

# Nucleotide sequence of the maltotetraohydrolase gene from *Pseudomonas saccharophila*

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

*Institute of Applied Biochemistry, University 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

Maltotetraohydrolase (EC 3.2.1.60,  $G_4$ -forming amylase) is one of the exo-amylases which catalyzes hydrolysis of  $\alpha$ -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  $\alpha$ -amylase [4–8],  $\beta$ -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

DNA were constructed by the method of Henikoff [18], using a kilo-sequence deletion kit (Takara Shuzo Co. Ltd). A 3.1-kbp *EcoRI/HindIII* 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 *HincII* 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 *G*<sub>4</sub>-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<sub>2</sub>PO<sub>4</sub>, 0.28% K<sub>2</sub>HPO<sub>4</sub>, 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*<sub>4</sub>-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*<sub>4</sub>-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

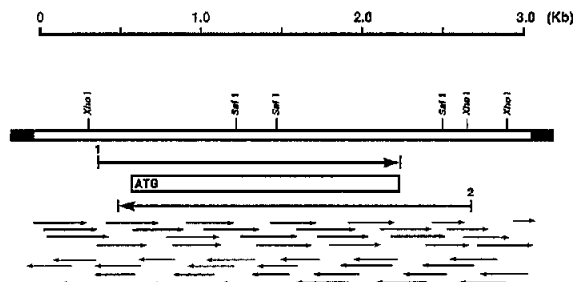


Fig.1. Restriction map and sequencing strategy for the *G*<sub>4</sub>-forming amylase (*mta*) gene subcloned into M13mp18. The dotted and open boxes represent the vector and insert DNAs, respectively. The open box with an ATG codon indicates the protein coding region. The open reading frames 1 and 2 (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.

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].

Protein sequence analysis indicated that N-terminal sequence of purified *G*<sub>4</sub>-forming amylase was Xaa-Gln-Ala-Gly-Lys-Ser-Pro-Ala-Gly-Val-, where 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*<sub>4</sub>-forming amylase. The DNA sequence of the *mta* gene and deduced amino acid sequence of the *G*<sub>4</sub>-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*<sub>4</sub>-forming amylase. The amino acid sequence at

Fig.2. Nucleotide and deduced amino acid sequences of *P. saccharophila* *G*<sub>4</sub>-forming amylase gene. The probable promoter 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.

GATGGGGT AGGTTTGGCA TTGGTTGGCC AGGGATATT TGGGGGTGC GCGACAGCC TGAAGCGAG CTGGGTGGCC GCGCGGGCC GTGGCGCGCA GCGCGGAGC CAGATAGCG	119
TGAAATGCA CGGCGAGGC CGGGCGGGG ACCAGCAGG CGGCGAGCAG GAGGGGGT TTAGGAGCA ACAGGGGGT GGGGGTGC TTGATAGCA GGTCTGTGT TTCTTGTTA	239
ATGGGAATC GATAGCGCT TGGTGGTG TGGAGGGG CAGCTGGTG GGGAGGCT GGGGATGC TGGCTGGG GATCTCTCC GACGAGAT TTGCTGGG CAGCTGGG	359
GCGTAATAG GATAGTGG GCGTAATCC TGGGGGGG CAGGCGGAG GAGGCGGAG ATGATTGCA GGGGCTTG GCTTGGCC TACGCGCT GCGCTGGG GGGGGAGTT	479
GGTGGGGG GGTGGAGG GAGGCTGG GGTGGGGT GAGGCGG CCGGCTTCA TCGTGGG GGGGCTTG CCGGATAC CGGAGAGC ACGAGAGC GAGTATGG	599
-21 ATG AGC CAC ATC CTG GGT GGC GGC GTA TTG GGG GGG GTC CTG CTG CCG TTT CCG GCA CTG GGC GAT CAG GGC GGC AGC AGC CCG GGC GGG GTG GGC TAC	698
Met Ser His Ile Leu Arg Ala Ala Val Leu Ala Ala Val Leu Leu Pro Phe Pro Ala Leu Ala Asp Gln Ala Gly Lys Ser Pro Ala Gly Val Arg Tyr	12
CAC GGC GGC GAC GAA ATC ATC CTC CAG GGC TTC CAC TGG AAC GTC GTC GGC GAA GGG CCG AAC GAC TGG TAC AAC ATC CTC GGC CAA CAG GGC TGG AGG	797
His Gly Gly Asp Glu Ile Ile Leu Gln Gly Phe His Trp Asn Val Val Arg Glu Ala Pro Asn Asp Trp Tyr Asn Ile Leu Arg Gln Gln Ala Ser Thr	45
ATC GGG GGC GAC GGC TTC TGG GCA ATC TGG ATG CCG GTG CCG TGG GGT GAC TTC TCG AGC TGG ACC GAC GGC GGC AGC TCG GGC GGC GAA GGC TAC	896
Ile Ala Ala Asp Gly Phe Ser Ala Ile Trp Met Pro Val Pro Trp Arg Asp Phe Ser Ser Trp Thr Asp Gly Gly Lys Ser Gly Gly Gly Glu Gly Tyr	78
TTC TGG CAC GAC TTC AAC AGC AAC GGC GGC TAC GGC AGC GAC GGC CAG CTG GGC CAG GGC GGC GGC GCA CTC GGT GGC GGC GGG GTG AAG GTG CTC TAC	995
Phe Trp His Asp Phe Asn Lys Asn Gly Arg Tyr Gly Ser Asp Ala Gln Leu Arg Gln Ala Ala Gly Ala Leu Gly Gly Ala Gly Val Lys Val Leu Tyr	111
GAT GTG GTG CCG AAT CAC ATG AAC GGC GGC TAC CCG GAC AAG GAG ATC AAC CTG CCG GGC GGC CAG GGC TTC TGG GGC AAC GAC TGC GGC GAC CCG GGC	1094
Asp Val Val Pro Asn His Met Asn Arg Gly Tyr Pro Asp Lys Glu Ile Asn Leu Pro Ala Gly Gln Gly Phe Trp Arg Asn Asp Cys Ala Asp Pro Gly	144
AAC TAC CCG AAC GAC TGC GAC GAC GGT GAC GGC TTC ATC GGC GGC GAG TGG GAC CTG AAC ACC GGC CAT CCG CAG ATT TAC GGC ATG TTT GGC GAC GAG	1193
Asn Tyr Pro Asn Asp Cys Asp Asp Gly Asp Arg Phe Ile Gly Gly Glu Ser Asp Leu Asn Thr Gly His Pro Gln Ile Tyr Gly Met Phe Arg Asp Glu	177
CTT GGC AAC CTG GGC AGC GGC TAC GGC GGC GGC TTC GGC TTC GAC TTC GTT GGC GGC TAT GGG CCG GAG CCG GTC GAC AGC TGG ATG AGC GAC AGC	1292
Leu Ala Asn Leu Arg Ser Gly Tyr Gly Ala Gly Gly Phe Arg Phe Asp Phe Val Arg Gly Tyr Ala Pro Glu Arg Val Asp Ser Trp Met Ser Asp Ser	210
GCC GAC AGC AGC TTC TGC GTT GGC GAG CTG TGG AAA GGC CCT TCT GAA TAT CCG AGC TGG GAC TGG GGC AAC AGC GGG AGC TGG CAG CAG ATC ATC AAG	1391
Ala Asp Ser Ser Phe Cys Val Gly Glu Leu Trp Lys Gly Pro Ser Glu Tyr Pro Ser Trp Asp Trp Arg Asn Thr Ala Ser Trp Gln Gln Ile Ile Lys	243
GAC TGG TCG GAC CCG GGC AAG TGC CCG GTG TTC GAC TTC GGT CTC AAG GAG GGC ATG CAG AAC GGC TGG GTC GGC GAC TGG AAG CAT GGC CTC AAT GGC	1490
Asp Trp Ser Asp Arg Ala Lys Cys Pro Val Phe Asp Phe Ala Leu Lys Glu Arg Met Gln Asn Gly Ser Val Ala Asp Trp Lys His Gly Leu Asn Gly	276
AAC CCG GAC CCG GGC TGG GGC GAG GTG GGG GTG ACC TTC GTC GAC AAC CAC GAC ACC GGC TAT TGG CCG GGG CAG AAC GGC GGC CAG CAC CAC TGG GGG	1589
Asn Pro Asp Pro Arg Trp Arg Glu Val Ala Val Thr Phe Val Asp Asn His Asp Thr Gly Tyr Ser Pro Gly Gln Asn Gly Gly Gln His His Trp Ala	309
CTG CAG GAC GGG CTG ATC GGC CAG GGC TAC GGC TAC ATC CTC ACC AGC CCG GGC AGC CCG GTG GTG TAC TGG TGG CAC ATG TAC GAC TGG GGC TAC GGC	1688
Leu Gln Asp Gly Leu Ile Arg Gln Ala Tyr Ala Tyr Ile Leu Thr Ser Pro Gly Thr Pro Val Val Tyr Trp Ser His Met Tyr Asp Trp Gly Tyr Gly	342
GAC TTC ATC GGC CAG CTG ATC CAG GTG CCG GGC ACC GGC GGC GTG GGC GGC GAT TGG GGG ATC AGC TTC CAT AGC GGC TAC AGC GGT CTG GTC GGT ACC	1787
Asp Phe Ile Arg Gln Leu Ile Gln Val Arg Arg Thr Ala Gly Val Arg Ala Asp Ser Ala Ile Ser Phe His Ser Gly Tyr Ser Gly Leu Val Ala Thr	375
GTC AGC GGC AGC CAG CAG ACC CTG GTG GTG GGG CTC AAC TCG GAT CTG GGC AAC CCG GGC CAG GTT GGC AGC GGC AGC TTC AGC GAG GGG GTC AAC GGC	1886
Val Ser Gly Ser Gln Gln Thr Leu Val Val Ala Leu Asn Ser Asp Leu Ala Asn Pro Gly Gln Val Ala Ser Gly Ser Phe Ser Glu Ala Val Asn Ala	408
AGC AAC GGC CAG GTG GGC GTC TGG GGC AGC GGT AGC GGC GAT GGC GGC GGG AAT GAC GGC GGC GAG GGT GGC TTG GTC AAT GTG AAC TTT GGC TGC GAC	1985
Ser Asn Gly Gln Val Arg Val Trp Arg Ser Gly Ser Gly Asp Gly Gly Gly Asn Asp Gly Gly Glu Gly Gly Leu Val Asn Val Asn Phe Arg Cys Asp	441
AAC GGC GTG AGC CAG ATG GGC GAC AGC GTC TAC GGG GTG GGC AAC GTC AGC CAG CTC GGC AAC TGG AGC CCG GGC TCC GGG GTA CCG CTG ACC GAC ACC	2084
Asn Gly Val Thr Gln Met Gly Asp Ser Val Tyr Ala Val Gly Asn Val Ser Gln Leu Gly Asn Trp Ser Pro Ala Ser Ala Val Arg Leu Thr Asp Thr	474
AGC AGC TAT CCG ACC TGG AAG GGC AGC ATC GGC CTG CCG GAC GGT CAG AAC GTG GAA TGG AAG TGC CTG ATC GGC AAC GAG GGG GAC GGG AGC CTG GTG	2183
Ser Ser Tyr Pro Thr Trp Lys Gly Ser Ile Ala Leu Pro Asp Gly Gln Asn Val Glu Trp Lys Cys Leu Ile Arg Asn Glu Ala Asp Ala Thr Leu Val	507
GCT CAG TGG CAA TGG GGC GGC AAC CAG GTC CAG GGC GGC GGC GGC GGG AGC ACC AGC GGC TGG TTC TGA CGACATGGG GCGGGGCTC GGTACGGCT	2284
Arg Gln Trp Gln Ser Gly Gly Asn Asn Gln Val Gln Ala Ala Ala Gly Ala Ser Thr Ser Gly Ser Phe ***	530
AGCGGGGG GCTCTCTGG ACCAGGGTG GCGAGGAG AGCGGGGGA CGGGCGGG CCGGATGCT GCGAGGACA CCATAAAGC CTTGGGCTG CGCTGGGTA TCGAGGCTG	2404
TTGATGTGG CCGAGAGCG CTGAGCGCT TTGGGCTTG GCTTCTGG CCGGCTGAC CTGCTGATG CCGGCTGCT GCGCTTCTG ATCTGCTAG CCGGAGCGC CCGGTTGGC	2524
ATGGCGGC TCGAGGCA TCGAGGCA ATCTGGGA CGGGGTGGG TCTGCTGTC AGCGAGCT TCTTGGTAC GTTGCGAGC CTGAGGAGA ACCTTGGGA CGGCTGGC	2644
GAGGAGGC CTGAGGCT CGAGGCTAT GTGGGGGC ATGGAGCT GCGAGGAG GCGCTGAGC TGTGGGGA GCTGAGGG GTGAGCGCG CAGATGGGA GAGGAGCA	2764
GCTGGGGC GCTGTTGG GAGCTGAG GCGGAGCT GCGCTGATC GATGGGAG CCGCTGTC GCGGGGGG CCGAGCGG TCGCTGGC GATCTGAG TCGATTCT	2884
GCGGCTAG CAGGAGCTC TCGAGGGA GTTGGGAG GCGAGGAG TGGTGGCTA TTGATGAG CAGTTCATC TCGGCTGA CAGGAGAG CAGCTGCTT CAGTGGATC	3004
GCGGCTCT CAGTAAAGC ACTGATGA GCGGCTGC AGATC	3049

Table 1

Amino acid composition of *Pseudomonas saccharophila* G<sub>4</sub>-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. <sup>a</sup>	23 ( 4.3)

<sup>a</sup> Not determined

residues -21 to -1 preceding the N-terminal amino acid (Asp-1) of the G<sub>4</sub>-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<sub>4</sub>-forming amylase is secreted, it is likely that this 21-residue sequence is a genuine signal peptide.

The amino acid composition of the G<sub>4</sub>-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<sub>4</sub>-forming amylase.

When the entire primary sequence of the G<sub>4</sub>-forming amylase was compared with those of other amylolytic enzymes (various  $\alpha$ -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  $\alpha$ -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<sub>4</sub>-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<sub>4</sub>-forming amylase. A 3.1-kbp DNA fragment encoding the G<sub>4</sub>-forming amylase [17] contains two open reading frames, ORF-1 and ORF-2, the former of which actually encodes the G<sub>4</sub>-forming amylase gene (figs 1 and

	I	II	III
G <sub>4</sub> -forming amylase ( <i>P. saccharophila</i> )	112 DVVPNE	189 GFRDFVRG	288 TFVDNEDT
G <sub>6</sub> -forming amylase ( <i>B. sp. #707</i> )	103 DVVANE	233 GFRDVAKE	328 TFVDNHDS
$\alpha$ -amylase (human)	89 DAVINE	196 GFRDASKH	286 VFVDNEDN
$\alpha$ -amylase (barley)	101 DIVINE	208 DGRLONGPE	317 TFVDNEDT
$\alpha$ -amylase ( <i>A. oryzae</i> )	117 DVVANE	202 GLRIDTVKE	291 TFVNEEDN
$\alpha$ -amylase ( <i>B. amyloliquefaciens</i> )	98 DVVINE	226 GFRDAAKH	322 TFVNEHDT
$\alpha$ -amylase ( <i>B. subtilis</i> )	97 DAVINE	172 GFRDAAKH	263 TFVSEHDT
CGTase ( <i>B. stearothermophilus</i> )	131 DFAPNE	221 GFRDAVKE	318 TFIDNEDM
CGTase ( <i>K. pneumoniae</i> )	130 DYAPNE	219 ATRDAIKH	327 VTMDNEDM
isoamylase ( <i>P. amyloclavus</i> )	291 DVVINE	370 GFRDLASV	501 NFIDVHDG
pullulanase ( <i>K. aerogenes</i> )	400 DVVINE	471 GFRDLMGY	526 NYVSKEDN

Fig.3. Sequence comparison of G<sub>4</sub>-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<sub>6</sub>-forming amylase [14]; human saliva gland  $\alpha$ -amylase [4]; barley  $\alpha$ -amylase [5]; *Aspergillus oryzae*  $\alpha$ -amylase [6]; *Bacillus amyloliquefaciens*  $\alpha$ -amylase [7]; *Bacillus subtilis*  $\alpha$ -amylase [8]; *Bacillus stearothermophilus* cyclomaltodextrin glucanotransferase (CGTase) [12]; *Klebsiella pneumoniae* CGTase [13]; *Pseudomonas amyloclavus* 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<sub>4</sub>-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<sub>4</sub>-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 promoter sequences of other *Pseudomonas* genes [26], and is positioned appropriately relative to the Shine-Dalgarno sequence [25]. Thus, this sequence is probably the promoter of the *mta* gene. Nucleotide sequences similar to well-known *E. coli* promoters (<sup>-35</sup>TTGACA and <sup>-10</sup>TATAAT, see [27]) are not present upstream of the coding region. The possible sequences which may serve as the transcriptional promoter of the *mta* gene in *E. coli* cells are presumed to be <sup>536</sup>TTCAT and <sup>562</sup>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<sub>4</sub>-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<sub>4</sub>-forming amylase are not responsible for the difference in specificity between G<sub>4</sub>-forming amylase and other amylolytic enzymes. In other work, a truncated form of the G<sub>4</sub>-forming amylase, lacking its C-terminal portion (including domain III), has been obtained by digestion with *Staphylococcus aureus* V<sub>8</sub> 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., Kashiwagi, 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., Momma, 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.