

Letter to the Editor

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

Jimin Pei^{1,*} and Nick V. Grishin^{1,2}

¹Howard Hughes Medical Institute; and ²Department 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.¹ 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*.² 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.² Clinical studies suggested that PrtH is a virulence factor, as the level of PrtH is associated with periodontal attachment loss.³ 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.⁴ Cloning of the *fdf* gene revealed that the original *prtH*

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.⁴ 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 Pfam⁵ and CDD⁶ did not yield significant predictions of known domains in this protein. A PSI-BLAST⁷ 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 gi1166367720 of *Microcystis aeruginosa*, a close homolog of PrtH, found a *Desulfovibrio salexigens* protein (gi1218151631) 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,⁸ 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,⁹ 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 database¹⁰ 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 HeiF.¹¹ 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.¹² 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 β -sheet and surrounding α -helices on both sides.¹¹ The structure of the cysteine protease domain from *V. cholerae* RTX toxin¹³ (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 β -strand preceding the catalytic histidine show high sequence divergence. Compared to other structures with a Rossmann-like fold, a distinct feature of caspases and gingipains lies in their

*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>

A caspase-like fold in *Tannerella forsythia* PrtH

		eeeeeee	hhhhhhhhh	eeee	eeeeeee	hhhhhhh	eeeeeeee
38707242 (PrtH)	Tf	30 IYGACSIIEDFL	12 AEGFLGYPTS	5 FIPANFWRKDG	20 DAVKVFYHSGHGNML	18 ALSKMAFANEK	12 LRYLFWSTCLS
108743329	Sk	59 WVGAFSIQKVF	11 ATGWLAYLQQ	4 FYDRNFWFADG	20 DAVCAVYHSGHGMD	18 AISNRMALGNEK	12 VNVVFWSTCDS
166367720	Ma	48 IYGGCSIEVGC	13 AQQWLDYVKK	6 FTPLNFWYQDC	20 DAVMAFYHSGHGMD	21 SSSDKLWVANER	12 VRYIFLSTCLS
119386630	Pd	71 IYGACSVETYSR	11 AGGFLDAVDR	4 FATPDFWRDDG	20 DSVRVFYHAGHGMD	19 LTRSDRMRLGSGA	12 LRYLFWSTCQS
222528938	At	41 PNGSCTYGDRA	GAHYVYNLAD	5 AWVTAYLIKLN	14 SQADLLVYSGHGLCF	29 ARTNEVRFHGSN	12 LKWIIMYTCNW
146297187	Cs	41 PNGSCTYGDRA	GAHYVYNLAD	5 AWVTAYLIKLN	14 SQADLLVYSGHGLCF	29 ARTNEVRFHGSN	12 LKWIIMYTCNW
78044552	Ch	39 LYPDCTYGDRA	GASVYNNLYN	4 KFTLGFYKLDN	13 NNVDFFSFAGHGKNY	30 ARTNETRFHGNK	12 LKVVNMYCQW
162452479	Sc	428 VVGAEWVGLCG	8 VDGLLKSFRE	1 GIDSRFNWGDQ	20 DSVDLTFYTGANGV	12 LHFSDAARWGAN	12 LEWMVVAACGP
118048019	Cau	475 SFGAEGNWDYQ	15 VTGFRSGMLG	1 GYTORFYWTNA	21 DRAAFVYVYAGHGPG	10 TWFSGANARYQN	12 LRWVGFASCTP
163848691	Cag	474 YFIEGNDWYD	15 VNGLRFGMLS	1 GYTORFFWSNS	21 DRAAFVYVYAGHGPG	10 TWFDTNARYQS	12 LRWVGFASCTP
84497076	Js	4 YVGSWWRIDTV	6 AEGFASTLGG	1 LSEFDWTQNHG	19 QDQVAMQMSHGPNP	9 NASEAIDFGKND	12 LEIFATHACDL
194337677	Pph	33 EGGFVFCREES	2 VDETYWVFLKH	1 SYEQYVWAVET	10 DNMDTAFFSFGHGNT	13 FASGAVSLGDVD	12 LEFLTIDACSV
194337435	Pph	37 EGGVYVSKTEG	2 TTPVWNFINN	1 SYDQYVWAEFF	10 DKMDFSIFVGHGSPW	14 GSTTHKGGWDYN	12 AEFVTLHACNV
194334999	Pa	37 EGGVYVSKTEG	2 TTPVWNFINN	1 DYNQYVWAEFF	10 DAMDFSIFVGHGSPW	14 GSFDAEGWDTN	12 SEFVTLHACNV
149924090	Ppa	18 EGTLYYVNDP	16 VGYLEDLDD	1 GFDSVYVYHGNR	25 DSADIGMYSHGKGN	26 AAGNDSWGGDT	12 LNAMIIDTCS
154150967	Cmb	599 TYSMTMIQNPG	19 HGVVQNWLET	2 GWSNTFYHTED	14 DNATFFYHFGHGHL	32 PSDVYKKGWQQ	12 NKWVFLDACSL
20089394	Mac	44 EVGVEVWYNDT	12 ANGLYNRLGN	1 GWTKSFNKGNS	17 DAVDIALYSGHGSD	10 VYHDEAEWGDYD	12 LEWVGLDCLCA
126178074	Mm	31 YCCAEEWNNYP	11 AQQFFYQLGG	2 SWSWGFYVYDGG	22 DDTHFAFSGHGAP	9 LSYSDALWNGT	12 MDWITIDACTV
11499161	Af	35 DVGAAVAFSED	8 VSQFVEIIS	2 TWVQHLNDYP	19 DYTELSLVLGHGVV	18 ALPDDIRLGYAS	6 SVWTFIICQSV
18977023	Pf	30 DVGGVMTFTDD	8 VKNFVEIIS	2 HWYQGLFKDYP	20 DYTELSLVLGHGCVY	13 ALPEKVRLLGYS	6 AIWTFIICQSV
73669560	Mb	28 SYSVTIIENYS	12 YGIVENLHD	2 GWNEYFYDSET	17 DDADFHVHLGHGVD	20 ASQVYKQWDTN	12 NEWVLLHACSI
169846526	Cc	435 RVGRYVVRNDI	3 TASARSLSS	10 DNTHFYWSRPF	10 NNVDIALTEAHGFSH	16 DIPAGGLNGYGA	9 LSYWVHSCSEV
119390377	2j32A Hs	10 EMGLCIIINNK	14 DVDAAANRET	2 NLKYEVRNRKN	20 RSSFVVCVLLSHGEG	9 DLKKITNFFRGD	8 PKLFIICQCRG
42543031	1m72A Sf	31 HRGMAIIFNHE	13 NVDSDNLSKV	2 TLGFKVTVFPN	20 ADCLLVAVLTHGELG	9 KPDNLWYFTAD	8 PKLFFIICQCRG
7245522	1cvrA Pg	141 WLQALCIASA	12 IQHENVIANL	2 QYGYTKIICKY	14 GGSLVNYTGHGSET	7 GTTHVKQLTNSN	1 LPFIFDVACVN
162330276	3bijA Gs	2 PKGIALLALGLN	17 EADAE DMAAI	2 ERGFAVTTLMT	21 GDIFMLSYSGHGQV	22 IDDELVALLGKF	4 RVLVFFSDSCHS
209870506	3eebA Vc	30 RFDGQIIVQME	5 AKAAANLAGK	-HAESSVVVQL	16 DGKLRWQLVGHGRDH	12 ADELAVKLAKFQ	12 PDHISIVGCSL

		hhhhhhh	eeeeeee	hhhhhhhhhhh	hhhhhhhhhhh	eeeeeee
38707242 (PrtH)	Tf	10 RTWRLANK	1 GLRMIIFGYESYDS	ARYGSEFWKQWKKG	KSFSDAFIEASWSLFRNQT	PVVCACGN 224 [536]
108743329	Sk	10 RTWAGPNI	GFRMIFGFETSIDSD	GDYGKKEWKEWRAG	QTYCDAWLNASWDIYHGQA	PSVAAVGT 251 [572]
166367720	Ma	10 RTWHPSNR	GWRMLFCYETVSWDN	PNYGKFFWEENKRN	KPLSTAWLDASWRIAHQDA	PSVACGA 245 [561]
119386630	Pd	10 HSWVRANQ	GLRMLFGFDSICWDS	GRYGANFHHWQMG	KPFAQAWLDGAWDIAPDQS	PVACACAP 264 [584]
222528938	At	10 ENIYKTFE	GATLVMGFASTMYLD	SREAVDFVKFLTGE	1 DSFVKAFVKAARIYQPRQ	4 SIVRIMGY 237 [281]
146297187	Cs	10 ENIYKTFE	GATLVMGFASTMYLD	SREAVDFVKFLTGE	1 LSFKEAFIKAASIYQPRK	4 SLVRIMGY 237 [282]
78044552	Ch	10 QNFYKMEF	GATLMLGFSVVMYLD	SREGTEFGQKLNVG	YTIKTAFLAEAAKKYQPRQ	3 SIATVIGY 232 [275]
162452479	Sc	9 YRWAGAFD	GLHLLLGYATESFD	TTEGSMFARTLLDD	3 APVRQAWVTAIEVQPDDE	VIYAVMGA 614 [653]
118048019	Cau	13 REWFNAFQ	GAHMLLGFNSNMADV	-AFGGNLAANMKMP	11 LTIQAQAWKTAAPDMHAGK-	PAYIYARS 677 [711]
163848691	Cag	13 REWFNAFQ	GAHMLLGFNSNMADV	-AFGGQLMNNMRMP	11 LTIQAQAWKTAAPDMNAGK-	PAYIYARS 676 [710]
84497076	Js	9 GRWIPAFQ	RLHYMLGFHNHSYSG	9 FAMYSAWLYWNSGS	6 IPVREAWAEANEIVEGNSV	TWAYLRAE 195 [227]
194337677	Pph	11 SGWGWVPH	RLHQLLSFRRTGWYD	GRVEDQYAINLLSG	KQYIDAWFNAANSVRSRSGVY	3 CAVIWAYP 206 [238]
194337435	Pph	7 ADWYSAWT	7 GVHVQVCGFRNTASMS	2 QDISAYFGSRIKSG	AAVWQAWFDAISLYDGGSE	RGAAMVYP 213 [240]
194334999	Pa	7 GDWYSAWT	7 GVHVQVCGFRNTASQS	2 QDISNYFGIRMKDG	GAVWQSWFDAIWMYNGSGE	RGAAMVYP 213 [240]
149924090	Ppa	8 SAYFHADG	NFATFLGFHGLSYDS	RKHTNNEFEDFVDS	2 NGLGDNWVDELTRRPIGSN	3 CAVAIIFE 232 [275]
154150967	Cmb	3 KRNGGALN	TSHGILGYSQSSYYS	TDLPNAPFFYNVAVR	1 ESIVQSYEATIASQAGY-	APAVIFAT 801 [840]
20089394	Mac	2 GDFSSSLN	GAHLILGFSTNCDYDS	-DLGLHWANLVVDD	6 YTVKNSWFYGDVDVQPSGV	TAKVFGET 224 [256]
126178074	Mm	6 TNWYNARF	GLHSMTGFHEKCHDV	SDRGSNFEVHRMVTG	3 QPIITAWFIAAKDTEPSTT	YAAALARE 220 [247]
11499161	Af	6 DDWLQTLQ	GVHMLVGFASATIS	NADFPPELAYRLTGT	5 EKVESAFFSTFLKSDGVHD	2 RARIIAEN 234 [263]
18977023	Pf	6 NNWLEALA	GIHMLGFAKEAKIA	LGDATELAYRLTGT	5 ESVDHAFEDTYVAYDGTSH	2 IARIIAEN 225 [255]
73669560	Mb	4 NDWAGALE	YSHGILGFTPEAFN	EELVDFKFTYTIID	1 DEICDANWFATVETTESPV	RAAMIADT 216 [249]
169846526	Cc	21 APWWDVFN	GLHAVLSYRTQMYIQ	DHATAEFGRSVALG	VFVVEAWMKAVSGDRIYRG	RKTYYSRN 637 [679]
119390377	2j32A Hs	16 MACHKIPV	EADFLYAYSTAPGY	8 SWFTQSLCAMLKQY	3 LEFMHILTRVNRKVAEFE	14 PCIVSMLT 242 [250]
42543031	1m72A Sf	18 STSYRIPV	HADFLIAFSTVPGYF	8 SWFMQALCEELRYA	3 RDILTLTLTFVQCQKVALDFE	15 PCITSMLT 260 [272]
7245522	1cvrA Pg	7 PCFAEALM	8 TGTVAIIASTIDQYV	7 DEMNEILCEKHPNN	2 RTFGGVTMNGMFAMVEKYK	9 WTVFGDPS 343 [435]
162330276	3bijA Gs	45 YDTIQKKT	10 KASILLISGGQDNQL	7 GAFTGQLLRVWKN	3 GSYRSEHKALVRRMPPDQT	PNFFTAGT 264 [285]
209870506	3eebA Vc	6 KGFHQFI	6 GLRVDSVSRSELAVD	-----	-----	21 KVSLSWDA 207 [209]

Figure 1. Multiple sequence alignments of PrtH homologs and several structures of caspase-like fold. This alignment was made using PROMALS3D¹⁴ 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 residue 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, *Anaerococcus thermophilus*; Cag, *Chloroflexus aggregans*; Cau, *Chloroflexus aurantiacus*; Cc, *Coprinopsis cinerea*; Ch, *Carboxydotherrus hydrogeniformans*; 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, *Methanococcus marisnigri*; Pa, *Prosthecochloris aestuarii*; Pd, *Paracoccus denitrificans*; Pf, *Pyrococcus furiosus*; Pg, *Porphyrionomonas 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.

C-termini that include two consecutive α -helices and one β -strand that is anti-parallel to the other β -strands in the central β -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 caspase-like fold. The other PrtH homologs are longer and could contain additional domains. For example, HHpred searches suggested that gi118048089 and gi1154150967 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., gi1194337677 and gi1212717887), 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 gi119386630, 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* (gi1166367720) 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* (gi1146297187) neighbors a gene encoding with a hypothetical protein (gi1146297186). 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 I1505 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.