Evaluation and characterization of catabolite-responsive elements (cre) of Bacillus subtilis

Yasuhiro Miwa, Atsuo Nakata, Atsushi Ogiwara, Mami Yamamoto and Yasutaro Fujita*

Faculty of Engineering, Fukuyama University, Fukuyama 729-0292, Japan and Computer Laboratory, National Institute for Basic Biology, Okazaki 444-8585, Japan

Received October 11, 1999; Revised December 23, 1999; Accepted January 4, 2000

ABSTRACT

A global mechanism of catabolite repression of the genus Bacillus comprises negative regulation exerted through the binding of the CcpA protein to the catabolite-responsive elements (cre) of the target genes. We searched for cre sequences in the Bacillus subtilis genome using a query sequence, WTGNAANCGNTNWCAWW (N and W stand for any base and A or T, respectively), picking out 126 putative cre sequences. To examine their cre function, we integrated spac promoter (Pspac)-cre-lacZ and known cre sequences. To examine their cre and A or T, respectively), picking out 126 putative WTGNAANCGNWNNCW (N and W stand for any base base degeneracy is allowed. Furthermore, the alignment of 22 cre that function in vivo implicated a consensus sequence, WWTGNAARCNGWWWCAWW (R stands for G or A). Interestingly, in the case where cre sequences are located in the protein-coding regions of the target genes, their conserved bases are preferentially the third bases of codons where base degeneracy is allowed.

INTRODUCTION

Bacilli as well as low-GC Gram-positive bacteria likely possess a common negative regulatory mechanism of catabolite repression, which is completely different from the positive regulatory one operating in enteric bacteria. This negative regulation of transcription of catabolite-repressive genes, which has been extensively studied in Bacillus subtilis, is exerted through the binding of the CcpA protein (1), which interacts with allosteric effectors, such as P-ser-HPr (2) and P-ser-Crh (3), to their cis-acting catabolite-responsive elements (cre) (4).

The B. subtilis cre was firstly identified in the promoter region of the amyE gene, the consensus sequence of which was deduced by means of site-directed mutagenesis to be TGWAANCGNTNWCA, where N and W stand for any base, and A or T, respectively (5). Another cre was found in the protein-coding region of gntR (6,7). Since then, various cre, including rather classical ones of xylA (8), hutP (9), ackA (10) and ackA (11), have been identified in either the promoter or protein-coding regions of the target genes. The sequences of these cre closely match the consensus sequence described above.

After Hueck et al. (4) had searched and analyzed cre-like sequences among the deposited nucleotide sequences of Gram-positive bacteria, the complete sequence of the B. subtilis genome was reported by Kunst et al. (12). Thus, it was thought to be very interesting to search for cre-like sequences in the B. subtilis genome, and to evaluate and characterize them in more detail. We chose another consensus sequence, WTGNAANCNGWNWCW, as a query sequence for searching for cre-like sequences in the genome, after repeated trials. This sequence is essentially the same as the cre consensus sequence proposed by Weickert and Chambliss (5), but is somewhat degenerate and one base longer. Our in vivo test for the cre function of various cre-like sequences, which had been revealed by our search, led us to find some interesting features of the cre sequence of B. subtilis.

MATERIALS AND METHODS

Bacterial strains and plasmids

The B. subtilis strains constructed in this work were derived from strain GM122 (trpC2 sacB′-lacZ) (13). Plasmid pCRE-test (Fig. S1, Supplementary Material) was constructed as follows. A region of plasmid pAG58 containing a spac promoter (Pspac) (14) was amplified by PCR using a primer pair designed to produce flanking EcoRI and BamHI sites. In addition, a region of plasmid pMUTIN1 containing a Shine–Dalgarno sequence and the 5′-portion of lacZ (12 codons) (15) was amplified using another primer set designed to produce flanking BamHI and HindIII sites. The resulting PCR products were digested with the respective endonucleases, and then

*To whom correspondence should be addressed at: Department of Biotechnology, Faculty of Engineering, Fukuyama University, 985 Sanzo, Higashimura-cho, Fukuyama-shi, Hiroshima 729-0292, Japan. Tel: +81 849 36 2111; Fax: +81 849 36 2459; Email: yfujita@bt.fubt.fukuyama-u.ac.jp

Present address:
Atsushi Ogiwara, Tsukuba Research Laboratories, Glaxo Wellcome K.K., Tsukuba 300-4247, Japan
ligated with the EcoRI–HindIII arm of plasmid pTRpBGI (16). The ligated DNA was used for the transformation of Escherichia coli strain JM1109 (17) to ampicillin resistance. The correct construction of plasmid pCRE-test was confirmed by sequencing.

**cre search of the B.subtilis genome**

cre-like sequences in the B.subtilis genome were searched for with an originally developed Perl program on a workstation (Sun SPARC station 20) with the query sequence of WTGNAANCNGNWNCW.

**Integration of the Pspac-cre-lacZ fusion into amyE**

An appropriate region containing each cre-like sequence (15 bp) and its upstream and downstream flanking sequences (each ∼30 bp long) was amplified by PCR using chromosomal DNA of B.subtilis strain Marburg 168 (trpC2) as a template and a primer pair designed to generate two flanking BamHI sites. The PCR products were digested with BamHI and then ligated with DNA of plasmid pCRE-test, which had been cleaved with the same enzyme. The ligated DNAs were used for the transformation of E.coli strain JM1109 to ampicillin resistance. The sequence and orientation of the cloned fragments were determined by sequencing. The constructed plasmids carrying each cre-like sequence in the same direction with respect to the transcription were linearized with PstI or ScaI, and then used for the integration of Pspac-cre-lacZ into the amyE locus of B.subtilis strain GM122 through a double crossover event by selecting chloramphenicol-resistant transformants (Fig. S1).

**Examination of catabolite repression of β-galactosidase (β-Gal) synthesis in integrants**

The integrants were grown to an optical density at 600 nm ($OD_{600}$) = 0.6 in S6 medium (18) containing 0.5% Casamino Acids (Difco), which was supplemented with tryptophan ($OD_{600}$) = 0.6 in S6 medium (18) containing 0.5% Casamino Acids (Difco), which was supplemented with tryptophan (10 mM glucose). The cells ($OD_{600}$ x ml = 3.6) were harvested, and then lysed by lysozyme treatment and brief sonication as described previously (19). The β-Gal activity was spectrophotometrically assayed as described by Atkinson et al. (20).

**RESULTS AND DISCUSSION**

**Search for cre-like sequences in the B.subtilis genome**

Firstly, we used a well-known cre consensus sequence of 14 bases, TGWAANCGNTNWNCW, proposed by Weickert and Chambliss (5) to search for cre-like sequences in the B.subtilis genome. During this search, we picked out 31 cre-like sequences which show no mismatching to this query sequence. However, this number was much lower than we expected, because rough estimation of glucose-repressive protein spots on a two-dimensional gel as well as a search for glucose-repressive genes using hundreds of plasmid pMUTIN-integrants (15) suggested that there might be at least 150 cres in B.subtilis (data not shown). In addition, well-characterized cres such as those located in gntR (6,7), xylA (8) and hutP (9) were not included in these 31 sequences.

We attempted to find a more suitable cre query sequence for a computer search for cre-like sequences in the genome, mainly by degenerating the consensus sequence, TGWAANCGNTNWNCW. After repeated trials, we finally chose a 15-base sequence, WTGNAANCNGNWNCW, for a search of cre-like sequences, which is partially palindromic. Our search with this query sequence led us to find 108 cre-like sequences, which is partially palindromic. Weickert and Chambliss (5) to search for glucose-like sequences out of 32 were found to function. The cre-like sequences carry one mismatch at the first and second bases to be G or C and A, respectively, the mismatch numbers are underlined. Among cre-like sequences carrying one mismatch in both directions, those which have at least one underlined mismatch are also listed in Table 1.

**Examination of cre function of putative cre sequences**

In order to determine whether or not the cre-like sequences function as cre in vivo, we constructed plasmid pCRE-test (Fig. S1). After cloning an appropriate region containing each cre-like sequence into the BamHI site of plasmid pCRE-test, the constructed plasmids were linearized with PstI or ScaI, and then used for the integration of Pspac-cre-lacZ into the amyE locus through a double crossover event. β-Gal synthesis in the integrant of the Pspac-cre-lacZ fusion without cre was almost constitutive in the presence and absence of glucose in the medium (Table 2). Thus, we were able to test the cre function by examining catabolite repression of β-Gal synthesis in the Pspac-cre-lacZ integrants, which was most likely evoked by the transcription roadblock owing to a complex of CcpA and P-ser-HPr (or another factor) bound to cre (2,7).

Among the 126 cre-like sequences listed in Table 1, 32 were tested for their ability as cres (Table 2). We chose them in a fashion to cover various kinds of mismatches in both directions and to include several known cres. As shown in Table 2, the β-Gal activities fluctuated by 6-fold in various integrants grown without glucose, probably because of the different stabilities of mRNAs carrying each of the 32 cre-like sequences between Pspac and lacZ. However, it is considered that the catabolite repression ratio for each of the cre-like sequences reflects its ability to cause a transcription roadblock. In Table 2, 22 cre-like sequences out of 32 were found to function in vivo (catabolite repression ratio >1), and are listed in the order of their strength from cre-acoA to cre-yxkJ. All of the well known cres tested [cre-ibglP (21), cre-gntR (6,7), cre-hutP (9), cre-amyE (5), cre-lacK (11) and cre-xylA (8)] were active in our in vivo cre test system, which indicates that this test system is highly reliable.

As shown in Table 2, we tested the repression ability of 28 cre-like sequences which exhibited no mismatching to the consensus sequence, at least in the same or inverse direction as
that of transcription of the target genes. All cre\textsubscript{s} classified as \([0,0]\), \([0,1]\), and \([0,1]\) exhibited repression ability, whereas among cre\textsubscript{s} classified as \([0,2]\) and \([1,0]\), some functioned but others did not. But, cre\textsubscript{s} classified as \([0,3]\) and \([1,0]\) did not function. Furthermore, we also tested four cre\textsubscript{-like} sequences which exhibited one mismatch in both the same and inverse directions as that of transcription of the target genes; a cre\textsubscript{ classified as \([1,1]\) functioned, but the others classified as \([1,1]\) and \([1,1]\) did not. These results imply that lower mismatching of cre\textsubscript{s} to the query sequence, especially in the same direction as that of transcription of their target genes, is required for their function, and that a more palindromic nature of cre\textsubscript{s} is desirable for a better function. The requirements of lower mismatching of cre\textsubscript{s} to the query sequence and their palindromic nature for their function can be explained by the cre binding strength of CcpA interacting with some effector, which likely depends on their mismatch levels in both directions with respect to that of the transcription of their target genes, because CcpA is supposed to be dimerized in vivo (22).

The results of the above cre\textsubscript{ tests implied that low mismatching of cre\textsubscript{s} in the same direction as that of the transcription of their target genes is likely more critical for their function than that in the inverse direction. To confirm this, we oppositely placed cre\textsubscript{-ibglP} and cre\textsubscript{-gntR} between \(P\textsubscript{spac}\) and \(lacZ\), and then examined the catabolite repression of \(\beta\)-Gal synthesis in the constructed integrants carrying \(P\textsubscript{spac}\)-\(cre\textsubscript{r-ibglP}\)\(-\)\(lacZ\) and \(P\textsubscript{spac}\)-\(cre\textsubscript{r-gntR}\)\(-\)\(lacZ\), respectively (Table 2). Thus, these inversed cre\textsubscript{s} are classified as \([1,0]\) instead of \([0,1]\) for the original cre\textsubscript{s} (Table 2). As shown in Table 2, the inversion of the ibglP- and gntR-cre\textsubscript{s} decreased their catabolite repression ratios from 8.0 to 1.3 and 6.0 to 5.0, respectively. These results as well as those of the above tests suggest that not only the binding strength of CcpA as to cre\textsubscript{s}
but also lower mismatching in the same direction as RNA polymerase moves might be determinants for a transcription roadblock to occur.

Although we did not test all the cre\(\text{\textsuperscript{a}}\)s listed in Table 1, we could predict from the above results whether or not the cre\(\text{\textsuperscript{a}}\)s not tested might function in vivo. cre\(\text{\textsuperscript{a}}\)s classified as \([0,0]\), \([0,1]\), \([0,1]\) and \([1,0]\) are expected to function in vivo, but those classified as \([0,3]\), \([2,0]\) and \([3,0]\) might not function. However, it is hard to predict whether or not the other cre\(\text{\textsuperscript{a}}\)s listed might function. Furthermore, we do not think that the cre\(\text{\textsuperscript{a}}\)s listed in Table 1 include all the \(B.\text{\textit{subtilis}}\) cre\(\text{\textsuperscript{a}}\)s. For example, \(\text{inaJ-cre}\) \([1,1]\) for the \(\text{inaJ-xynB}\) operon, which is known to function in vivo (3), is not listed in Table 1. Therefore, it is likely that we have to find more cre\(\text{\textsuperscript{a}}\)-like sequences which exhibit high mismatch numbers through careful consideration of their palindromic nature in order to cover all the cre\(\text{\textsuperscript{a}}\)s of \(B.\text{\textit{subtilis}}\).

Alignment of cre\(\text{\textsuperscript{a}}\)s and their location in target genes

As shown in Table 2, we aligned 22 cre\(\text{\textsuperscript{a}}\) sequences together with the surrounding ones, which were found to function in vivo with our cre test system. From this alignment, we found that the outside bases of these 15-base cre\(\text{\textsuperscript{a}}\) sequences were also conserved. When the first base of cre\(\text{\textsuperscript{a}}\), which corresponds to the first ‘W’ in the cre query sequence, is assigned as +1, the bases at positions –1, +16 and +17 are W with high probabilities of 21, 18 and 19 bases out of 22, respectively. The importance of the flanking AT-rich sequences of a 14-base cre\(\text{\textsuperscript{a}}\) consensus sequence proposed by Weickert and Chambliss (5) was also pointed out by Zalieckas et al. (23). In addition, preferable bases for the +7, +12, +13 and +15 positions are R (G or A), W, W and A, with high probabilities of 20, 19, 18 and 20 bases out of 22, respectively. Therefore, a consensus sequence for cre and its surrounding region (bases –1 to +17) was deduced to be WWTGNAARCGNWWWCAWW.

Among 22 cre\(\text{\textsuperscript{a}}\)s that function in vivo, 15 are located in the protein-coding regions of the target genes, so a very interesting question arose. Where are these 15 cre\(\text{\textsuperscript{a}}\) sequences localized in the three possible protein-coding frames? Thus, we examined the relative localization of the cre\(\text{\textsuperscript{a}}\) sequences in the protein-coding frames of the target genes, finding that in eight and seven genes, positions +1, +4, +7, +10 and +13 are the first or second bases of codons. However, most of these genes were not studied in detail.
allowed, whereas they are the first or second base of codons in
the protein-coding frames of the target genes in which
relatively strict bases are required. In other words, the other
bases of the cre consensus sequence are conserved, so there is
a relatively high probability that these positions are the third
bases of codons in the protein-coding frames where base
degeneracy is allowed. This fact implies the elegant harmony
between the establishment of a cre sequence and the evolution
of a functional protein encoded by a catabolite-repressive gene.

SUPPLEMENTARY MATERIAL

See Supplementary Material available at NAR Online for a
figure showing the in vivo test system.

ACKNOWLEDGEMENTS

We thank S. Eguchi, S. Kawahara, K. Okamura, T. Aoki,
M. Kou, S. Iijima and K. Kawai for their help in the experiments.
This work was supported by a grant, JSPS-RFTF96L00105,
from the Japan Society for the Promotion of Science.

REFERENCES