APPLIED AND ENVIRONMENTAL MICROBIOLOGY, May 1991, p. 1554–1559
0099-2240/91/051554-06$02.00/0

Vol. 57, No. 5

α-Amylase of *Clostridium thermosulfurogenes* EM1: Nucleotide Sequence of the Gene, Processing of the Enzyme, and Comparison to Other α-Amylases

HUBERT BAHL,[1]* GERHARD BURCHHARDT,[1]† ANDREAS SPREINAT,[1] KLAUS HAECKEL,[1] ANJA WIENECKE,[1] BERNHARD SCHMIDT,[2] AND GARABED ANTRANIKIAN[1]‡

Institut für Mikrobiologie, Grisebachstrasse 8,[1] and Zentrum Biochemie,[2] Georg-August-Universität Göttingen, W-3400 Göttingen, Federal Republic of Germany

Received 21 November 1990/Accepted 22 February 1991

The nucleotide sequence of the α-amylase gene (*amyA*) from *Clostridium thermosulfurogenes* EM1 cloned in *Escherichia coli* was determined. The reading frame of the gene consisted of 2,121 bp. Comparison of the DNA sequence data with the amino acid sequence of the N terminus of the purified secreted protein of *C. thermosulfurogenes* EM1 suggested that the α-amylase is translated from mRNA as a secretory precursor with a signal peptide of 27 amino acid residues. The deduced amino acid sequence of the mature α-amylase contained 679 residues, resulting in a protein with a molecular mass of 75,112 Da. In *E. coli* the enzyme was transported to the periplasmic space and the signal peptide was cleaved at exactly the same site between two alanine residues. Comparison of the amino acid sequence of the *C. thermosulfurogenes* EM1 α-amylase with those from other bacterial and eucaryotic α-amylases showed several homologous regions, probably in the enzymatically functioning regions. The tentative Ca^{2+}-binding site (consensus region I) of this Ca^{2+}-independent enzyme showed only limited homology. The deduced amino acid sequence of a second obviously truncated open reading frame showed significant homology to the *malG* gene product of *E. coli*. Comparison of the α-amylase gene region of *C. thermosulfurogenes* EM1 (DSM3896) with the β-amylase gene region of *C. thermosulfurogenes* (ATCC 33743) indicated that both genes have been exchanged with each other at identical sites in the chromosomes of these strains.

Starch is an abundant carbon source in nature, and it can be degraded by many bacteria under aerobic as well as under anaerobic conditions. α-Amylase (1,4-α-D-glucan glucanohydrolase, EC 3.2.1.1.), which hydrolyzes α-1,4 glucosidic linkages in starch and starch fragments, is one of several enzymes involved in starch degradation and is widespread among microorganisms (30).

Studies of α-amylase aimed at the elucidation of its protein structure and function, its secretion through the cell membrane, and its industrial application have been performed mostly with members of the genus *Bacillus*, and quite a number of α-amylase genes have been cloned and sequenced, including those from *B. subtilis* (31), *B. stearothermophilus* (18), and *B. amyloliquefaciens* (29).

In contrast, little information is available on the synthesis, structure, and secretion of this enzyme from thermophilic anaerobic bacteria, which are a promising source of thermostable enzymes needed for industrial applications (1). Recently, *Clostridium thermosulfurogenes* EM1, which produces and secretes high levels of α-amylase, pullulanase, and α-glucosidase, was isolated (15). Interestingly, α-amylase and pullulanase were stable in the absence of metal ions, e.g., Ca^{2+}, under aerobic conditions, in contrast to most enzymes currently used in industry (2). The independence of these amylolytic enzymes from metal ions was documented by the addition of 10 mM EDTA; under these conditions 70 to 80% of the enzyme activity was still measurable. The genes coding for the α-amylase and pullulanase from *C. thermosulfurogenes* EM1 have been cloned and expressed in *Escherichia coli* (5, 7). To analyze the special features of the α-amylase of *C. thermosulfurogenes* EM1, this paper presents the nucleotide sequence of the gene, analysis of the amino acid sequence, and analysis of processing of the enzyme for secretion.

* Corresponding author.
† Present address: Department of Microbiology and Cell Science, University of Florida, Gainesville, FL 32611.
‡ Present address: Technische Universität Hamburg-Harburg, W-2100 Hamburg 90, Federal Republic of Germany.

MATERIALS AND METHODS

Bacterial strains, plasmids, and growth conditions. *C. thermosulfurogenes* EM1 (DSM3896) (14) was used for the purification of α-amylase. *E. coli* JM105 [*supE* Δ(*lac-proAB*) *hsdR4* F′(*traD36 proAB*$^+$ *lacI*^{q} *lacZ*ΔM15)] (32) was used as a host for pCT2 (7), and *E. coli* HB101 (*hsdS recA*)(pCT2) (4) was used to determine the intracellular localization of the *C. thermosulfurogenes* α-amylase in *E. coli*. *C. thermosulfurogenes* EM1 was grown in a continuous culture in mineral medium supplemented with 1% (wt/vol) starch as described by Antranikian et al. (2). *E. coli* was routinely grown in Luria-Bertani medium (26) supplemented with ampicillin (100 μg/ml) and isopropyl-β-D-thiogalactopyranoside (IPTG) (40 μg/ml) if required. For preparation of cell fractions, *E. coli* HB101(pCT2) was grown in the low-phosphate medium of Neu and Heppel (20) supplemented with 0.6% (vol/vol) glycerol and ampicillin (100 μg/ml).

Plasmids for sequencing reactions were isolated by a modification (3a) of the alkaline lysis procedure originally described by Birnboim and Doly (3). Standard methods for restriction endonuclease digestion and DNA electrophoresis procedures were used (26). Restriction endonucleases were obtained from GIBCO/BRL GmbH (Eggenstein, Federal Republic of Germany), Boehringer GmbH (Mannheim, Fed-

eral Republic of Germany), and Pharmacia LKB GmbH (Freiburg, Federal Republic of Germany). All enzymes were used according to the manufacturers' instructions.

DNA sequencing. DNA sequencing was carried out by the dideoxy-chain termination method of Sanger et al. (27) using [^{35}S]dATP and a T7 sequencing kit from Pharmacia LKB. Single-stranded templates were prepared from both strands of plasmid pCT2. Sequencing was started using commercially available M13/pUC universal sequencing forward primer and reverse primer. Synthetic oligonucleotides (17-mers) complementary to the ends of already sequenced templates were prepared with a Gene Assembler Plus (Pharmacia LKB) according to the instructions of the manufacturer and used as primers for continued sequencing. The dideoxy-terminated fragments were separated on wedge-shaped thick gradient gels (0.2 to 0.4 mm, 6% [wt/vol] polyacrylamide) by using a Macrophor sequencing unit from Pharmacia LKB.

Computer programs. Analysis of the DNA fragment sequenced and the deduced amino acid sequence was performed with the DNA Strider Program (16) on a Macintosh SE computer (Apple Computer Inc., Cupertino, Calif.). Sequence comparisons with all respective data from EMBL and GenBank were performed with the Wisconsin Genetics Computer Group sequence analysis software package Version 6.0 (Genetics Computer Group, University of Wisconsin Biotechnology Center, Madison).

Preparation of *E. coli* fractions. *E. coli* HB101(pCT2) was grown in low-phosphate medium to an optical density at 600 nm of 4 to 5. To separate the periplasmic from the inner membrane-cytoplasmic fraction, cells were subjected to osmotic shock by the following procedure (20). Cells from 1-ml batches were resuspended in 0.15 ml of 20% (wt/vol) sucrose–0.01 M Tris-HCl (pH 8.0) at 0°C and treated with 5 μl of 0.5 M Na-EDTA (pH 8.0). After incubation for 10 min on ice the suspension was centrifuged for 15 min at 13,000 × *g*, resuspended in 100 μl of distilled water, and incubated for 20 min on ice. Centrifugation for 5 min at 13,000 × *g* yielded the periplasmic fraction (supernatant) and inner membrane-cytoplasmic fraction (pellet). The latter was suspended in 100 μl of distilled water.

Determination of α-amylase activity. α-Amylase activity was measured in the supernatant of the culture broth of *C. thermosulfurogenes* EM1 prepared by centrifugation and in crude extracts and cell fractions of *E. coli*, prepared after rupture by French press treatment and removal of cell debris by centrifugation and as described above, respectively. α-Amylase was assayed by detection of the reducing sugar molecules that were liberated by hydrolysis of starch (2). The activity of 1 U of α-amylase is defined as that amount of enzyme which liberates 1 μmol of reducing sugar from starch per min, with maltose as a standard. Protein was determined by the method of Lowry et al. (13).

Purification of α-amylase and determination of the N-terminal amino acid sequence. The extracellular α-amylase of *C. thermosulfurogenes* EM1 was purified by the method of Spreinat and Antranikian and was used to determine the N-terminal amino acid sequence. Purification of the heterologous enzyme expressed in *E. coli* was done by the following protocol. *E. coli* JM105(pCT2) was grown in 10 liters of Luria-Bertani medium supplemented with ampicillin (100 μg/ml) to an optical density at 600 nm of 2. The cells were harvested by centrifugation, resuspended in 50 ml of 50 mM sodium acetate buffer (pH 5.0), and disrupted by two passages through a French press cell (SLM Instruments, Hannover, Federal Republic of Germany) at 1,300 lb/in^2. The crude extract was incubated for 30 min at 60°C to precipitate the majority of *E. coli* proteins. The thermostable *C. thermosulfurogenes* EM1 α-amylase remained soluble under these conditions. After centrifugation (15,000 × *g*, 15 min, 4°C) 3.2 U of α-amylase was obtained in 20 ml (800 mg of total protein). This solution was loaded on a Q-Sepharose Fast flow column (200 ml, 5 by 40 cm) (Pharmacia LKB) and equilibrated with 50 mM sodium acetate buffer (pH 5.0). By washing the column with equilibration buffer, α-amylase was detectable in the eluate. Fractions with α-amylase activity were pooled (52 ml) and concentrated to 720 μl by using an Amicon cell with a PM10-membrane (Amicon, Lexington, Mass.) under N_2 pressure; the final protein concentration was 0.3 mg/ml, with a specific α-amylase activity of 3.3 U/mg of protein.

Aliquots of the concentrated α-amylase fraction were analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) (12). α-Amylase was still active under these conditions (6) and was identified by running two parallel lanes of the gel. Proteins of one lane were stained with Serva Blue G (Serva GmbH, Heidelberg, Federal Republic of Germany). The other lane of the gel was soaked for 60 min at 0°C in sodium acetate buffer (pH 5.0) containing 2% (wt/vol) soluble starch and then incubated for 10 min at 65°C. Finally the gel was incubated in 0.15% (wt/vol) iodine–1.5% (wt/vol) potassium iodide solution for 2 min at room temperature. The appearance of a colorless band in the vicinity of the dark-colored starch gel indicated the presence of amylase activity (7).

The final preparation of the α-amylase (68,000-Da apparent molecular mass) contained only a few contaminating proteins with low molecular masses. Therefore, this preparation could be used to determine the N-terminal amino acid sequence. For this purpose, 100 μg of the partially purified α-amylase expressed in *E. coli* and 400 μg of the purified extracellular enzyme from *C. thermosulfurogenes* EM1 were subjected to SDS-PAGE and blotted on a polyvinylidene difluoride membrane by using a Fast Blot apparatus (Biometra, Göttingen, Federal Republic of Germany) (30 min, constant current [5 mA/cm^2 of gel]). Direct sequencing of the blotted α-amylases was performed using a 477A pulsed liquid-phase protein-peptide sequencer and a 120A on-line phenylthiohydantoin amino acid analyzer (both from Applied Biosystems, Foster City, Calif.) according to the instructions of the manufacturer.

Nucleotide sequence accession number. The DNA sequence data reported here (see Fig. 1) have been submitted to GenBank and assigned the accession number M57580.

RESULTS AND DISCUSSION

Nucleotide sequence of the *amyA* gene of *C. thermosulfurogenes* EM1. It was previously shown that the α-amylase gene (*amyA*) of *C. thermosulfurogenes* EM1 is localized on a 2.8-kbp *Sau*3A DNA fragment which was cloned in the *Bam*HI site of pUC18 to yield plasmid pCT2 (7). The nucleotide sequence of this region was determined and is shown in Fig. 1. Analysis of the sequence revealed a complete open reading frame (ORF) starting from an ATG at position 659 and terminating in a TAG at position 2779. The ORF coded for a protein of 706 amino acid residues with a calculated molecular mass of 78,083 Da. The N-terminal amino acid sequence of the extracellular α-amylase of *C. thermosulfurogenes* EM1 was determined (underlined in Fig. 1), and was in agreement with the DNA-deduced amino acid sequence from residue 28 onwards. This confirmed the

```
GATCCCGTTTCAATCAGTTATGATTCCATTAGTAGCGGAATTTGGGAAATTTCATTTTCTTACAAGGTCAGGGCTTGTATTTATGTACTTGGGATTTGGTTCAAGCTTAGGAGTGTTT   118
  IleProPheGlnSerValMetIleProLeuValAlaGluPheGlyLysPheHisPheLeuThrArgSerGlyLeuValPheMetTyrLeuGlyPheGlySerSerLeuGlyValPhe

TTATACTATGGTGCATTAAAAGGAATACCGACATCTTTAGATGAAGCAGCATTAATAGATGGTTGTAGTAGATTTAGAATTTACTGGAATATCATATTGCCTTTATTAAATCCTACTACG   238
LeuTyrTyrGlyAlaLeuLysGlyIleProThrSerLeuAspGluAlaAlaLeuIleAspGlyCysSerArgPheArgIleTyrTrpAsnIleIleLeuProLeuLeuAsnProThrThr

ATTACTTTGGCTGTATTGGATATCATGTGGATATGGAATGACTACTTATTGCCATCTTTAGTCATAAACAAAGTGGGTTCCAGGACTCTCCCATTAATGATTTTTTACTTCTTTAGTCAA   358
IleThrLeuAlaValLeuAspIleMetTrpIleTrpAsnAspTyrLeuLeuProSerLeuValIleAsnLysValGlySerArgThrLeuProLeuMetIlePheTyrPhePheSerGln

TACACAAAGCAATGGAATCTCGGTATGGCAGGGCTAACTATAGCAATTTTACCCGTTGTAATTTTCTACTTCTTGGCGCAGAGAAAATTGGTTACAGCCATAATAGCTGGTGCTGTTAAA   478
TyrThrLysGlnTrpAsnLeuGlyMetAlaGlyLeuThrIleAlaIleLeuProValValIlePheTyrPheLeuAlaGlnArgLysLeuValThrAlaIleIleAlaGlyAlaValLys

CAGTAAGATGAATATTTGTATTTAAAACAAGTTAATTTAAAATTATAGACGATATGCAATCGATATGAAAGTATTTTGACATTAATATCGATTGCTTTTATATAACAGAAAGGAGTGATT   598
Gln***                                                                    -35                  -10

GAAAACCATAAATTATTAAACAACTCGCTAATCAAAAAATAATTAAAAGGAGGGTCTTACATGAAAAAAACGTTTAAATTGATATTGGTGCTGATGCTTTCACTTACACTTGTTTTTGGA   718
                                                  S/D        MetLysLysThrPheLysLeuIleLeuValLeuMetLeuSerLeuThrLeuValPheGly

TTGACAGCACCAATACAGGCAGCTTCTGATACAGCAGTAAGTAATGTTGTCAATTACTCGACGGATGTAATTTATCAGATTGTCACTGATAGATTTGTTGATGGAAATACCAGCAACAAT   838
LeuThrAlaProIleGlnAlaAlaSerAspThrAlaValSerAsnValValAsnTyrSerThrAspValIleTyrGlnIleValThrAspArgPheValAspGlyAsnThrSerAsnAsn

CCAACTGGTGATTTGTATGATCCAACACATACAAGTCTGAAAAAGTACTTTGGTGGAGACTGGCAGGGTATTATTAACAAGATAAATGATGGTTATCTAACAGGAATGGGTGTAACAGCA   958
ProThrGlyAspLeuTyrAspProThrHisThrSerLeuLysLysTyrPheGlyGlyAspTrpGlnGlyIleIleAsnLysIleAsnAspGlyTyrLeuThrGlyMetGlyValThrAla

ATTTGGATATCTCAGCCTGTAGAAAATATTTATGCAGTCTTACCTGATTCAACTTTTGGTGGAAGTACATCGTACCATGGTTGGGCTCGTGATTTAAGAGAACAATCATACTTTGGAAGC  1078
IleTrpIleSerGlnProValGluAsnIleTyrAlaValLeuProAspSerThrPheGlyGlySerThrSerTyrHisGlyTrpAlaArgAspLeuArgGluGlnSerTyrPheGlySer

TTTACTGATTTTCAAAATTTGATTAATACAGCTCATGCACATAACATAAAAGTAATTATCGATTTTGCACCAAATCATACTTCACCTGCATCTGAAACCGATCCGACTTATGCTGAAAAT  1198
PheThrAspPheGlnAsnLeuIleAsnThrAlaHisAlaHisAsnIleLysValIleIleAspPheAlaProAsnHisThrSerProAlaSerGluThrAspProThrTyrAlaGluAsn

GGAAGGCTATACGACAATGGAACATTACTTGGTGGGTATACAAATGATACAAACGGGTATTTTCATCATTATGGAGGAACAGATTTTTCATCTTATGAAGATGGCATTTATCGTAACTTA  1318
GlyArgLeuTyrAspAsnGlyThrLeuLeuGlyGlyTyrThrAsnAspThrAsnGlyTyrPheHisHisTyrGlyGlyThrAspPheSerSerTyrGluAspGlyIleTyrArgAsnLeu

TTTGATTTAGCAGATTTAAATCAACAGAATAGTACAATCGATTCATACTTAAAATCAGCAATTAAGGTGTGGCTCGATATGGGAATAGACGGTATACGTCTAGATGCTGTAAAACATATG  1438
PheAspLeuAlaAspLeuAsnGlnGlnAsnSerThrIleAspSerTyrLeuLysSerAlaIleLysValTrpLeuAspMetGlyIleAspGlyIleArgLeuAspAlaValLysHisMet

CCGTTTGGATGGCAGAAGAATTTTATGGATAGCATATTGAGCTATAGGCCTGTATTTACATTTGGAGAGTGGTTTCTTGGAACGAATGAAATTGATGTAAATAACACATACTTTGCAAAT  1558
ProPheGlyTrpGlnLysAsnPheMetAspSerIleLeuSerTyrArgProValPheThrPheGlyGluTrpPheLeuGlyThrAsnGluIleAspValAsnAsnThrTyrPheAlaAsn

GAAAGCGGTATGAGTCTTTTGGACTTTAGGTTTTCTCAAAAAGTGAGGCAGGTATTTAGGGATAATACGGATACTATGTATGGACTTGATTCAATGATTCAATCAACTGCATCTGATTAT  1678
GluSerGlyMetSerLeuLeuAspPheArgPheSerGlnLysValArgGlnValPheArgAspAsnThrAspThrMetTyrGlyLeuAspSerMetIleGlnSerThrAlaSerAspTyr

AATTTTATAAATGATATGGTTACTTTTATTGATAATCATGATATGGATAGATTCTATAATGGTGGTTCTACTCGTCCAGTTGAACAAGCATTAGCATTTACTTTGACTTCGAGGGGAGTA  1798
AsnPheIleAsnAspMetValThrPheIleAspAsnHisAspMetAspArgPheTyrAsnGlyGlySerThrArgProValGluGlnAlaLeuAlaPheThrLeuThrSerArgGlyVal

CCTGCTATATATTATGGAACAGTATATGACAGGCAATGGAGACCCTATAATAGAGCTATGATGACCTCATTTAATACATCAACTACAGCATATAATGTAATTAAAAAACTTGCTCCTTTG  1918
ProAlaIleTyrTyrGlyThrValTyrAspArgGlnTrpArgProTyrAsnArgAlaMetMetThrSerPheAsnThrSerThrThrAlaTyrAsnValIleLysLysLeuAlaProLeu

CGCAAATCTAATCCTGCAATTGCTTATGGTACTACTCAGCAACGTTGGATAAACAATGATGTTTATATATATGAAAGAAAATTTGGGAATAATGTTGCTCTTGTTGCTATAAATAGAAAT  2038
ArgLysSerAsnProAlaIleAlaTyrGlyThrThrGlnGlnArgTrpIleAsnAsnAspValTyrIleTyrGluArgLysPheGlyAsnAsnValAlaLeuValAlaIleAsnArgAsn

CTTTCAACAAGTTATAATATAACAGGACTGTATACAGCGCTTCCTGCTGGTACTTATACTGATGTTCTTGGTGGACTTTTAAATGGTAATAGTATTTCTGTCGCGAGTGATGGTTCTGTA  2158
LeuSerThrSerTyrAsnIleThrGlyLeuTyrThrAlaLeuProAlaGlyThrTyrThrAspValLeuGlyGlyLeuLeuAsnGlyAsnSerIleSerValAlaSerAspGlySerVal

ACACCATTTACACTTAGTGCCGGTGAAGTTGCAGTATGGCAGTATGTAAGTTCAAGTAATTCTCCGTTGATAGGACATGTTGGGCCAACAATGACAAAAGCAGGGCAAACTATAACAATA  2278
ThrProPheThrLeuSerAlaGlyGluValAlaValTrpGlnTyrValSerSerSerAsnSerProLeuIleGlyHisValGlyProThrMetThrLysAlaGlyGlnThrIleThrIle

GATGGAAGAGGTTTTGGTACGACATCAGGACAAGTTTTGTTTGGTTCGACGGCTGGTACTATTGTCTCTTGGGATGATACAGAGGTAAAAGTAAAAGTTCCTTCGGTGACACCTGGTAAA  2398
AspGlyArgGlyPheGlyThrThrSerGlyGlnValLeuPheGlySerThrAlaGlyThrIleValSerTrpAspAspThrGluValLysValLysValProSerValThrProGlyLys

TATAATATATCATTAAAAACTTCAAGTGGTGCTACAAGCAATACTTATAATAATATTAACATTCTTACAGGCAATCAGATATGTGTTAGATTTGTTGTAAATAATGCAAGTACAGTGTAT  2518
TyrAsnIleSerLeuLysThrSerSerGlyAlaThrSerAsnThrTyrAsnAsnIleAsnIleLeuThrGlyAsnGlnIleCysValArgPheValValAsnAsnAlaSerThrValTyr

GGAGAGAATGTTTATCTTACTGGTAATGTAGCTGAATTGGGAAACTGGGATACGTCAAAAGCAATAGGACCAATGTTTAATCAGGTAGTATATCAGTATCCTACATGGTATTATGATGTA  2638
GlyGluAsnValTyrLeuThrGlyAsnValAlaGluLeuGlyAsnTrpAspThrSerLysAlaIleGlyProMetPheAsnGlnValValTyrGlnTyrProThrTrpTyrTyrAspVal

AGTGTTCCAGCGGGTACAACAATACAATTTAAGTTTATAAAGAAAAATGGCAATACGATAACATGGGAAGGTGGTAGTAACCATACTTATACAGTTCCATCATCTAGTACAGGTACTGTA  2758
SerValProAlaGlyThrThrIleGlnPheLysPheIleLysLysAsnGlyAsnThrIleThrTrpGluGlyGlySerAsnHisThrTyrThrValProSerSerSerThrGlyThrVal

ATTGTAAACTGGCAACAATAGATTAAAATAATAAGGGTCAAGAGATC                                                                          2805
IleValAsnTrpGlnGln***
```

FIG. 1. Nucleotide sequence of the α-amylase gene. Only the antisense strand of the α-amylase gene together with its flanking regions are shown. The deduced amino acid sequences of the precursor α-amylase and of an upstream truncated ORF are indicated below the DNA sequence. The possible −35 and −10 sequences in the promoter region and the putative ribosome-binding site (S/D) are marked. The chemically determined amino acid sequence of the N terminus of the purified extracellular *C. thermosulfurogenes* EM1 α-amylase is underlined. The signal peptidase cleavage site is marked by an arrowhead.

ORF of the *amyA* gene and indicated the presence of a leader peptide. The N-terminal end of the deduced *amyA* gene product showed the characteristics typical of signal peptides of secretory precursors (23): two positively charged amino acid residues adjacent to the N terminus, a hydrophobic core, a conserved signal peptidase cleavage site (Ile-X-Ala), and a helix-breaking proline residue preceding the cleavage site. It can be deduced that the mature *C. thermosulfurogenes* EM1 α-amylase is composed of 679 amino acid residues with a molecular mass of 75,112 Da. This is not in good agreement with the molecular mass estimated by SDS-PAGE of the *amyA* product in *E. coli* and of the extracellular enzyme from *C. thermosulfurogenes* EM1 (68,000 Da) (7). Since this enzyme is released from the cell surface to the culture medium by *C. thermosulfurogenes* EM1 only under certain growth conditions (2), one could speculate about a membrane anchor sequence in the mature protein which keeps the protein attached to the surface and which is cleaved off for secretion. Although a hydropathy plot according to Kyte and Doolittle (11) of the α-amylase showed some hydrophobic regions in the C-terminal end, they do not seem to have the necessary length and hydrophobicity to cross or stick deep enough in the membrane. A definite answer could be expected from the determination of the C-terminal amino acid sequence of the released enzyme.

Seven base pairs upstream the structural gene was preceded by the sequence 5′-AGGAGG-3′, which shows complementarity to the 3′ end of the 16S rRNAs of *E. coli* and *B. subtilis*. This sequence likely forms the ribosome-binding site (28). In *E. coli* the *amyA* gene was apparently transcribed by its own promoter (7). A sequence which resembled the consensus promoters of *E. coli* and *B. subtilis* was observed at bases 554 to 581. The −35 sequence was TTGACA, and the −10 sequence was TTATAT, with 16 bp in between. However, it is not clear whether this region functions as a promoter in *C. thermosulfurogenes* EM1 in order to regulate the expression of α-amylase. No unequivocal transcriptional terminator (25) was found upstream of the *amyA* gene, but the C-terminal end of another obviously truncated ORF (TAA stop codon at position 484) coded for a polypeptide with significant homology (30% identity in a 122-amino-acid overlap) to the *malG* gene product of *E. coli*. Therefore, it is possible that the α-amylase is translated from a polycistronic mRNA which is initiated further upstream of the *amyA* gene and encodes proteins related to starch degradation. The stretch of 50 bp downstream of the *amyA* gene, which is present on the cloned fragment, is too short to make a prediction about a possible transcription stop.

The *amyA* gene had a G+C content of 35 mol%, which agreed with the value of 37 mol% reported for the total genomic DNA of *C. thermosulfurogenes* EM1 (14). Codon usage of the *amyA* gene showed a strong bias towards codons using predominantly A and T, as expected from the relatively low G+C content.

Processing of the α-amylase in *E. coli*. The α-amylase of *C. thermosulfurogenes* EM1 when expressed in *E. coli* is not secreted into the culture medium (7). However, one might suggest that in *E. coli* this enzyme is transported across the inner membrane by a leader peptide-dependent pathway, as has been observed for several *Bacillus* extracellular enzymes cloned in *E. coli*. Furthermore, export of the α-amylase to the periplasmic space in *E. coli* could be expected, since in *C. thermosulfurogenes* EM1 the leader peptide was cleaved at a site similar to conserved signal peptidase cleavage sites of exported *E. coli* proteins. The localization of α-amylase in *E. coli* was determined by measuring the enzyme activity in the periplasmic and cytoplasmic-inner membrane fractions, prepared as described in Materials and Methods. The activities of β-galactosidase, a cytoplasmic marker protein, and of alkaline phosphatase, a periplasmic marker, were also determined to judge the purity of fractions (Table 1). The majority of the α-amylase activity was found in the periplasmic space. To analyze the processing site of the α-amylase during its export to the periplasmic space in *E. coli*, the heterologously expressed enzyme was purified and the N-terminal amino acid sequence was determined. The results indicated that the leader peptide was cleaved at exactly the same site in *E. coli* as in *C. thermosulfurogenes* EM1, between two alanine residues. It has been reported that the *B. subtilis* endo-β-1,3-1,4-glucanase is processed in *E. coli* at a site two amino acid residues distant from the processing site used in *B. subtilis* (6).

TABLE 1. Distribution of enzyme activities in *E. coli*(pCT2)

Enzyme	% Enzyme activity in:	
	Periplasmic space	Intracellular-inner membrane fraction
α-Amylase	81 (0)[a]	19 (0)
β-Galactosidase	1 (1)	99 (99)
Alkaline phosphatase	85 (81)	15 (19)

[a] Values in parentheses represent the enzyme activities in *E. coli*(pUC18).

Comparison of the α-amylase region of *C. thermosulfurogenes* EM1 (DSM3896) with the β-amylase region of strain ATCC 33743. The type strain of *C. thermosulfurogenes* (ATCC 33743) was reported to produce β-amylase and glucoamylase during growth on starch (8), in contrast to strain EM1 (DSM3896), which induces α-amylase, pullulanase, and α-glucosidase under these conditions (2). The β-amylase gene has also been cloned and sequenced (10). A comparison of the sequence data for the α-amylase and β-amylase gene regions of both strains gave some interesting results. As expected, both enzymes are clearly different, thus confirming the biochemical data (2, 8). Surprisingly, upstream of the genes a region of high homology was found, i.e., 100% identity on the amino acid level (Fig. 2). This region corresponded to the truncated ORF described above (homologous to *malG* of *E. coli*) and ended immediately downstream of the stop codon at position 484 of pCT2 (Fig. 2). Therefore, the genetic information, probably coding for a protein similar to MalG, is also located upstream of the β-amylase gene of strain ATCC 33743. However, in this case a frameshift occurs in the truncated ORF at a position marked in Fig. 2. These results indicate that both genes have

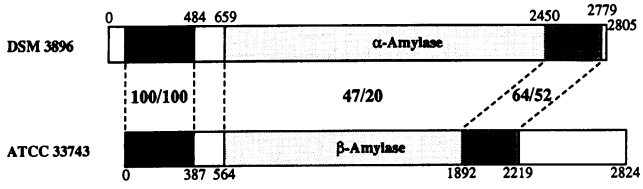

FIG. 2. Comparison of the α-amylase region of *C. thermosulfurogenes* EM1 (DSM3896) with the β-amylase region of *C. thermosulfurogenes* (ATCC 33743). The numbers above the top bar and below the bottom bar indicate nucleotide position. The bold numbers in between the bars represent the similarity/identity of the amino acid sequences. The vertical line in 'ORF' of strain ATCC 33743 marks a frameshift.

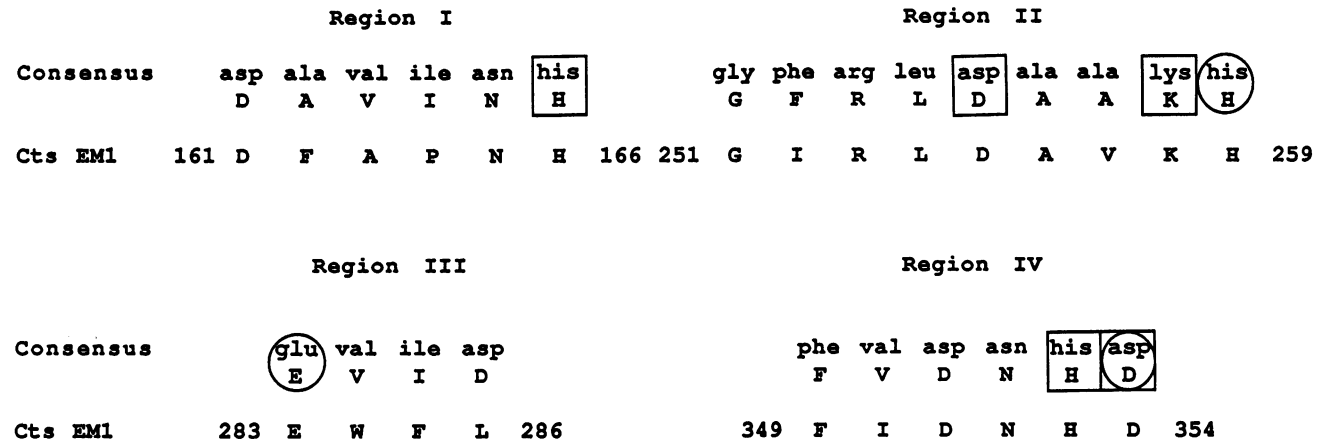

FIG. 3. Conserved regions (I to IV [19]) in the α-amylase of *C. thermosulfurogenes* EM1. Active-site residues and those involved in substrate binding proposed by Matsuura et al. (17) for Taka-amylase A from *A. oryzae* are surrounded by a circle and a rectangle, respectively. Cts EM1, *C. thermosulfurogenes* EM1.

been exchanged with each other at identical sites in the chromosomes of these two strains. Whether the inverted repeats found by Kitamoto et al. (10) near the end of the homologous region played a role in this genetic exchange can only be speculated.

Similarities between the α-amylase from *C. thermosulfurogenes* EM1 and other α-amylases. α-Amylases are known to contain four conserved regions (19, 30). On the basis of the known three-dimensional structure of the *Aspergillus oryzae* α-amylase it has been proposed that these regions are necessary for the catalytic activity of α-amylase and that they form the active center, the substrate-binding sites, and the binding site for the stabilizing calcium ion. Thus, it was not surprising that at least three of the four regions could also be identified in the α-amylase of *C. thermosulfurogenes* EM1 (Fig. 3), which has a similarity of 54% and an identity of 27% on the amino acid level to the α-amylase of *B. stearothermophilus*.

The consensus region I, which is nearest to the N-terminal end, has homology to the calcium-binding site of calmodulin, tropionin C, and the myosin light chain (24). Since the α-amylase of *C. thermosulfurogenes* EM1 was found to be Ca^{2+} independent, the structure of region I in this enzyme is of special interest. As is obvious from Fig. 3, the homology of region I in the *amyA* product of *C. thermosulfurogenes* EM1 to the consensus region is limited. This could explain the unusual behavior of this enzyme with respect to calcium dependence. It is possible that this protein contains no Ca^{2+} or that Ca^{2+} is bound very tightly at this altered binding site, which would explain the only slightly reduced activity in the presence of EDTA (2).

Region III is composed of four amino acid residues, only one of which is conserved in all α-amylases. The consensus sequence (E-V-I-D) is mainly found in α-amylases of eucaryotic origin (19). On the other hand, we found that in procaryotes the region preceding the E seems to be conserved; the sequence is F-T-V(F)-G(A)-E.

The thermostability of the α-amylase of *C. thermosulfurogenes* EM1 (up to 70°C) is between those of *B. subtilis* (thermolabile) and *B. stearothermophilus* (thermostable) (19). It has been suggested that the distance between region I and region II corresponds to the degree of thermostability (*B. subtilis*, 74 amino acids; *B. stearothermophilus*, 129 amino acids). In *C. thermosulfurogenes* EM1 both regions were 90 amino acid residues apart, which is in agreement with this hypothesis.

Increased numbers of cysteine residues have been shown to be responsible for heat stabilization of T4 lysozyme (22) and subtilisin BPN (21) and have been proposed as a possible factor in the thermostability of the β-amylase of *C. thermosulfurogenes* (ATCC 33743) (10). The α-amylase of *C. thermosulfurogenes* EM1 contains only one cysteine, which probably in this case plays no role in thermostability.

Aliphatic bonding may contribute to the thermostability of this enzyme. It can be evaluated by the aliphatic index proposed by Ikai (9), which is defined as the relative volume of a protein occupied by aliphatic side chains (alanine, valine, leucine, and isoleucine) and which is significantly higher for thermophilic proteins than for comparable mesophilic proteins. Thermostable enzymes like α-amylase with a low aliphatic index, i.e., an index of 66, have high contents of serine and threonine (hydrogen bond-forming amino acids). This was confirmed with the α-amylase of *C. thermosulfurogenes* EM1. The aliphatic index of this enzyme, 73, was relatively high, and of 679 amino acid residues 58 were serine and 74 were threonine. To what extent the obviously increased aliphatic and hydrogen bondings contribute to the thermostability of the α-amylase of *C. thermosulfurogenes* EM1 cannot be evaluated.

The data presented on the α-amylase of *C. thermosulfurogenes* EM1 are valuable for understanding the secretion and structure-function relationship of this enzyme and might add to the understanding of the complex nature of thermophily.

ACKNOWLEDGMENTS

This research was funded by grants from the "Bundesminsterium für Forschung und Technologie" to G.A. and H.B.

We thank C. Marck for a copy of the DNA Strider program and G. Gottschalk for generous support.

REFERENCES

1. **Antranikian, G.** 1990. Physiology and enzymology of thermophilic anaerobic bacteria degrading starch. FEMS Microbiol. Rev. **75**:201–218.
2. **Antranikian, G., C. Herzberg, and G. Gottschalk.** 1987. Production of thermostable α-amylase, pullulanase, and α-glucosidase in continuous culture by a new *Clostridium* isolate. Appl.

Environ. Microbiol. **53**:1668–1673.
3. **Birnboim, H. C., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7**:1513–1523.
3a. **Boehringer GmbH.** 1986. Guidelines for quick and simple plasmid sequencing. Boehringer GmbH, Mannheim, Federal Republic of Germany.
4. **Boyer, H. W., and D. Roulland-Dussoix.** 1969. A complementation analysis of the restriction and modification of DNA in *Escherichia coli*. J. Mol. Biol. **41**:459–472.
5. **Burchhardt, G., A. Wienecke, and H. Bahl.** 1991. Isolation of the pullulanase gene from *Clostridium thermosulfurogenes* (DSM 3896) and its expression in *Escherichia coli*. Curr. Microbiol. **22**:91–95.
6. **Gormley, E. P., B. A. Cantwell, P. J. Barker, R. S. Gilmour, and D. J. McConneil.** 1988. Secretion and processing of the *Bacillus subtilis* β-1,3-1,4-glucanase in *Escherichia coli*. Mol. Microbiol. **2**:813–819.
7. **Haeckel, K., and H. Bahl.** 1989. Cloning and expression of the thermostable α-amylase from *Clostridium thermosulfurogenes* (DSM3896) in *Escherichia coli*. FEMS Microbiol. Lett. **60**:333–338.
8. **Hyun, H. H., and J. G. Zeikus.** 1985. General biochemical characterization of thermostable extracellular β-amylase from *Clostridium thermosulfurogenes*. Appl. Environ. Microbiol. **49**: 1162–1167.
9. **Ikai, A.** 1980. Thermostability and aliphatic index of globular proteins. J. Biochem. **88**:1895–1898.
10. **Kitamoto, N., H. Yamagata, T. Kato, N. Tsukagoshi, and S. Udaka.** 1988. Cloning and sequencing of the gene encoding thermophilic β-amylase of *Clostridium thermosulfurogenes*. J. Bacteriol. **170**:5848–5854.
11. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**:105–132.
12. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **270**:680–685.
13. **Lowry, O. H., N. J. Rosebrough, A. L. Farr, and R. J. Randall.** 1951. Protein measurement with the Folin phenol reagent. J. Biol. Chem. **193**:265–275.
14. **Madi, E., and G. Antranikian.** 1989. Identification of a starch degrading anaerobic thermophile producing thermostable α-amylase and pullulanase. Appl. Microbiol. Biotechnol. **30**: 422–425.
15. **Madi, E., G. Antranikian, K. Ohmiya, and G. Gottschalk.** 1987. Thermostable amylolytic enzymes from a new *Clostridium* strain. Appl. Environ. Microbiol. **53**:1661–1667.
16. **Marck, C.** 1988. 'DNA Strider': a 'C' program for the fast analysis of DNA and protein sequences on the Apple Macintosh family of computers. Nucleic Acids Res. **16**:1829–1836.
17. **Matsuura, Y., M. Kusunoki, W. Harada, and M. Kakudo.** 1984. Structure and possible catalytic residues of Taka-amylase A. J. Biochem. **95**:697–702.
18. **Nakajima, R., T. Imanaka, and S. Aiba.** 1985. Nucleotide sequence of the *Bacillus stearothermophilus* α-amylase gene. J. Bacteriol. **163**:401–406.
19. **Nakajima, R., T. Imanaka, and S. Aiba.** 1986. Comparison of amino acid sequences of eleven different α-amylases. Appl. Microbiol. Biotechnol. **23**:355–360.
20. **Neu, H. C., and L. A. Heppel.** 1965. The release of enzymes from *Escherichia coli* by osmotic shock and during the formation of spheroplasts. J. Biol. Chem. **240**:3685–3692.
21. **Pantoliano, M. W., R. C. Ladner, P. N. Bryan, M. L. Rollence, J. F. Wood, and T. L. Poulos.** 1987. Protein engineering of subtilisin BPN′: enhanced stabilization through the introduction of two cysteines to form a disulfide bond. Biochemistry **26**: 2077–2082.
22. **Perry, L. J., and R. Wetzel.** 1984. Disulfide bond engineered into T4 lysozyme: stabilization of the protein toward thermal inactivation. Science **226**:555–557.
23. **Pugsley, A. P., and M. Schwartz.** 1985. Export and secretion of proteins by bacteria. FEMS Microbiol. Rev. **32**:3–38.
24. **Rogers, J. C.** 1985. Conserved amino acid sequence domains in alpha-amylases from plants, mammals and bacteria. Biochem. Biophys. Res. Commun. **128**:470–476.
25. **Rosenberg, M., and D. Court.** 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. Annu. Rev. Genet. **13**:319–353.
26. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
27. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
28. **Shine, J., and L. Dalgarno.** 1974. The 3′-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome-binding sites. Proc. Natl. Acad. Sci. USA **71**:1342–1346.
29. **Takkinen, K., R. F. Pettersen, N. Kalkkinen, I. Palva, H. Söderlund, and L. Kääriäinen.** 1983. Amino acid sequence of α-amylase from *Bacillus amyloliquefaciens* deduced from the nucleotide sequence of the cloned gene. J. Biol. Chem. **258**: 1007–1013.
30. **Vihinen, M., and P. Mäntsälä.** 1989. Microbial amylolytic enzymes. Crit. Rev. Biochem. Mol. Biol. **24**:329–418.
31. **Yang, M., A. Galizzi, and D. Henner.** 1983. Nucleotide sequence of the amylase gene from *Bacillus subtilis*. Nucleic Acids Res. **11**:237–249.
32. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33**:103–119.