Characterization of the 5′ subtilisin (aprE) regulatory region from Bacillus subtilis

Janet Jan a, Fernando Valle b, Francisco Bolivar a, Enrique Merino a,*

a Departamento de Microbiología Molecular, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca, Morelos 62210, Mexico
b Genencor Internacional Inc., Palo Alto, CA 94304, USA

Received 14 October 1999; received in revised form 17 November 1999; accepted 22 November 1999

Abstract

The aprE gene of Bacillus subtilis encodes the major serine alkaline protease known as subtilisin. It is expressed during the transition state and transcribed by the σA form of the RNA polymerase (RNAP). In this work, we characterized the regulatory region of the aprE gene (rraprE) from B. subtilis. By computer analysis and site-directed mutagenesis, we localized the aprE promoter sequence 7 bp upstream from its transcription initiation site (TIS). We also characterized the static curvature properties of the rraprE DNA and found two different areas of DNA bending, within the first 400 bp upstream of its TIS. We postulate that these particular curved DNA regions could play a role in the interaction with some regulatory proteins and discuss possible implications related to aprE transcription regulation. © 2000 Federation of European Microbiological Societies. Published by Elsevier Science B.V. All rights reserved.

Keywords: aprE; Curved DNA; Site-directed mutagenesis; Promoter sequence; Bacillus subtilis

1. Introduction

The aprE gene encodes the extracellular alkaline protease subtilisin from Bacillus subtilis. It is expressed during the stationary phase and reaches its highest expression level approximately 2 h after the onset. It has been established that AprE contributes to more than 70% of the total extracellular protease activity [22]. The transcription of aprE gene has been used as a model to understand the regulatory network that controls the gene expression at the transition state, when the cell ceases exponential growth and sporulation is initiated [16]. Alkaline protease production, as well as other extracellular proteases, are subject to regulation by a complex group of positive and negative regulators, such as: DegU/DegS, AbrB, Hpr and SinR. These proteins act at the transcriptional level and require at least 400 bp upstream of the aprE transcription initiation site (TIS) to present their normal mode of control [20].

Although the basic and applied importance of the aprE gene is widely accepted, and its TIS has been determined [4], there is no clear description of its promoter sequence. Previous studies by Park et al. [11] reported a putative aprE promoter 7 bp upstream of its TIS, that had 19 bp between their 3′10 and 3′35 boxes. Conflicting with these data, Strauch and Hoch [18] suggested a different promoter region, 12 bp upstream of its TIS and 16 bp between their −10 and −35 boxes. Taking into account these conflicting data, and the lack of experimental results to evaluate the relevant nucleotides (nt) for RNA polymerase (RNAP) recognition, in this work, we performed site-directed mutagenesis in the regulatory region of the aprE gene (rraprE) to determine the functional promoter sequence. Our results indicate that the functional aprE promoter does not correspond to any of the previous reported promoter sequences, at least under the growth conditions assayed in this study.

In addition, the curved DNA geometry of the rraprE has been proposed to be involved in aprE gene expression. The presence of a static curvature within this region has previously been reported [17]; however, there are no experimental studies that characterize the exact position of the region(s) involved in such curvature. Using electrophoretic analyses of circular permuted DNA fragments containing rraprE, as well as curvature analysis based on com-
puter programs, we have found that the major determinants of anomalous gel migration are located around 100 and 200 bp away from its TIS. The results of these studies are discussed in the light of the aprE transcription regulation.

2. Materials and methods

2.1. Computer analysis directed to identify putative $\sigma^4$ promoters

Based on the statistical algorithm proposed by Mulligan et al. [10], we developed a program to search $\sigma^4$ promoter sequences. In this algorithm, each base of the $-35$ and $-10$ boxes as well as the ones near to them are weighed according to a score matrix. This score matrix was built up from the frequency of occurrence of each base on a list of 122 B. subtilis $\sigma^4$ promoters compiled by Helmann [7]. The space between the $-35$ and $-10$ regions contributes in geometrical terms to the final score. The maximum possible score of a putative perfect promoter is 100. A cut-off value was used to select only the best promoters within a DNA region.

2.2. Computer analysis of the DNA static curvature at the rraprE

To visualize the DNA three-dimensional trajectory, we used the DNAstar computer program (DNASTAR, Madison, WI, USA), which is based on the nearest-neighbor wedge model [3]. The analysis was made with 536 bp of the rraprE sequence previously reported by Valle and Ferrari [20].

2.3. Site-directed mutagenesis of aprE promoter and plasmid construction

Plasmid pT7-aprE was constructed by subcloning a 509-bp EcoRI-BamHI fragment derived from plasmid pSG35.1 [4] into pT7 Blue (Novagen, Madison, WI, USA) and contains the aprE promoter and the first eight codons. pT7-aprE was used as a template for oligonucleotide-directed PCR mutagenesis according to the protocol described by Merino et al. [9]. PCRs were carried out with Taq DNA polymerase (Promega, Madison, WI, USA). The nt substitutions were as follows: $-7$ (T$\rightarrow$G), $-31$ (A$\rightarrow$C), $-37$ (T$\rightarrow$G) (see Fig. 1). The PCR products were double-digested with EcoRI and BamHI, cloned into the same restriction sites of pSG-PLK, and verified by DNA sequencing [14] to yield plasmids p-7G, p-31C and p-37G, respectively. pSG-PLK is a pSG35.1 derivative that carries a promoterless lacZ gene. In this plasmid, the aprE promoter has been replaced with the pUC19 poly-linker region. This modification makes the selection of transformants easier, because colonies that carry the mutated aprE promoter are blue when they are plated on X-Gal (5-bromo-4-chloro-3-indolyl-$\beta$-d-galactopyranoside) agar.

2.4. Electrophoretic analyses of circular permuted DNA fragments

The sequences of the primers used for the PCR analysis to map the curved loci are shown in Table 1. We utilized a PCR-based method [21] to construct circular permutations of the rraprE. These fragments were analyzed by 8% polyacrylamide gel electrophoresis (PAGE) at 4°C at a constant voltage of 8 V cm$^{-1}$ in 89 mM Tris-borate (pH 8.3), 2 mM EDTA buffer.

2.5. Bacterial strains and transformations

Escherichia coli JM101 [(lac-pro) supE thiF’ traD30 proAB lacI 9 ZM15] was used as the host for plasmid constructions. Plasmids were obtained as described in [2]. B. subtilis strains constructed in this study were derived from BSR1 [5]. Plasmids p-7G, p-31C and p-37G were linearized with PstI and used to transform B. subtilis BSR1, to construct the strains listed in Table 2. The transformation protocol was followed according to Anagnostopoulos and Spizizen [1]; the selection was made for chloramphenicol-resistance (5 $\mu$g ml$^{-1}$). Integration of plasmid DNA was confirmed by testing the amylase production by iodine staining, as described previously [15].

2.6. $\beta$-Galactosidase assays

B. subtilis cells carrying the mutagenized rraprE were grown at 37°C in Shafer’s sporulation media [16]. At regular time intervals, samples of 1.5 ml were taken and assayed for $\beta$-galactosidase specific activity as described in [5].

3. Results and discussion

3.1. Computer analysis of the aprE promoter

To identify and evaluate putative $\sigma^4$ promoters within the 5’ control region of the aprE gene (rraprE), we developed a computer program based on the statistical algorithm derived by Mulligan et al. [10]. With this program, a maximum value of 100 is obtained when all bases match the promoter extended consensus aaaaattgtTGTGACAt-11 bp-tatgATAATaaaa (see Section 2). When a cut-off value of 60 was used, the program only found two putative promoters within the first 100 bp upstream the aprE TIS. One of these promoters had a score value of 63 and corresponded to that previously proposed by Park et al. [11], located 7 bp upstream to the aprE TIS with a 19-bp space region between the $-10$ and $-35$ boxes. A second pro-
To verify our theoretical results and discern between the two potential promoter sequences, we accomplished site-directed mutagenesis at relevant nt in the regulatory region of the aprE promoter. For a given putative promoter sequence, we replaced the nt at position −7, which is a highly conserved T base in the B. subtilis σ^70 promoters, with a G. We constructed plasmid p-7G, which carries this modified nt in the aprE-lacZ translational fusion. The aprE-lacZ translational fusion was integrated by double homologous recombination into the amyE locus of B. subtilis strain BSR1, generating the BS-7G strain (see Section 2). This strain expresses only 4% of the β-galactosidase activity in relation to its parental BSR6 strain that carries the wild-type trpE. This result is presented in Table 3. The dramatic reduction of the β-galactosidase activity in the BS-7G strain clearly indicates that the sequence TACAAT, identified by our computer analysis as the −10 box, is in fact functional in vivo for RNAP recognition.

In order to discern which of the two −35 boxes identified by our computer program corresponded to the functional −35 box, we mutagenized nt −31 (A→C) and nt −37 (T→G); generating BS-31C and BS-37G strains, respectively. In both cases, we considered substitutions of a highly conserved base in the σ^70 promoter consensus, to the least conserved nt in that specific position. To avoid ambiguous results, the mutagenesis was designed in such a way that changes introduced in the DNA sequence would

**Table 2**

<table>
<thead>
<tr>
<th>Strain</th>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>BSR1</td>
<td>Δapr hisA glyB</td>
<td>Lab stock</td>
</tr>
<tr>
<td>BSR6</td>
<td>BSR1 amyE::pSG351.1 cat</td>
<td>This work</td>
</tr>
<tr>
<td>BS-7G</td>
<td>BSR1 amyE::p-7G cat</td>
<td>This work</td>
</tr>
<tr>
<td>BS-31C</td>
<td>BSR1 amyE::p-31C cat</td>
<td>This work</td>
</tr>
<tr>
<td>BS-37G</td>
<td>BSR1 amyE::p-37G cat</td>
<td>This work</td>
</tr>
</tbody>
</table>

Abbreviations: cat, chloramphenicol acetyltransferase gene.
only affect one of the two analyzed −35 boxes (see Fig. 1A). The BS-37G strain showed a 69% decrease in the aprE–lacZ transcription rate. On the other hand, BS-31C did not yield a significant variation on its aprE–lacZ transcription rate, as it can be inferred from their L-galactosidase activity, 81% with respect to the BSR6 strain (see Table 3). Therefore, our results indicate that, under the conditions tested, the aprE promoter has 17 bp between the −10 and −35 boxes and it is located 7 bp upstream of its TIS (Fig. 1A). Nevertheless, we could not discard the possibility that both promoter sequences may be functional and transcribe the aprE gene at particular conditions not assayed in our study. However, the conditions used in the present work are the same used by several other groups to study aprE regulation as well as many sporulation-associated processes.

3.3. Electrophoretic analysis of circular permuted DNA fragments of the rraprE

The presence of a static curvature DNA region on the rraprE has been reported [20]. However, the exact position of the sequence(s) responsible for such curvature was not identified. We used a PCR-based method to map curvature loci in this region. Using different sets of primers, we generated several DNA fragments of the same size, but with a specific DNA region permuted at different distances from the ends of the DNA fragments. This procedure allows the scanning for the contribution of different DNA regions to gel migration anomalies [21]. Because preliminary results suggested that the 160 bp preceding the −35 region was involved in DNA bending [20], we focused our efforts on that region.
Diagrams of the probes used, as well as the gel migration of these PCR-generated DNA fragments, are presented in Fig. 2. As can be seen, all the DNA fragments migrated anomalously, different from the expected size of 536 bp (Fig. 2B). Furthermore, when the TaqI-103 site was almost in the middle of the DNA segment (fragment CJ, Fig. 2A), the slowest migration was observed, suggesting that the center of bending was at, or very close to, this TaqI-103 site. This interpretation was supported by the migration behavior of the BI and DK DNA segments, where the TaqI-103 site is 208 or 209 bp from either end, and the fragments migrated almost identically.

In the case of the AH and EL DNA fragments, the TaqI-103 site is located at the same distance from each end, and therefore we would expect to observe the same migration. Nevertheless, it was not the case. Fragment AH migrated more slowly and therefore we can assume it has an extra curved element that is not present in the fragment EL (Fig. 2B). Furthermore, if we consider the important migration differences between fragments DK and EL, we can circumscribe this second curved locus to the DNA region located between 209 and 109 bp upstream the TaqI-103 site. These data locate the major determinants of anomalous gel migration of the DNA fragments containing the rrapE around 100 and 200 bp upstream the TIS and more than five helix-turns away from the AbrB-binding site, consistent with a previous report by Strauch and Ayazifar [17].

### 3.4. Computer analysis of the static DNA curvature

Considering the complexity of the rrapE, revealed by the permutation assay described before, we decided to utilize a computer-based program to predict the three-dimensional trajectory of the rrapE DNA sequence. The predictions were made with the DNAStar computer programs (see Section 2). This computer analysis showed that the DegU-binding site is surrounded by two curved loci, localized around 100 and 200 bp upstream the rrapE TIS (Fig. 3). This result is in good agreement with our circular permutation assay.

There are numerous examples of curvature loci associated with regulatory elements/regions for transcription in both prokaryotes and eukaryotes [6,19]. Plaskon and Wartell [13] analyzed the distribution of sequence elements associated with DNA curvature for prokaryotic promoters and found a strong correlation between the presence of putative elements of curvature and the promoter strength for regions upstream of the associated genes. The localization of the intrinsically curved DNA upstream the RNAP-binding site might vary, ranging from −235 bp in the distal activating region of streptococcal promoter pctII to −40 bp in the early promoter of T5 and T7 phages (reviewed in [12]). In most of the known cases, only one curved upstream DNA region was present, but examples can be found in which two regions placed in phase contribute to promoter activation [8]. In our case, we postulate that the two curved regions presented in the rrapE could play an active role in the formation of the transcription open complex, by mediating the interactions between the RNAP and the DegU regulatory protein. DegU has been involved in the transcriptional control of several genes that are expressed during the transition state, such as: levansucrase, neutral protease, α-amylose, β-gluc-
canase, serine protease (reviewed in [20]). However, there is not a clear consensus sequence for its binding. We consider that the curved loci present in the $rr_{aprE}$ might form a specific three-dimensional domain, that could modulate the interactions between the set of proteins that control $aprE$ transcription. Currently, we are testing this hypothesis.

Acknowledgements

We thank Edmundo Calva, Guillermo Gosset and Beatriz Palmeros for critical reading of the manuscript and Mercedes Enzaldo for her excellent technical assistance. Janet Jan was supported by scholarship from Consejo Nacional de Ciencia y Tecnología (CONACyT) México and Dirección General de Estudios de Posgrado (DGEP) UNAM-México.

References