Volume 239, number 2, 179–184 FEB 06438 November 1988

Similar short elements in the 5′ regions of the *STA2* and *SGA* genes from *Saccharomyces cerevisiae*

J.M. Pardo, E. Iáñez, M. Zalacaín, M.G. Claros and A. Jiménez

Centro de Biología Molecular, Universidad Autonoma, Cantoblanco, 28049 Madrid, Spain

Received 15 July 1988; revised version received 31 August 1988

The 5′ regions of the *SGA* and *STA2* genes, encoding the intra- and extracellular glucoamylases, respectively, from *Saccharomyces cerevisiae* have been sequenced. In addition, the transcription initiation sites have been determined. Four distinct short elements (named I to IV) were found in both genes. Element III has the consensus sequence PuCATTTAPiG with a bilateral symmetry around the central T, and is present in both genes as a direct repeat. This motive seems responsible for the coregulation of *STA2* and *SGA* by the repressor *STA10* gene of *S. cerevisiae*.

DNA sequencing; Transcription initiation; Regulatory element; *STA2* gene; *SGA* gene; (Yeast)

1. INTRODUCTION

The yeast *Saccharomyces cerevisiae* contains the *SGA* gene for a sporulation-specific intracellular glucoamylase [1–3]. In addition, certain varieties of *S. cerevisiae* (previously known as *S. diastaticus*) carry any of the STA1, STA2 or STA3 genes which code for extracellular glucoamylases, and are located at different genomic loci [4]. Restriction, transcription and sequencing analysis showed that the *STA1* to *STA3* genes are derived from the *SGA* gene by at least one recombinational event with another, not yet characterized gene. This unknown gene provides the export domain of the STA1 to STA3 gene products [2,5,6] and is known to transcribe a mRNA of 5.4 kb [2]. The *SGA* gene provides the catalytic domain.

The *STA1* to *STA3* and the *SGA* genes were shown to be coregulated negatively at the transcriptional level by the *STA10* gene of *S. cerevisiae* [2,7,8]. Interestingly, the restriction maps of the 5′ regions of *STA1* to *STA3* genes are totally different from that of the *SGA* gene [2,3]. However, it would be expected that they contain possibly short, homologous regions that would account for their coregulation by *STA10*. In order to investigate this possibility we have sequenced the 5′ regions of the *STA2* and *SGA* genes.

Correspondence address: A. Jiménez, Centro de Biología Molecular, Universidad Autonoma, Cantoblanco, 28049 Madrid, Spain

2. MATERIALS AND METHODS

The 1481 bp *Bam*HI-*Bam*HI DNA fragment (fig.1) from plasmid JMp40 [2] corresponding to the 5′ region of the *STA2* gene and the 2045 bp *Eco*RI-*Sal*I DNA fragment (fig.2) from plasmid JMp79 [2] comprising the 5′ region of *SGA* gene were sequenced by the dideoxynucleotide chain termination method [9] as modified by others [10]. *Saccharomyces cerevisiae* strains MCCX1-5d (*a, leu2-3, 112, his4, sta°, SGA, sta10°*), M1-2b (*α, trp1-289, ura3-52, gal2, sta°, SGA, STA10*) and 5206-1b (*a, arg4, SGA, STA2, sta10°*) were described elsewhere [2]. Plasmids JMp40 and JMp79 containing *STA2* and *SGA*, respectively, were described previously [2].

Transcription initiation sites were determined by the high-resolution S_1 mapping procedure described elsewhere [11]. For *STA2*, 150 ng of the 1600 bp *Pst*I-*Hin*dIII DNA fragment from plasmid JMp40 (fig.1) was labeled with [γ-^{32}P]ATP and then hybridized to 30 ng of poly(A$^+$) RNA from *S. cerevisiae* 5206-1b or M1-2b. For *SGA*, 140 ng of the 170 bp *Bam*HI-*Xho*I DNA fragment (fig.2) was labeled with [γ-^{32}P]ATP and then hybridized separately with 15 ng of poly(A$^+$) RNA from *S. cerevisiae* MCCX1-5d or M1-2b. The S1 digestion products were loaded on a polyacrylamide sequencing gel in parallel with sequencing reactions performed on a DNA fragment of known sequence.

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/88/$3.50

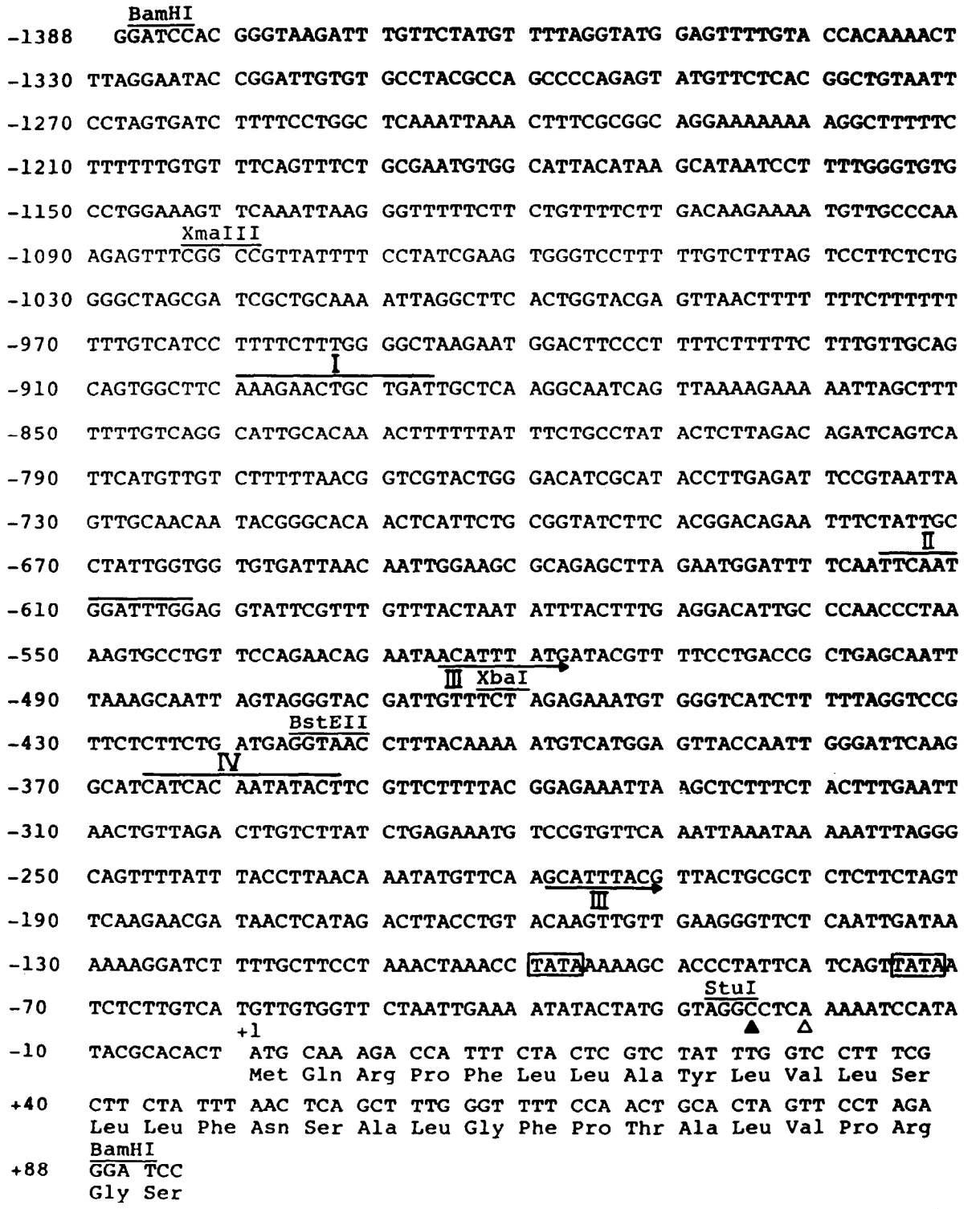

Fig.1. Nucleotide sequence of the 5′ region of the *STA2* gene. Motives I, II and IV are overlined. Motives III are underlined. The triangles indicate the sites of transcription initiation. TATA elements are boxed.

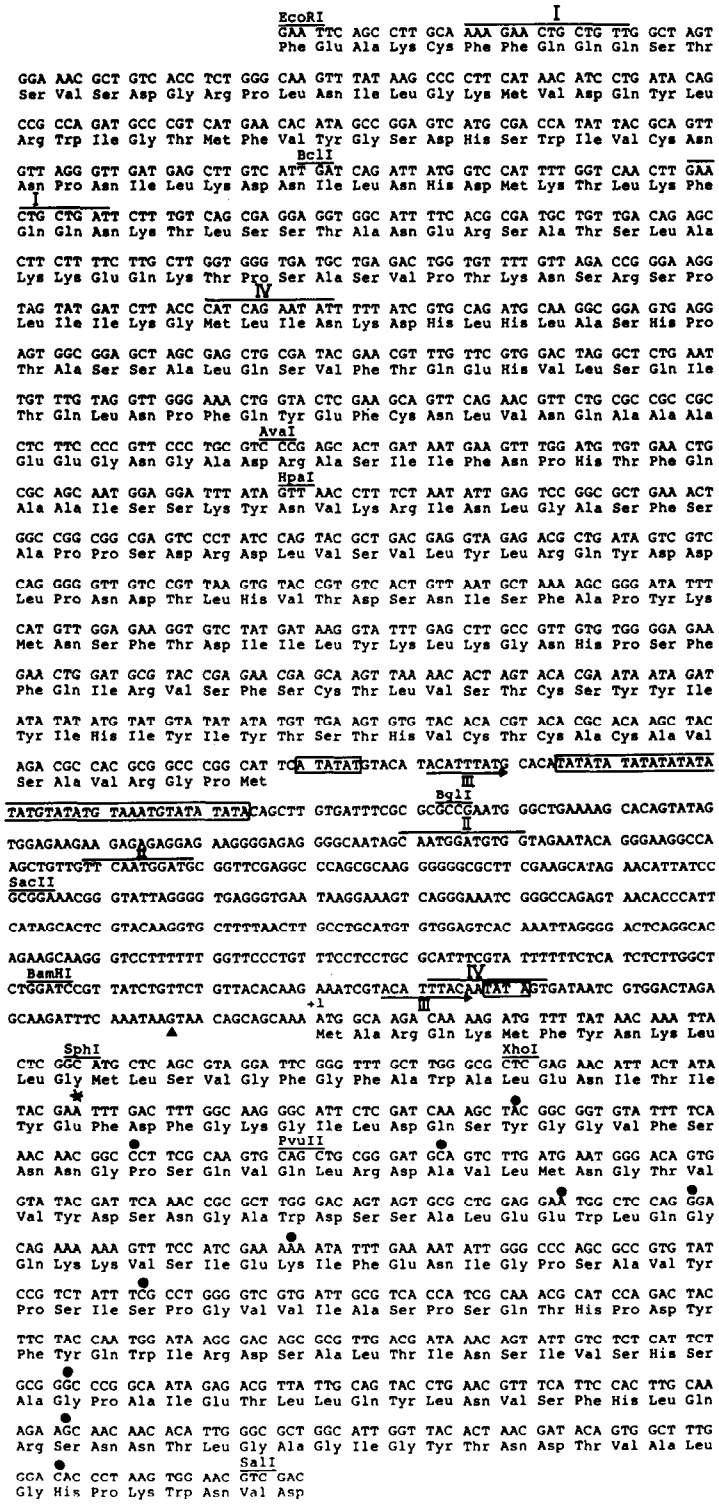

Fig.2. Nucleotide sequence of the 5′ region of the *SGA* gene. (*) The first nucleotide from which the *SGA* and *STA*1 [14] are identical; (●) nucleotide mismatches in the *SGA* and *STA*1 [14] sequences. Other symbols as in fig.1.

3. RESULTS AND DISCUSSION

3.1. *Nucleotide sequence and transcription initiation sites of the 5′ regions of* STA2 *and* SGA

The nucleotide (nt) sequences of the 5′ regions of *STA2* and *SGA* are outlined in figs 1 and 2, respectively. The sequence from *STA2* shows an open reading frame (ORF) starting by an ATG codon (nt +1; fig.1). This triplet is preceded by a CACACT sequence highly homologous to the consensus T/A A C/A A C A, which precedes the ATG initiator codon of yeast mRNAs, and is practically identical to that (CACAAT) of the *HML*a11 and *MAT*a11 genes [12,13]. The sequence of the 5′ region of the *STA2* gene is identical to that reported for the *STA1* gene [14] and spans from nt −126 to nt −1388. This finding agrees with the identical restriction maps of the *STA1*, *STA2* and *STA3* genes [2,8,14] indicating that, despite their scattered location in the yeast genome, they have a common origin and encode identical glucoamylases.

A high resolution S_1 mapping experiment showed that RNA from strain M1-2b, which contains both *STA2* and the *STA2*-repressor STA10 genes, lacked any S_1-protected DNA band (fig.3B, track 2), as expected. In contrast, RNA from strain 5206-1b (*STA2*, *sta*10°), which expresses the extracellular glucoamylase, protected two fragments from the DNA probe; a main one of 183 nt and a secondary one of 178 nt (fig.3B, track 1). This result indicates that transcription from *STA2* starts at the C residue located at −25, with a secondary site at the A residue at −21 (fig.1). This transcription is probably promoted from either (or both) of the TATA boxes located at positions −100 and −75 (fig.1).

The *SGA* gene gives rise to two transcripts of 2.4 and 1.95 kb, according to one report [2], or to only one transcript of 2.0 kb according to others [3,15]. The reasons for these differences are not known, although there is certain evidence suggesting that the synthesis of the 2.4 kb transcript is dependent on the growth conditions of the cells [16]. The nucleotide sequence of the *SGA* gene 5′ region only has a single ORF from *Eco*RI towards *Sal*I starting at nt +1 (fig.2). The first ATG of this ORF may be considered as the initiator codon of the intracellular glucoamylase. The DNA sequence

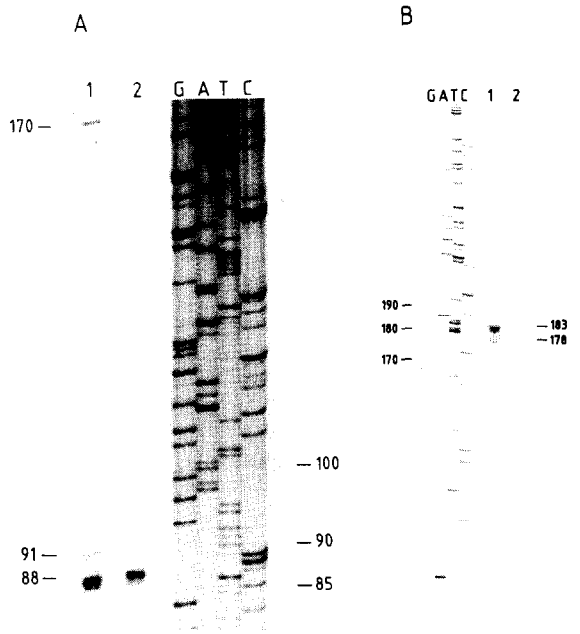

Fig.3. Determination of the transcription initiation sites from the *STA2* and *SGA* genes. (A) *SGA* gene. Poly(A^+) RNA was prepared from strains MCCX1-5d (track 1) and M1-2b (track 2) (as a negative control). Numbers on the right indicate the length (in nt) of the standards (GATC) from the sequencing reactions. Numbers on the left indicate the size (in nt) of the S_1-protected DNA fragments. The 170 nt fragment, representing full protection of the probe, should be derived from the 2.4 kb mRNA [2]. (B) *STA2* gene. Poly(A^+) RNA was prepared from strains 5206-1b (track 1) or M1-2b (track 2) (as a negative control). Numbers on the left indicate the size (in nt) of the standards (GATC) from the sequencing reactions. Numbers on the right indicate the length (in nt) of the S1-protected DNA fragments. For other details see section 2.

from nt +96 is identical to that encoding the catalytic domain of the extracellular glucoamylase of the *STA1* gene [14]. Both glucoamylases show cross-immunoreactivity and, therefore, should share similar amino acid sequences [5,17]. In the same reading frame, other putative ATG initiator triplets are present at nt −405 and −204. However, the presence of nonsense codons at positions −399, −186, −51, −48 and −45 indicate that they cannot act as the starting codons for the SGA gene product. No other significantly long ORF is present in the other two reading frames. The long AT-rich region (nt −496 to −546) preceding the site where the 2.4 kb mRNA starts (around nt −460 [2]; fig.2) could promote its

transcription if it is recognized as a promoter by the RNA polymerase II.

The complementary chain to that containing the *SGA* sequence carries an ORF (ORF2) starting by an ATG sequence at nt −563 (fig.2). This sequence has not previously been reported, as indicated by a data bank search. Whether it could correspond to an indispensable gene is being examined by gene disruption experiments.

To determine the transcription initiation site for the 1.95 kb mRNA from the *SGA* gene, poly(A^+) RNA was obtained from the yeast strains MCCX1-5d, which constitutively expresses this transcript [2], and M1-2b (as a negative control) which contains the *SGA*-repressor *STA10* gene and, therefore, lacks the 1.95 kb transcript [2]. Both RNAs protected a 88 nt fragment from the DNA probe (fig.3A, tracks 1 and 2). However, only the RNA from that strain expressing the 1.95 kb mRNA protected a fragment of 91 nt (fig.3A, track 1), indicating that it derives from this mRNA. Therefore, transcription for the 1.95 kb transcript starts at the G residue at −14 (fig.2), and is most likely promoted from the TATA box located at −53 (fig.2). The origin of the transcript which protects the 88 nt fragment in those strains is not yet clear. One possibility is that it corresponds to ORF2 (see above), although this has to be rigorously examined.

3.2. *Comparison of the nucleotide sequences of* STA2 *and* SGA

The finding that the *SGA* and *STA2* are co-regulated at the transcriptional level by the *STA10* gene indicates that they contain homologous sequence(s) in their 5′ regions. A computer search showed that there are four distinct homologous elements which were named I to IV (figs 1 and 2). A comparative schematic representation is shown in fig.4. Elements I, II and IV are present twice in *SGA* and only once in *STA2*. Most interesting is homology III of 9 nt, which is present as a direct repeat in both genes. Each repeat has a bilateral symmetry (PuCATTTAPiG) taking the central T as the axis. Similar geometric motives have been found in other regulatory regions of yeast genes. In both *STA2* and *SGA* genes one of the direct repeats is located next to the TATA boxes (figs 1,2 and 4) and, therefore, is a good candidate to be the motif for the regulation by *STA10*. Indeed, initial results with promoter-probe plasmids indicate that the DNA region downstream from the *Bst*EII site

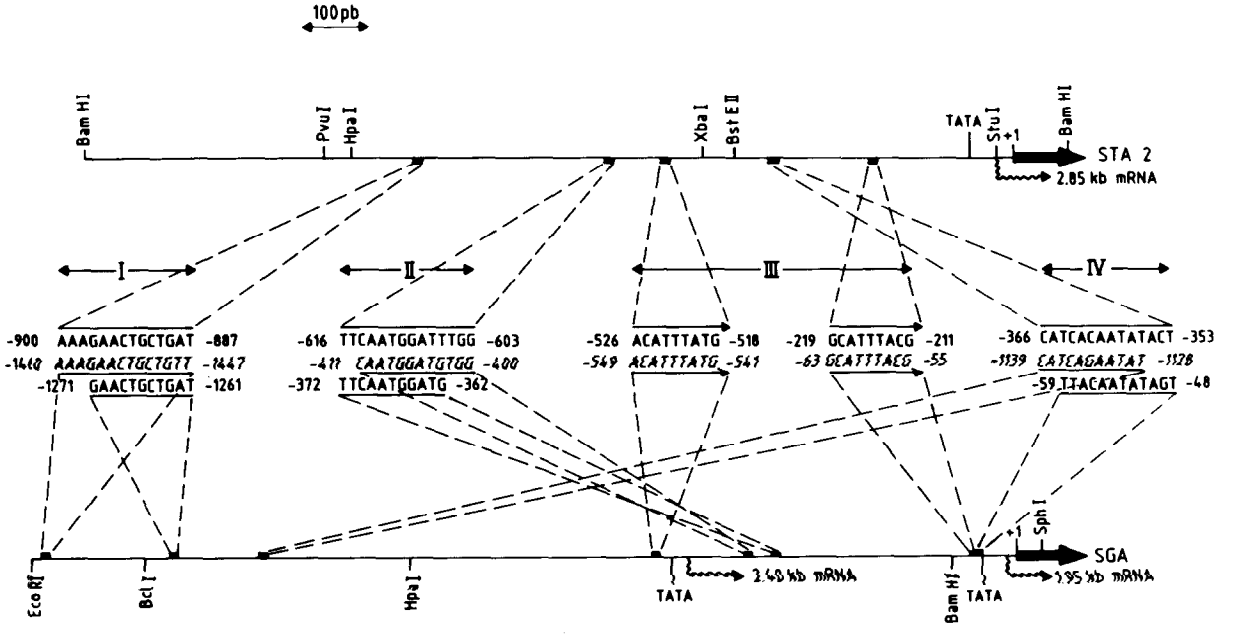

Fig.4. Schematic representation of the *STA2* and *SGA* 5′ regions. Only relevant TATA boxes are indicated.

(figs 1 and 4) of the *STA2* gene maintains both promoter activity and repressor capacity by *STA10* (Claros, M.G. and Jiménez, A., unpublished). Regulation of the expression of *STA* genes is of a complex nature, being not only repressed by the *STA10* gene, but also by the mating type locus and the presence of mitochondria; it also may be subjected to catabolite repression [8]. The possible roles that homologies I to IV play in these regulatory mechanisms is presently being studied.

Acknowledgements: We are grateful to Drs M.A. Delgado and J. Conde for their interest and comments during the course of this work and to Mrs A. Martín for expert technical assistance. This research was supported by grants from the CAYCIT (497), Plan Concertado CAYCIT-Cruzcampo SA (64/84) and by an Institutional Grant from the Fondo de Investigaciones Sanitarias to the Centro de Biología Molecular.

REFERENCES

[1] Colonna, W.J. and Magee, P.T. (1978) J. Bacteriol. 134, 844–853.
[2] Pardo, J.M., Polaina, J. and Jiménez, A. (1986) Nucleic Acids Res. 14, 4701–4718.
[3] Pretorius, I.S., Modena, D., Vanoni, M., Englard, S. and Marmur, J. (1986) Mol. Cell. Biol. 6, 3034–3041.
[4] Tamaki, H. (1978) Mol. Gen. Genet. 164, 205–209.
[5] Yamashita, I., Naemura, E., Hatano, T. and Fukui, S. (1985) J. Bacteriol. 161, 574–582.
[6] Pretorius, I.S., Chow, T. and Marmur, J. (1986) Mol. Gen. Genet. 203, 36–41.
[7] Polaina, J. and Wigss, M.Y. (1983) Curr. Genet. 7, 109–112.
[8] Pretorius, I.S., Chow, T., Modena, D. and Marmur, J. (1986) Mol. Gen. Genet. 203, 29–35.
[9] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 54–63.
[10] Hattori, M. and Sakaki, Y. (1986) Anal. Biochem. 152, 232–238.
[11] Favarolo, J., Freisman, R. and Kamen, R. (1980) Methods Enzymol. 65, 718–749.
[12] Hamilton, R., Watanabe, C.K. and De Boer, H.A. (1987) Nucleic Acids Res. 15, 3581–3593.
[13] Cigan, A.M. and Donahue, T.F. (1987) Gene 59, 1–18.
[14] Yamashita, I., Suzuki, K. and Fukui, S. (1985) J. Bacteriol. 161, 567–573.
[15] Yamashita, I. and Fukui, S. (1985) Mol. Cell. Biol. 5, 3069–3073.
[16] Pardo, J.M. (1987) PhD Thesis, Facultad de Ciencias, Universidad Autónoma, Madrid.
[17] Modena, D., Vanoni, M., Englard, S. and Marmur, J. (1986) Arch. Biochem. Biophys. 248, 138–150.