Volume 255, number 1, 37-41 FEBS 07569 September 1989

Nucleotide sequence of the maltotetraohydrolase gene from *Pseudomonas saccharophila*

Jianhua Zhou, Tadashi Baba, Toshiya Takano*, Shoichi Kobayashi* and Yuji Arai

Institute of Applied Biochemistry, Universtity of Tsukuba and **National Food Research Institute, Tsukuba City, Ibaraki 305, Japan*

Received 9 June 1989

The nucleotide sequence of the *Pseudomonas saccharophila* gene encoding maltotetraohydrolase (G_4-forming amylase) has been determined. The coding region for the G_4-forming amylase precursor contained 1653 nucleotides. The deduced precursor protein included an N-terminal 21-residue putative signal peptide; the deduced mature form of G_4-forming amylase contains 530 amino acid residues with a calculated molecular mass of 57 740 Da. Sequence similarities between the G_4-forming amylase and other amylolytic enzymes of species ranging from prokaryotes to eukaryotes are quite limited. However, three regions, which are involved in both the catalytic and substrate-binding sites of various amylolytic enzymes, are highly conserved in the G_4-forming amylase of *P. saccharophila*.

Maltotetraohydrolase; Gene cloning; Nucleotide sequence; Amino acid sequence; (*Pseudomonas saccharophila*)

1. INTRODUCTION

Maltotethraohydrolase (EC 3.2.1.60, G_4-forming amylase) is one of the exo-amylases which catalyzes hydrolysis of α-glucans to form maltotetraose from their nonreducing ends. G_4-forming amylase was originally discovered in a culture of *Pseudomonas stutzeri* [1], and was purified as two active forms with different isoelectric points [2,3]. These two forms had similar molecular masses (55–57 kDa) and enzymatic properties [3]; these forms both rapidly hydrolyzed starch, glycogen, and maltooligosaccharides larger than maltopentaose to produce maltotetraose, but hydrolyzed maltotetraose, maltotriose, and maltose only very slowly [2]. Recently, we have found that another bacterium, *P. saccharophila* IAM1504, produces an extracellular G_4-forming amylase (Kobayashi et al., manuscript in preparation). This G_4-forming amylase seems to possess enzymatic properties similar to those of the *P. stutzeri* enzyme.

The amino acid sequences of various amylolytic enzymes such as α-amylase [4–8], β-amylase [9], glucoamylase [10], cyclomaltodextrin glucanotransferase [11–13], maltohexaohydrolase [14], isoamylase [15], and pullulanase [16], have been determined by direct protein analysis or deduced from the DNA sequences. It is of great interest to study the molecular basis for the difference between the specificity of the G_4-forming amylase and those of the other amylolytic enzymes. To facilitate further studies of the G_4-forming amylase, it is necessary to establish its primary structure. We have recently cloned the G_4-forming amylase gene (*mta*) from the *P.saccharophila* IAM1504 [17]. Here, we report the DNA sequence of the cloned *mta* gene and compare the amino acid sequence of the G_4-forming amylase with those of other amylolytic enzymes.

2. MATERIALS AND METHODS

2.1. *DNA sequence analysis*

To sequence the *mta* gene, two sets of partial deletions of this

Correspondence address: T. Baba, Institute of Applied Biochemistry, University of Tsukuba, Tsukuba City, Ibaraki 305, Japan

Published by Elsevier Science Publishers B. V. (Biomedical Division)

DNA were constructed by the method of Henikoff [18], using a kilo-sequence deletion kit (Takara Shuzo Co. Ltd). A 3.1-kbp *Eco*RI/*Hin*dIII DNA fragment containing the *mta* gene was separated from the recombinant plasmid pGF-11 [17], blunt-ended with mung bean nuclease and Klenow fragment, and then cloned into the *Hin*cII site of M13mp18 in both orientations. These RF DNAs were digested with both *Bam*HI and *Sac*I, and then treated with nuclease Exo III. A series of increasingly large deletions extending from the *Bam*HI site was obtained by terminating the digestion after various time intervals. After treatment with mung bean nuclease and Klenow fragment, the DNAs with deletions of the desired size were recircularized using T4 DNA ligase and transformed into *E. coli* JM109. The nucleotide sequences of the subcloned inserts were determined by the dideoxy chain-termination method [19], using a 7-deaza sequencing kit (Takara).

2.2. *Purification of G4-forming amylase*

P. saccharophila IAM1504 was obtained from the Institute of Applied Microbiology, University of Tokyo. Cells were grown aerobically for 2 days at 30°C with constant shaking in 0.1% KH_2PO_4, 0.28% K_2HPO_4, 1% soluble starch, and 1% polypeptone. After removal of the cells, the culture supernatant was fractionated with ammonium sulfate (30–50% saturation). The enzyme-containing fraction was applied to a DEAE-Toyopearl 650S column (Tosoh Co.) previously equilibrated with 10 mM potassium phosphate, pH 7.0, and proteins were eluted with a linear gradient of 0–0.5 M NaCl in the same buffer. Fractions exhibiting enzyme activity were combined, concentrated, and then filtered through a Toyopearl HW-55F column (Tosoh Co.). The specific activity of the purified G_4-forming amylase was 434 IU/mg of protein. Details of the purification of this enzyme and its enzymatic properties will be reported elsewhere.

2.3. *Protein analysis*

Protein was hydrolyzed in vacuo in 6 M HCl for 24, 48, and 72 h at 110°C. The amino acid compositions of the resulting hydrolyzates were determined with a Dinex 502 amino acid analyzer. N-terminal sequence analysis of the protein was carried out using an Applied Biosystems pulse-liquid sequencer (model 477A/120A) with an on-line PTH analyzer. SDS-PAGE was performed by the method of Laemmli [20].

3. RESULTS

The gene encoding G_4-forming amylase from *P. saccharophila* IAM1504 was originally cloned as a 12.4-kbp DNA fragment, and was further localized to a 3.1-kbp fragment by subcloning [17]. The entire region of the 3.1-kbp DNA containing the *mta* gene was sequenced using the dideoxy chain-termination method [19] according to the strategy shown in fig.1. Two open reading frames, termed ORF-1 and ORF-2, which contained 1833 and 2211 nucleotides, respectively, were found in the 3.1-kbp DNA fragment. ORF-1 and ORF-2 overlapped each other extensively, but both were predicted to be coding regions by the algorithm of Fickett [21].

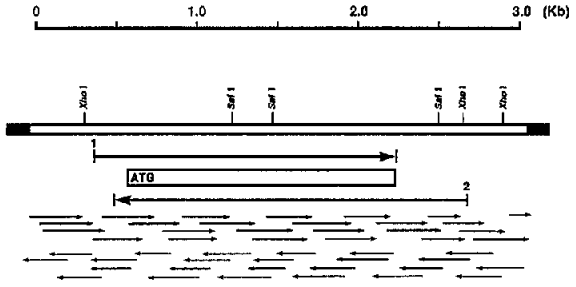
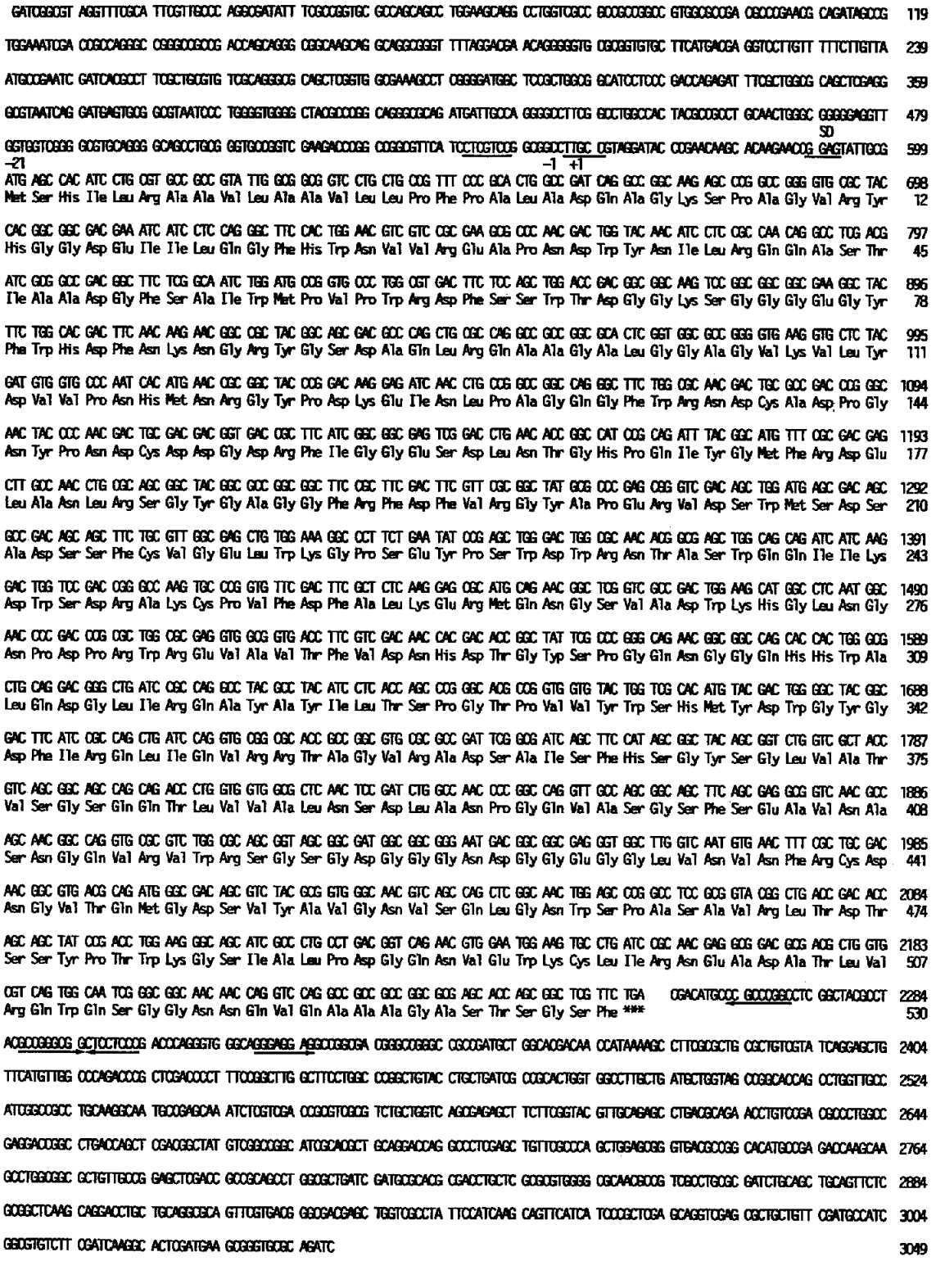


Fig.1. Restriction map and sequencing strategy for the G_4-forming amylase (*mta*) gene subcloned into M13mp18. The dotted and open boxes represent the vector and insert DNAs, respectively. The open box with an ATG codon indicates the protein coding region. The open reading frames 1 and 2 (ORF-1 and 2) are indicated as 1 and 2, respectively, and the arrows represent the directions of the reading. Two sets of arrows indicate the direction and extent of each sequencing determination.

Protein sequence analysis indicated that N-terminal sequence of purified G_4-forming amylase was Xaa-Gln-Ala-Gly-Lys-Ser-Pro-Ala-Gly-Val-, were Xaa was not identified (data not shown). This sequence matched residues 1–10 derived from the nucleotide sequence in ORF-1 (fig.2). No identical sequence was found in ORF-2. These results indicate that ORF-1 encodes the G_4-forming amylase. The DNA sequence of the *mta* gene and deduced amino acid sequence of the G_4-forming amylase are presented in fig.2. The mature enzyme contains 530 amino acid residues with a calculated molecular mass of 57 740 Da. When the purified enzyme was analyzed by SDS-PAGE, the apparent molecular mass was estimated to be 56 kDa, using a Bio-Rad SDS-PAGE standard as a molecular mass marker (not shown). This value agreed well with the calculated molecular mass of the G_4-forming amylase. The amino acid sequence at

Fig.2. Nucleotide and deduced amino acid sequences of *P. saccharophila* G_4-forming amylase gene. The probable promotor sequences are underlined. SD indicates the Shine-Dalgarno sequence. The N-terminal amino acid of the mature protein is indicated as residue +1, and residues −21 to −1 are the putative signal peptide.



39

Table 1

Amino acid composition of *Pseudomonas saccharophila* G_4-forming amylase

Amino acid	Residues/molecule (mol%)	
	Protein	Gene
Asx	69 (13.1)	75 (14.1)
Thr	19 (3.5)	17 (3.2)
Ser	47 (8.8)	46 (8.7)
Glx	47 (8.8)	43 (8.1)
Pro	24 (4.5)	23 (4.3)
Gly	72 (13.5)	71 (13.4)
Ala	46 (8.6)	45 (8.5)
Val	35 (6.6)	38 (7.2)
Cys	6 (1.2)	6 (1.1)
Met	7 (1.4)	7 (1.3)
Ile	18 (3.3)	17 (3.2)
Leu	27 (5.1)	26 (4.9)
Tyr	20 (3.7)	20 (3.8)
Phe	20 (3.7)	20 (3.8)
Lys	12 (2.3)	12 (2.3)
His	10 (1.9)	11 (2.1)
Arg	30 (5.7)	30 (5.7)
Trp	n.d.[a]	23 (4.3)

[a] Not determined

residues −21 to −1 preceding the N-terminal amino acid (Asp-1) of the G_4-forming amylase contains two positively charged residues (residues −19 and −16), a strongly hydrophobic sequence (residues −15 to −1), and an Ala-Xaa-Ala sequence (residues −3 to −1), which are all common features of bacterial signal peptides [22,23]. Since the G_4-forming amylase is secreted, it is likely that this 21-residue sequence is a genuine signal peptide.

The amino acid composition of the G_4-forming amylase is shown in table 1 along with that deduced from the DNA-derived sequence. These compositions are closely similar to each other, which further supports our conclusion that the *mta* gene encodes the G_4-forming amylase.

When the entire primary sequence of the G_4-forming amylase was compared with those of other amylolytic enzymes (various α-amylases [4–8], two cyclomaltodextrin glucanotransferases [12,13], maltohexaohydrolase [14], isoamylase [15], and pullulanase [16]), the sequence similarities were extremely weak (data not shown); the percent identity was 26.3% for barley α-amylase [5], but was less than 9% for all other amylolytic enzymes examined. However, three regions (domains I, II, and III), which constitute both the catalytic and substrate-binding sites of the amylolytic enzymes [24], are highly conserved in the sequence of the G_4-forming amylase (residues 112–117, 189–197, and 288–295, respectively, see fig.3).

4. DISCUSSION

This paper describes the nucleotide sequence of the *mta* gene and the deduced amino acid sequence of the *P. saccharophila* G_4-forming amylase. A 3.1-kbp DNA fragment encoding the G_4-forming amylase [17] contains two open reading frames, ORF-1 and ORF-2, the former of which actually encodes the G_4-forming amylase gene (figs 1 and

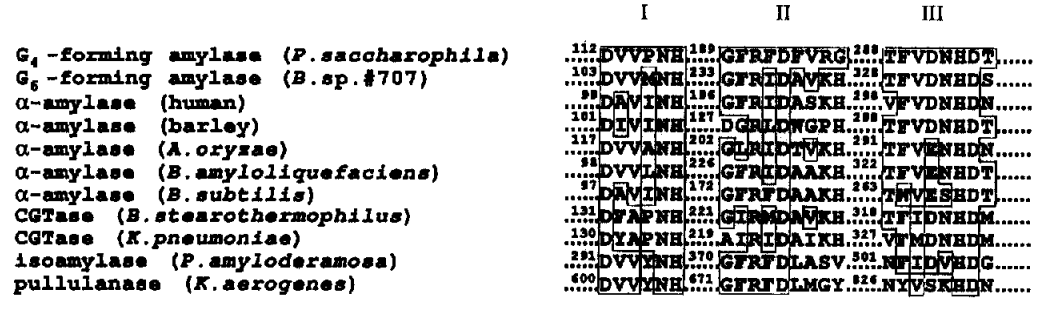


Fig.3. Sequence comparison of G_4-forming amylase with other amylolytic enzymes. Amino acid residues are numbered from the N-terminus of the mature protein. Identical amino acids are boxed. The amylolytic enzymes examined are as follows: *Bacillus* sp. no.707 G_6-forming amylase [14]; human saliva gland α-amylase [4]; barley α-amylase [5]; *Aspergillus oryzae* α-amylase [6]; *Bacillus amyloliquefaciens* α-amylase [7]; *Bacillus subtilis* α-amylase [8]; *Bacillus stearothermophilus* cyclomaltodextrin glucanotransferase (CGTase) [12]; *Klebsiella pneumoniae* CGTase [13]; *Pseudomonas amyloderamosa* isoamylase [15]; *Klebsiella aerogenes* pullulanase [16].

2). It is not clear at the present time what, if any, protein is coded by ORF-2. The G_4-forming amylase is synthesized as a single-chain polypeptide of 551 amino acids including the 21-residue signal peptide.

The Shine-Dalgarno sequence [25] for the G_4-forming amylase, GGAG, is located 8 nucleotides upstream from the ATG initiation codon (fig.2). The sequence CTCGTCC-(7 nucleotides)-TTGCC at positions 542–560 is similar to the promotor sequences of other *Pseudomonas* genes [26], and is positioned appropriately relative to the Shine-Dalgarno sequence [25]. Thus, this sequence is probably the promotor of the *mta* gene. Nucleotide sequences similar to well-known *E. coli* promotors ($^{-35}$TTGACA and $^{-10}$TATAAT, see [27]) are not present upstream of the coding region. The possible sequences which may serve as the transcriptional promotor of the *mta* gene in *E. coli* cells are presumed to be ^{536}TTCAT and ^{562}TAGGAT (fig.2).

Most of the amylolytic enzymes described so far contain three conserved regions (domains I, II and III, see fig.3) which are involved in the catalytic and substrate-binding sites [24]. X-ray crystallographic analysis of *Aspergillus oryzae* α-amylase (Taka-amylase A) has shown that His-122 in domain I, Asp-206, Lys-209 and His-210 in domain II, and His-296 and Asp-297 in domain III served as the active site residues [28]. The amino acid sequences of these three domains are highly conserved in the G_4-forming amylase, although two active site residues (Lys and His) in domain II are replaced by Arg and Gly, respectively (fig.3). It is likely that these replacements in the G_4-forming amylase are not responsible for the difference in specificity between G_4-forming amylase and other amylolytic enzymes. In other work, a truncated form of the G_4-forming amylase, lacking its C-terminal portion (including domain III), has been obtained by digestion with *Staphylococcus aureus* V_8 protease (unpublished data). This fragment still produces maltotetraose from various α-glucans. Therefore, the N-terminal region of this enzyme is probably responsible for the specificity of formation of maltotetraose.

Acknowledgements: We are grateful to Dr K. Haraguchi for performing the N-terminal sequencing of protein, and to Dr C. Jones for critical reading of the manuscript.

REFERENCES

[1] Robyt, J.F. and Ackerman, R.J. (1971) Arch. Biochem. Biophys. 145, 105–114.
[2] Sakano, Y., Kashiwagi, Y. and Kobayashi, T. (1982) Agric. Biol. Chem. 46, 639–646.
[3] Sakano, Y., Kashiyama, E. and Kobayashi, T. (1983) Agric. Biol. Chem. 47, 1761–1768.
[4] Nakamura, Y., Ogawa, M., Nishide, T., Emi, M., Kosaki, G., Himeno, S. and Matsubara, K. (1984) Gene 28, 263–270.
[5] Rogers, J.C. and Milliman, C. (1983) J. Biol. Chem. 258, 8169–8174 .
[6] Toda, H., Kondo, K. and Narita, K. (1982) Proc. Jpn. Acad. Ser. B Phys. Biol. Sci. 58, 208–212.
[7] Takkinen, K., Pettersson, R.F., Kalkkinen, N., Palva, I., Söderlund, H. and Kääriäinen, L. (1983) J. Biol. Chem. 258, 1007–1013.
[8] Yamazaki, H., Ohmura, K., Nakayama. A., Takeichi, Y., Otazai, K., Yamasaki, M., Tamura, G. and Yamane, K. (1983) J. Bacteriol. 156, 327–337.
[9] Kawazu, T., Nakanishi, Y., Uozumi, N., Sasaki, T., Yamagata, H., Tuskagoshi, N. and Udaka, S. (1987) J. Bacteriol. 169, 1564–1570.
[10] Ashikari, T., Nakamura, N., Tanaka, Y., Kiuchi, N., Shibano, Y., Tanaka, T., Amachi, T. and Yoshizumi, H. (1986) Agric. Biol. Chem. 50, 957–964.
[11] Takano, T., Fukuda, M., Monma, M., Kobayashi, S., Kainuma, K. and Yamane, K. (1986) J. Bacteriol. 166, 1118–1122.
[12] Sakai, S., Kubota, M., Yamamoto, K., Nakada, T., Torigoe, K., Ando, O. and Sugimoto, T. (1987) J. Jpn. Soc. Starch Sci. 34, 140–147.
[13] Binder, F., Huber, O. and Böck, A. (1986) Gene 47, 269–277.
[14] Tsukamoto, A., Kimura, K., Ishii, Y., Takano, T. and Yamane, K. (1988) Biochem. Biophys. Res. Commun. 151, 25–31.
[15] Amemura, A., Chakraborty, R., Fujita, M., Noumi, T. and Futai, M. (1988) J. Biol. Chem. 263, 9271–9275.
[16] Katsuragi, N., Takizawa, N. and Murooka, Y. (1987) J. Bacteriol. 169, 2301–2306.
[17] Zhou, J., Takano, T. and Kobayashi, S. (1989) Agric. Biol. Chem. 53, 301–302.
[18] Henikoff, S. (1984) Gene 28, 351–359.
[19] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[20] Laemmli, U.K. (1970) Nature (Lond.) 227, 680–685.
[21] Fickett, J.W. (1982) Nucleic Acid Res. 10, 5303–5318.
[22] Oliver, D. (1985) Annu. Rev. Microbiol. 39, 615–648.
[23] Kreil, G. (1981) Annu. Rev. Biochem. 50, 317–348.
[24] Rogers, J.C. (1985) Biochem. Biophys. Res. Commun. 128, 470–476.
[25] Shine, J. and Dalgarno, L. (1974) Proc. Natl. Acad. Sci. USA 71, 1342–1346.
[26] Johnson, K., Parker, M.L. and Lory, S. (1986) J. Biol. Chem. 261, 15703–15708.
[27] Rosenberg, M. and Court, D. (1979) Annu. Rev. Genet, 13, 319–353.
[28] Matsuura, Y., Kusunoki, M., Harada, W. and Kakudo, M. (1984) J. Biochem. (Tokyo) 62, 62–67.