Protease Complement of the Thermophilic Bacterium *Coprothermobacter proteolyticus*

Hong Cai[1], Jianying Gu[2, *], Yufeng Wang[1, *]

[1] Department of Biology, South Texas Center for Emerging Infectious Diseases
University of Texas at San Antonio, San Antonio, TX 78249, USA
hong.cai@utsa.edu, yufeng.wang@utsa.edu (*corresponding authors)
[2] Department of Biology, College of Staten Island
City University of New York, Staten Island, NY 10314, USA
jianying.gu@csi.cuny.edu

***Abstract*—Thermal bacteria that live in higher temperature have been considered as good candidates for bioremediation and processing of protein-rich wastewater. However, very little is known about the proteases, the enzymes that digest the protein wastes in these organisms. In this study, we present a comparative genomic analysis of the protease complement in a thermal bacterium *Coprothermobacter proteolyticus*. The proteases common to a group of thermophilic bacteria have been identified, providing a short list of important enzymes for experimental characterizations.**

Keywords- genome, protease, Coprothermobacter, degradome, bioinformatics, gene family

1. Introduction

Bacteria, a member of the Domains of life, mediates the fundamental geochemical cycles that sustain life on earth. These microorganisms live in diverse habitats and environments. Although some bacteria are human pathogens, the majority of bacteria species are harmless and some have important applications in biotechnology. For example, bacteria that are capable of degrading organic compounds have been used in bioremediation and waste processing in industry.

The advent of high throughput genomic technology and the development of effective bioinformatics data mining approach have provided an unprecedented opportunity to investigate the adaption and evolution of bacteria. Previously uncharacterized organisms can now be explored at a genome level. This study is focused on a bioinformatics characterization of gene families in an understudied bacterium *Coprothermobacter proteolyticus* (strain DSM 5265). This bacterium is anaerobic. Its most important feature is the high growth temperature (about 63°C). It was first isolated from a thermophilic digestor for fermenting water wastes and animal manure. Waterwaste often contains proteins. *Coprothermobacter proteolyticus* was found to have strong protease activity to degrade proteins and peptides [1, 2]. Here we report a comprehensive survey of the protease complement (or degradome) in the genome of *C. proteolyticus*, which may be a good candidate for facilitating waste water processing under high temperature.

2. Methods

A total of over 34,000 sequences of characterized and predicted proteases were obtained from the Merops database (http://www.merops.ac.uk) [3]. These sequences were searched against the *C. proteolyticus* predicted protein sequences using BLASTP with default settings and an E-value cutoff of less than 10-5 for defining protease homologs. Partial sequences (less than 80% of fulllength) and redundant sequences were excluded. The domain/motif organization of predicted *C. proteolyticus* proteases was revealed by an InterPro search. For each putative protease, the known protease sequence or domain with the highest similarity was used as a reference for annotation; the catalytic type and protease family were predicted in accordance with the classification in Merops, and the enzyme was named in accordance with SWISS-PROT enzyme nomenclature (http://www.expasy.ch/cgi-bin/lists?peptidas.txt) and the literature.

3. Results

One of the most prominent physiological features of the anaerobic thermophilic *Coprothermobacter proteolyticus*, formerly *Thermobacteroides proteolyticus*, is its well-documented proteolytic activity [1, 2]. Although proteolytic activity is common in the anaerobic bacteria that are mesophilic, it is observed in only a few thermophiles [4-7]. *C. proteolyticus* has attracted the attention of researchers interested in its potential applications in high temperature environments, including the treatment of protein-rich wastewater, for example. Despite this interest, however, not a single protease in *C. proteolyticus* has been systematically characterized at the biochemical and molecular level to date.

Our comparative genomic analysis revealed that its proteolytic repertoire (degradome) consists of a total of 59 protease homologs, which account for approximately 1.9% of the proteome (Table 1). The fraction of proteases in the Ich genome is close to the average observed in the 1,569 organisms with completed genomes (2.6%). Using the Merops protease nomenclature, which is based on intrinsic evolutionary and structural relationships [3], the *C. proteolyticus* proteases were divided into four known and one unknown catalytic classes that encompass 38 families. These families include: Two aspartic protease families and five cysteine protease families, each represented by a single member; 24 metalloproteases belonging to 17 families, 23 serine proteases belonging to 12 families, and two families (five proteases) with unknown catalytic types. Clearly, gene duplication occurred at a very small scale during the evolution of *C. proteolyticus* proteases, which accounts for the large number of singletons.

A glance at the *C. proteolyticus* degradome reveals some significant features. The entire catalytic class of proteasome-specific threonine proteases is missing, which is consistent with the observation that the proteasome is absent. *C. proteolyticus* has an abundant catalog of metalloproteases (40.7%) and serine proteases (40.0%), compared to aspartic (3.4%) and cysteine proteases (8.5%). The most abundant protease family, serine protease subtilisin (S8), has 6 members. Interestingly, many subtilisins that have been characterized are thermostable [8-10]. The lineage specific expansion of subtilisins in *C. proteolyticus* is likely to be adaptive: at least two subtilisins (COPRO5265_1473 and COPRO5265_1474) are the products of one tandem gene duplication event. Specifically, two sublitisins (COPRO5265_1474 and COPRO5265_1431) are extracellular Vpr peptidases. Vpr was previously only found in a number species from the *Bacillales* [11]; the homologs found in *C. proteolyticus* expand the range of Vpr to the *Clostridiales*.

Table 1. Protease complements in *Coprothermobacter proteolyticus* and other model organisms.

Organism	Catalytic Class					Total	Percentage of the Proteome[a]
	Aspartic	Cysteine	Metallo	Serine	Threonine		
Coprothermobacter proteolyticus	2 (3.4%) [b]	5 (8.5%)	24 (40.7%)	23 (40.0%)	0 (0%)	59[c]	1.9
Neurospora crassa	19 (8.1%)	41 (17.4%)	81 (34.5%)	75 (31.9%)	19 (8.1%)	235	2.4
Saccharomyces cerevisiae	19 (11.1%)	41 (24.0%)	57 (33.3%)	38 (22.2%)	16 (9.4%)	171	2.4
Caenorhabditis elegans	27 (5.6%)	125 (25.9%)	190 (39.4%)	115 (23.9%)	25 (5.2%)	482	2.4
Drosophila melanogaster	46 (6.2%)	86 (11.5%)	207 (27.7%)	373 (49.9%)	35 (4.7%)	747	5.4
Homo sapiens	320 (29.3%)	190 (17.4%)	252 (23.0%)	291 (26.6%)	41 (3.7%)	1,094	4.5
Arabidopsis thaliana	233 (27.6%)	162 (19.2%)	112 (13.3%)	306 (36.2%)	31(3.7%)	849	3.1

[a]. The percentage of the whole genome that encodes putative proteases.
[b]. Percentage of individual catalytic class in the protease complement is included in parentheses.
[c]. The total proteases in *Coprothermobacter proteolyticus* includes 5 protease homologs with unknown classifications.

C. proteolyticus possesses a core degradome structure that may be common in the thermophilic bacteria, as shown by comparison with *Moorella thermoacetica* and *Thermoanaerobacter tengcongensis*, which are the most closely related sequenced species in the family *Thermoanaerobacteriaceae* to have a detailed analysis of its proteases published in Merops [12]. Nineteen protease families are present in all the three organisms. For example, at least three proteases may be actively involved in the secretion system: signal peptidase I (S26) typically processes newly-synthesized secreted proteins by removing the hydrophobic signal peptides when the precursors are translocating the membrane; the bacteria-specific signal peptidase II (A8) is membrane bound and it plays an important role in the production of cell wall by removing the signal peptide from the murein prolipoprotein; type IV prepilin peptidase (A24) processes prepilins by removing leader peptides. Fifteen protease families found in *C. proteolyticus* are also present in either *Moorella thermoacetica* or *Thermoanaerobacter tengcongensis*, but not both. Four protease families are uniquely present in *C. proteolyticus*. They are papain (C1), dipeptidase A (C69), RTX toxin (M6), and carboxypeptidase Taq (M32). Among them, Taq (M32), by its presence in a variety of thermophiles and hyperthermophiles [13], has a demonstrated ability to tolerate high temperatures. While the RTX toxin was implicated in several bacterial pathogens to be a virulence factor as host immune inhibitor, its role in the non-pathogenic *C. proteolyticus* remains unclear [14].

2. Conclusions

We performed a comparative genomic study of the proteases in thermophilic bacterium *Coprothermobacter proteolyticus*. These enzymes play important roles in digesting and breaking down proteins and peptides into smaller fragments. Functional characterization of these enzymes in this bacterium may provide a better understanding of the mechanisms of physiological adaptation to hot temperature and a better assessment of its potential application to wastewater processing.

Acknowledgment

This work is supported by NIH grant AI067543 to YW, and the PSC-CUNY Research Award PSCREG-39-497 to JG.

REFERENCES

[1] Ollivier BM, Mah RA, Ferguson TJ, Boone R, Garcia JL, Robinson R (1985) Emendation of the genus Thermobacteroides: Thermobacteroides proteolyticus sp. nov., a proteolytic acetogen from a methanogenic enrichment. Int J Syst Bacteriol. 35: 425-428.

[2] Rainey FA, Stackebrandt E (1993) Transfer of the type species of the genus Thermobacteroides to the genus Thermoanaerobacter as Thermoanaerobacter acetoethylicus (Ben-Bassat and Zeikus 1981) comb. nov., description of Coprothermobacter gen. nov., and reclassification of Thermobacteroides proteolyticus as Coprothermobacter proteolyticus (Ollivier et al. 1985) comb. nov. Int. J. Syst Bacteriol. 43: 857-859

[3] Rawlings ND, Barrett AJ, Bateman A. (2009) MEROPS: the peptidase database. Nucleic Acids Res. 2010 Jan;38(Database issue):D227-33.

[4] Toda Y, Saiki T, Uozumi T, Beppu T (1988) Isolation and characterization of a protease-producing, thermophilic anaerobic bacterium, Thermobacteroides leptospartum sp. nov. Agric Biol Chem 52: 1339–1344.

[5] Engle M, Li Y, Rainey F, DeBlois S, Mai V, Reichert A, Mayer F, Messner P, Wiegel J (1996) Thermobrachium celere gen. nov., sp. nov., a rapidly growing thermophilic, alkalitolerant, and proteolytic obligate anaerobe. Int J Syst Bacteriol 46: 1025–1033.

[6] Tarlera S., Mux L, Soubes M, Stams AJM (1997) Caloramator proteoclasticus sp. nov., a new moderately thermophilic anaerobic proteolytic bacterium. Int J Syst Bacteriol 47: 651–656.

[7] Etchebehere C, Pavan ME, Zorzopulos J, Soubes M, Muxi L. (1998) Coprothermobacter platensis sp. nov., a new anaerobic proteolytic thermophilic bacterium isolated from an anaerobic mesophilic sludge. Int J Syst Bacteriol. 48:1297-1304.

[8] Siezen RJ, Leunissen JA (1997) Subtilases: the superfamily of subtilisin-like serine proteases. Protein Sci. 6:501-523.

[9] Toogood HS, Smith CA, Baker EN, Daniel RM. (2000) Purification and characterization of Ak.1 protease, a thermostable subtilisin with a disulphide bond in the substrate-binding cleft. Biochem J. 350:321-328.

[10] Godde C, Sahm K, Brouns SJ, Kluskens LD, van der Oost J, de Vos WM, Antranikian G. (2005) Cloning and expression of islandisin, a new thermostable subtilisin from Fervidobacterium islandicum, in Escherichia coli. Appl Environ Microbiol. 71:3951-3958.

[11] Corvey C, Stein T, Dusterhus S, Karas M, Entian KD (2003) Activation of subtilin precursors by Bacillus subtilis extracellular serine proteases subtilisin (AprE), WprA, and Vpr. Biochem Biophys Res Commun. 304: 48-54.

[12] G. Bao Q, Tian Y, Li W, Xu Z, Xuan Z, Hu S, Dong W, Yang J, Chen Y, Xue Y, Xu Y, Lai X, Huang L, Dong X, Ma Y, Ling L, Tan H, Chen R, Wang J, Yu J, Yang H. (2002) A complete sequence of the T. tengcongensis genome. Genome Res. 12:689-700.

[13] Motoshima H, Kaminogawa S (2004) Carboxypeptidase Taq. In Handbook of Proteolytic Enzymes, 2nd edition (Barrett,A.J., Rawlings,N.D. & Woessner,J.F. eds), p.776-778, Elsevier, London.

[14] Dalhammar G, Steiner H (1984) Characterization of inhibitor A, a protease from Bacillus thuringiensis which degrades attacins and cecropins, two classes of antibacterial proteins in insects. Eur J Biochem 139: 247-252.