1206–1210 *Nucleic Acids Research, 2000, Vol. 28, No. 5* © 2000 Oxford University Press

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

Yasuhiko Miwa, Atsuo Nakata, Atsushi Ogiwara[1], Mami Yamamoto and Yasutaro Fujita*

Faculty of Engineering, Fukuyama University, Fukuyama 729-0292, Japan and [1]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*s) of the target genes. We searched for *cre* sequences in the *Bacillus subtilis* genome using a query sequence, WTGNAANCGNWNNCW (N and W stand for any base and A or T, respectively), picking out 126 putative and known *cre* sequences. To examine their *cre* function, we integrated *spac* promoter (P*spac*)-*cre*-*lacZ* fusions into the *amyE* locus. Examination of catabolite repression of β-galactosidase synthesis in the integrants led us to the following conclusions: (i) lower mismatching of *cre* sequences to the query sequence is required for their function; (ii) although *cre* sequences are partially palindromic, low mismatching in the same direction as that of transcription of the target genes is more critical for their function than that in the inverse direction; and (iii) yet, a more palindromic nature of *cre* sequences is desirable for a better function. Furthermore, the alignment of 22 *cre*s that function *in vivo* implicated a consensus sequence, WWTGNAARCGNWWWCAWW (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 (*cres*) (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*s, including rather classical ones of *xylA* (8), *hutP* (9), *acsA* (10) and *ackA* (11), have been identified in either the promoter or protein-coding regions of the target genes. The sequences of these *cre*s 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, WTGNAANCGNWNNCW, 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 (P*spac*) (14) was amplified by PCR using a primer pair designed to produce flanking *Eco*RI and *Bam*HI 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 *Bam*HI and *Hin*dIII 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 *Eco*RI–*Hin*dIII arm of plasmid ptrpBGI (16). The ligated DNA was used for the transformation of *Escherichia coli* strain JM109 (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 WTGNAANCGNWNNCW.

Integration of the P*spac*-*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 *Bam*HI sites. The PCR products were digested with *Bam*HI 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 JM109 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 *Pst*I or *Sca*I, and then used for the integration of P*spac-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 (50 µg/ml) and chloramphenicol (5 µg/ml), with or without 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, TGWAANCGNTNWCA, 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 *cre*s in *B.subtilis* (data not shown). In addition, well-characterized *cre*s 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, TGWAANCGNTNWCA. After repeated trials, we finally chose a 15-base sequence, WTGNAANCGNWNNCW, 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 in the genome exhibiting no mismatching to it, which are included in a list of known and putative *cre*s (Table 1).

The *cre*-like sequences were located in the protein-coding or intergenic regions of the putative target genes. In the latter cases, the names of the genes, the 5′-ends of which are closer to the sequences, are preceded by 'i' (Table 1). As described below, orientation of *cre* with respect to the direction of transcription of its target gene was found to be important for *cre* to function. The *cre*-like sequences are partially palindromic, so their mismatch numbers as to the query sequence with regard to the same and inverse directions as that of transcription of the target genes, that is, those on the anti-sense and sense strands, are given as the first and second numbers in brackets, respectively (Table 1). As discussed below, the first and second bases, W and T, of the query sequence were found not to be strictly conserved. So, when 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 *Bam*HI site of plasmid pCRE-test, the constructed plasmids were linearized with *Pst*I or *Sca*I, and then used for P*spac-cre-lacZ* integration into the *amyE* locus through a double crossover event. β-Gal synthesis in the integrant of the P*spac-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 P*spac-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 *cre*s (Table 2). We chose them in a fashion to cover various kinds of mismatches in both directions and to include several known *cre*s. 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 P*spac* 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 *cre*s tested [*cre*-i*bglP* (21), *cre-gntR* (6,7), *cre-hutP* (9), *cre*-i*amyE* (5), *cre*-i*ackA* (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

Table 1. List of known and putative *cre*s

cre	Position[a]	Direction[b]	*cre*	Position[a]	Direction[b]	*cre*	Position[a]	Direction[b]
spoVG [1,0]	55912	F	*ykqB* [0,2]	1520504	F	*ytnA* [1,1]	3124473	R
iybaL [0,2]	157388	F	*ctaD* [1,1]	1560977	F	*iytkA* [1,1]	3136502	R
ybbD [1,1]	188287	R	*pksF* [2,0]	1789117	F	*gbsB* [0,2]	3184075	R
ybbI [1,0]	192096	R	*xylA* [0,2]	1891196	F	*iyuxG* [0,1]	3203042	R
ybyB [0,2]	230927	R	*iyncC* [0,0]	1895568	F	*yugH* [0,2]	3225096	R
igltP [0,1]	254855	R	*ynzB* [2,0]	1906371	F	*yuxJ* [0,1]	3232020	F
iamyE [0,1]	327046	F	*yneN* [0,2]	1928865	F	*yuiB* [0,1]	3298666	R
iydbH [0,1]	499659	F	*yobO* [0,2]	2076638	F	*iyuiA* [2,0]	3299021	R
yddN [1,0]	550323	R	*iyojA* [0,2]	2125513	R	*citG* [2,0]	3388814	R
ydeF [3,0]	564622	R	*yodO* [1,1]	2138932	R	*iyvfO* [0,1]	3502039	R
ydhO [0,1]	627658	F	*yodT* [1,0]	2144522	R	*iyvfH* [1,0]	3509792	F
iydjk [0,0]	675986	F	*cgeB* [1,0]	2149058	F	*iyvdG* [0,1]	3554930	R
ydjK [1,0]	676847	F	*yonH* [0,2]	2228798	F	*yvpA* [0,1]	3589943	F
yeaA [2,0]	682441	R	*sunT* [1,0]	2266867	R	*irbsR* [0,0]	3700434	F
gabP [1,1]	685638	R	*kdgA* [0,1]	2322487	R	*rbsK* [0,1]	3701699	F
yesO [3,0]	761744	R	*ikduI* [0,0]	2324900	F	*rbsA* [0,1]	3703084	F
iyesW [1,1]	769722	F	*ikduI* [0,1]	2325031	F	*nrgA* [1,0]	3755983	F
icitM [0,1]	833859	F	*resB* [1,0]	2418905	R	*ywhK* [0,1]	3843974	R
treP [1,1]	850192	F	*ansA* [1,1]	2455924	R	*iywcE* [0,2]	3912770	F
acoA [0,1]	878941	F	*mmgB* [2,0]	2511394	R	*galT* [0,2]	3919217	R
iglvA [0,1]	889500	F	*icccA* [0,1]	2599349	R	*ywcC* [1,1]	3921504	R
glvC [0,2]	891982	F	*iyrkC* [1,0]	2714924	R	*iepr* [0,0]	3938499	F
iyfiG [1,0]	899485	F	*azlB* [0,3]	2729294	F	*dltB* [1,1]	3953404	F
yhcR [0,2]	991482	F	*yrbE* [0,1]	2838472	R	*ilicB* [0,0]	3960745	R
glpP [1,0]	1001824	F	*iysnE* [1,0]	2896634	F	*sigY* [0,3]	3969365	R
glpF [1,0]	1002087	F	*sdhC* [1,1]	2907542	R	*yxkJ* [0,1]	3978572	F
yhfL [1,0]	1100887	F	*lcfA* [0,1]	2919019	R	*imsmX* [1,0]	3984057	R
iyhjM [1,1]	1129143	F	*pheT* [3,0]	2926904	R	*iyxjJ* [0,1)	3995214	F
yisY [2,0]	1169249	R	*cstA* [1,0]	2937280	R	*iyxjC* [0,1]	4002964	R
iyjdC [0,1]	1270023	F	*araQ* [2,0]	2939120	R	*ibglS* [1,0]	4011393	R
yjdD [2,0]	1272217	R	*araB* [0,0]	2945795	R	*iyxiM* [1,0]	4017476	R
yjdD [1,0]	1272625	F	*iabnA* [0,1]	2949170	R	*ibglP* [0,1]	4034649	R
yjhB [1,1]	1289812	F	*iphoP* [0,1]	2977756	R	*hutP* [0,0]	4040441	F
yjmA [0,2]	1299956	F	*ytmK* [0,3]	3006485	R	*inupC* [0,1]	4050346	R
yjmA [0,1]	1301095	F	*iackA* [0,1]	3015526	R	*dra* [0,2]	4051031	R
yjmC [0,1]	1302892	F	*acsA* [0,2]	3039000	R	*yxeB* [1,1]	4066021	F
ykoX [2,0]	1409840	F	*amyX* [1,1]	3060807	R	*iolB* [0,1]	4080971	R
ykvD [3,0]	1431722	R	*ytdP* [1,0]	3083455	F	*yxnB* [3,0]	4097589	F
ykvZ [2,0]	1455163	F	*iytcP* [1,0]	3087305	R	*gntR* [0,1]	4112731	F
ptsG [0,2]	1456982	F	*imsmR* [0,1]	3095777	F	*gntK* [1,0]	4113782	F
ptsG [1,1]	1457296	F	*iytqA* [0,2]	3119049	F	*iyydK* [2,0]	4122278	R
ifruR [1,0]	1506954	F	*ytnA* [0,3]	3124181	R	*iyycO* [1,1]	4138860	R

[a]Position of the first base of *cre* is indicated. Base positioning is the same as that of Kunst *et al.* (12).
[b]F and R indicate the location of *cre* candidates in the forward and reverse strands of the chromosome (12), respectively.

that of transcription of the target genes. All *cre*s classified as [0,0], [0,1] and [0,1] exhibited repression ability, whereas among *cre*s classified as [0,2] and [1,0], some functioned but others did not. But, *cre*s classified as [0,3] and [1,0] did not function. Furthermore, we also tested four *cre*-like sequences which exhibited one mismatch in both the same and inverse directions as that of transcription of the target genes; a *cre* 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* sequences 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* sequences is desirable for a better function. The requirements of lower mismatching of *cre* sequences 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* tests implied that low mismatching of *cre* sequences 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*-i*bglP* and *cre*-*gntR* between P*spac* and *lacZ*, and then examined the catabolite repression of β-Gal synthesis in the constructed integrants carrying P*spac*-(*cre*-r-i*bglP*)-*lacZ* and P*spac*-(*cre*-r-*gntR*)-*lacZ*, respectively (Table 2). Thus, these inversed *cre*s are classified as [1,0] instead of [0,1] for the original *cre*s (Table 2). As shown in Table 2, the inversion of the i*bglP*- and *gntR*-*cre*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*s

Table 2. Catabolite repression of β-Gal synthesis exerted by various *cre*s and their sequence alignment

cre[a]	Region[b]	β-Gal activity[c] (nmol/min per mg protein)		Glucose repression ratio	*cre* and its flanking sequence	Location of *cre* position +1 in codon base[d]
		– Glc	+ Glc	(– Glc/+ Glc)		
none		248.8	297.2	0.8	+1 +15	
acoA [0,1]	–36/+47	134.7	8.4	16.0	ACAGACTAAAA ATGTAAGCGTTTGCT TTTTCGGGGAC	2
yjmA [0,1]	–21/+38	171.4	15.9	10.8	TGCTCGACCAA ATGAAAGCGTTATCA AATGTTGGCCT	1
iyxjC [0,1]	–19/+35	170.8	20.1	8.5	CGGACAGGTTT TTGTAAACGCTTTCT AGTTCGCGGAA	–
ibglP [0,1]	–22/+34	227.8	28.4	8.0	CAACACCAAAA ATGAAAGCGTTGACA TCTCACGAATC	–
hutP [0,0]	–25/+51	91.1	12.7	7.2	AGTAGCCGCAA TTGAAACCGCTTCCA AAAAGAGCGGT	2
gntR [0,1]	–31/+50	214.6	35.8	6.0	TTCAGGTCTGA TTGAAAGCGGTACCA TTTTATCAGAA	2
yobO [0,2]	–33/+45	130.5	23.6	5.5	TAAAAACGTTA ATGTAAGCGGATTCA CAGCGATTGGT	2
treP [1,1]	–22/+46	305.8	56.5	5.4	TGCAGCGTGCT GTGAAAACGCTTGCA GATATTTTTAT	1
iamyE [0,1]	–11/+64	256.0	48.9	5.2	GATAATTTTAA ATGTAAGCGTTAACA AAATTCTCCAG	–
icitM [0,1]	–22/+49	47.6	9.6	5.0	GGGAGAAAAAA ATGTAAGCGGATTCA TTTAAGGGGGA	–
iackA [0,1]	–83/+50	144.2	30.7	4.7	ACGACTTCTTA TTGTAAGCGTTATCA ATACGCAAGTT	–
araB [0,0]	–36/+50	226.5	49.9	4.5	TAAATCCTTCA ATGAAAACGATTACA AAGGACAAATT	1
lcfA [0,1]	–32/+58	118.7	26.4	4.5	AGGCAATAAAA ATGAAAACGTTATCA ATAGTCGACCA	1
iolB [0,1]	–22/+34	282.8	62.4	4.5	GACTGTGGGAA ATGAAAACGTTGTCA TCGTTCCTGCG	2
kdgA [0,1]	–21/+37	131.3	29.2	4.5	CGAGCGAAATT ATGGAAGCGCTGACA TTCGGTTTTAC	1
xylA [0,2]	–30/+145	204.7	63.4	3.2	AATCAACTATT TTGGAAGCGCAAACA AAGTGGTTTAC	2
imsmX [1,0]	–24/+38	72.6	23.4	3.1	ACACTCAATAT AAGAAAGCGTTTACA ATAACAAAGGG	–
iydbH [0,1]	–30/+48	193.6	73.6	2.6	CCCCAGACTGT ATGAAAACGCTATCA TTCTAGTAAGA	–
galT [0,2]	–36/+50	218.6	101.0	2.2	AAACCGGCCGA ATGGAAGCGGATACA GATACGTACCG	1
ydhO [0,1]	–35/+41	245.6	128.3	1.9	GCTTGTCGGAC TTGGAAGCGGTATCA TTCCAACACTG	2
dra [0,2]	–31/+43	294.8	154.8	1.9	ATCATACAGCT TTGAAACCGCATACA CAAAAAGCGGA	1
yxkJ [0,1]	–29/+45	165.3	95.7	1.7	CACACATGCAA TTGCAAACGGATACA ATTCACGAAGG	1
r-*ibglP* [1,0]		248.3	191.7	1.3	WTGNAANCGNWNNCW (consensus for *cre* search)	
r-*gntR* [1,0]		146.1	29.2	5.0	W WTGNAARCGNWWWCA WW (newly deduced consensus)	

[a]Among 32 *cre* candidates examined for their catabolite repressive ability, only functional *cre*s are listed. *cre* candidates (*ybyB* [0,2], *resB* [1,0], *ansA* [1,1], *sdhC* [1,1], *sunT* [1,0], *ptsG* [1,1], *ptsG* [0,2], *spoVG* [1,0], *azlB* [0,3] and *sigY* [0,3]) did not function; their catabolite repression ratios were between 0.7 and 1.0.

[b]*cre* regions which were cloned into plasmid pCRE-test are indicated; +1 is the first base of each *cre*-like sequence.

[c]The P*spac-cre-lacZ* integrants were grown with and without glucose (Glc). The cells were disrupted, and β-Gal was assayed as described in the text. The β-Gal activities (1 and 0.5 nmol/min/mg protein) in cells of strain GM122 grown with and without glucose, respectively, have been subtracted from the corresponding activities in the integrants.

[d]+1 of each *cre* in the coding region is the first or second base of codons (1 or 2), which means that not only position +1 but also +4, +7, +10 and +13 are the first or second bases of codons.

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*s listed in Table 1, we could predict from the above results whether or not the *cre*-like sequences not tested might function *in vivo*. *cre*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*s listed might function. Furthermore, we do not think that the *cre*s listed in Table 1 include all the *B.subtilis cre*s. For example, i*ynaJ-cre* [1,1] for the *ynaJ-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*-like sequences which exhibit high mismatch numbers through careful consideration of their palindromic nature in order to cover all the *cre*s of *B.subtilis*.

Alignment of *cre*s and their location in target genes

As shown in Table 2, we aligned 22 *cre* 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* sequences were also conserved. When the first base of *cre*, 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* 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*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* sequences localized in the three possible protein-coding frames? Thus, we examined the relative localization of the *cre* sequences in the protein-coding frames of the target genes, finding that in eight and seven genes, positions +1, +4, +7, +10 and +13 of WTGNAANCGNWNNCW are the first and second bases of codons, respectively, but in no case are they their third bases. The bases at these positions are W or N in the *cre* query sequence, in which relative randomness of base species is

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

1. Henkin,T.M., Grundy,F.J., Nicholson,W.L. and Chambliss,G.H. (1991) *Mol. Microbiol.*, **5**, 575–584.
2. Fujita,Y., Miwa,Y., Galinier,A. and Deutscher,J. (1995) *Mol. Microbiol.*, **17**, 953–960.
3. Galinier,A., Deutscher,J. and Martin-Verstraete,I. (1999) *J. Mol. Biol.*, **286**, 307–314.
4. Hueck,C.J., Hillen,W. and Saier,M.H.Jr (1994) *Res. Microbiol.*, **145**, 503–518.
5. Weickert,M.J. and Chambliss,G.H. (1990) *Proc. Natl Acad. Sci. USA*, **87**, 6238–6742.
6. Miwa,Y. and Fujita,Y. (1990) *Nucleic Acids Res.*, **18**, 7049–7053.
7. Miwa,Y. and Fujita,Y. (1993) *J. Biochem.*, **113**, 665–671.
8. Jacob,S., Allmansberger,R., Gärtner,D. and Hillen,W. (1991) *Mol. Gen. Genet.*, **229**, 189–196.
9. Wray,L.V.Jr, Pettengill,F.K. and Fisher,S.H. (1994) *J. Bacteriol.*, **176**, 1894–1902.
10. Grundy,F.J., Waters,D.A., Takova,T.Y. and Henkin,T.M. (1993) *Mol. Microbiol.*, **10**, 259–271.
11. Grundy,F.J., Waters,D.A., Allen,S.H.G. and Henkin,T.M. (1993) *J. Bacteriol.*, **175**, 7348–7355.
12. Kunst,F., Ogasawara,N., Moszer,I., Albertini,A.M., Alloni,G., Azevedo,V., Bertero,M.G., Bessières,P., Bolotin,A., Borchert,S. *et al.* (1997) *Nature*, **390**, 249–256.
13. Miwa,Y., Nagura,K., Eguchi,S., Fukuda,H., Deutscher,J. and Fujita,Y. (1997) *Mol. Microbiol.*, **23**, 1203–1213.
14. Jaacks,K.J., Healy,J., Losick,R. and Grossman,A.D. (1989) *J. Bacteriol.*, **171**, 4121–4129.
15. Vagner,V., Dervyn,E. and Ehrlich,S.D. (1998) *Microbiology*, **144**, 3097–3104.
16. Shimotsu,H. and Henner,D.J. (1986) *Gene*, **43**, 85–94.
17. Yanisch-Perron,C., Vieira,J. and Messing,J. (1985) *Gene*, **33**, 103–119.
18. Fujita,Y. and Freese,E. (1981) *J. Bacteriol.*, **145**, 760–767.
19. Fujita,Y. and Fujita,T. (1986) *Nucleic Acids Res.*, **14**, 1237–1252.
20. Atkinson,M.R., Wray,L.V.Jr and Fisher,S.H. (1990) *J. Bacteriol.*, **172**, 4758–4765.
21. Krüger,S. and Hecker,M. (1995) *J. Bacteriol.*, **177**, 5590–5597.
22. Miwa,Y., Saikawa M. and Fujita,Y. (1994) *Microbiology*, **140**, 2567–2575.
23. Zalieckas,J.L., Wray,L.V.Jr and Fisher,S.H. (1998) *J. Bacteriol.*, **180**, 6649–6654.