**amyP**, a reporter gene to study strain degeneration in *Clostridium acetobutylicum* ATCC 824

Fabrice Sabaté, Christian Croux, Emmanuel Cornillot, Philippe Soucaille

* Centre de Bioingénierie Gilbert Durand, UMR-CNRS 5504, Lab. Ass. INRA, Institut National des Sciences Appliquées, 135 avenue de Rangueil, 31077 Toulouse, France
* Genencor International, 925 Page Mill Road, Palo Alto, CA 94304, USA

Received 4 December 2001; received in revised form 22 February 2002; accepted 28 February 2002

First published online 20 March 2002

**Abstract**

*Clostridium acetobutylicum* produces an extracellular α-amylase when grown on glucose as the sole carbon source. This enzyme was previously characterized from a biochemical point of view but its encoding gene was never identified. The 2283-bp *amyP* gene encodes a 830Da mature protein with an N-terminal domain that exhibits strong identity to the family 13 glycosyl hydrolases such as the *Bacillus* α-amylases. Transcriptional analysis revealed that *amyP* is transcribed in solventogenic but not in acidogenic chemostat cultures. These results are in agreement with the extracellular α-amylase activities indicating that the expression of *amyP* is regulated at the transcriptional level. *amyP* is located on the pSOL1 megaplasmid that carries all the genes involved in the final steps of solvent formation. Degeneration of *C. acetobutylicum* has been associated to the loss of pSOL1. We demonstrate here that *amyP* can be used as a reporter system to quantitatively follow this phenomenon. © 2002 Published by Elsevier Science B.V. on behalf of the Federation of European Microbiological Societies.

**Keywords:** Megaplasmid; *amyP* amylase; Transcription; Degeneration; *Clostridium acetobutylicum*

1. Introduction

Degeneration is the process by which *Clostridium acetobutylicum* ATCC 824 loses its ability to produce acetone and butanol after serial subcultures or in continuous culture [1]. This phenomenon typically involves a gradual decrease in the amount of solvents produced [2]. The most important advance regarding the molecular basis for degeneration was the identification of a megaplasmid (pSOL1) that carries all the genes involved in the final steps of solvent formation and was shown to be lost in two degenerate strains [3–5]. It would be very useful to understand how pSOL1 is lost but there is currently no simple and rapid method to follow the loss of pSOL1 in a cell population of *C. acetobutylicum*.  

*C. acetobutylicum* ATCC 824 produces an extracellular α-amylase when grown on glucose as the sole carbon source. This enzyme has been purified previously and its N-terminal amino acid sequence determined [6]. DG1, a degenerated strain of *C. acetobutylicum*, which has lost pSOL1, does not produce an α-amylase when grown on glucose, suggesting that the encoding gene might be localized on the megaplasmid. Consequently, the previously characterized α-amylase might be a good reporter to follow the degeneration process.  

In the following report we have cloned and characterized from a molecular point of view *amyP*, the gene encoding the extracellular α-amylase produced during growth on glucose. After demonstrating that *amyP* is localized on pSOL1, a simple and rapid method to quantitatively follow the degeneration process of *C. acetobutylicum* during subculture is presented.

2. Materials and methods

2.1. Organisms and culture conditions

*C. acetobutylicum* ATCC 824 was obtained from the American Type Culture Collection. For DNA preparation, *C. acetobutylicum* was grown under anaerobic con-
ditions in batch mode at 37°C on tryptase–glucose–yeast extract medium (2YTG) [7]. For Northern (RNA) experiments and amylolytic assays, continuous cultures (dilution rate of 0.05 h\(^{-1}\)) were run at 35°C in phosphate-limited synthetic medium as previously described [8]. \textit{C. acetobutylicum} DG1 is a spontaneous mutant of \textit{C. acetobutylicum} ATCC 824 that has lost the pSOL1 megaplasmid [9]. \textit{Escherichia coli} ER2275 containing pUC18 (New England Biolabs, Beverly, MA, USA) was grown on Luria–Bertani broth supplemented with 100 mg l\(^{-1}\) ampicillin.

2.2. Nucleic acid isolation and manipulation

For cloning and PCR purposes, standard recombinant DNA techniques were used [10]. For pulsed field gel electrophoresis, unsheared total DNA was prepared by the method of Wilkinson and Young [11] with the modifications of Cornillot et al. [4]. Total RNA was extracted by the hot phenol–chloroform method as previously [12,13].

2.3. Hybridization

For the identification of small restriction fragments containing the amylose gene (amy\(P\)), total DNA of \textit{C. acetobutylicum} was digested to completion with \textit{PstI} and \textit{BglII} and separated by electrophoresis. For the genomic localization of amy\(P\), agarose embedded DNA was digested by \textit{SmaI}, subjected to pulsed field gel electrophoresis and Southern blotting. For Northern blot experiments, total RNA was separated on 1% (w/v) denaturing formaldehyde agarose gels and transferred to nylon membranes (Schleicher and Schuell Inc., Keene, NH, USA) as described by Sambrook et al. [10]. The hybridization probe used for Southern and Northern was synthesized by PCR using two degenerate oligonucleotides (Amy1

\[ \text{5'GGWGTTATGYTWCATGCWTTTGATGG-3'} \]

Amy2

\[ \text{5'-ATCWAWCKAAWCCATCWGCWCC-3'} \]

derived from the \textit{C. acetobutylicum} \(\alpha\) amylase N-terminal amino acid sequence (DGVMHLAFDW) and a conserved internal amino acid sequence (GADGFR(F/I)ID) of various \(\alpha\) amylases respectively. The DNA probe was radiolabeled with [\(\alpha\)-\(32\)P]dATP by the random-priming method with the Megaprime DNA labeling system (Amersham, Les Ulis, France). Hybridization and washing were done according to the manufacturer’s instructions for Southern or Northern blots (Schleicher and Schuell).

2.4. Construction and screening of genomic banks

Total and partial genomic banks were constructed with completely digested chromosomal DNA of \textit{C. acetobutylicum}. For partial libraries, the 8–10-kb fraction of \textit{PstI} fragments and the 10–12-kb fraction of \textit{BglII} fragments were isolated by gel electrophoresis on low-melting point agarose. Once ligated to \textit{PstI-} or \textit{BamHI}-digested (and dephosphorylated) pUC18, the preparations were used to transform competent \textit{E. coli} ER2275. The recombinant colonies were transferred to Hybond-N1 membranes (Amersham) by replica plating, after which they were lysed by alkaline treatment as recommended by the manufacturer. DNA was fixed by heating at 80°C for 2 h. Membranes were then screened by hybridization as for Southern blot experiments.

2.5. Determination of the transcription start site

Primer extension analysis was performed with Superscript reverse transcriptase (BRL Life Technologies). The 21-mer oligonucleotide AMY-PE2 (5’-CCGGAAGTTTCCTATTCTCAC-3’) complementary to the 5’ end of the amy\(P\) transcript was end-labeled with [\(\gamma\)-\(32\)P]dATP (specific activity, 5000 Ci mmol\(^{-1}\); Amersham) using T4 polynucleotide kinase (Amersham). Annealing was done by mixing 1 pmole of labeled AMY-PE2 with 3 \(\mu\)g of total RNA in a volume of 6 \(\mu\)l, incubating at 70°C for 10 min followed by immediate cooling on ice. After addition of 2 \(\mu\)l of 5\(\times\)buffer, 1 \(\mu\)l of 0.1 M dithiothreitol, 0.5 \(\mu\)l of 10 mM dATP, 10 mM dCTP, 10 mM dGTP, 10 mM dTTP, and 0.5 \(\mu\)l Rтасe Superscript, the reverse transcription reaction was incubated at 42°C for 50 min. Inactivation of the reverse transcriptase was done at 70°C for 15 min and after cooling on ice, 5 \(\mu\)l of stop buffer from the sequencing kit were added. To map the exact transcriptional start site, sequencing reactions were performed on the PCR amplified DNA region using the same primer that was used for the primer extension reaction. Sequencing and reverse transcription reactions were analyzed on a 5% polyacrylamide sequencing gel.

2.6. Nucleotide sequence accession number

The sequence data reported here have been submitted to GenBank (assigned accession No. AF164199) before the release on the Internet and the publication [14] of the entire sequence of the genome of \textit{C. acetobutylicum}.

3. Results and discussion

3.1. Cloning of the DNA region encoding the \(\alpha\)-amylase

The \(\alpha\)-amylase produced during growth on glucose was previously purified [6]. Two degenerate oligonucleotides were designed (see Section 2) from the N-terminal amino acid sequence and an internal conserved sequence of various \(\alpha\)-amylases and used to amplify the encoding gene. Sequencing of the 0.5-kb PCR product confirmed the N-terminal sequence of the purified clostridial \(\alpha\)-amylase. Southern blotting using this PCR probe showed single hybridization signals with various restriction enzyme digests of \textit{Clostridium} genomic DNA (9 and 11 kb for \textit{PstI}}
and BglII respectively). To clone large fragments of chromosomal DNA containing the whole α-amylase gene and surrounding regions, total and partial PstI and BglII genomic libraries were constructed in pUC18. However, this approach was not successful and no positive clones were isolated from either of these two libraries. These results tended to suggest that the entire K-amylase gene or other flanking regions were toxic for E. coli, as was already found with the cloning of the K-amylase genes from Bacillus subtilis [15,16] and Lactobacillus plantarum A6 [17].

An inverse-PCR strategy was then used to determine the entire sequence of the 9-kb PstI fragment. PstI fragments between 8 and 10 kb were agarose-purified and self-ligated. PCR amplification was then performed on these self-ligated fragments using two divergent oligonucleotide primers derived from the sequence of the 0.5-kb PCR product. An 8.5-kb fragment was successfully amplified and sequences upstream and downstream from the initial 0.5-kb PCR fragment were determined by gene walking.

To avoid errors due to misincorporation by DNA polymerase in the earlier cycles of PCR amplification, a pool of five independently PCR-amplified fragments was used as a matrix for sequencing.

3.2. Nucleotide sequence analysis of amyP

A total of 4421 bp were sequenced on both strands of the 8.5-kb fragment. This region contained two complete open reading frames (ORFs) (ORF1 and ORF2), transcribed in opposite directions, and the 5' region of a truncated ORF3 (GenBank accession No. AF164199).

Sequence homology allowed the unambiguous identification of the ORF1 as the gene encoding the α-amylase of C. acetobutylicum produced during growth on glucose. This gene of 2283 nucleotides was named amyP. A putative ribosome-binding site (GGAGA) was located 9 bp upstream from the UUG start codon. A 30-bp region with a stem loop structure and a calculated free energy of −17.8 kcal mol⁻¹ was identified immediately downstream from its UAA stop codon. This structure, followed by a short T-rich sequence, probably functions as a Rho-independent transcriptional terminator [13]. It may also serve as a transcriptional terminator of ORF2, since it is also located immediately downstream from the UAA stop codon.

Comparison with the deduced amino acid sequence and the N-terminus sequence of the mature extracellular protein [6] indicated that the protein is initially translated with a cleavable N-terminal peptide of 45 residues. The sequence of the first 35 residues presents the typical characteristics of bacterial signal peptides [18]. The sequence (10 amino acids (aa) long) between the signal peptide and the mature protein may correspond to a pro-peptide, as already described for other Gram-positive exoproteins [19]. It can then be deduced that the mature clostridial α-amylase is composed of 715 amino acid residues, with a calculated molecular mass of 77904 Da and an isoelectric point of 4.36. These values are in good agreement with the experimental data determined from the purified mature protein, molecular mass of 84 kDa, and pI of 4.7 [6]. The N-terminal amino acid sequence determined from the purified mature protein showed only one difference with that deduced from the nucleotide sequence, a proline residue (+3) substituting for an arginine residue. However, as these amino acids are eluted at very similar retention times on the Edman sequencer it was probably due to a protein sequencing error.

Fig. 1. Alignment of putative motifs essential for raw starch binding in the glucoamylase from Aspergillus awamori var. kawachi [23], the α-amylase from B. subtilis sp. 1018, the CGTase from Bacillus stearothermophilus, the G4-forming amylase from Pseudomonas stutzeri MO-1, the glucoamylase from Aspergillus niger [24] and the three repeated modules of the C-terminal part of the clostridial α-amylase. Identical matches (vertical lines) and conservative substitutions (double dots) are indicated. Asterisks indicate the positions of the invariable amino acids. The identical amino acids in four or more of the eight motifs are in bold type. The number in brackets in the right margin indicates the distance between the two invariable W residues in each motif.

| A. awamori | 650-Π L W Y V T V L P A G-57L | W-589 (27) |
| B. subtilis 1018 | 630-Π N W Y V D S V P A G-645 | W-662 (28) |
| B. stear. | 627-Π T W K I D V S V P E G-638 | W-656 (27) |
| Pseudomonas | 474-Π T W K G S A L P A G-485 | W-506 (30) |
| A. niger | 562-Π L W Y V T V L P A G-573 | W-591 (27) |
| AmyP, mod1 | 480-Π K Q T T T P E-491 | W-506 (24) |
| AmyP, mod2 | 580-Q F W E T A I-591 | W-605 (23) |
| AmyP, mod3 | 674-Π R L A G T L P G-685 | W-700 (24) |

Fig. 2. amyP is located on the pSOL1 megaplasmid. Unrestricted and SmaI-digested DNA from C. acetobutylicum ATCC 824 (lane 1) and DG1 (lane 2) was separated by PFGE and either stained with ethidium bromide (A) or hybridized with the PCR-amplified amyP gene as a probe (B).
3.3. Amino acid sequence comparison of AmyP

Sequence comparisons revealed that the N-terminal part, corresponding to the first 450 amino acids, has all the typical features of the well characterized \( \alpha \)-amylase family, currently known as the glycosyl hydrolase family 13 [20,21]. The eight invariant amino acid residues displayed by the amylolytic enzymes are all conserved in the AmyP sequence. Particularly D-177, E-217 and D-276, corresponding to D-206, E-230 and D-297 respectively in TAKA amylase A from \textit{Aspergillus oryzae} [22], are the three crucial invariant carboxylic acid residues from the catalytic triad involved in glycosidic bond cleavage.

A detailed analysis of the C-terminal amino acid sequence of AmyP revealed that this region might have originated from the fusion of three repeated modules, two complete of about 100 aa (positions 451–550 and 551–643 respectively) and one truncated of about 70 aa (position 644–715). This particular structure has recently been described for the C-terminal amino acid sequences of the \textit{L. plantarum A6} and \textit{Lactobacillus amylovorus} \( \alpha \)-amylases [17], consisting respectively of four 104 aa and five 91 aa perfectly homologous repeats. Interestingly, each repeat module of the clostridial \( \alpha \)-amylase contained sequences homologous to the typical motifs assigned to be essential for raw starch digestion [23,24] (Fig. 1). Although the AmyP protein secreted by \textit{C. acetobutylicum} shared typical motifs shown to play an important role in raw starch hydrolysis, purified AmyP has no detectable activity on starch azure and the activity on raw starch was less than 2% of the activity on soluble starch. All of these data suggest that this \( \alpha \)-amylase is not a raw starch hydrolyzing enzyme.

3.4. \textit{pSOL1} localization of \textit{amyP}

Interestingly, homology searches allowed the identification of the shorter ORF2 and the 3’ truncated ORF3 as the previously described unknown ORF2 and the 5’ part of the \textit{adc} gene respectively [12], later demonstrated to be located immediately downstream from the \textit{sol} operon [2]. As it was recently demonstrated that the \textit{sol} operon is located on the \textit{pSOL1} megaplasmid [3,5], it was highly probable that \textit{amyP} was also located on \textit{pSOL1}. To unambiguously establish this, an internal radiolabeled fragment of \textit{amyP} was used as a probe for Southern analysis of \textit{SmaI} digestions of total DNA from \textit{C. acetobutylicum} ATCC 824 and DG1, a degenerated strain that no longer contains the \textit{pSOL1} megaplasmid [3,5]. The results shown in Fig. 2 demonstrated that \textit{amyP} was absent from DG1 and that a hybridization signal was only obtained with the 190-kb \textit{SmaI} band corresponding to the \textit{pSOL1} megaplasmid. To our knowledge this is the first description of an \( \alpha \)-amylase gene carried by a plasmid.

3.5. Transcriptional analysis of \textit{amyP}

Northern blot analyses were conducted with total RNA isolated from cells grown in phosphate-limited chemostat cultures (dilution rate 0.05 h\(^{-1}\)) producing acids (pH 6.5) and solvents (pH 4.4). The fermentation product profiles of the two cultures were within the experimental error margin (\( \pm 5\% \)) of our previously published data [25]. Us-

![Fig. 3. Northern blot analysis of total RNA isolated from \textit{C. acetobutylicum} ATCC 824 in acidogenic conditions (lane 1) and solventogenic conditions (lane 2).](https://academic.oup.com/femsle/article-abstract/210/1/93/601690)
ing an internal radiolabeled fragment of the \textit{amyP} gene as a probe, a single hybridization signal of 2.5 kb was detected for solvent producing cells while no signal was obtained for acid producing cells (Fig. 3). These Northern blot results were in agreement with the \(\alpha\)-amylase activities of 0.05 and 0.41 U ml\(^{-1}\) detected in acidogenic and solventogenic cultures respectively. These data demonstrate that the expression of \textit{amyP} is regulated at the transcriptional level.

Primer extension analyses were performed on RNA extracted from cells in the two different physiological conditions (Fig. 4A). Extension products were only obtained for solvent producing cells, confirming the results of Northern blots. Two major transcription start sites were detected. P1 was located 56 bp upstream from the UUG start codon of \textit{amyP} while P2 was located 46 bp downstream from P1 immediately after the ribosome-binding site. The deduced -10 and -35 sequences of both promoters showed low homology to the consensus region for the \(\sigma^{37}\) sigma factor of Gram-positive bacteria (Fig. 4B).

\subsection*{3.6. Use of \textit{amyP} to study the degeneration process of \textit{C. acetobutylicum}}

We previously proposed that the loss of pSOL1 would explain the apparent characteristics of the degeneration process of \textit{C. acetobutylicum}: segregational pSOL1 loss by a fraction of the cell population would lead to decreased solvent production until all cells in the population lose pSOL1 [5].

In order to follow the degeneration process, we investigated the possibility of using \textit{amyP} as a reporter gene for the presence of pSOL1. Wild-type \textit{C. acetobutylicum} and the DG1 mutant (that has lost pSOL1) were spread on an agar plate containing starch (2\%) and subjected to iodine staining. The results showed that both strains produced a halo of starch hydrolysis although it was larger for the wild-type strain than for DG1 (Fig. 5A). This indicated that \textit{C. acetobutylicum} possesses at least one more \(\alpha\)-amylase encoded on the chromosome. In a second experiment, \textit{C. acetobutylicum} ATCC 824 and DG1 were spread on an agar plate containing starch (2\%) and glucose (0.2\%). The results in Fig. 5B show that DG1 did not produce a halo of starch hydrolysis whereas the ATCC 824 strain did. This result demonstrated that the other(s) \(\alpha\)-amylase gene(s) was/were subjected to catabolic repression by glucose while \textit{amyP} is not. To follow strain degeneration, \textit{C. acetobutylicum} was subcultured every 24 h in a synthetic medium and regularly spread on starch+glucose agar plates. The results showed that after 12 subcultures, 10\% of the clones have lost the amylolytic activity. At the 20th subculture, only 5\% of the clones retained an amylolytic activity. To confirm the absence of pSOL1 in the strains with a negative amylase phenotype, total DNA of these strains was extracted and analyzed by pulsed field gel electrophoresis (data not shown). All of the clones with a negative amylase phenotype had lost pSOL1, demonstrating that \textit{amyP} can be used as a pSOL1 marker.

In conclusion, the \textit{amyP} gene located on the megaplasmid pSOL1 is not sensitive to catabolic repression unlike the other(s) amylase gene(s) located on the chromosome. Taken together, these results show that \textit{amyP} is a very good reporter gene to quantitatively study the degeneration phenomenon of \textit{C. acetobutylicum}.

\section*{References}


