S S S S V S A N Q	GDYGI PNYGI GRYGJ SREAN SREAN SREG7 TTEG -AFGG -AFGG -AFGG QDISI QDISI RKHT TDLPI -DLGI SDGG NADFI LGDA1 EELVI DHAT	ARE WEAK AN FWHEW AN FWHEW VDFVKFL' IEFCGKL' SMFARTL' SNLAANM SOLMNNM SOLMNNM SOLANNM VFGSRI VYFGSRI V	RAG NKN QMG TGY TGE VNG LDD KMP SGS SGS KAG DNS FVR VDD VGT TGT TGT		QTYCD KFLST KFFAQ DSFVK LSFKEJ YTIKT APVRQ LTIAQ LTIAQ IPVREJ QKYID AAVWQ GAVWQ GAVWQ GAVWQ GAVWQ YTVKN QPIIT EKVESJ DEICD VPVVEJ	AWLDAS AWLDAS AWLDAS AWLDAS AFIKAA AFIKAA AFIKAA AWFIKAA AWVKTA AWVKTA AWVKTA AWVKTA AWFDAI SWFDAI SWFDAI SWFDAI SWFYGA AWFIAA AFFSTE AFFDTY AWWFAI	WDIY SWRIA ARIY ASIY ASIY ASIY ASIY ASIY ASIY ASIY AS	YHGQA AHDQA AHDQA APDQS 2PQRQ 2PQRK 2PQRA 2PQRK 2PQRA 2PQRK 2PQRA 2PQRK 2PQRA	PPP 4 SS 4 SS PPT 3 CRR 3 A TYR 2 L R R	SVAAVG SVACACA VACACA IVRIMG IVRIMG IIVRIMG AYIYAR AYIYAR WAYLRA AVIYAR GAAVMY GAAVMY GAAVMY AVAIIF PAVIFA AAVAIIF AAVAIAR ARIIAE ARIIAE AAMIAD	A 245 P 264 Y 237 Y 237 Y 237 Y 237 Y 232 A 614 S 677 S 676 S 676 S 676 S 195 P 206 P 213 P 208 F 201 F 201	[561] [584] [281] [275] [653] [711] [275] [238] [240] [240] [275] [240] [275] [840] [256] [247] [256] [247] [256] [247] [256] [247] [256] [247]
108743329 166367720 119386630 222528938 146297187 78044552 162452479 118048019 163848691 84497076 194337435 194337435 194334999 149924090 154150967 20089394 126178074 11499161 18977023 73669560 169846526 119390377 2j3:	Sk Ma Pd Cs Ch Sc Cau Cau Cau Cau Dp Pph Pa Ppa Ppa Ma CMb Ma Cc 2A Hs	100 100 100 100 100 100 100 100 100 100	NTWACHNIN RTWACHNIN RTWACHNIN ENIYKTFE ENIYKTFE QRYYKTFE YRWACAFD REWFNAFQ SGWHCYFA ADWYSAWT GDWYSAWT GDWYSAWT GDWYSAWT GDFSSSLN TNWYNAFG DDWIQTLQ NNWLEALA NDWACALE APWWDYFN	77	GFF GWI GUL GAN GAN GAN GAN GAN GAN GAN GAN GUL GAN GUL GAN GUL GAN GAN GAN GAN GAN GAN GAN GAN GAN GAN	MIFGFET RMLFGYET RMLFGFDS TLVMGFAS TLMGFSS TLLGYAT HILLGFNS HILLGFNS HILLGFNS HULLGFNS HULGFNS HULGFST HSMTGFHI HNUGFAS HMILGFAS	TTSID VVSWD BICWD BICWD TTMYLL TTMYLL TSFMYL TSFM VNNAD UHSYS TTSFM TASM VNASQ UHSYS TTASM VNASQ USSYD VNCYD SKYV VNCYD SKYV VNCYD TATII TATII TATII TATI TATI TATI TATI	S N S D D D D V V G D S S S S V S A N Q Y	GDYGI PNYGI GRYGJ SREAN SREAN SREG TTEG -AFG -AFG -AFG QDISJ 2 QDISJ 2 QDISJ 2 QDISJ 8 KHTT DLPI -DLGJ SDRG NADFI LGDZ EELVI DHATJ 8 SWFIQ	ARE WEAK AN FWHHW VDFVKFL' VDFVKFL' PFGQKL' SNHARTL' SNLAANMI SQLMNNM SAWLYYS SAWLYS	RAG NKN QMG TGY TGE VNG LDD KMP SGS SGS KSG KAG DNS FVR VDD VGT TGT TGT IDD ALG		QTYCD QTYCD XVPIAC XVPIAC DSFVKD LSFKE YTIKT APVRO LTIAQU LTIAQU LTIAQU QKYID AAVWQ GAVWQ: GAVWQ: GAVWQ: GAVWQ: GAVWQ: GAVWQ: CFIT ESIVQ: YTVKN QPIIT EKVESS ESVHD DEICD VPVVE LEFMH	AWLDAS AWLDAS AFUKAA AFIKAA AFIKAA AFIEAA AWVTTA AWVTTA AWVTAA AWVTAA AWVTAA AWVTAA AWVTAA AWFDAI SWFDAI SWFDAI SWFDAI AWFTAA AFFSTF AFFSTF AFFTTY AWMKAY	WDIY WWDIY WWDIA ARIYQ ASIYQ ASIYQ ASIYQ ASIYQ ASIYQ ASIYQ ISIY ISIY	YHGQA AHDQA AHDQA APDQS 2PQRQ 2PQRK 2PQRK 2PQRK 2PQRK 2PQRK 2PQRK 2PGR 2PGR 2PGR 2PGR 2PGR 2GSV 2PSGV 2PSGV 2PSGV 2PSGV 2PSGV 2PSGV 2PSTT 2GVHD 2GVHD 2GTHN 7ESPV XIYRG ATEFE	P P P P P P P P P P P P P P P P P P P	SVAAVG SVVACG. VACACA IVRIMG LIVRIMG LIVRIMG LIVRIMG AYIYAR. A	1   231     1   231     1   231     2   244     2   237     2   237     2   237     2   237     2   237     2   237     2   237     2   237     2   237     2   232     2   232     2   239     2   239     2   232     2   232     2   232     2   232     2   232     2   232     2   232     2   220     1   2     1   2     1   2     1   2     1   2     1   2     1   3     1   3     1   3     1   3	[561] [584] [281] [275] [653] [711] [710] [227] [238] [240] [240] [240] [256] [240] [256] [247] [263] [255] [263] [255] [249] [679] [250]

Figure 1. Multiple sequence alignments of PrtH homologs and several structures of caspase-like fold. This alignment was made using PROMALS3D<sup>14</sup> followed by manual adjustment. Catalytic histidines and cysteines are shaded in black. Non-polar residues in positions with mainly hydrophobic residues are shaded in yellow. Small residues (G, A, S, C, T, N, D, V and P) in positions with mainly small residues are colored blue. Starting and ending residues numbers are shown in italic numbers and sequence lengths are shown in brackets. Insertion regions in the alignment are replaced by the numbers of residues. Consensus secondary structure predictions are shown above the alignment (*h*: α-helix; e: β-strand). The proteins are identified by their NCBI gene identification (gi) numbers, followed by the species name abbreviations. The last five sequences have known structures with a caspase-like fold. Their Protein Data Bank ids and chain ids are shown after the gi numbers (2j32, caspase 3; 1m72, caspase 1; 1cvr, gingipain; 3bij, a bacterial caspase homolog; 3eeb, cysteine protease domain from V. cholerae RTX toxin). Species name abbreviations are: Af, Archaeoglobus fulgidus; At, Anaerocellum thermophilum; Cag, Chloroflexus aggregans; Cau, Chloroflexus aurantiacus; Cc, Coprinopsis cinerea; Ch, Carboxydothermus hydrogenoformans; CMb, Candidatus Methanoregula boonei; Cs, Caldicellulosiruptor saccharolyticus; Gs, Geobacter sulfurreducens; Hs, Homo sapiens; Js, Janibacter sp.; Mac, Methanosarcina acetivorans; Ma, Microcystis aeruginosa; Mb, Methanosarcina barkeri; Mm, Methanoculleus marisnigri; Pa, Prosthecochloris aestuarii; Pd, Paracoccus denitrificans; Pf, Pyrococcus furiosus; Pg, Porphyromonas gingivalis; Pph, Pelodictyon phaeoclathratiforme; Ppa, Plesiocystis pacifica; Sc, Sorangium cellulosum; Sf, Spodoptera frugiperda; Sk, Streptomyces kanamyceticus; Tf, Tannerella forsythia; Vc, Vibrio cholerae. Bacterial, archaeal, and eukaryotic species name abbreviations are shown in black, red and green respectively.

7 DEMNEILCEKHPNN

7 GAFTGOLLRVWKNG

2 RTFGGVTMNGMFAMVEKYK

3 GSYRSFHKAIVRRMPPDOT

9 WTVFGDPS

21 KVSLSWDA

PNFFTAGT

343 [435]

264 [285]

207 [209]

7245522

1cvrA Pg

162330276 3bijA Gs

209870506 3eebA Vc

7 PCFAEALM

6 KGFGHOFI

8 TGTVAIIASTIDOYW

6 GLRVDVSVRSELAVD

45 YDTIQOKT 10 KASILLISGCODNOL

C-termini that include two consecutive  $\alpha$ -helices and one  $\beta$ -strand that is anti-parallel to the other  $\beta$ -strands in the central  $\beta$ -sheet. The alignment suggests that these secondary structural elements are also present in PrtH homologs (Fig. 1, the last three blocks).

Close homologs of PrtH have a limited phyletic distribution with less than 30 proteins. In addition to bacterial homologs, several PrtH homologs are from archaeal species and one is from eukaryotic species Coprinopsis cinerea (Fig. 1). Except PrtH, these proteins have not been experimentally characterized, and they are annotated as 'hypothetical protein' or 'predicted protein'. PrtH homologs with less than 300 residues contain only one domain with the caspaselike fold. The other PrtH homologs are longer and could contain additional domains. For example, HHpred searches suggested that gil118048089 and gil154150967 contain a PKD domain (with an immunoglobulin-like fold) and a beta-propeller domain respectively. The caspase-like domain in PrtH lies at the N-terminus, and the C-terminal region of PrtH (about 300aa) has unknown structure and function. A HHpred search using the C-terminal region of PrtH found weak similarity to Pfam family 'Fungalysin/Thermolysin Propeptide Motif' (Pfam accession number: PF07504, with a HHpred probability score of 88.2). This motif was found in some bacterial metalloproteases and was linked to assisting folding and inhibiting catalytic activity. The C-terminal region of PrtH could have similar functions. Some PrtH homologs contain predicted signal peptides (e.g., gil194337677 and gil212717887), suggesting that they are secreted exoproteins or located between the inner and outer membranes of Gram-negative bacteria. Conservation of the catalytic cysteine (except for gil119386630, where the cysteine is replaced by a serine) and histidine residues indicates that PrtH homologs also function as active proteolytic enzymes. The putative protease function of these proteins is also supported by genome context mining. For example, the gene of the PrtH homolog from M. aeruginosa (gil 166367720) shares an operon with another gene encoding a protein annotated as 'secreted metalloprotease', suggesting concerted proteolytic functions for them. The gene of another PrtH homolog from Caldicellulosiruptor saccharolyticus (gil146297187) neighbors a gene encoding with a hypothetical protein (gil146297186). Homologs of this hypothetical protein found by PSI-BLAST also frequently co-occur with proteins containing various protease domains.

## Acknowledgements

We would like to thank Lisa Kinch for critical reading of the manuscript and helpful suggestions. This work was supported in part by NIH grant GM67165 and Welch foundation grant 11505 to N.V.G.

### References

- 1. Tanner AC, et al. Periodontol 2000 2006; 42:88-113.
- 2. Saito T, et al. Infect Immun 1997; 65:4888-91.
- 3. Hamlet SM, et al. J Periodontol 2008; 79:144-9.
- 4. Nakajima T, et al. Biochem Biophys Res Commun 2006; 351:133-9.
- 5. Finn RD, et al. Nucleic Acids Res 2008; 36:281-8.
- 6. Marchler-Bauer A, et al. Nucleic Acids Res 2008.
- 7. Altschul SF, et al. Nucleic Acids Res 1997; 25:3389-402.
- 8. Soding J, et al. Nucleic Acids Res 2005; 33:244-8.
- 9. Ginalski K, et al. Bioinformatics 2003; 19:1015-8.
- 10. Rawlings ND, et al. Nucleic Acids Res 2008; 36:320-5.
- 11. Aravind L, et al. Proteins 2002; 46:355-67.
- 12. Pei J, et al. Protein Sci 2009; in press DOI:10.1002/pro.78.
- 13. Lupardus PJ, et al. Science 2008; 322:265-8.
- 14. Pei J, et al. Nucleic Acids Res 2008; 36:2295-300.