The Type II Pullulanase of *Thermococcus hydrothermalis*: Molecular Characterization of the Gene and Expression of the Catalytic Domain

MARTA ERRA-PUJADA,[1,2] PHILIPPE DEBEIRE,[1] FRANCIS DUCHIRON,[2]
AND MICHAEL J. O'DONOHUE[1]*

*Unité de Physicochimie et Biotechnologie des Polymères, Institut National de la Recherche
Agronomique,[1] and Laboratoire de Microbiologie Industrielle, Université
de Reims Champagne-Ardenne,[2] 51687 Reims Cedex 02, France*

The gene encoding a hyperthermostable type II pullulanase produced by *Thermococcus hydrothermalis* (Th-Apu) has been isolated. Analysis of a total of 5.2 kb of genomic DNA has revealed the presence of three open reading frames, one of which (*apuA*) encodes the pullulanase. This enzyme is composed of 1,339 amino acid residues and exhibits a multidomain structure. In addition to a typical N-terminal signal peptide, Th-Apu possesses a catalytic domain, a domain bearing S-layer homology-like motifs, a Thr-rich region, and a potential C-terminal transmembrane domain. The presence of these noncatalytic domains suggests that Th-Apu may be anchored to the cell surface and be O glycosylated.

Pullulanases (EC 3.2.1.41) cleave the α-1,6 glucosidic bonds in pullulan (2) and can be classified according to the additional ability (type II) or inability (type I) to degrade the α-1,4 glucosidic bonds of other polysaccharides (1). In recent years, a considerable number of type II pullulanases (also termed amylopullulanases) have been isolated from a wide variety of microorganisms, particularly thermophilic ones, since scientific interest in this class of enzymes is motivated by industrial applications (11, 18, 23, 25, 26, 39). To date, four members of the archaeal order *Thermococcales*, *Pyrococcus furiosus*, *Thermococcus litoralis* (3), *P. woesei* (35), and *T. celer* (4), have been described as pullulanase producers. In each case, the pullulanase activity was described as type II and was localized in the culture medium, indicating that these enzymes are secreted. The enzymes from *P. furiosus* and *T. litoralis* have been described as glycosylated and may be present in multiple forms. In addition, the production of these enzymes appears to be inducible, with malto-oligosaccharides being the principal inducers. Among these enzymes, only those of *P. furiosus* and *P. woesei* have been studied at the genetic level. In both cases, a gene encoding a type II pullulanase has been isolated, cloned, and characterized, although the actual nucleotide sequence of the *P. woesei* type II pullulanase is not available for analysis.

Previously, we have characterized the amylolytic activities of *T. hydrothermalis* and have shown that this archaebacterium produces several amylolytic enzymes, including at least one pullulanase (Th-Apu) (13, 19). Purification and characterization of this enzyme have revealed that it is highly thermostable and is capable of hydrolyzing the α-1,6 glucosidic bonds of pullulan, producing maltotriose, and both the α-1,6 and α-1,4 glucosidic bonds of starch, producing oligosaccharides (degree of polymerization, as low as 4) (12). In this work, we report the isolation and characterization of the gene encoding Th-Apu, as

well as two other partial open reading frames (ORFs) which encode *mal*-like operon elements.

Isolation and characterization of the Th-Apu gene and its surrounding DNA environment. In order to isolate the gene encoding Th-Apu, we first performed N-terminal microsequencing of the native protein purified from the medium of a *T. hydrothermalis* AL662 culture. This yielded a sequence of 11 amino acid residues (AEPKPLNVIIV) which was compared to the N-terminal sequence of the mature pullulanase from *P. furiosus* (Pf-Apu) (12). The high degree of similarity between these two small regions (9 of 11 amino acid residues are identical) indicated that the two proteins might share a high degree of overall homology. Therefore, using the primary sequence of Pf-Apu, three degenerate oligonucleotides were designed for use in standard PCRs using *T. hydrothermalis* AL662 genomic DNA as the template. In this way, specific sequences were amplified and then sequenced, thus providing the necessary sequence information to perform genome crawling in the upstream and downstream directions. Likewise, a 5,200-bp sequence was generated which contains the entire Th-Apu gene (designated *apuA*) flanked by two partial ORFs. Translation of the partial ORFs generated two polypeptides of 118 and 261 amino acids, respectively. The first of these proteins is highly similar (81% identity) to the C-terminal part of the hypothetical MalG-like protein from *P. furiosus* (10) and indeed displays identity with several known MalG proteins, including that of *Escherichia coli* (41%) (9). Similarly, the second polypeptide exhibits strong identity with several MalK proteins, including those of *E. coli* (57% identity) (14), *Salmonella typhimurium* (56% identity), and *Enterobacter aerogenes* (57% identity) (8), leading us to the conclusion that this protein may be a MalK homologue. The apparently promoterless *apuA* gene, surrounded by these *mal*-like genes, is composed of 4,011 bp, 2,562 bp of which show some similarity (~40%) to the sequence encoding Pf-Apu. However, the *P. furiosus* and *T. hydrothermalis* sequences diverge after this point. Whereas the Pf-Apu ORF ends at this position with a stop codon, the *T. hydrothermalis* ORF has an additional 1,449 bp before the stop codon (TGA) is found.

* Corresponding author. Mailing address: Unité de Physicochimie et Biotechnologie des Polymères, Institut National de la Recherche Agronomique, Université de Reims Champagne-Ardenne, Bât. 18, Moulin de la Housse, BP 1039, 51687 Reims Cedex 02, France. Phone: (33) 3 26 91 32 24. Fax: (33) 3 26 91 38 87. E-mail: odonohue@lille.inra.fr.

FIG. 1. Multidomain structure of Th-Apu and comparison of this protein to Pf-Apu. R1 and R2 are two homologous, 230-amino-acid (aa) residue repeats. SLD1 and SLD2 (checkered) are SLH motif-bearing domains (containing approximately two and a half SLH motifs per domain). SP is a signal peptide, PET is a Thr-rich domain, and TM is a hydrophobic C-terminal domain.

Primary structure analysis of Th-Apu. Translation of *apuA* revealed a polypeptide sequence comprised of 1,339 amino acids which are arranged into several well-defined domains (Fig. 1). Like Pf-Apu, Th-Apu possesses an N-terminal sequence (amino acids 2 to 27) which bears the characteristics of a signal peptide (33). The ensuing sequence (PUL), composed of 827 amino acids, presents a high degree of sequence identity (79%) with Pf-Apu and, as such, probably represents the catalytic domain. Beyond this point, three other domain types, absent in Pf-Apu, can be distinguished. The first of these consists of two almost identical sequence repeats (R1 and R2), each of which possesses a repeated motif which, when compared to the PROSITE database, shows similarity to the S-layer homology (SLH) signature. Alignment of these repeated motifs with the prealigned sequences of ProDom domain 1624 (7) confirmed this finding while indicating that there is no complete consensus between the Th-Apu motifs (Fig. 2). Indeed, the latter part of the SLH consensus motif, defined as **ILLA TS R ASQ EDQ**, where consensus amino acid residues are in boldface (24) is absent in the first SLH motif of both Th-Apu S-layer motif-bearing domains (SLD1 and SLD2), while being barely distinguishable in the second SLH motif of each of these domains. Although this sequence appears to be a well-conserved element of the SLH motif, its absence does not necessarily indicate that the Th-Apu-derived domains are not SLDs, since a large sequence diversity exists, even among the known eubacterial SLH motifs. Indeed, a similar, incomplete SLH motif in the SlpA protein of *Clostridium thermocellum* has recently been described (21). Furthermore, apart from this apparent divergence from the canonical sequence, the alignment revealed that the thermococcal domains possess most of the other pertinent features of a eubacterial SLD. In contrast, comparison of the SLH motifs of Th-Apu with those of several proteins from methanogenic archaebacteria (6, 28, 29) failed to reveal anything more than superficial similarity.

The presence of two SLDs in Th-Apu is interesting since such domains have already been observed in a variety of extracellular polysaccharide-degrading enzymes, including three from *T. thermosulfurigenes*, a type II pullulanase (26), a polygalacturonate hydrolase and a xylanase (27), a xylanase from *C. thermocellum* (16), and an α-amylase–pullulanase from *Bacillus* sp. (18). Previous studies have indicated that these enzyme-associated SLDs may be responsible for the anchoring of proteins to the cell surface, possibly by interacting with peptidoglycan in certain bacteria or with the S-layer itself via SLH-SLH interactions (22, 31, 36).

A remarkable Thr-rich region (PET) follows domains R1 and R2 and precedes a smaller domain (TM) which shows a

FIG. 2. Alignment of the two Th-APu SLDs with a variety of eubacterial SLDs. The alignment is based on ProDom domain 1624 (INRA, Toulouse, France), which was constructed from 15 eubacterial sequences aligned with the MultiAlin alignment tool. The proteins used to define ProDom 1624 are as follows: ANCA_CLOTM, cellulosome anchoring protein from *C. thermocellum*; GUN_BACS6, endoglucanase from *Bacillus* sp. strain KSM-635; SLAP_ACEKI, S-layer glycoprotein from *A. kivui*; SLPH_BACBR, SLP from *Bacillus brevis*; APU_THETU, type II pullulanase from *T. thermosulfurigenes*; XYNX_CLOTM, exoglucanase from *C. thermocellum*; XYNA_THESA, endoxylanase A from *Thermoanaerobacterium saccharolyticum*; SLAP_BACSH, SLP from *Bacillus sphaericus*; OMPA_THEMA, outer membrane protein A from *Thermotoga maritima*; SLAP_THETH, SLP from *Thermoanaerobacterium thermosulfurigenes*; SLPM_BACBR, middle cell wall protein from *B. brevis*; SLAP_BACAN, SLP protein from *Thermoanaerobacter ethanolicus*; SLAP_BACLI, SLP from *Bacillus licheniformis*. The Th-APu sequences were manually fitted to the ProDom alignment by introducing uniform gaps where necessary. White characters on a black background indicate areas where sequence identity is equal to or greater than 50%, while gray shading indicates areas where sequence similarity is equal to or greater than 50%. The groups of similar residues were considered to be I, L, V, M, D, E; N, Q, S, T; R, K; and F, Y, W. Dots indicate gaps in the sequences. SLD1_THYDRO and SLD2_THYDRO are the two Th-APu SLDs.

striking resemblance to an inverted lipoprotein signal peptide (34). Database searching revealed that this arrangement is very similar to that found in the C termini of S-layer proteins (SLPs) of *Haloferax halobium* (17) and *H. volcanii* (38). In these proteins, the Thr-rich regions have been shown to be targets for intensive O glycosylation (glucose-galactose disaccharides) while the shorter, hydrophobic domains have been proposed as transmembrane anchors. In Th-Apu, therefore, it is possible that the PET domain is O glycosylated and that the shorter, C-terminal TM domain serves as either a transmembrane anchor or, indeed, as a hydrophobic cell wall anchor, rather like the C-terminal domain of the SLP of *Corynebacterium glutamicum* (5).

Interestingly, several SLH-bearing enzymes exhibit variations of the Thr-rich region which are usually described as linker regions. Indeed, in the case of the *T. thermosulfurigenes* pullulanase, it has been suggested that this region, due to the probable extended, flexible nature of its structure, would allow optimal orientation of the enzyme's catalytic site toward the substrate (26). In another carbohydrate-degrading enzyme, the glucoamylase from *Aspergillus awamori*, it has been shown that this region is subject to O glycosylation (hypermannosylation) and that its presence in this enzyme increases the efficiency of degradation of insoluble starch granules (37). In the same way, the glycosylated-linker regions of two glycanases from *Cellulomonas fimi* appear to increase the affinity of these enzymes for microcrystalline cellulose and may favor the disruption of cellulose fibers (32).

Expression of the catalytic domain in *E. coli*. Having failed in our attempts to obtain a plasmid containing an intact *apuA* gene, we constructed a plasmid containing only the PUL domain of Th-Apu (pAPUΔ1). Expression trials using pAPUΔ1 led to the production of a significant amount of recombinant protein in *E. coli* cells which was present in a soluble form in the cytoplasmic fraction after cell lysis. Crude purification of this protein could be achieved by a simple heat treatment (80°C, 30 min) which provoked the precipitation of most of the contaminating proteins. Examination of the recombinant protein by sodium dodecyl sulfate-polyacrylamide gel electrophoresis revealed the presence of two isoforms (the two species have identical N-terminal sequences) which exhibit molecular masses of approximately 90 and 85 kDa. Zymogram analysis was then employed, the results of which showed that both species possess thermostable pullulanolytic activity, indicating that the catalytic determinants are indeed present within the PUL domain.

Conclusion. On the basis of our results, we propose that Th-Apu forms part of a maltose transport operon different from the one previously described for *T. litoralis* (15). This enzyme is probably secreted in *T. hydrothermalis* cells, although rather than being released into the extracellular medium, Th-APu may be anchored to the cell membrane or another hydrophobic component of the cell envelope via its TM domain, with its catalytic domain exposed to the surrounding medium. By analogy to other Pro-Thr-rich domains, it can be assumed that the PET domain has a rather extended, perhaps flexible, conformation which would be susceptible to proteolytic digestion (30). Thus, intense O glycosylation of this domain would reduce the vulnerability of this structure. With regard to its function, the PET domain may serve as a linker and/or perform a role similar to that of the Thr-rich region of the glucoamylase from *A. awamori* or may fulfill a cell wall-anchoring role rather like the Gly/Ser-rich region in SbsB (36). The role of the Th-Apu SLDs is unclear since, to our knowledge, nothing is known concerning the cell envelope of *T. hydrothermalis*. However, by analogy to other proteins which exhibit SLH motifs, we speculate that these domains interact with a hitherto unidentified component of the cell envelope.

Clearly, the idea that Th-Apu is localized at the cell surface appears to contradict previous experimental data from chemostat cultures which have indicated that Th-Apu is completely secreted and released into the extracellular medium. However, as others have already suggested (20, 26), it is possible that cellular attachment may be a transient state prior to complete release of the protein and/or that the prevailing conditions of a chemostat culture may provoke the degradation of the cell envelope, thus eliminating certain protein attachment points (e.g., SLH anchoring determinants).

Nucleotide sequence accession number. The GenBank accession number for the 5.2-kb DNA fragment described here is AF113969.

REFERENCES

1. **Antranikian, G.** 1990. Physiology and enzymology of thermophilic anaerobic bacteria degrading starch. FEMS Microbiol. Rev. **75**:201–218.
2. **Bender, H., and K. Wallenfels.** 1961. Untersuchungen an Pullulan. Biochem. Z. **334**:79–95.
3. **Brown, S. H., and R. M. Kelly.** 1993. Characterization of amylolytic enzymes, having both α-1,4 and α-1,6 activity, from the thermophilic archaea *Pyrococcus furiosus* and *Thermococcus litoralis*. Appl. Environ. Microbiol. **59**:2614–2621.
4. **Canganella, F., C. M. Antrade, and G. Antranikian.** 1994. Characterization of amylolytic and pullulytic enzymes from thermophilic archaea and from a new *Fervidobacterium* species. Appl. Microbiol. Biotechnol. **42**:239–245.
5. **Chami, M., N. Bayan, J.-L. Peyret, T. Gulik-Krzywicki, G. Leblon, and E. Schechter.** 1997. The S-layer protein of *Corynebacterium glutamicum* is anchored to the cell wall by its C-terminal hydrophobic domain. Mol. Microbiol. **23**:483–492.
6. **Conway de Macario, E., and A. J. L. Macario.** 1997. S-layer and ABC transporter genes in methanogenic archaea. FEMS Microbiol. Rev. **20**:59–64.
7. **Corpet, F.** 1988. Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res. **16**:10881–10890.
8. **Dahl, M. K., E. Francoz, W. Saurin, W. Boos, M. D. Manson, and M. Hofnung.** 1989. Comparison of sequences from the *malB* regions of *Salmonella typhimurium* and *Enterobacter aerogenes* with *Escherichia coli* K12: a potential new regulatory site in the interoperonic region. Mol. Gen. Genet. **218**:199–207.
9. **Dassa, E., and M. Hofnung.** 1985. Sequence of gene *malG* in *E. coli* K12: homologies between integral membrane components from binding protein-dependent transport systems. EMBO J. **4**:2287–2293.
10. **Dong, G., C. Vieille, and J. G. Zeikus.** 1997. Cloning, sequencing, and expression of the gene encoding amylopullulanase from *Pyrococcus furiosus* and biochemical characterization of the recombinant enzyme. Appl. Environ. Microbiol. **63**:3577–3584.
11. **Ganghofner, D., J. Kellermann, W. L. Staudenbauer, and K. Bronnenmeier.** 1998. Purification and properties of an amylopullulanase, a glucoamylase and an α-glucosidase in the amylolytic enzyme system of *Thermoanaerobacterium thermosaccharolyticum*. Biosci. Biotechnol. Biochem. **62**:302–308.
12. **Gantelet, H., and F. Duchiron.** 1998. Purification and properties of a thermoactive and thermostable pullulanase from *Thermococcus hydrothermalis*, a hyperthermophilic archaeon isolated from a deep-sea hydrothermal vent. Appl. Microbiol. Biotechnol. **49**:770–777.
13. **Gantelet, H., C. Ladrat, A. Godfroy, G. Barbier, and F. Duchiron.** 1998. Characteristics of pullulanases from extremely thermophilic archaea isolated from deep-sea hydrothermal vents. Biotechnol. Lett. **20**:819–823.
14. **Gilson, E., H. Nikaido, and M. Hofnung.** 1982. Sequence of the *malK* gene in *E. coli* K12. Nucleic Acids Res. **25**:7449–7458.
15. **Horlacher, R., K. B. Xavier, H. Santos, J. DiRuggiero, M. Kossman, and W. Boos.** 1998. Archaeal binding protein-dependent ABC transporter: molecular and biochemical analysis of the trehalose/maltose transport system of the hyperthermophilic archaeon *Thermococcus litoralis*. J. Bacteriol. **180**:680–689.
16. **Jung, K. H., K. M. Lee, H. Kim, K. H. Yoon, S. H. Park, and M. Y. Pack.** 1998. Cloning and expression of a *Clostridium thermocellum* xylanase gene in *Escherichia coli*. Biochem. Mol. Biol. Int. **44**:283–292.
17. **Lechner, J., and M. Sumper.** 1987. The primary structure of a prokaryotic glycoprotein. J. Biol. Chem. **262**:9724–9729.

18. **Lee, S.-P., M. Morikawa, M. Takagi, and T. Imanaka.** 1994. Cloning of the *aapT* gene and characterization of its product, an amylase-pullulanase (AapT), from thermophilic and alkaliphilic *Bacillus* sp. strain XAL601. Appl. Environ. Microbiol. **60:**3764–3773.

19. **Legin, E., A. Copinet, and F. Duchiron.** 1998. Production of thermostable amylolytic enzymes by *Thermococcus hydrothermalis*. Biotechnol. Lett. **20:** 363–367.

20. **Leibovitz, E., M. Lemaire, I. Miras, S. Salamitou, P. Béguin, H. Ohayon, P. Gounon, M. Matuschek, K. Sahm, and H. Bahl.** 1997. Occurrence and function of a common domain in S-layer and other exocellular proteins. FEMS Microbiol. Rev. **20:**127–133.

21. **Lemaire, M., I. Miras, P. Gounon, and P. Béguin.** 1998. Identification of a region responsible for binding to the cell wall within the S-layer protein of *Clostridium thermocellum*. Microbiology **144:**211–217.

22. **Lemaire, M., H. Ohayon, P. Gounon, T. Fujino, and P. Béguin.** 1995. OlpB, a new outer layer protein of *Clostridium thermocellum*, and binding of its S-layer-like domain to components of the cell envelope. J. Bacteriol. **177:** 2451–2459.

23. **Lin, L.-L., M.-R. Tsau, and W.-S. Chu.** 1996. Purification and properties of a 140 kDa amylopullulanase from the thermophilic and alkaliphilic *Bacillus* sp. strain TS-23. Biotechnol. Appl. Biochem. **24:**101–107.

24. **Lupas, A., H. Engelhardt, J. Peters, U. Santarius, S. Volker, and W. Baumeister.** 1994. Domain structure of the *Acetogenium kivui* surface layer revealed by electron crystallography and sequence analysis. J. Bacteriol. **176:**1224–1233.

25. **Mathupala, S. P., and J. G. Zeikus.** 1993. Improved purification and biochemical characterization of extracellular amylopullulanase from *Thermoanaerobacter ethanolicus* 39E. Appl. Microbiol. Biotechnol. **39:**487–493.

26. **Matuschek, M., G. Burchhardt, K. Sahm, and H. Bahl.** 1994. Pullulanase of *Thermoanaerobacterium thermosulfurigenes* EM1 (*Clostridium thermosulfurigenes*): molecular analysis of the enzyme and a common model for its attachment to the cell surface. J. Bacteriol. **176:**3295–3302.

27. **Matuschek, M., K. Sahm, A. Zibat, and H. Bahl.** 1996. Characterization of genes from *Thermoanaerobacterium thermosulfurigenes* EM1 that encode two glycosyl hydrolases with conserved S-layer-like domains. Mol. Gen. Genet. **252:**493–496.

28. **Mayerhofer, L. E., E. Conway de Macario, and A. J. Macario.** 1995. Conservation and variability in Archaea: protein antigens with tandem repeats encoded by a cluster of genes with common motifs in *Methanosarcina mazei* S-6. Gene **165:**87–91.

29. **Mayerhofer, L. E., E. Conway de Macario, R. Yao, and A. J. Macario.** 1998. Structure, organization and expression of genes encoding for envelope components in the archaeon *Methanosarcina mazei* S-6. Arch. Microbiol. **169:** 339–345.

30. **Moens, S., and J. Vanderleyden.** 1997. Glycoproteins in prokaryotes. Arch. Microbiol. **168:**169–175.

31. **Olabarria, G., J. L. Carrascosa, M. de Pedro, and J. Berenguer.** 1996. A conserved motif in S-layer proteins is involved in peptidoglycan binding in *Thermus thermophilus*. J. Bacteriol. **178:**4765–4772.

32. **Ong, E., D. G. Kilburn, R. C. Miller, and R. A. J. Warren.** 1994. *Streptomyces lividans* glycosylates the linker region of a β-1,4-glycanase from *Cellulomonas fimi*. J. Bacteriol. **176:**999–1008.

33. **Perlman, D., and H. O. Halvorson.** 1983. A putative signal peptidase recognition site and sequence in eukaryotic and prokaryotic signal peptides. J. Mol. Biol. **167:**391–409.

34. **Pugsley, A.** 1993. The complete general secretory pathway in Gram-negative bacteria. Microbiol. Rev. **57:**50–108.

35. **Rüdiger, A., P. L. Jorgensen, and G. Antrinikian.** 1995. Isolation and characterization of a heat-stable pullulanase from the hyperthermophilic archaeon *Pyrococcus woesei* after cloning and expression of its gene in *Escherichia coli*. Appl. Environ. Microbiol. **61:**567–575.

36. **Sàra, M., E. M. Egelseer, C. Dekitsch, and U. B. Sleytr.** 1998. Identification of two binding domains, one for peptidoglycan and another for a secondary cell wall polymer, on the N-terminal part of the S-layer protein SbsB from *Bacillus stearothermophilus* PV72/p2. J. Bacteriol. **180:**6780–6783.

37. **Semimaru, T., M. Goto, K. Furukawa, and S. Hayashida.** 1995. Functional analysis of the threonine- and serine-rich Gp-1 domain of glucoamylase I from *Aspergillus awamori* var. Kawachi. Appl. Environ. Microbiol. **61:**2885–2890.

38. **Sumper, M., E. Berg, R. Mengele, and I. Strobel.** 1990. Primary structure and glycosylation of the S-layer protein of *Haloferax volcanii*. J. Bacteriol. **172:** 7111–7118.

39. **Suzuki, Y., K. Hatagaki, and H. Oda.** 1991. A hyperthermostable pullulanase produced by an extreme thermophile, *Bacillus flavocaldarius* KP 1228, and evidence for the proline theory of increasing protein thermostability. Appl. Microbiol. Biotechnol. **34:**707–714.