α-Amylase from the Hyperthermophilic Archaeabacterium

Pyrococcus furiosus

CLONING AND SEQUENCING OF THE GENE AND EXPRESSI0N IN ESCHERICHIA COLI*

(Received for publication, August 24, 1992, and in revised form, July 28, 1993)

Kenneth A. Laderman†, K. Asada‡, T. Uemori§, H. Mukai§, Y. Taguchi§, I. Kato§, and
Christian B. Anfinsen††

From the ‡Department of Biology, The Johns Hopkins University, Baltimore, Maryland 21218 and §Takara Shuzo Co.,
Seta 3-4-1, Otsu, Shiga 520-21, Japan

A gene encoding a highly thermostable α-amylase from the hyperthermophilic archaeabacterium Pyrococcus furiosus was cloned and expressed in Escherichia coli. The nucleotide sequence of the gene predicts a 649-amino acid protein with a calculated molecular mass of 76.3 kDa, which corresponds well with the value obtained from purified enzyme using denaturing polyacrylamide gel electrophoresis. The NH₂ terminus of the deduced amino acid sequence corresponds precisely to that obtained from the purified enzyme, excluding the NH₂-terminal methionine. The amylase expressed in E. coli exhibits temperature-dependent activation characteristic of the original enzyme from P. furiosus, but has a higher apparent molecular weight which is attributed to the improper formation of the native quaternary structure. No homology was found with previously characterized pro-

Hyperthermophilic archaeabacteria provide an extraordinary opportunity to study the factors influencing protein thermostability. Unfortunately, due to the recentness of their discovery, the extent of the research completed in this area remains limited. Pyrococcus furiosus is an anacrobic marine heterotroph with an optimal growth temperature of 100 °C, isolated by Fiala and Stetter (1986) from solfataric mud off the coast of Vulcano island, Italy. α-Amylase activity has been reported in the cell homogenate and growth medium of P. furiosus (Brown et al., 1990; Koch et al., 1990), and the enzyme has been purified to homogeneity (Laderman et al., 1993). The amylase is a homodimer with a molecular mass of 130 kDa which exhibits optimal activity at the optimal growth temperature of the organism. In an attempt to better understand the mechanisms of the enzyme's inherent thermostability the gene coding for the α-amylase from P. furiosus was cloned and expressed in Escherichia coli, and the nucleotide sequence was determined. In addition the 3' and 5'-noncoding regions were analyzed in an attempt to identify sequences involved in transcriptional regulation, and a search for homology between the deduced amino acid sequence and other α-amylases was completed.

MATERIALS AND METHODS

Bacterial Strains— Cultures of P. furiosus (DSM 3638) were grown as described previously (Laderman et al., 1993). For cloning and expression of the α-amylase gene, E. coli strain MJ109 (rec A1, d(lac pro AB), end A1, gyr A86, thr -1, hsd R17, rel A1, sup E44, F' tra D 36, pro AB') was used.

Plasmids, Enzymes, and Chemicals—The vector pUC18 was used for cloning and DNA sequencing. For expression the vector pTV118N from Takara Shuzo Co. (Kyoto, Japan) was used.

Restriction endonucleases, alkaline phosphatase, DNA Blunting Kit, Random Primer DNA Labeling Kit, DNA Ligation Kit, T-DEAE Sequencing Kit, Mutan-K site-directed mutagensis system, and SeaKem Ultrapure Agarose, 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside were obtained from Takara Shuzo Co. (Kyoto, Japan). GeneCleave II Kit was obtained from Bio 101 (La Jolla, CA). PCR reagents and enzymes were from Perkin-Elmer Cetus. Hybond-N nylon hybridisation filters and [γ-32P]ATP were obtained from Amersham Corp. Ingredients for E. coli media were from Difco. Isopropyl-β-D-thiogalactopyranoside (IPTG) and ampicillin were obtained from Sigma. Ultrapure Urea was obtained from Bio-Rad.

Assay of Amylase Activity and Native Polyacrylamide Gel Electrophoresis—The dextrinizing activity of α-amylase was determined at 92 °C using the I/KI method as described previously (Laderman et al., 1993). One unit of the enzyme activity was defined as the amount which hydrolysed 1 mg of starch/min. Native gel electrophoresis and subsequent staining technique were performed as described elsewhere (Laderman et al., 1993).

Preparation of Chromosomal DNA from P. furiosus—The cells were harvested and approximately 0.1 g of cells (wet weight) was suspended in 0.5 ml of 0.05 M Tris-HCl (pH 8.0) containing 25% sucrose. To the suspension was added 0.1 ml of lysozyme (5 mg/ml). Incubation of the mixture for 1 h at 20 °C was followed by the addition 4 ml of SPT solution (150 mM NaCl, 1 mM EDTA, and 50 mM Tris-HCl (pH 8.0)), 0.5 ml of 5% SDS and 100 μl of proteinase K (10 mg/ml) were added, and the mixture was incubated for 1 h at 37 °C. The solution was extracted with phenol-chloroform, and the DNA was precipitated with 2 volumes of ethanol. DNA was recovered by washing and it was reared in 80% ethanol. The yield of genomic DNA was approximately 1.3 mg from 0.1 g of cells.

Preparation of a DNA Probe Using PCR—The NH₂-terminal sequence of the intact α-amylase as well as a peptide fragment, determined previously (Laderman et al., 1993), were used for the construction of three degenerate oligonucleotide probes (Fig. 1). The probes were synthesized using an Applied Biosystems model 380B DNA synthesizer.

The PCR reactions were performed using 500 ng of P. furiosus

1 The abbreviations used are: PCR, polymerase chain reaction; IPTG, isopropyl-β-D-thiogalactopyranoside; kb, kilobase(s)

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank℠/EMBL Data Bank with accession number(s) L22346.

† To whom correspondence should be addressed: Dept. of Biology, The Johns Hopkins University, 34th and Charles Sta., Baltimore, MD 21218. Tel: 410-516-8552; Fax: 410-516-5913

Printed in U.S.A.
Cloning of α-Amylase from P. furiosus

RESULTS

Cloning and Sequencing of the α-Amylase Gene. Characteristics of Coding and Noncoding Regions—The preparation of a probe using nested PCR resulted in the amplification of a DNA fragment of approximately 1 kilobase. The amplified DNA fragment was blunted using the DNA Blunting Kit (Takara Shuzo, Kyoto) and subcloned into the HindIII site of pUC18 (Sambrook et al., 1986). The cloned plasmid was sequenced by the dideoxy method.

When a Southern blot prepared with digested genomic DNA from P. furiosus was probed with the 32P-labeled PCR product, a 5.3-kb PstI fragment, a 3.1-kb HindIII fragment, a 5.3-kb XhoI fragment, and two EcoRI fragments of 0.7 and 2.4 kb were found to specifically hybridize to the probe.

Three clones carrying an identical 5.3-kb PstI fragment were identified by colony hybridization of an enriched gene bank of P. furiosus genomic DNA using the PCR product known to contain the coding region for the protein’s NH2-terminal sequence. One of these clones, shown to contain the α-amylase coding region in the same orientation as the vector-derived lac promoter, was digested with HindIII and allowed to self-ligate removing a portion of the 3′-noncoding region. The nucleotide sequence of the resulting 3.1-kb insert was determined in both orientations. The restriction map and sequencing strategy are shown in Fig. 2. The complete nucleotide sequence of the 3.1-kb insert is given in Fig. 3.

The α-amylase gene encompasses 1950 nucleotides, with the initiation codon GTG at position 715 (Fig. 3). There is no strong homology, preceding the coding region, with known archaebacterial, eukaryotic, or eubacterial consensus promoter sequences described previously. Immediately upstream of the coding region is the sequence GGTGGA, similar to the putative ribosome-binding site of the glyceraldehyde-3-phosphate dehydrogenase gene of Pyrococcus woesei (Zwickl et al., 1990). The G + C content of the α-amylase gene is 41.6%, slightly higher than the value reported for the total genome of 38% (Fiala and Stetter, 1986).

The five transcripts of SSV-1 (Reiter et al., 1988b) and the glyceraldehyde-3-phosphate dehydrogenase gene of P. woesei (Zwickl et al., 1990) include the sequence TTTTTT in a pyrimidine-rich region directly downstream of the termination codon. A pyrimidine-rich region exists 34 bases 3′ of the termination codon in the
Cloning of \( \alpha \)-Amylase from \( P. \) furiosus

Base pairs

\[
\begin{array}{cccc}
\text{P} & \text{E} & \text{Hc} & \text{E} \\
1000 & & & 2000 \text{ bp} \end{array}
\]

\[
\begin{array}{cccc}
\text{S} & \text{S} & \text{H} & \\
3000 & & & \\
\end{array}
\]

**Fig. 2.** Restriction map and sequencing strategy of the cloned \( P. \) furiosus \( H\)-HindIII insert carrying the \( \alpha \)-amylase gene of \( P. \) furiosus. Arrows indicate the individual sequence runs. ORF, open reading frame. 

\( P. \) PolI; \( E. \) EcoRI; \( H. \) HindIII; \( Hc. \) HincII; \( S. \) SacI.

\( \alpha \)-amylase gene, but unlike other archaeobacterial sequences examined, there is no pyrimidine-rich region immediately downstream from the stop codon.

Expression of \( P. \) furiosus \( \alpha \)-Amylase Gene in \( E. \) coli and Comparison with the Enzyme Purified from \( P. \) furiosus—For expression of the \( P. \) furiosus \( \alpha \)-amylase in \( E. \) coli an insert containing the gene flanked by archaeobacterial noncoding regions was transfected in the expression vector \( pTV118N \). This construct, with the \( P. \) furiosus 5' sequence intact, was found to not express any thermophilic \( \alpha \)-amylase activity. To prepare a more streamlined construct for expression in \( E. \) coli, a unique Ncol restriction site was created in the initiation codon, converting the initiation codon from GTG to ATG, and the \( P. \) furiosus noncoding region was removed. The resulting expression plasmid, denoted \( pKENF-NH \), placed the gene in the correct reading frame at an appropriate distance downstream of the vector promoter (Fig. 4).

IPTG-induced \( E. \) coli JM109 cells transformed with the plasmid \( pKENF-NH \) were found to produce 0.2 unit/ml of thermophilic amylase activity in the crude extract. The heterologously expressed amylase was compared with the enzyme purified from \( P. \) furiosus regarding molecular weight and temperature dependence of activity. When the apparent molecular weights of the recombinant protein and the isolated enzyme were compared on an activity stained native gel the protein produced in \( E. \) coli displayed a higher apparent molecular weight. The comparison of the temperature dependence of the \( \alpha \)-amylase activity between the purified and recombinant proteins displayed virtually identical relationships of relative activity as a function of temperature (Fig. 5).

When the complete amylase gene was analyzed, no protein or nucleotide sequence was found which displayed complete homology with the deduced NH\(_2\)-terminal sequence of the peptide fragment. This suggests that the peptide sequence, on which the synthesis of primer 3 was based, represents a contaminant and not a portion of the \( P. \) furiosus \( \alpha \)-amylase. It was therefore fortuitous that the degenerate primer prepared was able to act as a random primer for the PCR reaction.

To confirm that the \( \alpha \)-amylase expressed in \( E. \) coli was the enzyme purified from the bacterium, not a unique enzyme with a similar activity profile and a shared amino terminal
Cloning of α-Amylase from *P. furiosus*

sequence, the physical characteristics of the protein were examined. The gene product and the *P. furiosus* α-amylase were found to have comparable isoelectric points, specific activities, and pH of optimal activity (data not shown).

**Primary Structure of *P. furiosus* α-Amylase and Computer Aided Comparison with Enzyme Homologs from Mesophilic and Thermophilic Sources**—The deduced amino acid sequence of the *P. furiosus* α-amylase comprises 649 amino acids, with a calculated molecular mass of 76.3 kDa. This agrees well with the apparent molecular mass of the protein, determined by gel electrophoresis under denaturing conditions, of 66 kDa (Laderman et al., 1993).

It is known that the amylase of *P. furiosus* is secreted into the growth medium under native conditions (Koch et al., 1990). When the primary sequence of the protein was analyzed using the PC/GENE PSIGNAL program, no typical eubacterial or eukaryotic NH₂-terminal signal sequences were found. Using the standard activity assay, it was not possible to detect any thermophilic amylase activity in the extracellular media of JM109/pKENF-NH cell, confirming the absence of a signal sequence which would make the protein competent for export in *E. coli*. When the NH₂-terminal sequence of the native enzyme purified from *P. furiosus* was compared with the predicted NH₂-terminal sequence they were found to be identical, suggesting there is no NH₂-terminal processing.

The GenBank™ FASTA searches resulted in the acquisition of only a single strongly homologous protein sequence. The query with the protein sequence of the *P. furiosus* α-amylase, using the method of Pearson and Lipman (1988), identified a 41.9% identity in a 537-amino acid overlap in the sequence of one of the α-amylases from the extremely thermophilic bacterium *Dictyoglomus thermophilum* designated α-amylase A (Fukusumi et al., 1988). The complete sequence of the *D. thermophilum* α-amylase and a number of additional α-amylases from various sources were aligned using the PC/GENE CLUSTAL multiple sequence alignment program, and none displayed the high homology noted above. When a search for homology with previously identified consensus sequences, known to be located in the active center and participate in substrate binding in a number of amylases from a variety of sources (Bahl et al. 1991; Tsukagoshi et al., 1985), was performed no significant homology was found.

The codon usage of α-amylases from three thermophilic sources, *P. furiosus*, *D. thermophilum* (Fukusumi et al., 1988), and *Bacillus stearothermophilus* (Tsukagoshi et al., 1985), were compared. As noted previously, the higher the optimal temperature of activity the more extensive the bias against the usage of the dinucleotide CG (Zwickl et al., 1990). This trend is apparent not only for the arginine codons, but for the serine, proline, threonine, and alanine codons as well. No other anomalies in codon usage attributable to thermostability are apparent other than shifts inherent to the changes in amino acid composition.

The *D. thermophilum* α-amylase A displays physical characteristics similar to those observed with the *P. furiosus* α-amylase. The *D. thermophilum* enzyme exhibits optimal activity at 90 °C with approximately 70% residual activity following incubation for 1 h at this temperature (Fukusumi et al., 1988).
Cloning of α-Amylase from P. furiosus

A number of unusual characteristics of the P. furiosus α-amylase gene set it apart from a majority of the genes characterized to date. The gene utilizes the relatively rare initiation codon GTG, as does the glyceraldehyde-3-phosphate dehydrogenase gene from P. woei (Zwickl et al., 1990). It is possible that this represents a tendency in the usage of this initiation codon in hyperthermophilic archaea. Unfortunately the number of structural genes isolated from these sources are too limited to allow an accurate assessment of whether the initiation codon GTG is in fact a preferred initiation codon in these organisms, although this may be an intriguing possibility.

It is known that the production of α-amylase from P. furiosus is increased by the presence of starch (data not shown) indicating that the gene possesses an inducible promoter, a feature previously uncharacterized in hyperthermophilic archaea. When the 5′-noncoding region of the amylase gene was compared with the promotor sequences of other archaea, no homology with the consensus sequences previously identified in Sulfolobus and Methanococcus (Reiter et al., 1988a) was found. The lack of homology with the putative ribosome-binding site of the P. woei glyceraldehyde-3-phosphate dehydrogenase gene suggests either a difference in the 3′ terminus of the 16 S rRNA between the two closely related species, or a different promotor mechanism or both. In addition the P. furiosus amylase gene lacks the pyrimidine-rich region found immediatly downstream of the proteins 3′ termini in the forementioned archaebacterial genes. This lack of homology with previously investigated archaebacterial promoters and terminaion sequences may be a result of the local environment of the gene. The P. woei glyceraldehyde-3-phosphate dehydrogenase gene is flanked closely by a number of open reading frames, which would benefit from a translational coupling mechanism. In two instances with the SSV1 genes in Sulfolobus, shown to share similar termination sequence characteristics with the P. woei gene, this linkage between termination and re-initiation was observed. The P. furiosus gene exhibits no flanking open reading frames within the insert which was sequences, therefore precluding the necessity for translational coupling. This characteristic as well as the inducibility of the gene are unique among the archaebacterial genes investigated thus far.

The α-amylase from P. furiosus is one of a number of thermophilic proteins which have been expressed, in mesophilic hosts, in an active form. The three forms of α-amylase from D. thermophilum (Fukusumi et al., 1988), xylan-degrading enzymes from Caldoccum saccharolyticum (Luthi et al., 1990), and the glyceraldehyde-3-phosphate dehydrogenase from P. woei (Zwickl et al., 1990), produced under in vivo conditions at 73, 70, and 100 °C, respectively, have all been successfully expressed in E. coli. In these instances, the proteins produced by the transformed bacteria remained inactive until they were heated to the temperature appropriate to the enzyme’s native conditions. This suggests that the proper folding of the protein into a structure which can be temperature-activated is possible at temperatures other than those at which the protein is produced under native growth conditions.
The *P. furiosus* α-amylase which is a dimer differs from the aforementioned examples in that it additionally requires the formation of an appropriate quaternary structure. Although temperature-dependent amylase activity was observed in *E. coli*, the apparent native molecular weight of the enzyme was higher than the form purified from *P. furiosus*, suggesting improper subunit assembly. It is not possible, however, to determine whether this improper assembly is due to translation at lower temperature or to unidentified aspects of production in *E. coli*.

Perhaps the most interesting facet of this research are the evolutionary implications of the sequence homology between the α-amylases of *P. furiosus* and *D. thermophilum*. Based on molecular phylogeny using rRNA sequences, existing organisms are seen to fall into three coherent groups: eu-bacteria, and archaebacteria (Fox et al., 1980). Substantial physiological and structural differences exist between archaebacteria and eu-bacteria, which is evidence of their deep evolutionary separation (Woese, 1985). The phylogenetic tree prepared by Pace et al. (1986) places archaebacteria closer to the common ancestor of all the kingdoms than one or both of the other primary kingdoms, suggesting that archaebacteria are more primitive than one or both of the other lines. *D. thermophilum* is a Gram-negative, obligately anaerobic, extremely thermophilic bacterium (Saiki et al., 1985). It shares with *P. furiosus* a low G + C content and a tolerance for extremely thermal conditions, but is a member of a different phylogenetic kingdom. *D. thermophilum* has been shown to display a high degree of homology with the *P. furiosus* α-amylase, and therefore the *D. thermophilum* amylase A, may be an example of an archaic form of the enzyme which is well suited to extreme temperatures. In contrast, the *D. thermophilum* amylases B and C contain regions known to be well conserved in several *Bacillus* species, hog, mouse, and human amylases (Fukuosumi et al., 1988), and they represent the common form of the enzyme, various examples of which are active over a wide range of temperatures.

**REFERENCES**


