Office de la Propriété Intellectuelle du Canada
Un organisme d'Industrie Canada

Canadian Intellectual Property Office
An agency of Industry Canada

CA 2483006 A1 2003/10/30

(21) 2 483 006

(12) DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION

(13) A1

(86) Date de dépôt PCT/PCT Filing Date: 2003/04/18
(87) Date publication PCT/PCT Publication Date: 2003/10/30
(85) Entrée phase nationale/National Entry: 2004/10/19
(86) N° demande PCT/PCT Application No.: US 2003/012044
(87) N° publication PCT/PCT Publication No.: 2003/089604
(30) Priorités/Priorities: 2002/04/22 (60/374,627) US;
2002/04/22 (60/374,735) US

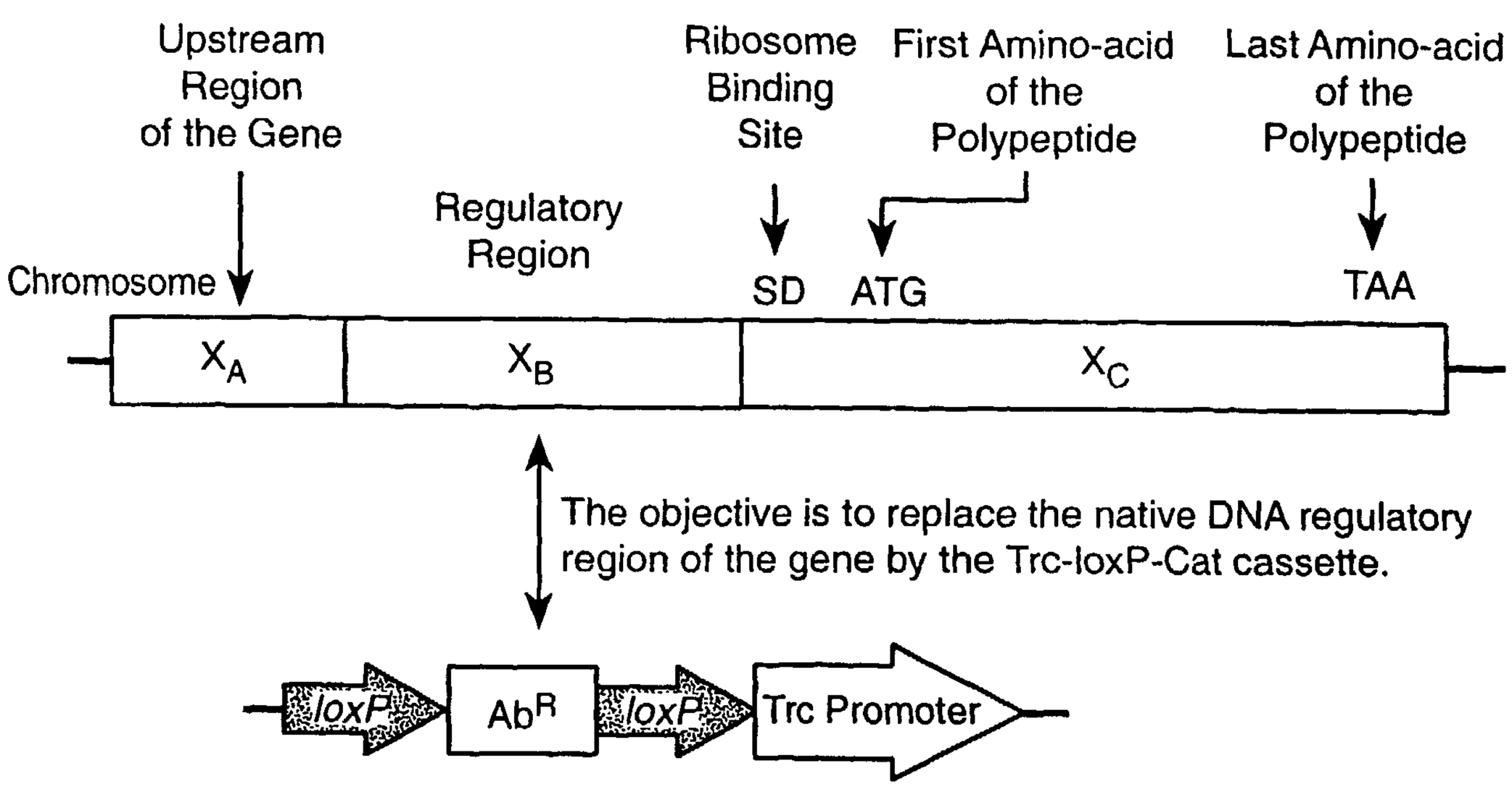
(51) Cl.Int.^{7}/Int.Cl.^{7} C12N 15/70, C12N 15/75, C12N 15/74, C12N 1/21

(71) Demandeur/Applicant:
GENENCOR INTERNATIONAL, INC., US

(72) Inventeurs/Inventors:
CERVIN, MARGUERITE A., US;
VALLE, FERNANDO, US

(74) Agent: BERESKIN & PARR

(54) Titre : METHODE DE CREATION DE PROMOTEURS MODIFIES PERMETTANT D'OBTENIR DIFFERENTS NIVEAUX D'EXPRESSION GENIQUE
(54) Title: METHODS OF CREATING MODIFIED PROMOTERS RESULTING IN VARYING LEVELS OF GENE EXPRESSION



(57) Abrégé/Abstract:
The present invention relates to a method of creating promoter cassettes that include modified precursor promoters and transforming a population of bacterial host cells with a promoter library comprising the promoter cassettes resulting in bacterial clones having a range of expression levels of a gene of interest. The invention further relates to selecting a transformed bacterial host cell which has an optimum level of gene expression.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
30 October 2003 (30.10.2003)

PCT

(10) International Publication Number
WO 03/089604 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US03/12044

(22) International Filing Date: 18 April 2003 (18.04.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/374,735	22 April 2002 (22.04.2002)	US
60/374,627	22 April 2002 (22.04.2002)	US

(71) Applicant *(for all designated States except US)*: **GENENCOR INTERNATIONAL, INC.** [US/US]; 925 Page Mill Road, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **CERVIN, Marguerite, A.** [CA/US]; 404 Palomar Dr., Redwood City, CA 94062 (US). **VALLE, Fernando** [MX/US]; 1015 Cadillac Way, Apt. 212, Burlingame, CA 94010 (US).

(74) Agent: **ITO, Richard, T.**; GENENCOR INTERNATIONAL, INC., 925 Page Mill Road, Palo Alto, CA 94304 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

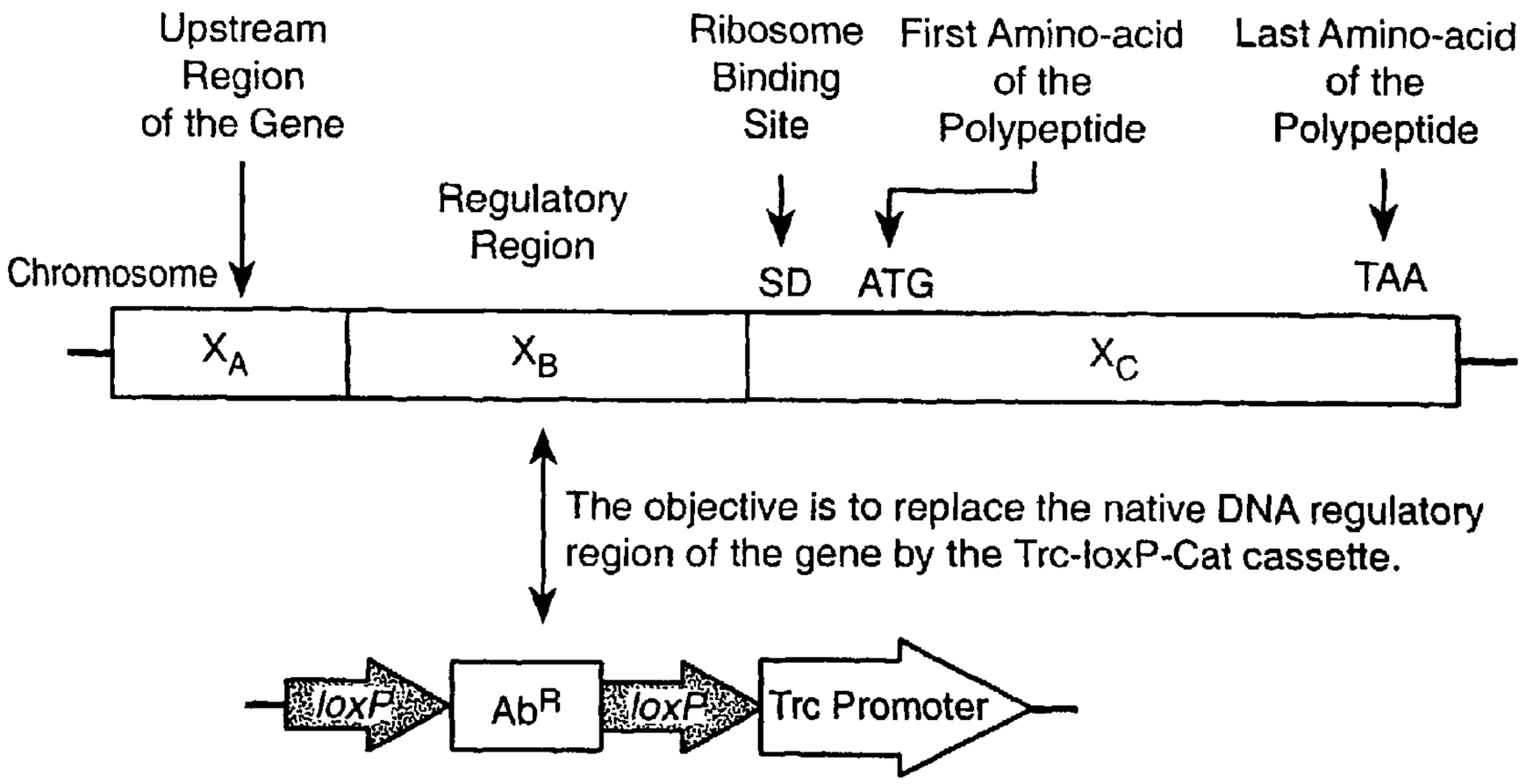
(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

[Continued on next page]

(54) Title: METHODS OF CREATING MODIFIED PROMOTERS RESULTING IN VARYING LEVELS OF GENE EXPRESSION



(57) Abstract: The present invention relates to a method of creating promoter cassettes that include modified precursor promoters and transforming a population of bacterial host cells with a promoter library comprising the promoter cassettes resulting in bacterial clones having a range of expression levels of a gene of interest. The invention further relates to selecting a transformed bacterial host cell which has an optimum level of gene expression.

WO 03/089604 A2

WO 03/089604 A2

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

METHODS OF CREATING MODIFIED PROMOTERS RESULTING IN VARYING LEVELS OF GENE EXPRESSION

FIELD OF INVENTION

The present invention relates to the genetic modification of bacterial cells. Particularly, the invention relates to a method of constructing a library of promoters that comprises precursor and modified precursor promoters, and use of the promoter library to replace the promoter of a chromosomal gene of interest in bacterial host cells, resulting in a population of bacterial host cells having a range of expression levels for the chromosomal gene of interest.

BACKGROUND OF THE INVENTION

For many years microorganisms have been exploited in industrial applications for the production of valuable commercial products, such as industrial enzymes, hormones, and antibodies. Despite the fact that recombinant DNA technology has been used in an attempt to increase the productivity of these microorganisms, the use of metabolic genetic engineering to improve strain performance, particularly in industrial fermentations, has been disappointing.

A common strategy used to increase strain performance is to alter gene expression, and a number of means have been used to achieve this end. One approach includes the cloning of a heterologous or a homologous gene in a multi-copy plasmid in a selected host strain. Another approach concerns altering chromosomal gene expression. This has been accomplished by various methods some of which include: 1) site-specific mutations, deletions or insertions at a predetermined region of a chromosome; 2) reliance on transposons to insert DNA randomly into chromosomes; and 3) altering of native regulatory regions of a gene at its chromosomal location. The alteration of regulatory regions can be accomplished for example, by changing promoter strength or by using regulatable promoters which are influenced by inducer concentration. Reference is made to Jensen and Hammer, (1998) *Biotechnology and Bioengineering* 58:193 -195; Jensen and Hammer, (1998) *Appl. Environ. Microbiol.* 64:82 – 87; and Khlebnikov et al. (2001) *Microbiol.* 147:3241. Other techniques used to replace regulatory regions of chromosomal genes

have been disclosed in Abdel-Hamid et al. (2001) *Microbiol.* 147:1483-1498 and Repoila and Gottesman (2001) *J. Bacteriol.* 183:4012 – 4023.

With respect to optimizing metabolic pathway engineering in a selected host, the above mentioned approaches have had limited success and each approach has certain disadvantages. Research has shown the expression level of a genetically modified gene on a plasmid is not necessarily correlated with the level of expression of the same modified gene located in the chromosome. (See Khlebnikov et al. (2001), *Microbiol.* 147:3241 and McCraken and Timms, (1999) *J. Bacteriol.* 181:6569).

Moreover, the effect of increasing expression of one gene in a metabolic pathway may have only a marginal effect on the flux through that metabolic pathway. This may be true even if the gene being manipulated codes for an enzyme in a rate-limiting step because control of a metabolic pathway may be distributed over a number of enzymes. Therefore, while a gene has been engineered to achieve a high level of expression, for example a 10 to 100 fold increase in expression, the overall performance of the engineered microorganism in a bioreactor may decrease. The decrease could be due to the balance of other factors involved in the metabolic pathway or the depletion of other substances necessary for optimum cell growth.

The above problem is addressed in part by Jensen and Hammer (WO 98/07846). The disclosure of WO 98/07846 describes the construction of a set of constitutive promoters that provide different levels of gene expression. Specifically, artificial promoter libraries are constructed comprising variants of a regulatory region that includes a –35 consensus box, a –10 consensus box and a spacer (linker) region that lies between these two consensus boxes. However, one of the drawbacks of the method described in WO 98/07846 is the extensive screening, which would be required of the promoter library. It is also disclosed in the reference that the modulation of promoter strength, by a few base pair changes in the consensus sequences or by changes in the length of the linker sequence, would result in a large impact in promoter strength, and therefore, it would not be feasible to achieve small steps in promoter strength modulation.

Therefore, a need still exists, in the area of metabolic pathway engineering, to develop a quick and efficient means of determining the optimum expression level of a gene in a metabolic pathway which in turn results in an optimization of strain performance for a desired product. The present invention satisfies this need by providing a method to characterize small changes in promoter strength of a modified precursor promoter and hence allowing for the selection of a cell providing an optimum level of gene expression.

SUMMARY OF THE INVENTION

In one aspect the invention relates to a method of creating a library of bacterial cells having a range of expression levels of a chromosomal gene of interest which comprises obtaining a promoter library which includes at least two promoter cassettes, wherein the promoter cassette comprises in sequential order a 5' sequence homologous to an upstream flanking region of a target site; a first recombinase recognition site; a selectable marker; a second recombinase recognition site; a precursor promoter or a modified precursor promoter comprising a –35 consensus region, a linker sequence and a –10 consensus region, wherein the modified promoter includes at least one nucleotide position that has been modified from the precursor promoter; and a 3' sequence homologous to a downstream flanking region of the target site; transforming bacterial host cells with the promoter library, wherein the promoter cassettes are integrated into the bacterial host cells by homologous recombination to produce transformed host cells; culturing the transformed host cells under suitable growth conditions; and obtaining a library of transformed bacterial cells, wherein the transformed bacterial cells exhibit a range of expression levels of a chromosomal gene of interest. In one embodiment the method further comprises selecting transformed bacterial cells from the library. In a further embodiment, the selected transformed host cells will have a higher level of expression of the gene of interest than bacterial cells comprising the precursor promoter. In a second embodiment the selected bacterial cells have a lower level of expression of the gene of interest than the bacterial cells comprising the precursor promoter. In a third embodiment the invention pertains to the transformed bacterial cells selected according to the method above. In further embodiments the promoter library comprises the Ptrc precursor promoter and modified Ptrc precursor promoters; the Ptac precursor promoter and modified Ptrc precursor promoters; and the P_{GI} precursor promoter and modified P_{GI} precursor promoters.

In a second aspect the invention relates to a promoter cassette comprising in sequential order a 5' sequence homologous to an upstream flanking region of a target site; a first recombinase recognition site; a selectable marker; a second recombinase recognition site; a modified precursor promoter comprising a –35 consensus region, a linker sequence and a –10 consensus region, wherein the modified promoter includes at least one modified nucleotide in a position corresponding to a –35 consensus region, a linker sequence or a –10 consensus region of a precursor promoter; and a 3' sequence homologous to a downstream flanking region of the target site. In one preferred embodiment the precursor promoter is selected from the sequences comprising base pairs –35 to +1 of the sequences in the group consisting of P_{trc}, P_{tacI}, $P_{D/E20}$, P_{H207}, P_{N25}, P_{G25}, P_{J5},

P_{A1}, P_{A2}, P_{A3}, P_L, P_{lac}, P_{lacUV5}, P_{con}, P_{GI} and P_{bla}, In a second embodiment the –35 region of the precursor promoter is selected from the group consisting of TTGACA, TTGCTA, TTGCTT, TTGATA, TTGACT, TTTACA and TTCAAA and the –10 region of the precursor promoter is selected from the group consisting of TAAGAT, TATAAT, TATACT, GATACT, AATAAT, TACGAT, TATGTT and GACAAT. In a further embodiment, a preferred precursor promoter is Ptrc wherein at least one nucleotide is modified in either the –35 box, the – 10 box or the linker region. In a further embodiment, the precursor promoter is P_{GI}, wherein at least one nucleotide is modified in either the –35 box, the – 10 box or the linker region.

In a further aspect, the invention relates to a promoter library comprising at least two promoter cassettes as defined above. In one embodiment, the invention relates to host cells transformed with a promoter cassette or a promoter library wherein the host cells are selected from the group consisting of *E. coli*, *Bacillus sp.* and *Pantoea sp.*

In yet another aspect, the invention relates to a method of modifying the regulatory function of a native promoter of a chromosomal gene of interest comprising, obtaining a promoter cassette according to the invention; transforming a host cell with the promoter cassette to allow homologous recombination between the promoter cassette and homologous flanking regions of a target site, wherein the promoter cassette replaces a native promoter region of a chromosomal gene of interest; and culturing the transformed host cells under suitable growth conditions. In one embodiment of this aspect, the selectable marker is excised from the transformed host cells, and in another embodiment the transformed host cells are further isolated.

In yet a further aspect, the invention concerns a method for altering the expression of a chromosomal gene of interest comprising obtaining a promoter cassette according to the invention; transforming a host cell with the promoter cassette; and allowing homologous recombination between the promoter cassette and homologous flanking regions of the target site, wherein the promoter cassette replaces a native promoter region of a chromosomal gene of interest and alters the expression of the chromosomal gene of interest as compared to the expression of the chromosomal gene of interest in a corresponding parent host cell.

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1A – 1K illustrate the nucleotide sequence of the pTrCm2 plasmid (SEQ ID NO. 1). The plasmid includes a first recombinase site and second recombinase site of loxP and a chloramphenicol (cat or Cm) marker gene flanked on each side by loxP. Further, the

plasmid includes the *trc* promoter (P_{trc}) region comprising the -35 consensus box, the -10 consensus box and the linker sequence and a bla coding region for the beta-lactamase enzyme that confers ampicillin resistance. Additionally the amino acid sequence of the coding regions is illustrated (SEQ ID NO. 2).

Figure 2 depicts a map of the TrCm2 plasmid.

Figure 3 depicts a map of the TrCm1 plasmid.

Figure 4 illustrates the nucleotide sequence of the promoter region and the relative promoter strength of the P_{tac} promoter and variants thereof having 1 or 2 base pair (bp) changes as disclosed in Sommer et al., (2000) *Microbiology* 146:2643-2653. P_{tac}, a chimeric bacterial promoter is represented by SEQ ID NO.3; variant M1 is represented by SEQ ID NO. 4; variant M2 is represented SEQ ID NO. 5; variant M3 is represented by SEQ ID NO. 6; variant M4 is represented by SEQ ID NO. 7; variant M12 is represented by SEQ ID NO. 8; variant M13 is represented by SEQ ID NO. 9; variant M14 is represented by SEQ ID NO. 10; variant M23 is represented by SEQ ID NO. 11 and variant M34 is represented by SEQ ID NO. 12.

Figure 5 illustrates the sequences of various well-characterized promoters and includes approximately 45 base pairs (bp) upstream of the transcriptional start site (+1), including the –35 consensus box, the linker sequence, and the -10 consensus box. The promoters are aligned with respect to the first T of the –35 consensus region and the last T of the –10 consensus region. The conserved regions are indicated by the boxes. $P_{D/E20}$ is represented by SEQ ID NO. 13; P_{H207} is represented by SEQ ID NO. 14; P_{N25} is represented by SEQ ID NO. 15; P_{G25} is represented by SEQ ID NO. 16; P_{J5} is represented by SEQ ID NO. 17; P_{A1} is represented by SEQ ID NO. 18; P_{A2} is represented by SEQ ID NO. 19; P_{A3} is represented by SEQ ID NO. 20; P_L is represented by SEQ ID NO. 21; P_{lac} is represented by SEQ ID NO. 22; P_{lacUV5} is represented by SEQ ID NO. 23; P_{tacI} is represented by SEQ ID NO. 24; P_{con} is represented by SEQ ID NO. 25; P_{bla} is represented by SEQ ID NO. 26; and P_{GI} is represented by SEQ ID NO. 34.

Figures 6A and 6B include a schematic representation according to the invention of a method used to replace the regulatory regions of a chromosomal gene of interest.

A. A promoter cassette is constructed including a loxP recombinase site, an Ab^R antibiotic marker, a second loxP recombinase site and a P_{trc} promoter (which may be a modified P_{trc} wherein segments A, B and C are from the same gene. However, the A segment could be from a different gene or region non-relevant to the regulatory segment C.

B. An upstream flanking region of homology (X_A) and a downstream flanking region of homology (X_C) to a chromosomal gene of interest is incorporated into the cassette. The flanking regions of homology are used for recombination by a double crossover event.

C. The promoter cassette replaces the native regulatory region of the chromosomal gene of interest (Xc). The selective marker is excised from the chromosome and the final chromosomal structure includes a precursor or modified precursor promoter.

Figure 7 is a schematic representation illustrating the replacement of the wild-type precursor lacZ promoter with a promoter cassette including an upstream nucleic acid fragment homologous to a 5' end of lacZ, a loxP recombinase site, a chloroamphenicol resistance gene, a second loxP site, a trc promoter and a nucleic acid fragment homologous to a downstream region of the lacZ gene. Nucleic acid sequences represent double stranded DNA regions relevant to designing PCR primers; lacZ1 (SEQ ID NO. 41) and lacZ2 (SEQ ID NO. 42) (see example 4).

Figure 8 illustrates the nucleotide sequence of the precursor promoter Ptrc (SEQ ID NO. 27) and 7 modified precursor promoters wherein the precursor promoter is Ptrc. The –35 box of Ptrc is represented by TTGACA and the –10 box is represented by TATAAT. Modified precursor promoters NF-T (SEQ ID NO. 28), NF-G (SEQ ID NO. 29), and NF-C (SEQ ID NO. 30) include nucleotide base changes in the –35 box. The –35 box of NF-T is TTGACT, the –35 box of NF-G is TTGACG, and the –35 box of NF-C is TTGACC. Modified precursor promoters NF-1T (SEQ ID NO. 32) and NF-2T (SEQ ID NO. 33) include nucleotide base additions of "T" and "TT" respectively between ATTAAT and CATCCGGCT…. of the 17 bp sequence of Ptrc. Promoter strength is determined by beta-galactosidase activity in the presence and absence of the inducer isopropyl-beta-D-thiogalactopyranoside (IPTG) measured relative to the promoter strength of the control, chromosomal promoter Plac using a standard beta-galactosidase assay (Miller J.H. (1972) EXPERIMENTS IN MOLECULAR GENETICS. Cold Spring Harbor Laboratory Press pp 352 – 355).

DETAILED DESCRIPTION OF THE INVENTION

One aspect of the present invention relates to the discovery that by modifying one or two nucleotides of a precursor promoter corresponding to nucleotides in the –35 consensus box, the –10 consensus box, or the linker region, promoter strength in terms of chromosomal gene expression could be changed to a level which allows quick identification of a range of gene expression. Furthermore, by constructing suitable promoter cassettes, the inventors were able to quickly test promoter efficiency at the chromosomal level.

A. Definitions

In this application, unless otherwise stated, illustration of the techniques used may be found in any of several well-known references such as Sambrook, J., et al., MOLECULAR CLONING: A LABORATORY MANUAL, Cold Spring Harbor Laboratory Press (1989); Goeddel, D., ed., GENE EXPRESSION TECHNOLOGY, METHODS IN ENZYMOLOGY, 185, Academic Press, San Diego, Calif. (1991); Deutshcer, M. P., ed., GUIDE TO PROTEIN PURIFICATION, METHODS IN ENZYMOLOGY, Academic Press, San Diego, Calif. (1989); and Innis, et. al., PCR PROTOCOLS: A GUIDE TO METHODS AND APPLICATIONS, Academic Press, San Diego, Calif. (1990).

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one or ordinary skill in the art to which this invention pertains. Both Singleton et al., DICTIONARY OF MICROBIOLOGY AND MOLECULAR BIOLOGY, 2D ED., John Wiley and Sons, New York (1994) and Hale and Marham, THE HARPER COLLINS DICTIONARY OF BIOLOGY, Harper Perennial New York (1991) provide one of skill in the art with general dictionaries of many of the terms used in this invention. One is also directed to Sambrook et al., MOLECULAR CLONING: A LABORATORY MANUAL, Cold Spring Harbor Laboratory Press (1989) for definitions and terms of the art.

For the purposes of the present invention, the following terms are used to describe the invention herein.

A "promoter" or "promoter region" is defined herein as a nucleotide sequence, which is recognized and bound by a DNA dependent RNA polymerase during the initiation of transcription. In the context of the present invention a promoter includes two consensus regions generally hexamers. The first consensus region is centered about 10 base pairs (bp) upstream from the start site of transcription initiation and is referred to as the –10 sequence, -10 box or *Pribnow* box. The second consensus region is centered about 35 bp upstream of the start site and is referred to as the –35 sequence or -35 box. It is these two regions of homology in which the E. coli RNA polymerase is believed to recognize and functionally bind most tightly.

A linker sequence extends between each consensus sequence and is comprised of about 14 to 20 base pairs. With the exception of nucleotides –15, -16, -17 and –18, linker regions in general do not appear to have significant regions of homology larger than two bp from one promoter to another.

A promoter may be a regulatable promoter such as the trc promoter which is induced by IPTG or a constitutive promoter. Other sequences which are considered part of the regulatory region of a gene include the ribosomal binding site and transcription start

site. The transcriptional start site means the first nucleotide to be transcribed and is designated +1. Nucleotides downstream of the start site are numbered +2, +3, +4 etc., and nucleotides in the opposite direction (upstream) are numbered –1, -2, -3 etc. The "ribosome binding site (RBS)" is a short nucleotide sequence usually comprising about 4 – 16 base pairs and is involved in the interaction of the mRNA with the ribosome for translation of an encoded protein.

A "precursor promoter" as used herein includes wild-type promoters and known variant promoters. A "wild-type" promoter is a naturally occurring promoter in either the plasmid or chromosome of a host organism (a non-mutant promoter), and the term is used interchangeability herein with "native" promoter. A variant promoter is a known mutant or modified wild type promoter including known hybrid promoters. A precursor promoter as used herein may be a wild-type promoter or a variant promoter.

A "modified precursor promoter" is a precursor promoter that has been modified by altering a nucleotide in at least one position corresponding to the –35 box, the -10 box or the linker region. A "promoter cassette" or "promoter construct" as used herein includes a precursor promoter or a modified precursor promoter. The term "promoter library" refers to a population of promoter cassettes wherein the population has at least two members. IA promoter library can be used, for example to generate a library of transformed host cells wherein members of the library (bacterial clones) have varying levels of promoter activity relative to the promoter activity of the precursor promoter. The varying levels of promoter strength result in a library of clones with different levels of expression for the same coding region of a gene of interest. Transformed host cells or bacterial clones may then be selected for optimal expression.

Modification or alteration may include addition (insertion), deletion or change (substitution) in at least one nucleotide base of a nucleic acid segment and particularly of a precursor promoter sequence. A "deletion" is defined as a change in one or more nucleotides wherein said nucleotides are absent. An "insertion" is the addition of one or more nucleotides as compared to the precursor promoter. A "substitution" results from the replacement of one or more nucleotides with a different nucleotide.

For the purpose of this invention a "tac promoter (Ptac)" also referred to as tacI in the literature is a precursor promoter comprising the nucleic acid sequence set forth in SEQ ID NO. 3 and SEQ ID NO. 24, wherein the –35 box is TTGACA, the linker is represented by 16 base pairs the –10 box is TATAAT (Brosius et al., *J. Biol. Chem*. 260:3539 (1985) and Deuschle et al., *EMBO J* 5:2987 – 2994 (1986)).

As used herein a "trc promoter (Ptrc)") is a precursor promoter comprising the nucleic acid sequence set forth in SEQ ID NO. 27. The nucleotide sequence of the -10 box and the -35 box is the same as Ptrc, but the linker section includes 17 bp. Ptrc differs from Ptac by the addition of a C nucleotide between nucleotides -18 and -19 of Ptac. Ptrc and Ptac are essentially identical in strength. (Russell and Bennett, *Gene* 20:231 (1982); Amann et al., (1983) *Gene* 25:167-178) and Mulligan et al., *J. Biol Chem*. 260:3529 (1985)).

A "gene" is defined herein as a sequence of nucleotides that code for a functional polypeptide or RNA molecule (regulatory RNA's, tRNA's, rRNA's). Genes may include both coding regions (exons), non-coding regions (introns) and regulatory regions such as promoters and enhancers.

The term "nucleic acid" includes RNA, DNA, and cDNA molecules. The term is used interchangeably with polynucleotide. An oligonucleotide is a short chain nucleic acid molecule. A primer is an oligonucleotide, whether occurring naturally as in a purified restriction digest or produced synthetically, which is capable of acting as a point of initiation of synthesis when placed under conditions in which synthesis of a primer extension product which is complementary to a nucleic acid strand is induced, (i.e. in the presence of nucleotides and an inducing agent such as DNA polymerase and at a suitable temperature and pH). The primer is preferably single stranded for maximum efficiency in amplification. The primer must be sufficiently long to prime synthesis of extension products in the presence of the inducing agent. In one embodiment of the invention primers are degenerate primers wherein one nucleotide base is modified relative to a sequence of a precursor promoter.

As used herein the term "polypeptide" refers to a compound made up of amino acid residues linked by peptide bonds. The terms "protein" and "polypeptide" are used interchangeably herein. It will be understood by one of skill in the art that as a result of the degeneracy of the genetic code, a multitude of nucleotides encoding a given protein may be produced.

A "DNA construct" refers to a sequence that is used to introduce a polynucleotide into a host cell. The definition of a DNA construct encompasses, for example a promoter cassette including a precursor promoter and/or a modified precursor promoter. In one embodiment, a DNA construct is used to integrate a polynucleotide into a chromosomal target site by homologous recombination. A DNA construct may include either homologous and/or heterologous sequences of a host cell gene. In one embodiment a DNA construct may be inserted into a vector.

The term "target site" is intended to mean a predetermined genomic location within a bacterial chromosome where the integration of a DNA construct is to occur.

The term "introduced" used in the context of inserting a nucleic acid into a cell means transfection, transformation, protoplast fusion, transduction or the like and includes reference to the incorporation of a nucleic acid into a prokaryotic cell where the nucleic acid may be incorporated into the genome of the cell, converted into an autonomous replicon or transiently expressed.

A "flanking region" or "flanking sequence" means any region or sequence that is either upstream or downstream of the sequence or region being discussed, e.g. for genes A, B, and C, gene B is flanked by A and C gene sequences. Homologous flanking regions are homologous (essentially identical) to a nucleic acid sequence in the host cell chromosome.

A "vector" refers to a nucleic acid construct designed for transfer between different host cells. A vector may include a DNA construct, plasmid, cloning vector, expression vector and bacteriophage.

A "recombinase recognition site" is a novel recombination site which facilitates directional insertion of nucleotide sequences into corresponding recombination sites at a target site in a chromosome.

As used herein a "selectable marker' or "selective gene" refers to a gene capable of expression in a host cell which allows for ease of selection of those host cells containing an introduced DNA construct with the selective marker. Typically a selective marker is a gene that confers antimicrobial resistance or a metabolic advantage on the host cell to allow cells containing an exogenous introduced DNA to be distinguished from cells that have not received the exogenous nucleic acid.

Chromosomal integration is a process wherein a DNA construct such as a promoter cassette according to the invention is introduced into a host chromosome. The homologous flanking regions of the promoter cassette will align with homologous regions at the target site of the host chromosome.

"Homologous recombination" means the exchange of nucleic acid fragments between two DNA molecules or pair chromosomes (during crossing over) at the site of identical nucleotide sequences. In the present invention chromosome integration is preferably by homologous recombination.

A "metabolic pathway" is a series of chemical reactions that either break down a large molecule into smaller molecules (catabolism) or synthesize more complex molecules from smaller molecules (anabolism). Most of these chemical reactions are catalyzed by a

number of enzymes. In many metabolic pathways there are rate-limiting enzymatic steps which serve to regulate the pathway. For example, in the glycolytic pathway wherein glucose is converted to pyruvate and ATP, phosphofructokinase is considered a key enzyme in regulation; and in the pentose phosphate pathway wherein NADPH and ribose-5-phosphate are generated, glucose -6-phosphate dehydrogenase and fructose 1,6-diphosphatase are considered key enzymes.

The term "homology" refers to sequence similarity or identity, with identity being preferred. Homology is determined using standard techniques known in the art. (Pearson et a., (1988) *PNAS* USA 85:2444 and Needleman et al., (1970) *Adv. Appl. Math*. 2:482).

As used herein the term "expression" refers to the process by which a polypeptide is produced based on the nucleic acid sequence of a gene. The process includes both transcription and translation. A “range of expression levels” means the plurality of expression levels of a gene of interest obtained from a library of bacterial clones transformed with a library of promoter cassettes.

As used herein, "optimal expression" refers to the cumulative conditions that provide an optimal level of gene expression for a particular coding region. Under certain conditions, optimal expression may mean a low level of gene expression.

"Operably linked" means that the nucleic acid sequences are functionally related. Generally operably linked means that the nucleic acid sequences being linked are contiguous. Linking is accomplished by ligation at convenient restriction sites. If such sites do not naturally exist, synthetic oligonucleotide adaptors or linkers are used in accordance with conventional practice.

"Isolated" as used herein refers to a nucleic acid or polypeptide that is removed from at least one component with which it is naturally associated.

"Host cell" means a cell that has the capacity to act as a host and expression vehicle for a promoter cassette or DNA construct according to the invention. The host cell may be a recombinant host cell. A "corresponding parent host cell" means a bacterial cell that has not been transformed with a promoter cassette comprising a modified precursor promoter according to the invention and which retains a precursor promoter. In general when a corresponding parent host cell and a transformed host cell comprising a modified precursor promoter are compared with respect to the level of gene expression, both cells will be grown under essentially the same growth conditions unless indicated otherwise.

As used herein the term polymerase chain reaction (PCR) refers to the methods of U.S. Patent Nos. 4,683,195; 4,683,202 and 4,965,188 which include methods for increasing the concentration of a segment of a polynucleotide in a mixture of DNA without

cloning or purification. This process for amplifying a target sequence or DNA fragment consists of introducing two oligonucleotide primers to the DNA mixture containing the target sequence, followed by a sequence of thermal cycling in the presence of DNA polymerase. The two primers are complementary to their respective strands of the target sequence.

The term" heterologous nucleic acid" or “heterologous polypeptide” as used herein refers to a nucleic acid or polypeptide sequence that does not naturally occur in a host cell. With respect to a heterologous nucleic acid, the sequence has a portion which is not native to the cell in which it is expressed.

A “homologous nucleic acid" or “homologous polypeptide” as used herein refers to a nucleic acid or polypeptide that naturally occurs in a host cell.

When numeric ranges are used herein they are inclusive of the numbers defining the range.

As used in the specification the singular “a”, “an” and “the” include the plural references unless the context clearly dictates otherwise. For example, the term “a cassette” may include a plurality of cassettes.

The published patent applications, issued patents and references cited herein are hereby incorporated by reference in the instant application.

B. Embodiments

Precursor promoters useful for creating a promoter cassette according to the invention include the sequences of the precursor promoters listed in Table 1 below. Figure 5 illustrates the sequences of these precursor promoters including the –35 region, the –10 region and the linker region. All promoters in the table are characterized with respect to the β-lactamase promoter Pbla and promoter strengths are given in “Pbla-units”. (Deuschle, et al., *EMBO J.*,5:2987-2994 (1986)).

///
///

Table 1

PROMOTER	Source	Relative Activity	SEQ ID NO.
β-lactamase (bla)	*E.coli* vector	1	26
Pconsensus (con)	Synthetic DNA	4	25
PTac I (Trc)	Hybrid of 2 promoters	17	24
PLacUV5	Mutant of Lac	3.3	23
PLac	*E.coli lacZ* gene	5.7	22
PL	Phage λ	37	21
PA1	Phage T7	22	18
PA2	Phage T7	20	19
PA3	Phage T7	76	20
PJ5	Phage T5	9	17
PG25	Phage T5	19	16
PN25	Phage T5	30	15
PD/E20	Phage T5	56	13
PH207	Phage T5	55	14

In general precursor promoter sequences useful in the invention include sequences of between 200 to 20 bp, preferably of between 150 to 25 bp, more preferably of between 30 to 100 bp and most preferably between 50 to 30 bp upstream from the transcriptional start site (+1).

In one embodiment a preferred precursor promoter is a Trc promoter (Ptrc) wherein the –35 box is TTGACA, the linker is represented by 17 base pairs, and the –10 box is TATAAT (SEQ ID NO. 27) (Amann et al., (1983) *Gene* 25:167-178).

In another embodiment a preferred precursor promoter is a tac promoter (Ptac) (SEQ ID NOs. 3 and 24). The nucleotide sequence of the -10 box and the -35 box is the same in Ptac and Ptrc, but the linker region differs by 1 bp.

In another preferred embodiment, a precursor promoter is a P_{GI}. This promoter is also known in the literature as a xylose isomerase promoter and the regulatory sequence encompassing the promoter is disclosed in Amore et al., (1989) *Appl. Microbiol. Biotechnol.* 30:351 - 357. The sequence of the short segment of the promoter (+50 to –7 of the –10 box) is illustrated in SEQ ID NO. 34.

5' CGAGCCGTCACGCCCTTGACA ATGCCACATCCTGAGCA AATAAT, 3'

wherein the –35 box is represented by TTGACA and the –10 box represented by AATAAT.

A precursor promoter may be determined by various exemplary methods. While not wanting to be limited, in one embodiment, sequencing of a particular genome may be performed and putative promoter sequences identified using computerized searching algorithms. For example, by using Neural Network for Promoter Prediction software, NNPP. NNPP is a time-delay neural network consisting mostly of two feature layers, one for recognizing TATA-boxes (-10 boxes) and one for recognizing so called "initiators", which are regions spanning the transcription start site. Both feature layers are combined into one output unit. These putative sequences may then be cloned into a cassette suitable for preliminary characterization in *E. coli* and/or direct characterization in *E. coli*

Promoter sequences can also be identified by homology analysis. For example, a homology study of a family of genomes may be performed and analyzed for the presence of putative consensus promoters using BLAST. These putative promoter sequences may then be cloned into a cassette suitable for preliminary characterization in *E. coli*. Some preferred precursor promoters are listed in Figures 4 and 5.

A modified precursor promoter according to the invention will comprise at least one modification to a nucleotide in a precursor promoter. In one embodiment, the modification will be to a nucleotide base positioned in the –35 consensus region. This modification may include a modification to one or more nucleotide bases at a position equivalent to the –30, -31, -32, -33, -34,and/or –35 position of a precursor promoter. Preferably the modification will be of one nucleotide or two nucleotides and preferably the modification will be a substitution. When two positions are to be modified four positions will be conserved, and when one position is modified five positions will be conserved. In a further embodiment the modified precursor promoter will include a change at a position corresponding to -30 and/or a change to a position corresponding to -35.

In another embodiment a modified precursor promoter is obtained from a precursor promoter having a –35 region represented by the following sequences, TTGACA, TTGCTA, TTGCTT, TTGATA, TTGACT, TTTACA and TTCAAA. Particularly preferred –35 consensus regions to be modified from a precursor promoter are TTTACA and TTGACA. As a non-limiting example when TTGACA is the –35 box of the precursor promoter, the nucleotide at position –30 which is A may be substituted with a C, T, or G; the nucleotide at position –31 which is C may be substituted with a G, A or T; the nucleotide at position –32 which is A may be substituted with a T, C, or G; the nucleotide at position –33 which is G may be substituted with a C, T, or A; the nucleotide at position –34 which is T may be substituted with a C, G, or A; and/or the nucleotide at position –35 which is T may be substituted with a C, G or A. In one particular embodiment, the modified precursor promoter will include a

modification of one to four nucleotides in the consensus region represented by TTGACA. In one embodiment four positions will be conserved and two positions will be modified. In another embodiment five positions will be conserved and one position will be modified. In a further embodiment the modified precursor promoter will include a change at a position corresponding to -30 or to a position corresponding to -35. When TTGACA is the precursor promoter it may be modified as follows, TTGACT, TTGACG, TTGACC or CTGACA.

In a further embodiment, the modification of the precursor promoter will be in the –10 region. This modification may include a modification to one or more nucleotides at a position equivalent to the –7, -8, -9, -10 –11 or –12 position of a precursor promoter. Preferably, the modification will be in one or two nucleotide positions. In a preferred embodiment, the modification will be a substitution at one or two nucleotide positions. Preferred precursor promoters include the following sequences in the –10 region: TAAGAT, TATAAT, TATACT, GATACT, TACGAT, AATAAT, TATGTT and GACAAT. Particularly preferred –10 regions include the sequences AATAAT, TATAAT, TATGTT and TAAGAT. In one embodiment the precursor promoter is Ptrc (SEQ ID NO. 27) and the modified precursor promoter will include at least one modification of a nucleotide in the –10 region represented by TAAGAT. For example the nucleotide at position –7 which is T may be substituted with G, C or A; the nucleotide at position –8 which is A, may be substituted with T, C or G; the nucleotide at position –9 which is G, may be substituted with C, T or A; the nucleotide at position –10 which is A may be substituted with T, C or G; the nucleotide at poison –11 which is A may be substituted with T, C or G and the nucleotide at position –12 which is T may be substituted with T, C or G. In one embodiment four positions will be conserved and two positions will be modified. In another embodiment five positions will be conserved and one position will be modified.

In some embodiments of the invention both the –35 region and the –10 region of the precursor promoter may have modifications. In one embodiment, the modification will include one nucleotide position in the - 35 box, wherein the other nucleotides remain conserved, and will include a modification at one nucleotide position in the -10 box wherein the other nucleotides remain conserved. In another embodiment, the modification will comprise a modification to a -35 region represented by TTGACA and a –10 region represented by TATAAT, and in another embodiment, the modification will comprise a modification to a -35 region represented by TTGACA and a –10 region represented by AATAAT. The total number of modifications in a –35 box and –10 box of a precursor promoter may include one position in each consensus box, wherein the other positions are conserved. The total number of modifications in a –35 box and –10 box of a precursor

promoter may also include one position in one consensus box and two positions in the other consensus box, wherein the other positions are conserved. Also the total number of modifications in a –35 box and –10 box of a precursor promoter may include two positions in each consensus box, wherein the other positions are conserved.

In a further embodiment a modified precursor promoter includes a modification to the linker sequence of the precursor promoter. Linker sequences are in general 14 to 20 nucleotides, and more typically 16 to 18 nucleotides. A modification may include the addition of 1, 2, 3, 4, or 5 base pairs to a linker sequence or the substitution of 1, 2, 3, 4, 5 or more base pairs. Preferably the modification is the addition of one or two base pairs in the linker region wherein the addition may be any one of the nucleotides of A, T, C or G. Preferably the addition comprises one or two base pairs to the linker sequence of Ptrc. Further the addition to Ptrc preferably occurs between base pair -23 and -24. Further embodiments include the addition of T (see SEQ ID NO. 32) and TT between base pairs -23 and -24 (see SEQ ID NO. 33). In a further embodiment the modification may include the substitution of one, two or three nucleotide bases in any position of the precursor linker sequence. Preferably the modification is the substitution of one nucleotide base in any position of the precursor linker.

In further embodiments a modified precursor promoter includes a modification to the –35 box, the linker region and the –10 box. The modification will include at least three nucleotide positions and no more than eight nucleotide positions wherein each region includes one modification. Preferably the modification will include three nucleotide positions wherein each region includes one modification.

The modified precursor promoter sequence may be generated by means well known in the art including but not limited to mutagenesis techniques including chemical mutagenesis, polymerase chain reaction and site-directed mutagenesis to one or more nucleotides. (see Miller, J. H. A. A SHORT COURSE IN BACTERIAL GENETICS, Cold Spring Harbor Laboratory Press 1992). In one embodiment degenerate oligonucleotides are synthesized for the host cell for which a promoter library is to be constructed. In a preferred embodiment, alteration to the precursor promoter is accomplished by site-directed mutagenesis using the QuickChange commercial kit (Stratagene, La Jolla, CA).

Individual promoters of the modified promoters defined herein are also comprised by the invention. In one embodiment specific promoters, which have been constructed according to the invention, include those modified P_{trc} promoters: NF-T (SEQ ID NO. 28); NF-G (SEQ ID NO. 29); NF-C (SEQ ID NO. 30); NF-1T (SEQ ID NO. 32) and NF-2T (SEQ

ID NO. 33). Another promoter comprised by the invention is the promoter designated MC-C3 having the sequence TCTGAAATGAGCTGCTGACA ATTAATCATCCGGCTCG TATAAT GTGTGG (SEQ ID NO. 31) wherein the -35 box is CTGACA, the linker region is ATTAATCATCCGGCTCG and the –10 box is TATAAT.

A promoter cassette according to the invention will include a precursor promoter and/or a modified precursor promoter as disclosed above. Further a promoter cassette according to the invention may include a 5' sequence homologous to an upstream flanking region of a target site wherein the target site is preferably a chromosomal gene of interest. A 5' sequence homologous to an upstream flanking region of a target site may include from 5 to 500 nucleotides, preferably from 10 to 200 nucleotides, also from 10 to 100 nucleotides and additionally 10 to 50 nucleotides, which are homologous to the nucleotides upstream of the target site.

The gene of interest may be any chromosomal gene. In one embodiment the gene of interest encodes a therapeutically significant protein such as growth factors, cytokines, hormones, ligands, receptors and antibodies. In another embodiment the gene of interest encodes a commercially important enzyme such as amylases, proteases, glucoamylases, dehydrogenases, esterase, cellulases, galactosidases, oxidases, reductases, kinases, xylanases, laccases, phenol oxidases, glucose oxidases, catalases, lipases and phytases. In further embodiments the gene of interest encodes transporter proteins, such as glucose and/or galactose permease (transporters). In other embodiments, the gene of interest may encode enzymes in a metabolic pathway, such as glucose dehydrogenase, pyruvate dehydrogenase and pyruvate oxidase. In particular embodiments the gene of interest will encode industrially important proteins such as lipases, esterases, hydrogenases and proteases. The chromosomal gene of interest or encoding region thereof may be heterologous or homologous to the host cell, but will be operably linked to a native promoter or precursor promoter which will be replaced by a library of promoters or by a modified precursor promoter according to the invention.

A promoter cassette according to the invention may also include two recombinase recognition sites and a selectable marker flanked by each recombinase recognition site. Examples of recombinase recognition sites are well known in the art. Recombinases generally fall into two distinct families that each use a different mechanism of catalysis. These are the tyrosine recombinases and the serine recombinases. Either type of recombinase system may be used in the present invention. The tyrosine recombinase family is also known as the lambda-integrase family and includes 100 or more identified

members. (Nunes-Duby, et al, *Nucleic Acids Research* 26:391 – 406 (1998)). There are more than 72 serine recombinases described in the literature and these include Tn3, Hin, SOPIVCA, ΦC31 and λ. Particularly preferred recombinases which could be used in the invention include Cre and Flp (Nunes-Duby, D, et al, *Nucleic Acids Research* 26:391- 406 (1998) and Huang et al., *Nucleic Acids Research* 19:443 (1991)); XerC-XerD (cer, parB dif and psi) (Blake et al., (1997) *Mol. Microbiol.* 23:387 – 398); P22 xis-int (AttP22 and ataA) (Cho et al. (1999) *J. Bact.* 181:4245 – 4249); SPOIVCA (SpoIVCB, SpoIIIC, AttPskin and AttBskin) (Straiger et al., (1989) *Sci.* 243:507-512); Resolvase (res) (Yang and Steitz (1995) *Cell* 82:193 – 207) and λInt (Att, attL and attR) (Hallet and Sherrat (1997) *FEMS Microbiol. Lett.* 21: 157 – 178).

Particularly preferred recombinases are Cre and Flp. In a most preferred embodiment, the first and second recombinase recognition sites include the bacteriophage P1 Cre/*loxP* recombination system, which comprises a Cre enzyme and two asymmetric 34 bp *loxP* recombination sites (See Sternberg and Hamilton (1981) *J. Mol. Biol.* 150:467 – 486; Van Duyne (2001) *Ann. Rev. Biophys. Biomol. Struct.* 30:87 – 104 and Palmeros, B, et al *Gene* 247:255 (2000)). A l*oxP* site comprises two 13 bp sequences, inverted and imperfectly repeated, which surround an 8 bp core asymmetric sequence, where crossing-over occurs. The Cre-dependent intramolecular recombination between two parallel *loxP* sites results is excision of any intervening DNA sequence as a circular molecule, producing two recombination products, each containing one *loxP* site. Preferably, the promoter cassette includes a selective marker flanked by two loxP sites. In a particularly preferred embodiment the recombinase sequence will include variants of the loxP site because if two LoxP sites are in the right orientation in the chromosome, they can promote the loop-out or inversion of big regions of the chromosome. Reference is made to Sauer E., *Curr. Opin. Biotech.* (1994), 5: 521-527; Palmeros et al. *Gene* 247 (2000) 255-264. and Hoess et al.,(1986) *NAR,* 14:2287-2230).

In a preferred embodiment a selectable marker is located between the two recombinase sites. While various selectable markers may be used a preferred selective marker is an antibiotic resistance gene. These are well known in the art. For example, the gene may be an erythromycin resistance gene (Em^r), an ampicillin resistance gene (Ap^r), a chloramphenicol resistance gene (Cm^r), gentamicin resistance gene (Gm^r) or a kanamycin resistance gene (Km^r). Preferably once the promoter cassette including the selectable marker is introduced into a host cell, the marker is removed for example by following the teaching of Palmeros et al. (2000) *Gene* 247 (2000) 255-264.

A promoter cassette may also include a 3' sequence homologous to a downstream flanking region of the target site. The 3' sequence may include from 5 to 500 nucleotides, also 10 to 200 nucleotides, also 10 to 100 nucleotides and further 10 to 50 nucleotides which are homologous to the nucleotides downstream of a target site.

The DNA constructs including the promoter cassettes of the invention may include various restriction sites facilitating the ligation of various fragments of the DNA construct. Restriction sites may include XbaI, EcoRI, BglII, BamHI, TaqI and the like. For example restriction sites may be used for ligation with the –35 sequence fragment and the translation start codon and gene sequences positioned downstream therefrom.

The promoter cassettes and modified precursor promoter nucleic acids may also include other sequence such as ribosome binding sites (RBS), mRNA stabilizing sequences, enhancers, silencing sequences, transcriptional terminators, transcriptional attenuators, operators and mRNA destabilizing sequences.

In some embodiments the precursor promoter and the chromosomal gene of interest are heterologous and in other embodiment the precursor promoter and the chromosomal gene of interest are homologous.

Promoter cassettes may be constructed using standard well-known recombinant engineering techniques such as PCR and as described in various references such as Sambrook *supra*., Palmeros et al. (2000) *Gene* 247:255 - 264; Datsenko and Wanner (2000) *Proc. Natl. Acad. Sci.* USA 10:640 - 6645). In general a promoter is cloned into a vector and a selected marker is linked to the promoter by cloning the marker upstream of the promoter in such a manner that the cloning step does not influence the function of the promoter. The resulting marker-promoter region is used as a promoter cassette. The cassette may be isolated by PCR or with restriction enzymes. To allow for homologous recombination, the cassette may be linked to regions of homology with the host cell chromosome using PCR by incorporating regions of homology in PCR primers. Or by ligating proper DNA restriction fragments (Datsenko and Wanner (2000) *Proc. Natl. Acad. Sci.* USA 10:6640 – 6645).

In other embodiments certain portions of the modified promoter may be deleted or excised from the promoter cassette, and the modified precursor promoter re-tested. In the event that expression is observed after this modification, and determination of whether expression has increased or decreased following modification, a positive or negative regulatory element of the promoter may be identified. In addition, specific regions of the precursor promoter may be isolated and tested in isolation. In this manner, specific elements may be identified that regulate gene expression in the host cell.

The invention further includes a library of promoters. A library will comprise at least two promoter cassettes. However, a promoter library may comprise 10^3 or more members. In preferred embodiments, the promoter library will comprise at least 2, at least 3, at least 4, at least 8, at least 16, at least 64 members. Preferably the library will comprise promoter cassettes wherein the modified precursor promoters are obtained from the same precursor promoter. In one embodiment a library of promoters may include promoter cassettes comprising modified Ptrc and precursor Ptrc. In another embodiment a library of promoters will include promoter cassettes comprising modified Ptac and precursor Ptac, and in another embodiment a library of promoters will include promoter cassettes comprising a modified P_{GI} and a precursor P_{GI}. Preferred precursor promoters include those mentioned above for the promoter cassette. In another embodiment the library will comprise promoter cassettes wherein the modified precursor promoter is obtained from different precursor promoters.

Promoter cassettes according to the invention may be used individually and introduced into a host cell or may be used in a promoter library. Figures 6 and 7 schematically illustrate the construction of a promoter cassette, the introduction of the promoter cassette into a host, the replacement of the host regulatory region with the promoter cassette and excision of the selective marker according to the invention.

In one embodiment a host cell is a bacterial cell. A bacterial host cell may be a gram-positive cell. Preferably the host cell is a *Bacillus* species. *Bacillus* species include but are not limited to *B. subtilis*, *B. licheniformis*, *B. lentus*, *B. brevis*, *B. stearothermophilus*, *B. alkalophilus*, *B. amyloliquefaciens*, *B. clausii*, *B. halodurans*, *B. megaterum*, *B. coagulans*, *B. circulans*, and *B. thuringiensis*.

In another embodiment a host cell is a gram-negative bacterial cell, such as an *Escherichia* species or *Pantoea* species. *E. coli* are the most preferred host cells. The genus of *Pantoea* includes all members known to those of skill in the art, including but not limited to *P. citrea, P. terrea, P. agglomerans, P. dispersa, P. punctata, P. ananas* and *P. stewartii.* It is recognized that the genus *Pantoea* continues to undergo taxonomical reorganization. Thus, it is intended that the genus include species that have been reclassified including but not limited to such microorganisms as *Erwinia herbicola*.

Preferably promoter cassettes are introduced into host cells by transformation. General methods for transformation are well known and reference is made to CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, vol. 1 eds. Ausubel et al., (1987) Chapter 7, John Wiley & Sons. Transformation techniques include electroporation, use of calcium chloride and rubidium chloride. (Maniatis et al., (1982) MOLECULAR CLONING : A LABORATORY

MANUAL, chapter 8. Cold Spring Harbor Laboratory, *supra*) and Potter, H (1988) *Anal. Biochem* 174:361 - 373).

Methods suitable for the growth and maintenance of host cells are also well known and reference is made to the MANUAL OF METHODS OF GENERAL BACTERIOLOGY, Eds. P. Gerhardt et al., America Society for Microbiology, Washington, DC (1994) and T.D. Brock in BIOTECHNOLOGY: A TEXTBOOK OF INDUSTRIAL MICROBIOLOGY 2Ed. (1989) Sinauer Associates, Sunderland, MA. Typically cells are grown at 35°C in appropriate media. Preferred growth media are common commercially prepared media such as Luria Bertani broth (LB) , Sabouraud Dextrose (SD) or other known growth media.

Transformed host cells comprising a precursor promoter or modified precursor promoter are selected based upon the phenotype response to a selectable marker, which was provide with the promoter cassette. In some embodiments of the invention the selective marker is excised from the transformed host cells. Reference is made to Figure 6B and Palmeros et al. *supra*. In specific embodiments the loxP site is left upstream of the promoter (for example in Figure 6B Ptrc. If the loxP site is left in the host it could become a problem if the marker excision process is repeatable. If two loxP sites are in the right orientation in the host chromosome, they may promote loop-out or inversion of large regions of the host chromosome (See Sauer, (1994) *Curr. Opin. Biotech* 5:521 – 527). To solve this problem, the present invention also discloses the construction of loxP-cat constructs that contain variants of the loxP site that do not recombine efficiently with the wild-type loxP site. It is known that the loxP wild-type and the loxP511 sites do not recombine with each other in a Cre-dependent manner (Hoess et al. (1986) *NAR*, 14:2287 – 2230). Other non-competitive loxP sites are known.

Transformed host cells having an optimum level of gene expression may be selected and further isolated. Optimization of gene expression in host cells is achieved by selecting transformed host cells having between about 1 to 250%, between about 5 to 200%, between about 10 to 150%, and between about 10 to 100% the strength or expression of the precursor promoter. That is about 1%, 5%, 10%, 15%, 20%, 25%, 40%, 50%, 60%, 80% 100%, 150%, 200% and 250% or more of the strength of the precursor promoter.

Promoter strength can be quantified using in vitro methods that measure the kinetics of binding of the RNA polymerase to a particular piece of DNA, and also allows the measurement of transcription initiation. Reference is made to Hawley D.K et al., Chapter 3: in: PROMOTERS: STRUCTURE AND FUNCTION IN R.L. Rodriguez and M.J. Chamberlin eds. Praeger Scientific. New York. In vivo methods may also be used to quantify promoter

strength. For example a promoter may be fused to a reporter gene and the efficiency of RNA synthesis measured. (Deuschle et al., (1986) *EMBO J.* 5: 2987-2994). The strength of *E. coli* promoters using different reporter genes was measured by Deuschle et al. and the data is presented in Table 1 above. Moreover, promoter strength may be defined in a number of ways. One common method is a relative one, wherein the mRNA or protein expressed by one gene per unit of time is compared to a control where the same gene is expressed by a different promoter. For example, the relative level of expression of the lacZ gene when it is transcribed by its native promoter (P_{lac}) or from the P_{lacUV5} promoter.

In an embodiment of the invention, promoter cassettes may be used in a method of modifying the regulatory function of a native promoter of a chromosomal gene of interest or by altering expression of a chromosomal gene of interest by transforming host cells with one or more promoter cassettes (a promoter library) and allowing homologous recombination between a promoter cassette and homologous flanking regions of a target site of the chromosomal gene of interest in the host cell, wherein the promoter cassette replaces a native promoter region of the chromosomal gene of interest.

Different promoter cassettes comprising modified precursor promoters derived from the same precursor promoter will produce transformed host cells having a varying range of gene expression levels. The promoter strength and/or gene expression may be determined by various known methods. Promoter strength or gene expression may be compared between different transformed host cells and a parent having a precursor promoter when cultured under essentially the same growth conditions. A transformed host having a desired level of expression may then be selected and isolated. The level of expression may be lower or higher than the expression of the same gene in a control parent having a native or precursor promoter.

Additionally, selected transformed host cells may be chosen from a promoter library wherein the expression of the gene of interest from the modified precursor promoter is between about 1 to 250%, between about 5 to 200%, between about 10 to 150%, and between about 10 to 100% the strength of the original precursor promoter.

Using a promoter library to create a population of bacterial cells having varying levels of expression of a gene of interest is particularly useful in a metabolic engineering pathway framework. Metabolic engineering is being used to optimize the metabolic pathways of strains to overproduce biomolecules. These biomolecules could be intermediates of cellular metabolism, polypeptides RNAs, carbohydrates, lipids and others.

Furthermore, the synthesis of complex molecules such as steroids, antibiotics, and other pharmaceuticals may require complicated and multiple catalytic pathways.

In an isolated system, each step in a particular metabolic pathway would need to be engineered. In contrast, the microorganism utilized in a whole cell system provides each of the required pathways. However, the use of certain promoters may incur problems in that a particular promoter may be too strong. As a result, the over expression of a particular gene may occur and be detrimental to a cell, for instance the gene may be expressed to the exclusion of other genes. The cell viability can thus be reduced and the production time may be limited.

The methods provided herein are utilized to provide a library of promoters to be introduced into bacterial host cells, which results in a population of transformed bacterial cells having a range of gene expression. The range of expression is useful because it allows the selection of specific bacterial clones having an optimum level of expression but still maintaining cell viability (e.g. the flux production of the desired end product relative the viability of the host cell in sustaining the desired level of production or sustaining the desired level of production). In certain embodiments the optimum level of expression of a gene will be high and in other embodiments the optimum level of gene expression will be low. A direct advantage of this method is that a bacterial clone may be selected based on the expression level obtained by the modified precursor promoter and then be ready for use in a fermentation process whereby cell viability is not negatively affected by expression of the gene of interest.

The following Examples are for illustrative purposes only and are not intended, nor should they be construed as limiting the invention in any manner. Those skilled in the art will appreciate that variations and modifications can be made without violating the spirit or scope of the invention.

EXAMPLES

EXAMPLE 1 - Construction of a Trc-loxP-Cat promoter cassette.

An excisable selectable marker was introduced upstream of the Trc promoter by the following method. The commercial plasmid pTrc99a (Pharmacia) was digested with the restriction enzymes Hind III and NcoI according to supplier instructions (New England Biolabs). The digested DNA was purified and then submitted to a fill-in reaction with T4 DNA polymerase as described by Maniatis et al. *supra.* The resulting blunt-end linear DNA was re-circularized according standard protocols (Sambrook et al., *supra*) and the resulting ligation mixture transformed into *E.coli* TOP-10 competent cells (Invitrogen). The cells were plated on LB-agar plates containing 50 micrograms/ml of carbenicillin. After 16 hrs of

incubation at 37 ^{0}C, a number of colonies appeared on the plate. Four of the colonies were chosen for further analysis.

Purified plasmid DNA was obtained from these colonies and subjected to restriction enzymes analysis. It was confirmed that the 4 colonies contained the same plasmid and that the DNA region between HindIII and Nco I was deleted. This plasmid was named pTrc1.

Plasmid pTrc1 contained only one recognition site for the restriction enzyme BspM1, located approximately 120 base pairs upstream of the –35 region of the Trc promoter. This location was selected to introduce the excisable selectable marker. pTrc1 was digested with the BspM1 enzyme according to the instructions of the supplier (New England Biolabs). The linear pTrc1 was gel-purified using a QIAquick gel extraction kit (QIAGEN), and submitted to a fill-in reaction with T4 DNA polymerase as described by Maniatis et al. *supra*. The resulting blunt-end linear DNA was ligated to a DNA construct including a chloramphenicol resistance gene (cat) flanked by loxP sites. This construct was obtained from plasmid pLoxCat2 (Palmeros et al., *Gene* 247 (2000) 255-264) by digestion with Ssp1 and Bam H1. The Ssp1–Bam H1 DNA fragment was gel purified and blunt ended. Linear pTRC1 and the Ssp1-BamHI fragments were ligated. The ligation mixture was transformed into *E.coli* TOP-10 competent cells (Invitrogen) and plated on LB-agar plates containing 50 micrograms/ml of carbenicillin and 20 micrograms/ml of chloramphenicol. After 16 hrs. of incubation at 37 ^{0}C, several colonies appeared on the plate. Some of these colonies were transferred to a fresh LB-plate containing Carbenicillin and Chloramphenicol. After plasmid purification and restriction enzyme analysis, two clones containing the loxP-cat cassette in both orientations were selected. These plasmids were named pTrCm1 and pTrCm2 (Figures 2 and 3).

An important consideration of the pLoxCat1 and 2 vectors described here, is that they still contain the lac operator that allows the binding of the LacI repressor and provide certain degree of regulation for the Trc promoter (Amann et al, (1988) *Gene*, 69 301-315.)

EXAMPLE 2. Construction of modified Trc promoters in the –35 box

There are numerous methods for DNA mutagenesis, the procedure described here is used to exemplify the process, but is not restricted to it. The QuickChange site-directed mutagenesis kit (Stratagene) was chosen to introduce a small number of modifications (mutations) in a defined region of DNA. This method is based on the use of polymerase chain reaction (PCR) to mutagenize a template (normally a plasmid) and the process requires two primers. After several PCR cycles many copies of the template are produced.

Each copy was primed by the mutagenic primers. The PCR reaction is then treated with the restriction enzyme Dpn1 that only cuts the original template (non-mutagenized) in many sites. The PCR-products are insensitive to Dpn1. After the Dpn1 treatment, the resulting DNA is used to transform *E.coli* competent cells. The recovered transformants normally are highly enriched for the mutants.

To change the second A (position –31) of the TTGACA sequence of the –35 region, 2 mutagenic primers were designed and synthesized by a commercial supplier (Operon technologies Inc.).

Primer A:

5'-CTGAATGAGCTGTTGACNATTAATCATCCGGCTCG-3' (SEQ ID NO. 35) and

Primer B:

5'-CGAGCCGGATGATTAATNCTCAACAGCTCATTTCAG 3' (SEQ ID NO. 36)

wherein "N" indicates T, G or C.

These two primers were used for the mutagenesis using the pTrCm2 plasmid (Figures 1 and 2) as a template, and the QuickChange kit, following the procedure recommended by the supplier. The QuickChange kit is provided with its own transformation protocol and competent cells (*E. coli* strain XL1 blue) that after transformation normally produce numerous transformants.

After transformation and plating on selective plates, plasmid DNA was purified from several clones and submitted for sequencing. In this manner modified precursor promoters were identified. i.e: promoters containing a –35 box with the sequence:

Clone NF-T (SEQ ID NO. 28) TTGACT= plasmid pTrcCm31T

Clone NF-C (SEQ ID NO. 30) TTGACC = plasmid pTrCm31C

Clone NF-G (SEQ ID NO. 29) TTGACG= plasmid pTrCm31G

Reference is made to Figure 8.

<u>EXAMPLE 3. - Construction of modified Trc promoters in the spacer region between the -10 and –35 boxes</u>.

To demonstrate that a small number of modifications are enough to provide a range of levels of promoter strength in the linker region, the normal spacing of the Trc promoter was increased 1 bp at the time. i.e., 1 or 2 bases were added using the QuickChange protocol described in Example 2 using the following primers:

Pair for pTrCm 18

5'-GACAATTAATTCATCCGGCTCG-3' (SEQ ID NO. 37)

5'-CGAGCCGGATGAATTAATTGTC-3 (SEQ ID NO. 38)

Pair for pTrCm19

5'-GACAATTAATTTCATCCGGCTCG-3' (SEQ ID NO. 39)

5'-CGAGCCGGATGAAATTAATTGTC (SEQ ID NO. 40)

The modified (mutant) plasmids were named as follow:

Clone NF-1T (SEQ ID NO 32) pTrCm18: spacer length =18

Clone NF-2T (SEQ ID NO. 33) pTrCm19: spacer length =19

EXAMPLE 4. - Use of pTrCm2 to replace chromosomal regulatory regions.

Replacement of the lacZ promoter

The method disclosed in Datsenko and Wanner (*Proc,. Natl. Acad. Sci.* USA, 10: 6640-6645, (2000)) was utilized to replace the native regulatory regions of the lacZ gene with Ptrc and modified Ptrc promoters This method utilizes 30 – 50 nucleotides as regions of homology to promote homologous recombination between PCR products and the *E. coli* host cell chromosome. Plasmid TrCm2 and its derivatives were used as templates for the PCR reactions,

Primers

lacZ1 (SEQ ID NO. 41) - 5' – AGCGCAACGCAATTAATGTGAGTTAGCTCAC TCATTAGGGATGCATATGGCGGCCGCA -3'

and

lacZ2 (SEQ ID NO. 42) – 5' GTCACGACGTTGTAAAACGACGGCCAGTGAATCCGTAATCATGGTCTGTTTCCTGTGT GAAA – 3'

were designed to contain 20 nucleotides complementary to the pTrcCm2 and 39 nucleotides complementary to the lacZ gene regulatory region (Figure 7). Using these primers, a 1333bp DNA fragment was generated by PCR.

E. coli strain, MG1655 was transformed with plasmid pKD46 as recommended by Datsenko and Wanner (*supra*). The resulting strain MG1655/pKD46 was used to prepare competent cells according to the method described in Datsenko and Wanner, supra. Competent cells (100 µl) were transformed by electroporation with 20 to 100 ng of the 1333 bp PCR product described above. After recovering the cells for 1 hr in 1.0ml SOC media

(sterile 10ml of1M $MgCl_2$ and sterile 20 ml of 1M glucose is added to autoclaved 20g tryptone, 5g yeast extract, 0.5g NaCl, and 2.5 ml 1M KCl), they were plated on 4 LA plates, containing 10µg/ml of chloramphenicol. Plates were incubated at 37C for at least 16 hrs. Cm^R colonies were transferred to fresh LA plates containing 10 µg/ml chloramphenicol. To verify that the native regulatory region of the lacZ gene has been modified chromosomal DNA from the MG1655 and some of the Cm^R transformants were purified using the UltraClean microbial DNA isolation kit (MO BIO Labs, Solana Beach, CA). These DNAs were used as substrates for PCR reactions using primers.

LacT1 (SEQ ID NO. 43)

5' GGCACGACAGGTTTCCCGAC – 3'

and

lacT2 (SEQ ID NO. 44)

5' GAGGGGACGACGACAGTATC 3'

These two primers hybridize with regions outside of where lacZ1 and lacZ2 hybridize and should generate PCR products of the following sizes: MG1655 – 425bp and MG1655:: Ptrc-Cm-lacz-1585 bp (and 1586 bp and 1587 bp for the pTrCm18 and pTrCm19 cassettes.

The PCR products were separated in a 2% agarose gel. Based on the results of the gel colonies with proper modifications to the lacZ regulatory region were identified.

The proper integration of the cassettes can be further corroborated by sequencing the PCR products.

Furthermore, plasmids pTrcCm31T, pTrcCm31C and pTrcCm31G described in example 2, can be also used with the same purpose. It is expected that the cassettes of pTrCm1 and pTrCm2 will provide higher levels of expression than the cassettes from pTrcCm31T, pTrcCm31C or pTrcCm31G. It is also expected that the pTrCm18 and pTrCm19 promoters are weaker than the wt trc promoter.

EXAMPLE 5. - Measurement of the expression of the lacZ gene.

The lacZ codes for the β-galactosidase enzyme in *E. coli* and this gene has been widely used as a reporter to quantify gene expression. The β-galactosidase activity is measured using the synthetic substrate ONPG (ortho-Nitrophenyl- β-D-galactoside) according to the procedure described by Miller, J.H., (A SHORT COURSE IN BACTERIAL GENETICS. Cold Spring Harbor Laboratory Press,1992), To quantify the level of expression of the *lacZ* gene in the strains described in example 3, the strains were grown overnight in Luria broth (LB) media. These overnights were used to inoculate 250 ml flask containing 50

ml of LB or LB+ 100uM IPTG. As a reference point, strain MG1655 was also inoculated. In the case of MG1655, lacZ will be expressed from its native promoter.

Flasks were incubated in a shaker at 37 ^{0}C until they reached early -exponential phase (~0.8 OD at 600 nm) and 1.5 samples were collected by centrifugation. The cells were resuspended in 1 ml of buffer Z and the β-galactosidase activity assayed as described by Miller J,H, (A SHORT COURSE IN BACTERIAL GENETICS. Cold Spring Harbor Laboratory Press,1992). After correcting for the volumes utilized in the assay, the relative activity of β-galactosidase per unit of optical density (600 nm) was calculated. The results of these measurements are presented in Table 2.

TABLE 2

Promoter controlling LacZ expression	SEQ ID NO.	LB Relative β-galactosidase activity	LB + 100 µM IPTG Relative β-galactosidase activity
wt lac	22	-	1
Trc	27	1.46	4.78
TrCm31T	28	0.81	4.13
TrCm31G	29	1.47	5.01
TrCm31C	30	0.50	3.82
TrCm18	32	0.82	2.23
TrCm19	33	0.038	0.40

Table 2. Relative β-galactosidase activity measured in strains containing the lacZ gene under the control of different promoters.

<u>EXAMPLE 6. - Construction of a LoxP-Cat cassette containing a LoxP variant.</u>

Sequences of the wild type and loxP511 are shown below. The 2 loxP sites differ in only one base pair.

LoxP wild type:

5' ATAACTTCGTATA.ATGTATGC.TATACGAAGTTAT 3' (SEQ ID NO. 45) and

LoxP511

5' ATAACTTCGTATA.ATGTAT**<u>A</u>**C.TATACGAAGTTAT 3' (SEQ ID NO. 46).

To construct the LoxP511-Cat cassette, two 2 primers containing the mutated base were used to amplify the loxP-Cat cassette by PCR. This approach (and others) can be used to construct any variant of the LoxP sequence.

The sequences of the PCR primers are:

LoxF1: 5'-GCTGGATCCATAACTTCGTATAATGTATACTATACG-3' (SEQ ID NO. 47)

and

LoxF2: 5'-GCATATGGCGGCCGCATAACTTCGTATAGTATACATT-3' (SEQ ID NO. 48).

PCR was performed and the PCR product was cloned in the vector pCR-Blunt II TOPO, following the instructions provided with the vector-kit (Invitrogen). E. coli cells were transformation and many colonies were obtained. After plasmid purification and restriction analysis, 3 colonies with the correct restriction pattern, were submitted for DNA sequencing and it was found that one plasmid presented the correct LoxP511 sequence. This plasmid was named pLoxCat27. In one embodiment this cassette would be used with a modified precursor promoter or precursor promoter according to the invention when a promoter cassette is being introduced into a bacterial host strain that may already have a LoxP site.

Those skilled in the art will recognize or be able to ascertain using not more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

CLAIMS

1. A method of creating a library of bacterial cells having a range of expression levels of a chromosomal gene of interest comprising,

a) obtaining a promoter library comprising at least two promoter cassettes, wherein the promoter cassettes comprise in sequential order a 5' sequence homologous to an upstream flanking region of a target site; a first recombinase recognition site; a selectable marker; a second recombinase recognition site; a precursor promoter or a modified precursor promoter comprising a –35 consensus region, a linker sequence and a –10 consensus region, wherein the modified precursor promoter includes at least one nucleotide position that has been modified from the precursor promoter; and a 3' sequence homologous to a downstream flanking region of the target site.

b) transforming bacterial host cells with the promoter library, wherein the promoter cassettes are integrated into the bacterial host cells by homologous recombination to produce transformed host cells;

c) culturing the transformed host cells under suitable growth conditions; and

d) obtaining a library of transformed bacterial cells, wherein the transformed bacterial cells exhibit a range of expression levels of a chromosomal gene of interest.

2. The method according to claim 1 further comprising selecting transformed bacterial cells from the library.

3. The method according to claim 1, wherein the host cells are selected from the group consisting of *E. coli*, *Bacillus* sp. and *Pantoea* sp.

4. The method according to claim 2, wherein the selected bacterial cells have a higher level of expression of the gene of interest than bacterial cells comprising the precursor promoter.

5. The method according to claim 2, wherein the selected bacterial cells have a lower level of expression of the gene of interest than the bacterial cells comprising the precursor promoter.

6. Transformed bacterial cells selected according to the method of claim 2.

7. The method according to claim 1, wherein the promoter library comprises the Ptrc precursor promoter and modified Ptrc precursor promoters.

8. The method according to claim 1, wherein the promoter library comprises the Ptac precursor promoter and modified Ptrc precursor promoters.

9. The method according to claim 1, wherein the promoter library comprises the P_{GI} precursor promoter and modified P_{GI} precursor promoters.

10. The method according to claim 1, wherein the promoter library comprises modified promoters having SEQ ID NO. 28, SEQ ID NO. 29 and SEQ ID NO. 30.

11. A promoter cassette comprising in sequential order

a) a 5' sequence homologous to an upstream flanking region of a target site;

b) a first recombinase recognition site;

c) a selectable marker;

d) a second recombinase recognition site;

e) a modified precursor promoter comprising at least one modified nucleotide in a position corresponding to a –35 consensus region, a linker sequence or a –10 consensus region of a precursor promoter; and

f) a 3' sequence homologous to a downstream flanking region of the target site.

12. The promoter cassette of claim 11, wherein the precursor promoter is selected from the group consisting of P_{trc}, P_{tacI}, $P_{D/E20}$, P_{H207}, P_{N25}, P_{G25}, P_{J5}, P_{A1}, P_{A2}, P_{A3}, P_L, P_{lac}, P_{lacUV5}, P_{con}, and P_{bla},

13. The promoter cassette of claim 11, wherein the –35 region of the precursor promoter is selected from the group consisting of TTGACA, TTGCTA, TTGCTT, TTGATA, TTGACT, TTTACA and TTCAAA.

14. The promoter cassette of claim 11, wherein the –10 region of the precursor promoter is selected from the group consisting of TAAGAT, TATAAT, AATAAT, TATACT, GATACT, TACGAT, TATGTT and GACAAT.

15. The promoter cassette of claim 11, wherein the –35 region of the precursor promoter is TTGACA and the –10 region of the precursor promoter is TATAAT.

16. The promoter cassette of claim 11, wherein the –35 region of the precursor promoter is TTGACA and the –10 region of the precursor promoter is AATAAT.

17. The promoter cassette of claim 11, wherein the linker sequence of the precursor promoter is modified.

18. The promoter cassette of claim 11, wherein said first and said second recombinase recognition sites are non-identical recombinase sites and selected from lox and mutant lox sites.

19. The promoter cassette of claim 11, wherein the modified precursor promoter is selected from the group consisting of SEQ ID NO. 28 (NF-T), SEQ ID NO. 29 (NF-G), SEQ ID NO. 30 (NF-C), SEQ ID NO. 32 (NF-1T) and SEQ ID NO. 33 (NF-2T).

20. A promoter library comprising at least two promoter cassettes of claim 11.

21. A promoter library comprising at least two promoter cassettes of claim 13.

22. A vector comprising the promoter cassette of claim 11.

23. A host cell transformed with a promoter cassette of claim 11.

24. The host cell of claim 23, wherein the host cell is selected from the group consisting of *E. coli*, *Bacillus sp.* and *Pantoea sp.*

25. A method of modifying the regulatory function of a native promoter of a chromosomal gene of interest comprising,

a) obtaining a promoter cassette according to claim 11;

b) transforming a host cell with the promoter cassette to allow homologous recombination between the promoter cassette and homologous flanking regions of a target site, wherein the promoter cassette replaces a native promoter region of a chromosomal gene of interest; and

c) culturing the transformed host cells under suitable growth conditions.

26. The method according to claim 25 further comprising, excising the selectable marker from the transformed host cell.

27. The method according to claim 25, further comprising, isolating the transformed host cells.

28. A method for altering the expression of a chromosomal gene of interest comprising,

a) obtaining a promoter cassette according to claim 11;

b) transforming a host cell with the promoter cassette; and

c) allowing homologous recombination between the promoter cassette and homologous flanking regions of the target site, wherein the promoter cassette replaces a native promoter region of a chromosomal gene of interest and alters the expression of the chromosomal gene of interest as compared to the expression of the chromosomal gene of interest in a corresponding parent host cell.

29. An isolated promoter comprising the sequence set forth in SEQ ID NO. 28.

30. An isolated promoter comprising the sequence set forth in SEQ ID NO. 29.

31. An isolated promoter comprising the sequence set forth in SEQ ID NO. 30.

32. An isolated promoter comprising the sequence set forth in SEQ ID NO. 32.

33. An isolated promoter comprising the sequence set forth in SEQ ID NO. 33.

34. An isolated promoter comprising the sequence set forth in SEQ ID NO. 31.

SUBSTITUTE SHEET (RULE 26)

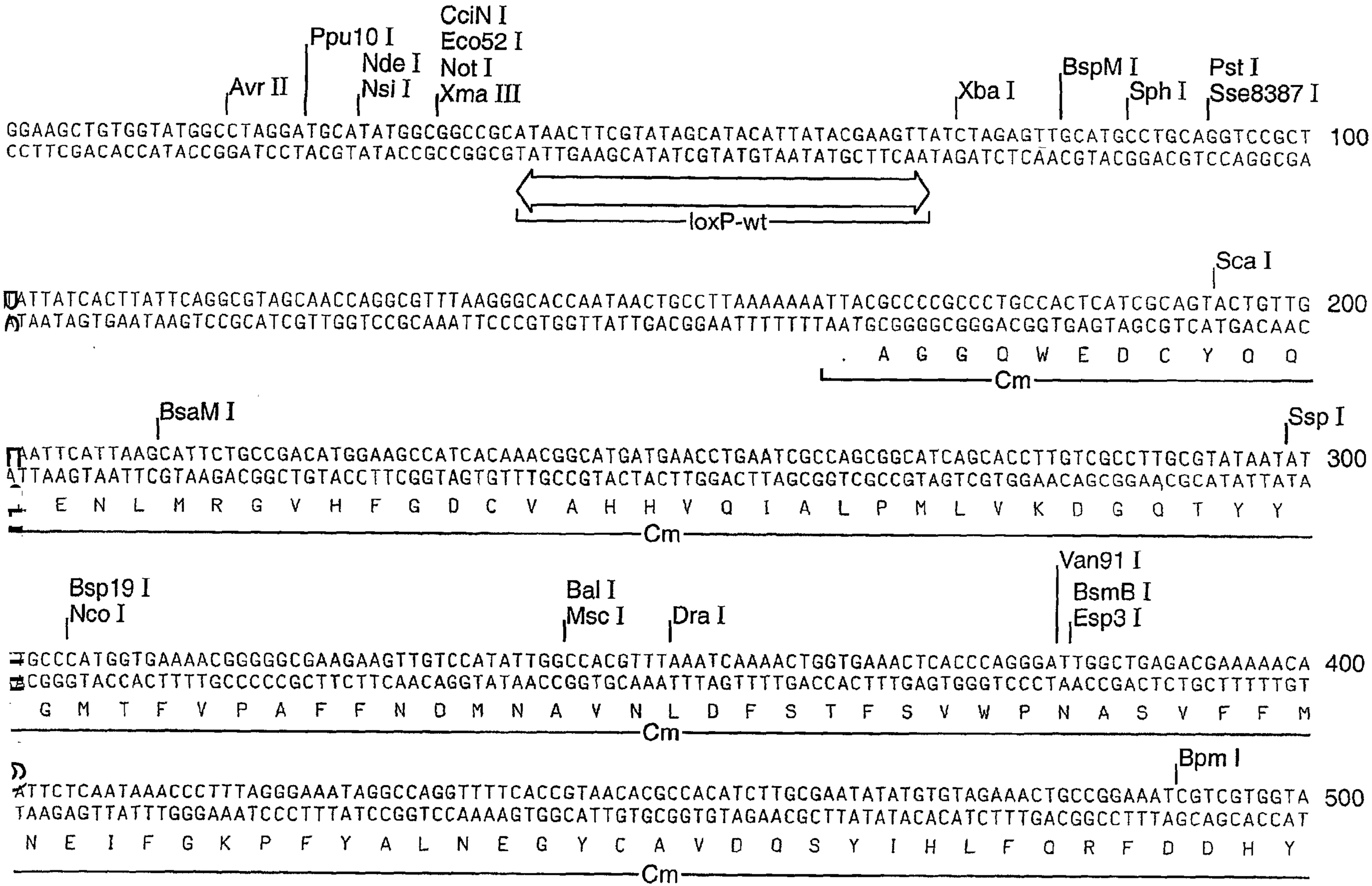


FIG._1A

Psp1406 I
BseM I
BsrD I

TTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAGGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCC 600
AAGTGAGGTCTCGCTACTTTTGCAAAGTCAAACGAGTACCTTTTGCCACATTGTTCCCACTTGTGATAGGGTATAGTGGTCGAGTGGCAGAAAGTAACGG
E S W L S S F T E T Q E H F V T Y C P H V S D W I V L E G D K M A
Cm

EcoR I
Acc II
BsaM I
Dra I

GTACGGAATTCCGGATGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTTTCTTTACGGTCTTTAAAAAGGCCG 700
CATGCCTTAAGGCCTACTCGTAAGTAGTCCGCCCGTTCTTACACTTATTTCCGGCCTATTTTGAACACGAATAAAAAGAAATGCCAGAAATTTTTCCGGC
R F E P H A N M L R A L I H I F A P Y F K H K N K K V T K L F A T
Cm

Pvu II

GAATATCCAGCTGAACGGTCTGGTTATAGGTACATTGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGGTGGT 800
CTTATAGGTCGACTTGCCAGACCAATATCCATGTAACTCGTTGACTGACTTTACGGAGTTTTACAAGAAATGCTACGGTAACCCTATATAGTTGCCACCA
I D L Q V T Q N Y T C Q A V S Q F A E F H E K R H W Q S I D V T T
Cm

Bpu10 I

TATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGATAACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGA 900
ATAGGTCACTAAAAAAAGAGGTAAAATCGAAGGAATCGAGGACTTTTAGAGCTATTGAGTTTTTTATGCGGGCCATCACTAGAATAAAGTAATACCACT
G T I K K E M
Cm
SD

BsmB I
Esp3 I
Van91 I

AGTTGGAACCTCTTACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACAGGGACACCAGGATTTATTTATTCTG 1000
TCAACCTTGGAGAATGCACGGCTAGTTGCAGAGTAAAAGCGGTTTTCAACCGGGTCCCGAAGGGCCATAGTTGTCCCTGTGGTCCTAAATAAATAAGAC

FIG._1B

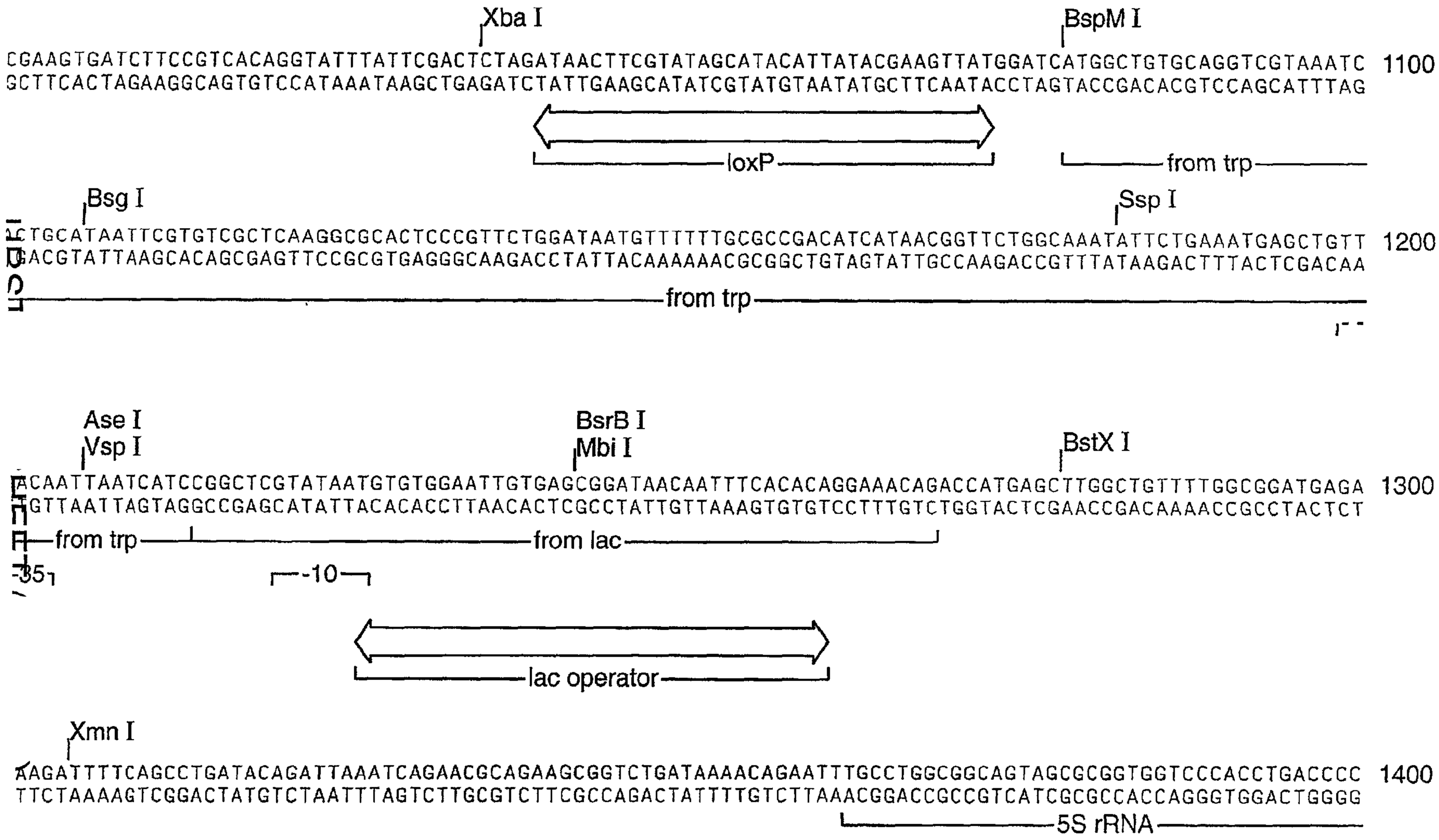

Xba I
BspM I
CGAAGTGATCTTCCGTCACAGGTATTTATTCGACTCTAGATAACTTCGTATAGCATACATTATACGAAGTTATGGATCATGGCTGTGCAGGTCGTAAATC 1100
GCTTCACTAGAAGGCAGTGTCCATAAATAAGCTGAGATCTATTGAAGCATATCGTATGTAATATGCTTCAATACCTAGTACCGACACGTCCAGCATTTAG
loxP
from trp
Bsg I
Ssp I
CTGCATAATTCGTGTCGCTCAAGGCGCACTCCCGTTCTGGATAATGTTTTTTGCGCCGACATCATAACGGTTCTGGCAAATATTCTGAAATGAGCTGTT 1200
GACGTATTAAGCACAGCGAGTTCCGCGTGAGGGCAAGACCTATTACAAAAAACGCGGCTGTAGTATTGCCAAGACCGTTTATAAGACTTTACTCGACAA
from trp
Ase I
Vsp I
BsrB I
Mbi I
BstX I
ACAATTAATCATCCGGCTCGTATAATGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGACCATGAGCTTGGCTGTTTTGGCGGATGAGA 1300
TGTTAATTAGTAGGCCGAGCATATTACACACCTTAACACTCGCCTATTGTTAAAGTGTGTCCTTTGTCTGGTACTCGAACCGACAAAACCGCCTACTCT
from trp
from lac
-35
-10
lac operator
Xmn I
AAGATTTTCAGCCTGATACAGATTAAATCAGAACGCAGAAGCGGTCTGATAAAACAGAATTTGCCTGGCGGCAGTAGCGCGGTGGTCCCACCTGACCCC 1400
TTCTAAAAGTCGGACTATGTCTAATTTAGTCTTGCGTCTTCGCCAGACTATTTTGTCTTAAACGGACCGCCGTCATCGCGCCACCAGGGTGGACTGGGG
5S rRNA

FIG._1C

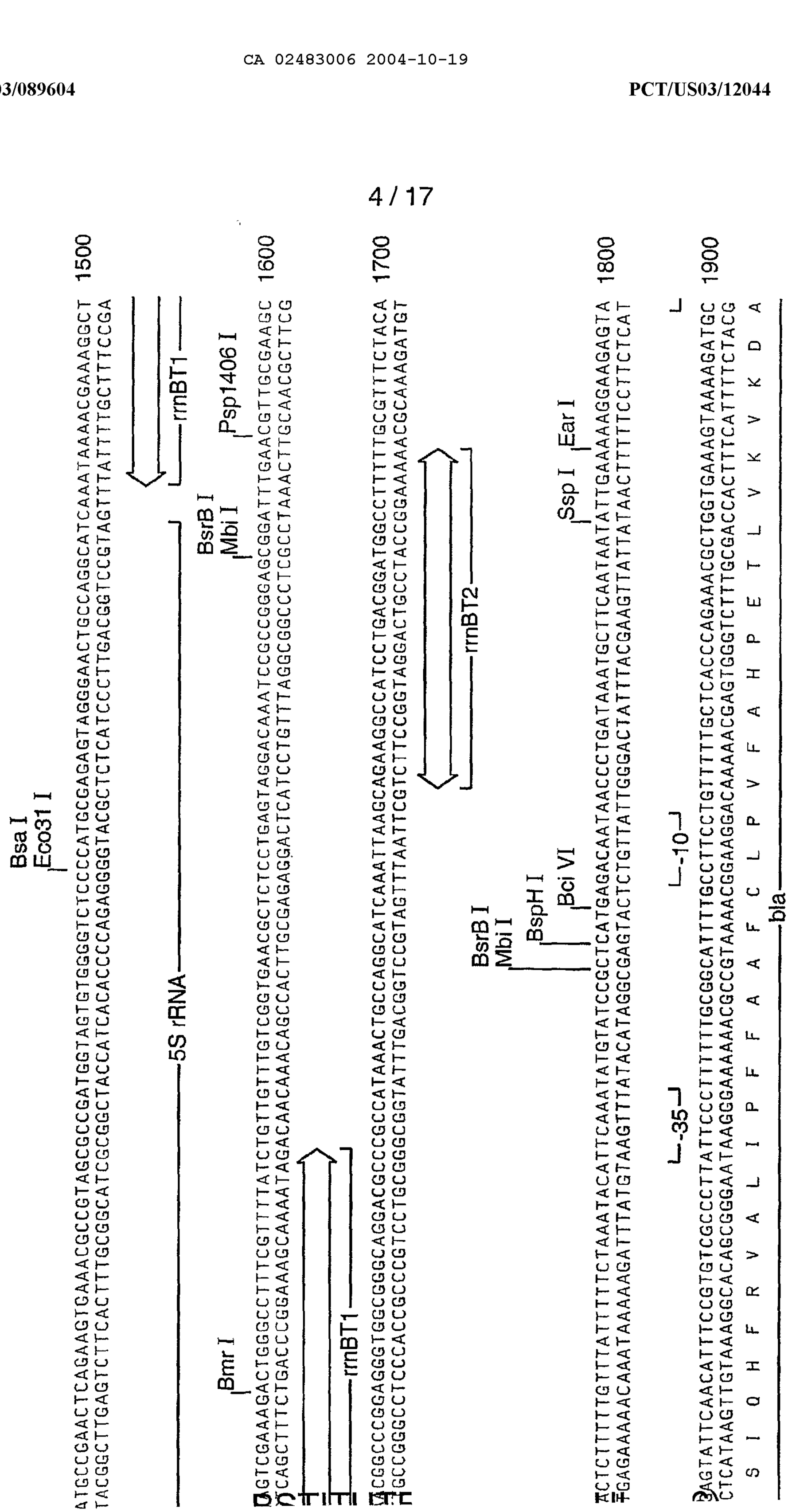

Bsa I
Eco31 I
5S rRNA
BsrB I
Mbi I
Psp1406 I
rrnBT1
rrnBT2
Ssp I
Ear I
BspH I
Bci VI
-35
-10
bla
Bmr I
M A L I P F F A A F C L P V
V F A H P E T L V K V K D A
S I Q H F R V A L P S L T N
L
K
P
Q
E
S

FIG._1D

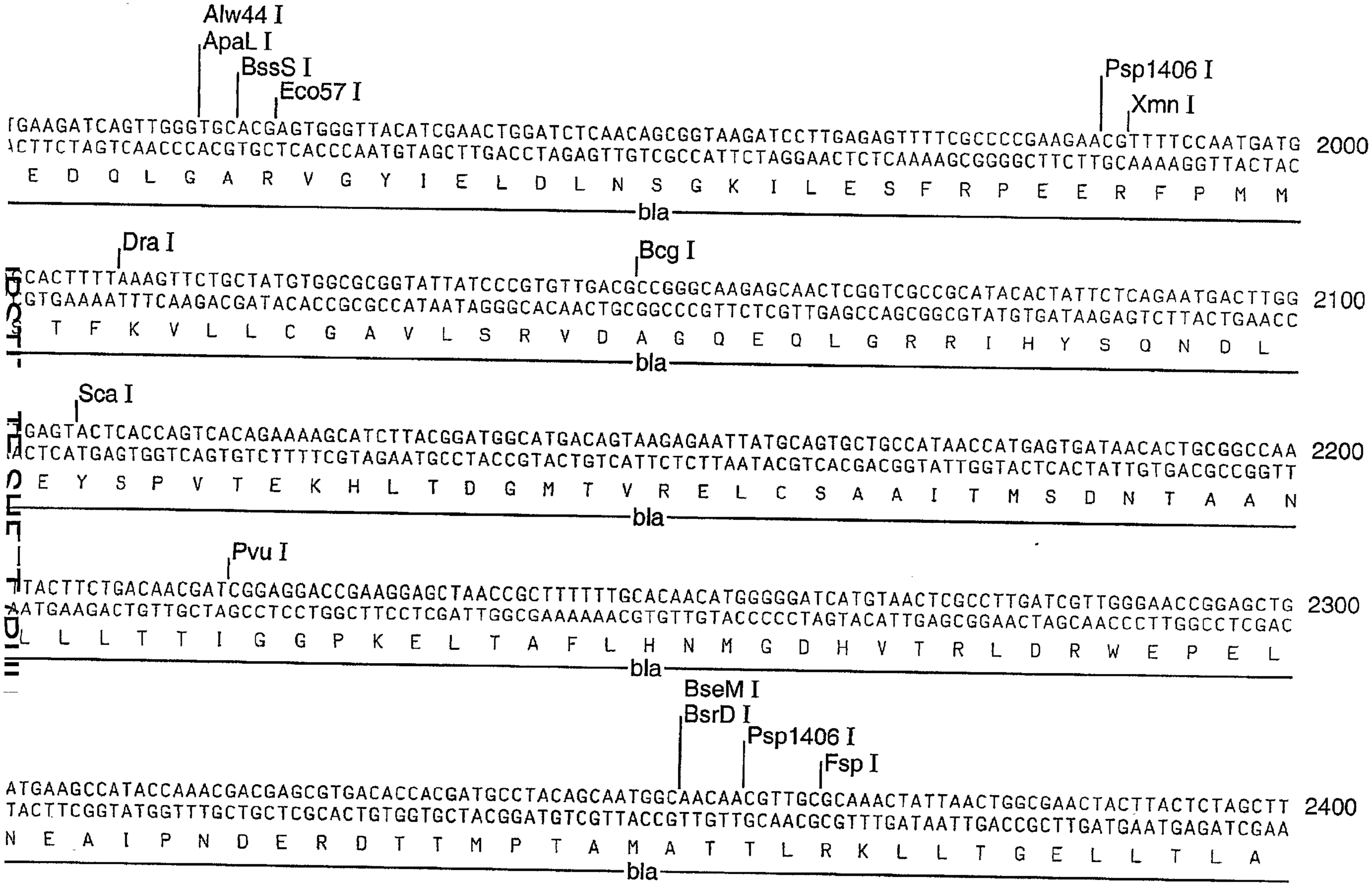

Alw44 I
ApaL I
BssS I
Eco57 I
Psp1406 I
Xmn I
TGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATG 2000
ACTTCTAGTCAACCCACGTGCTCACCCAATGTAGCTTGACCTAGAGTTGTCGCCATTCTAGGAACTCTCAAAAGCGGGGCTTCTTGCAAAAGGTTACTAC
E D Q L G A R V G Y I E L D L N S G K I L E S F R P E E R F P M M
bla
Dra I
Bcg I
CCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTGTTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGG 2100
GGTGAAAATTTCAAGACGATACACCGCGCCATAATAGGGCACAACTGCGGCCCGTTCTCGTTGAGCCAGCGGCGTATGTGATAAGAGTCTTACTGAACC
T F K V L L C G A V L S R V D A G Q E Q L G R R I H Y S Q N D L
bla
Sca I
TGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAA 2200
ACTCATGAGTGGTCAGTGTCTTTTCGTAGAATGCCTACCGTACTGTCATTCTCTTAATACGTCACGACGGTATTGGTACTCACTATTGTGACGCCGGTT
E Y S P V T E K H L T D G M T V R E L C S A A I T M S D N T A A N
bla
Pvu I
TACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTG 2300
AATGAAGACTGTTGCTAGCCTCCTGGCTTCCTCGATTGGCGAAAAAACGTGTTGTACCCCCTAGTACATTGAGCGGAACTAGCAACCCTTGGCCTCGAC
L L L T T I G G P K E L T A F L H N M G D H V T R L D R W E P E L
bla
BseM I
BsrD I
Psp1406 I
Fsp I
ATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTACAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTT 2400
TACTTCGGTATGGTTTGCTGCTCGCACTGTGGTGCTACGGATGTCGTTACCGTTGTTGCAACGCGTTTGATAATTGACCGCTTGATGAATGAGATCGAA
N E A I P N D E R D T T M P T A M A T T L R K L L T G E L L T L A
bla

FIG._1E

SUBSTITUTE SHEET (RULE 26)

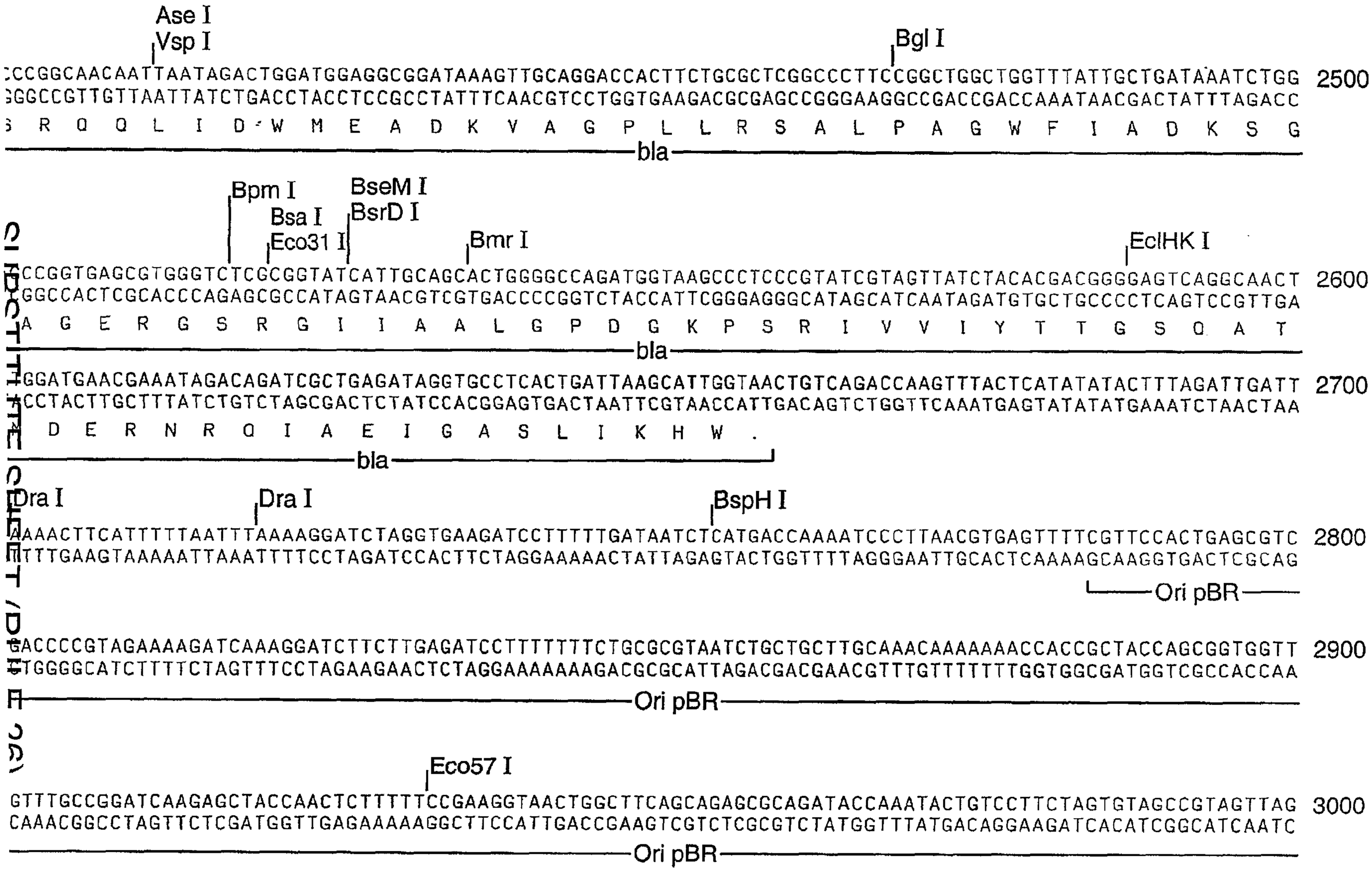


FIG._1F

AlwN I

GCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGG 3100
CGGTGGTGAAGTTCTTGAGACATCGTGGCGGATGTATGGAGCGAGACGATTAGGACAATGGTCACCGACGACGGTCACCGCTATTCAGCACAGAATGGCC

Ori pBR

Alw44 I
ApaL I

TTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAA 3200
AACCTGAGTTCTGCTATCAATGGCCTATTCCGCGTCGCCAGCCCGACTTGCCCCCCAAGCACGTGTGTCGGGTCGAACCTCGCTTGCTGGATGTGGCTT

Ori pBR

Bci VI

GGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGC 3300
CCTCTATGGATGTCGCACTCGATACTCTTTCGCGGTGCGAAGGGCTTCCCTCTTTCCGCCTGTCCATAGGCCATTCGCCGTCCCAGCCTTGTCCTCTCG

Ori pBR

BssS I

Drd I

CACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGG 3400
GTGCTCCCTCGAAGGTCCCCCTTTGCGGACCATAGAAATATCAGGACAGCCCAAAGCGGTGGAGACTGAACTCGCAGCTAAAAACACTACGAGCAGTCC

Ori pBR

BspLU11 I

GGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCT 3500
CCCGCCTCGGATACCTTTTTGCGGTCGTTGCGCCGGAAAAATGCCAAGGACCGGAAAACGACCGGAAAACGAGTGTACAAGAAAGGACGCAATAGGGGA

Ori pBR

BsrB I
Mbi I

Ear I
Sap I

ATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAG 3600
TAAGACACCTATTGGCATAATGGCGGAAACTCACTCGACTATGGCGAGCGGCGTCGGCTTGCTGGCTCGCGTCGCTCAGTCACTCGCTCCTTCGCCTTC

FIG._1G

SUBSTITUTE SHEET (RULE 26)

SUBSTITUTE SHEET (RULE 26)

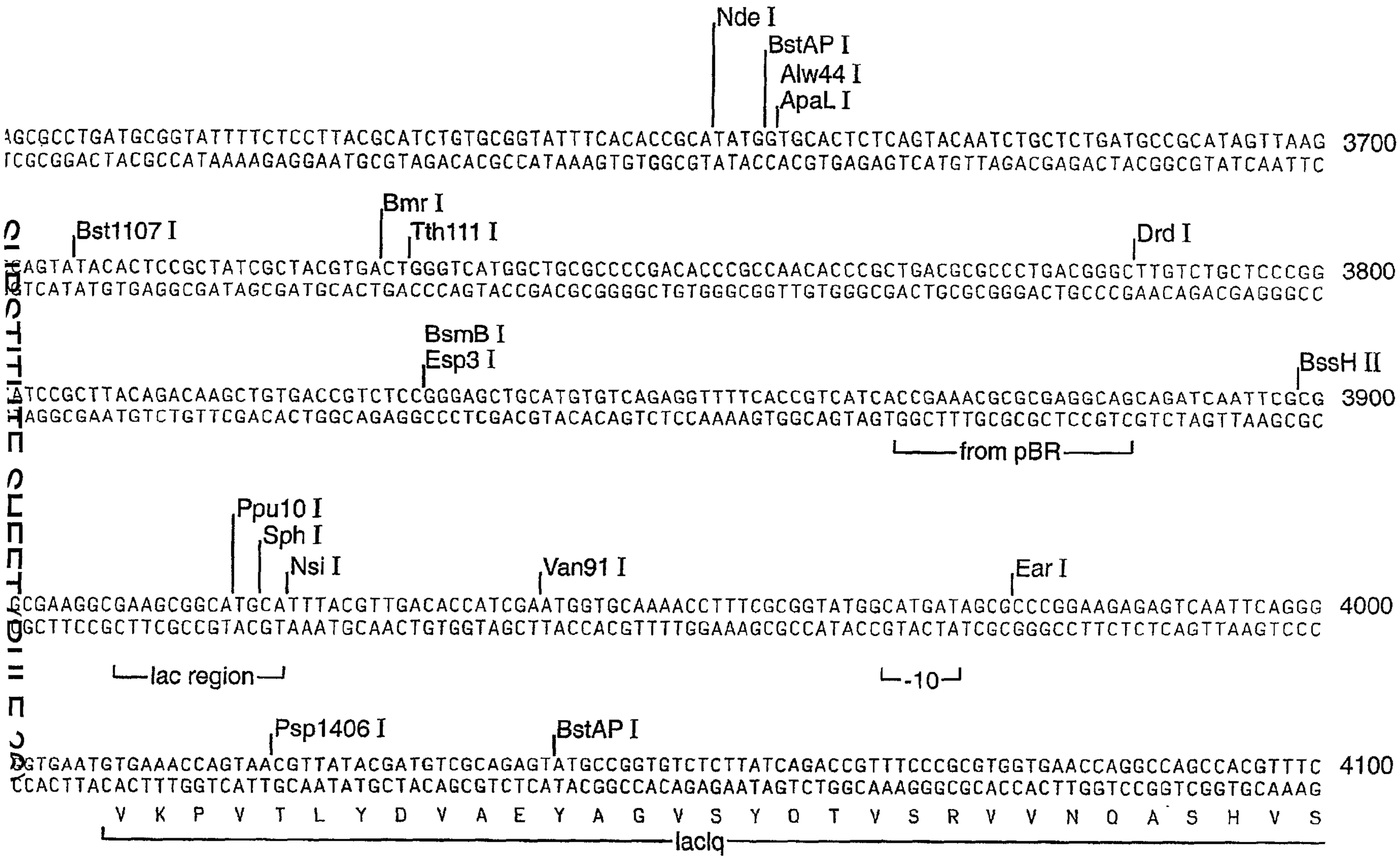


FIG._1H

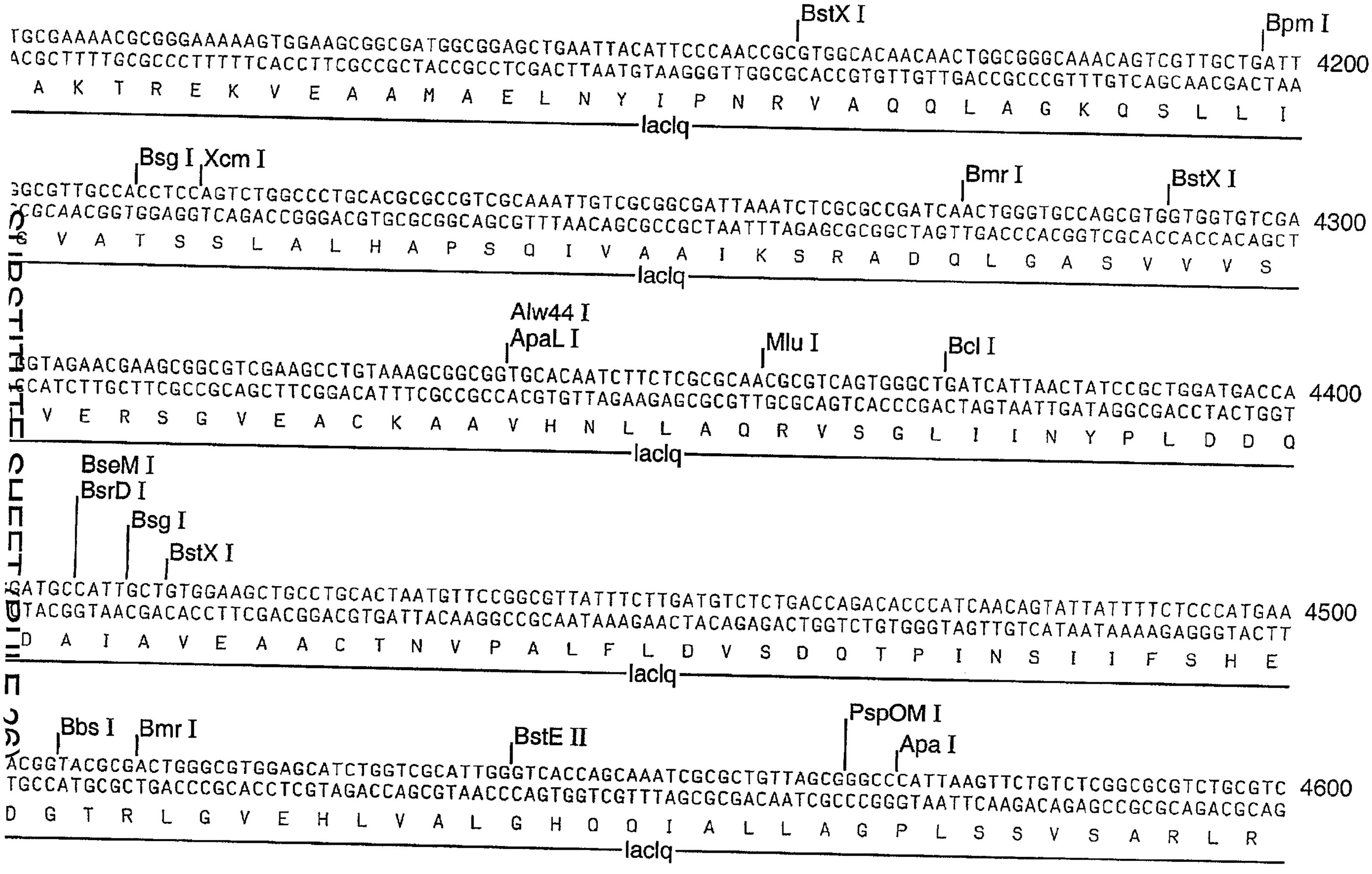

BstX I
Bpm I
TGCGAAAACGCGGGAAAAAGTGGAAGCGGCGATGGCGGAGCTGAATTACATTCCCAACCGCGTGGCACAACAACTGGCGGGCAAACAGTCGTTGCTGATT 4200
ACGCTTTTGCGCCCTTTTTCACCTTCGCCGCTACCGCCTCGACTTAATGTAAGGGTTGGCGCACCGTGTTGTTGACCGCCCGTTTGTCAGCAACGACTAA
A K T R E K V E A A M A E L N Y I P N R V A Q Q L A G K Q S L L I
lacIq
Bsg I
Xcm I
Bmr I
BstX I
GGCGTTGCCACCTCCAGTCTGGCCCTGCACGCGCCGTCGCAAATTGTCGCGGCGATTAAATCTCGCGCCGATCAACTGGGTGCCAGCGTGGTGGTGTCGA 4300
CCGCAACGGTGGAGGTCAGACCGGGACGTGCGCGGCAGCGTTTAACAGCGCCGCTAATTTAGAGCGCGGCTAGTTGACCCACGGTCGCACCACCACAGCT
G V A T S S L A L H A P S Q I V A A I K S R A D Q L G A S V V V S
lacIq
Alw44 I
ApaL I
Mlu I
BclI
GGTAGAACGAAGCGGCGTCGAAGCCTGTAAAGCGGCGGTGCACAATCTTCTCGCGCAACGCGTCAGTGGGCTGATCATTAACTATCCGCTGGATGACCA 4400
CCATCTTGCTTCGCCGCAGCTTCGGACATTTCGCCGCCACGTGTTAGAAGAGCGCGTTGCGCAGTCACCCGACTAGTAATTGATAGGCGACCTACTGGT
V E R S G V E A C K A A V H N L L A Q R V S G L I I N Y P L D D Q
lacIq
BseM I
BsrD I
Bsg I
BstX I
GATGCCATTGCTGTGGAAGCTGCCTGCACTAATGTTCCGGCGTTATTTCTTGATGTCTCTGACCAGACACCCATCAACAGTATTATTTTCTCCCATGAA 4500
CTACGGTAACGACACCTTCGACGGACGTGATTACAAGGCCGCAATAAAGAACTACAGAGACTGGTCTGTGGGTAGTTGTCATAATAAAAGAGGGTACTT
D A I A V E A A C T N V P A L F L D V S D Q T P I N S I I F S H E
lacIq
Bbs I
Bmr I
BstE II
PspOM I
Apa I
ACGGTACGCGACTGGGCGTGGAGCATCTGGTCGCATTGGGTCACCAGCAAATCGCGCTGTTAGCGGGCCCATTAAGTTCTGTCTCGGCGCGTCTGCGTC 4600
TGCCATGCGCTGACCCGCACCTCGTAGACCAGCGTAACCCAGTGGTCGTTTAGCGCGACAATCGCCCGGGTAATTCAAGACAGAGCCGCGCAGACGCAG
D G T R L G V E H L V A L G H Q Q I A L L A G P L S S V S A R L R
lacIq

FIG._1I

SUBSTITUTE SHEET (RULE 26)

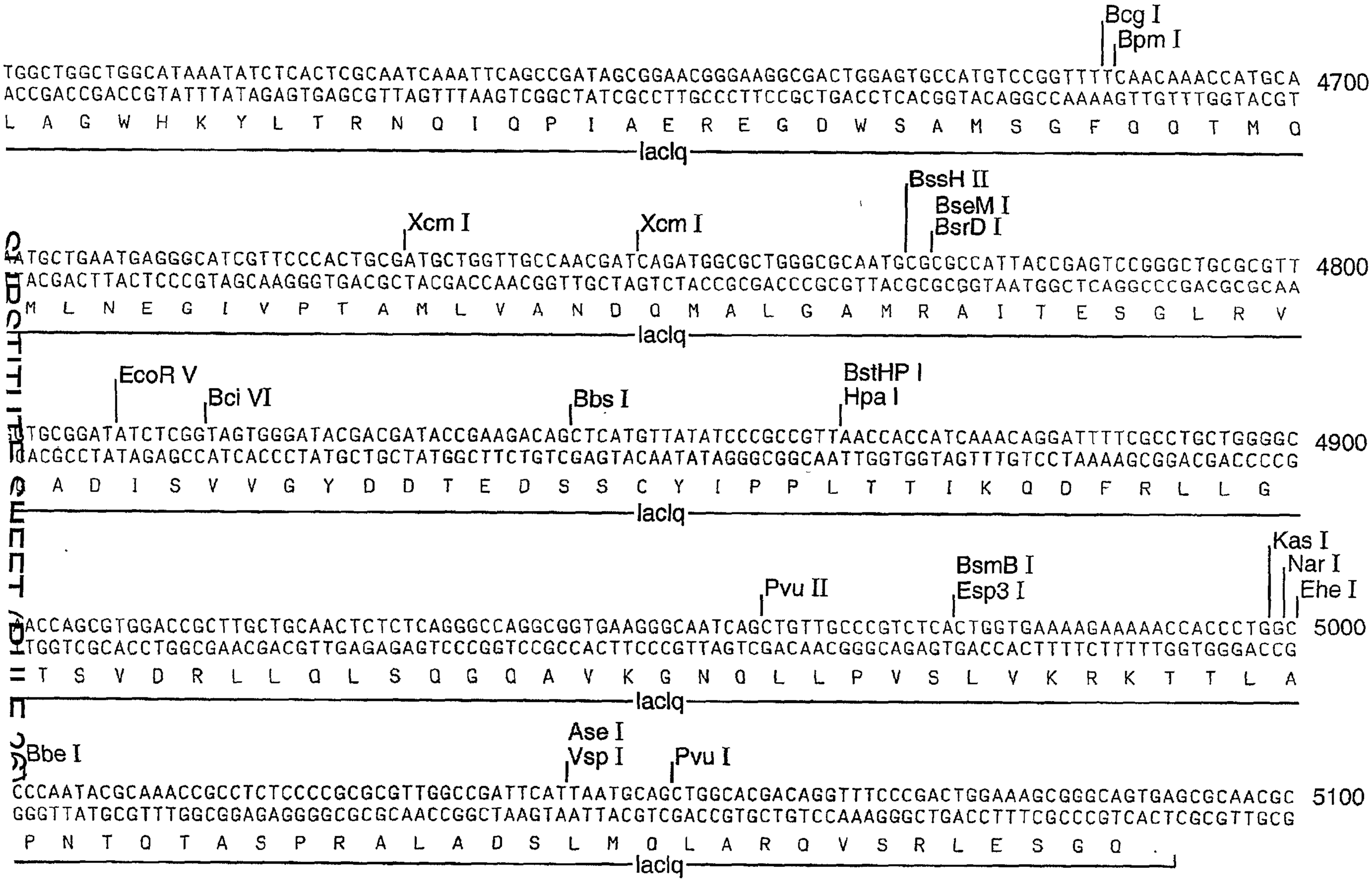


FIG._1J

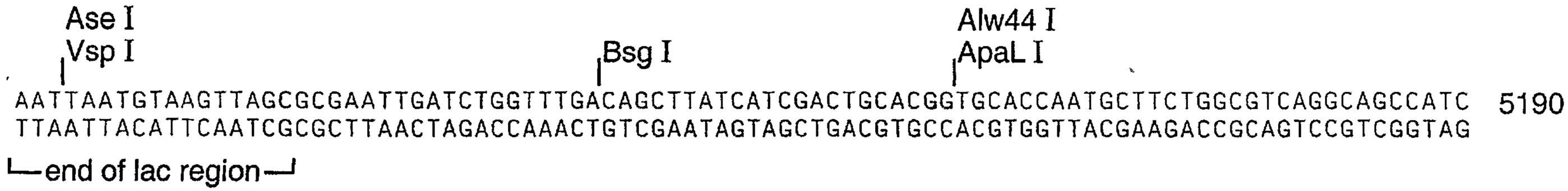

Ase I
Vsp I
Bsg I
Alw44 I
ApaL I
AATTAATGTAAGTTAGCGCGAATTGATCTGGTTTGACAGCTTATCATCGACTGCACGGTGCACCAATGCTTCTGGCGTCAGGCAGCCATC 5190
TTAATTACATTCAATCGCGCTTAACTAGACCAAACTGTCGAATAGTAGCTGACGTGCCACGTGGTTACGAAGACCGCAGTCCGTCGGTAG
end of lac region

FIG._1K

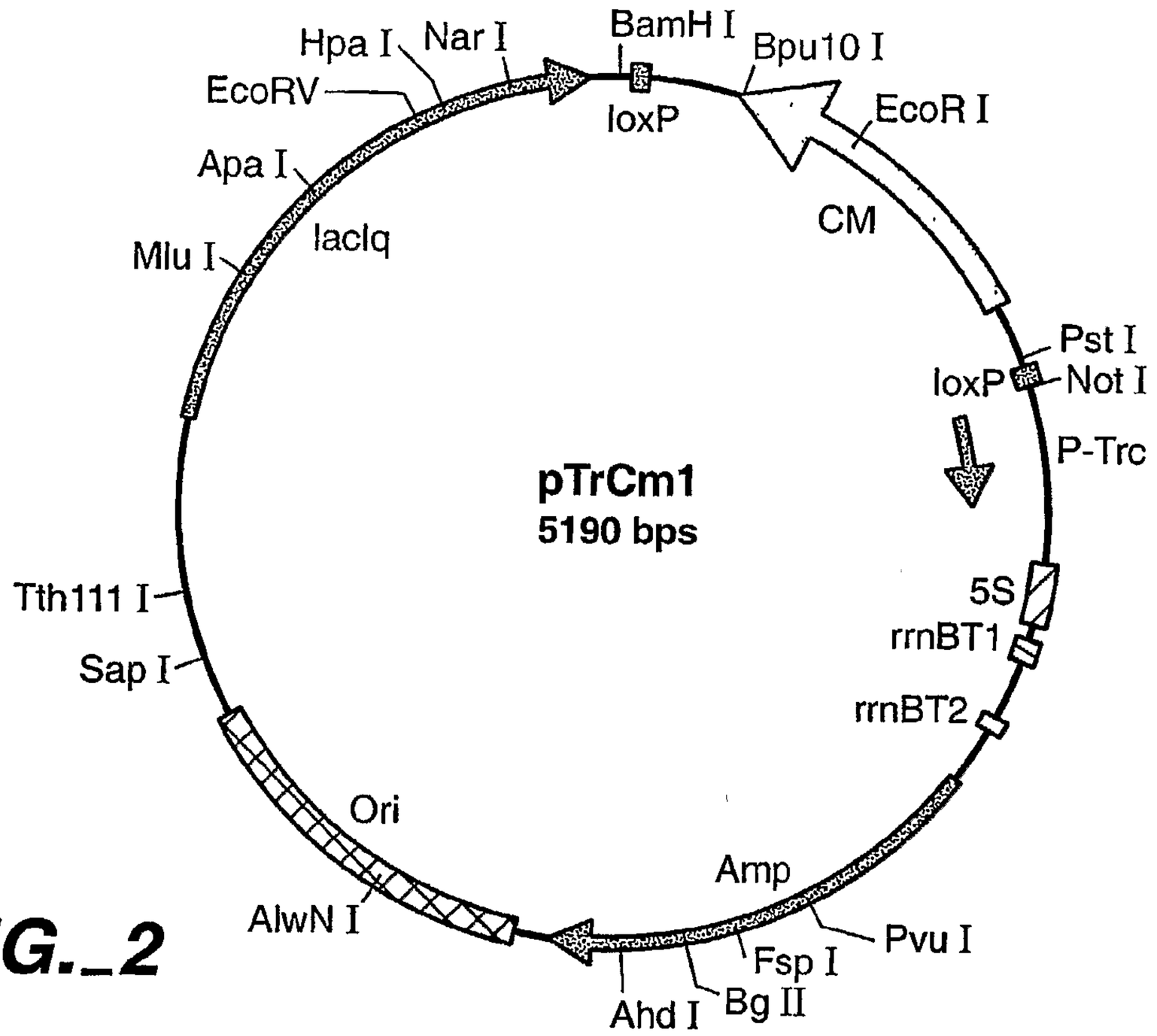

Hpa I
Nar I
BamH I
Bpu10 I
EcoRV
loxP
EcoR I
Apa I
CM
Mlu I
lacIq
Pst I
loxP
Not I
P-Trc
pTrCm1
5190 bps
5S
rrnBT1
Tth111 I
rrnBT2
Sap I
Ori
Amp
AlwN I
Pvu I
Fsp I
Ahd I
Bg II

FIG._2

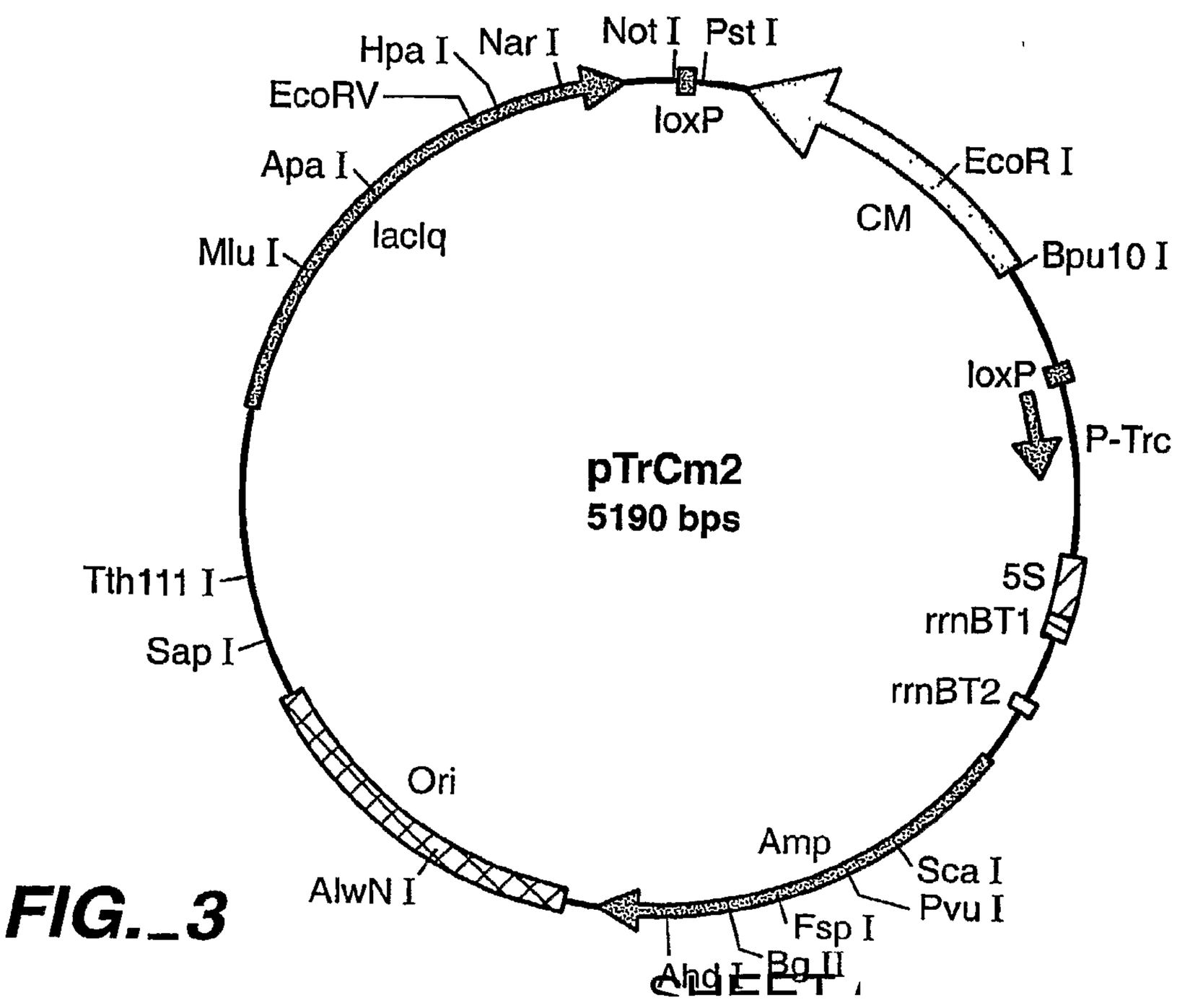

Hpa I
Nar I
Not I
Pst I
EcoRV
loxP
Apa I
EcoR I
CM
Mlu I
lacIq
Bpu10 I
loxP
P-Trc
pTrCm2
5190 bps
Tth111 I
5S
rrnBT1
Sap I
rrnBT2
Ori
Amp
Sca I
AlwN I
Pvu I
Fsp I
Ahd I
Bg II

FIG._3

FIG._4

	16 bp			
P_{tac}	TCTGAAATGAGCTGTTGACA ATTAATCATCGGCTCG TATAAT GTGTGG	1.00	SEQ ID NO.	3
M1	TCTGAAATGAGCTGTT**T**ACA ATTAATCATCGGCTCG TATAAT GTGTGG	0.25	SEQ ID NO.	4
M2	TCTGAAATGAGCTGTTGACA ATTAATCATCGG**G**TCG TATAAT GTGTGG	0.64	SEQ ID NO.	5
M3	TCTGAAATGAGCTGTTGACA ATTAATCATCGGCT**G**G TATAAT GTGTGG	0.77	SEQ ID NO.	6
M4	TCTGAAATGAGCTGTTGACA ATTAATCATCGGCTCG TA**C**AAT GTGTGG	0.70	SEQ ID NO.	7
M12	TCTGAAATGAGCTGTT**T**ACA ATTAATCATCGG**G**TCG TATAAT GTGTGG	0.31	SEQ ID NO.	8
M13	TCTGAAATGAGCTGTT**T**ACA ATTAATCATCGGCT**G**G TATAAT GTGTGG	0.32	SEQ ID NO.	9
M14	TCTGAAATGAGCTGTT**T**ACA ATTAATCATCGGCTCG TA**C**AAT GTGTGG	0.05	SEQ ID NO.	10
M23	TCTGAAATGAGCTGTTGACA ATTAATCATCGG**G**T**G**G TATAAT GTGTGG	1.03	SEQ ID NO.	11
M34	TCTGAAATGAGCTGTTGACA ATTAATCATCGGCT**G**G TA**C**AAT GTGTGG	0.75	SEQ ID NO.	12

SUBSTITUTE SHEET (RULE 26)

FIG._5

	-35		-10
$P_{D/E20}$	AACTGCAAAAATAGT	TTGACA	CCCTAGCCGATAGGCTT	TAAGAT
P_{H207}	TTTTAAAAAATTCAT	TTGACA	AACGCTTCAAATTCTCG	TATAAT
P_{N25}	TCATAAAAAATTTAT	TTGCTT	TCAGGAAAATTTTTCTG	TATAAT
P_{G25}	TGAAAAATAAAATTC	TTGATA	AAATTTTCCAATACTAT	TATAAT
P_{J5}	ATATAAAAACCGTTA	TTGACA	CAGGTGGAAATTTAGAA	TATACT
P_{A1}	TTATCAAAAAGAGTA	TTGACT	TAAAGTCTAACCTATAG	GATACT
P_{A2}	CACGAAAAACAGGTA	TTGACA	ACATGAAGTAACATGCAG	TAAGAT
P_{A3}	GGTGAAACAAAACGG	TTGACA	ACATGAAGTAAACACGG	TACGAT
P_{L}	TTATCTCTGGCGGTG	TTGACA	TAAATACCACTGGCGGT	GATACT
P_{lac}	TTAGGCACCCCAGGC	TTTACA	CTTTATGCTTCCGGCTGG	TATGTT
P_{lacUV5}	TTAGGCACCCCAGGC	TTTACA	CTTTATGCTTCCGGCTGG	TATAAT
P_{tacI}	TTCTGAAATGAGCTG	TTGACA	ATTAATCATCGGCTCG	TATAAT
P_{con}	AATTCACCGTCGTTG	TTGACA	TTTTTAAGCTTGGCGGT	TATAAT
P_{bla}	TTTTTTCTAAATACA	TTCAAA	TATGTATCCGCTCATGA	GACAAT
P_{GI}	CGAGCCGTCACGCCC	TTGACA	ATGCCACATCCTGAGCA	AATAAT

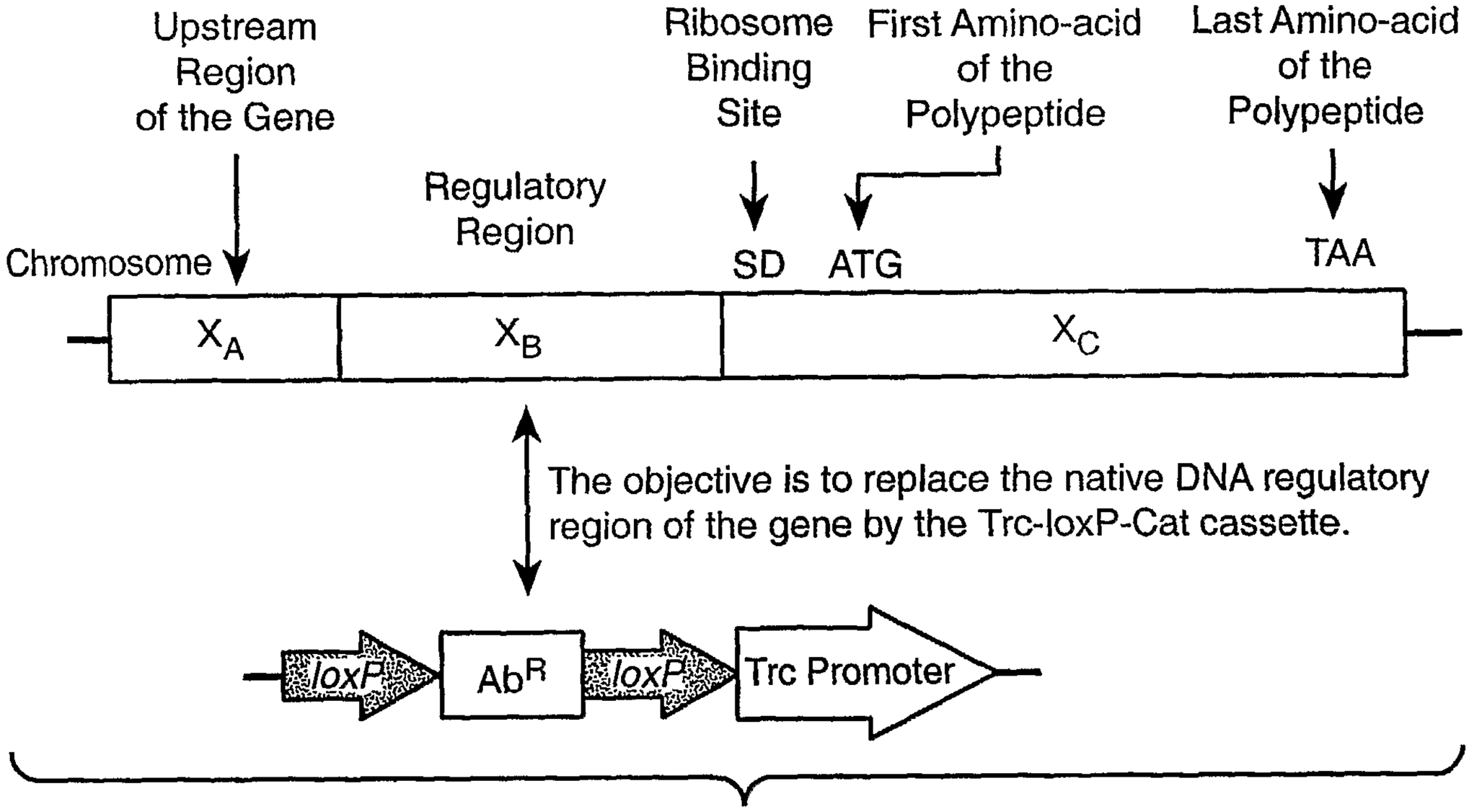

Upstream Region of the Gene
Regulatory Region
Ribosome Binding Site
First Amino-acid of the Polypeptide
Last Amino-acid of the Polypeptide
Chromosome
SD
ATG
TAA
X_A
X_B
X_C
The objective is to replace the native DNA regulatory region of the gene by the Trc-loxP-Cat cassette.
loxP
Ab^R
loxP
Trc Promoter

FIG._6A

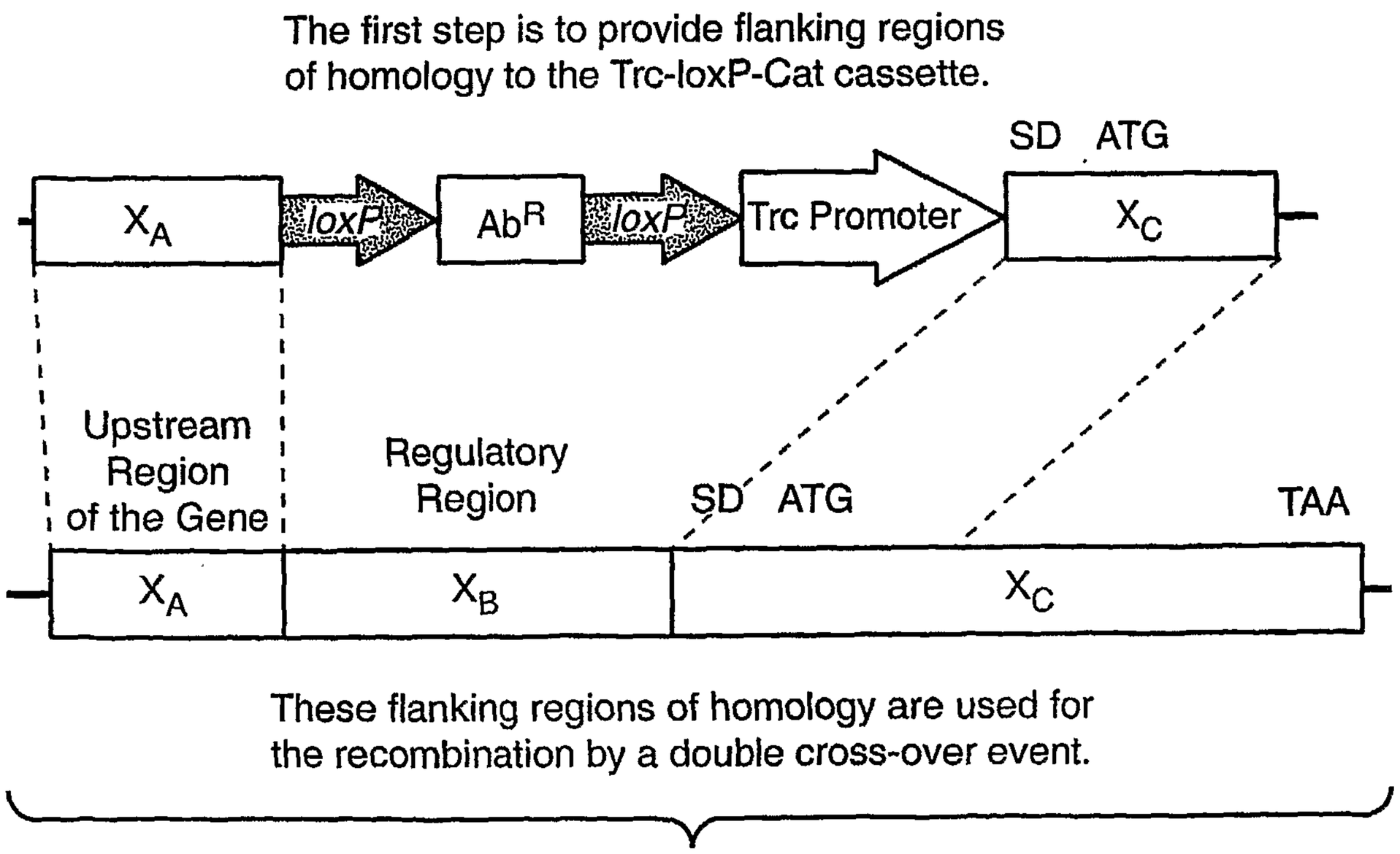

The first step is to provide flanking regions of homology to the Trc-loxP-Cat cassette.
SD
ATG
X_A
loxP
Ab^R
loxP
Trc Promoter
X_C
Upstream Region of the Gene
Regulatory Region
SD
ATG
TAA
X_A
X_B
X_C
These flanking regions of homology are used for the recombination by a double cross-over event.

FIG._6B

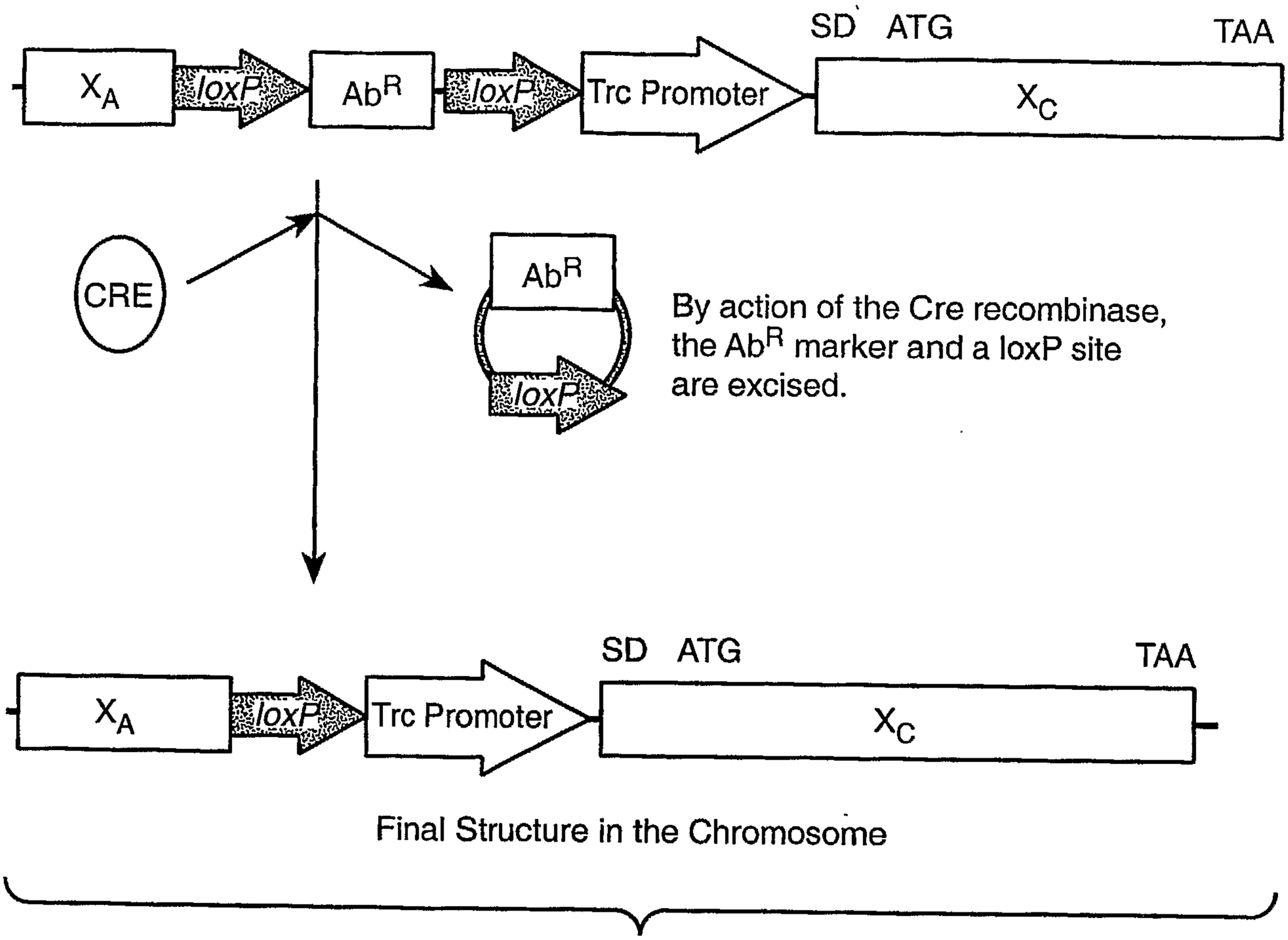

The Trc-loxP-Cat cassette replaces the native regulatory region.
SD
ATG
TAA
X_A
loxP
Ab^R
loxP
Trc Promoter
X_C
CRE
Ab^R
loxP
By action of the Cre recombinase, the Ab^R marker and a loxP site are excised.
SD
ATG
TAA
X_A
loxP
Trc Promoter
X_C
Final Structure in the Chromosome

FIG._6C

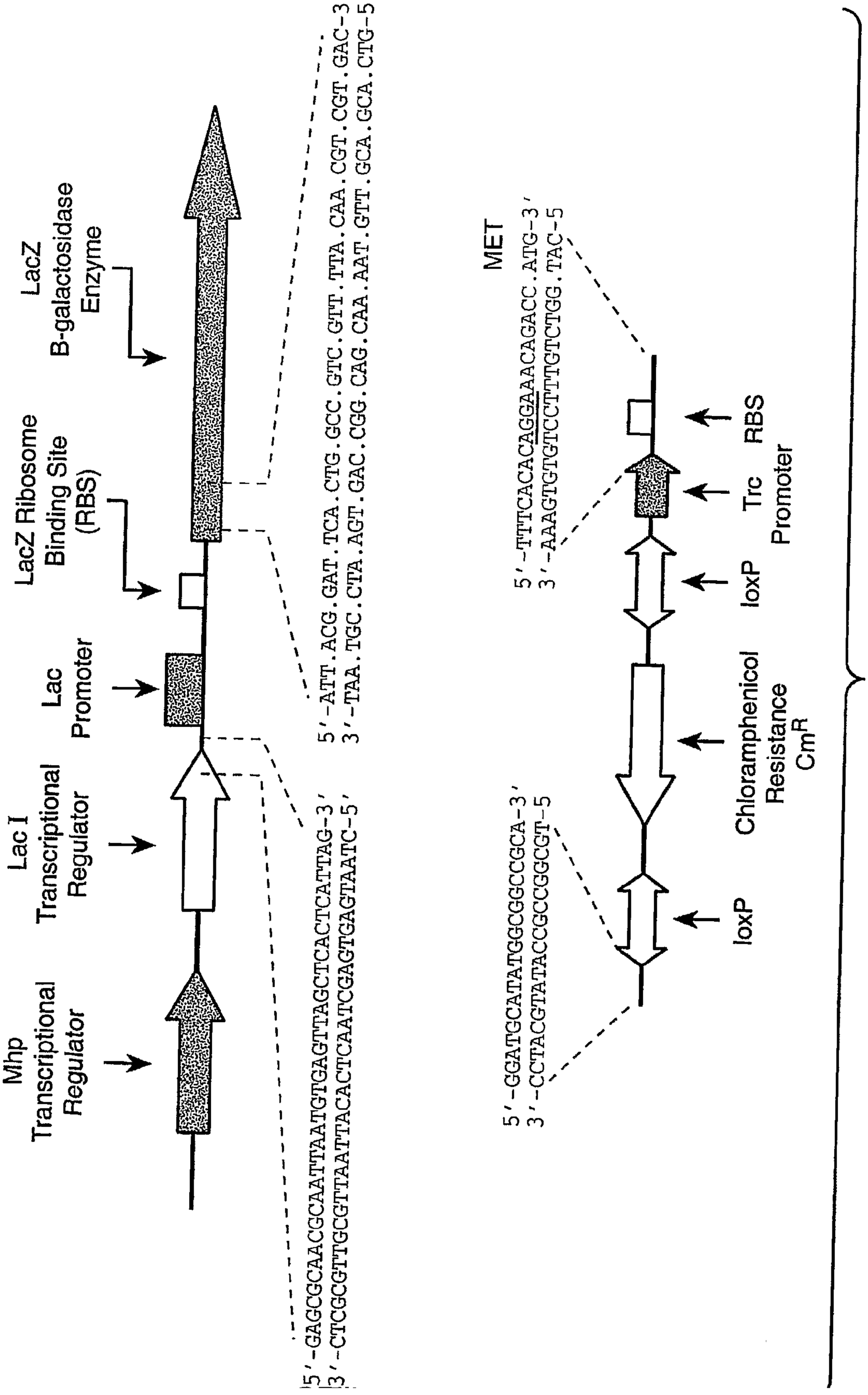

LacZ Ribosome Binding Site (RBS)
LacZ B-galactosidase Enzyme
Mhp Transcriptional Regulator
Lac I Transcriptional Regulator
Lac Promoter
Trc Promoter
RBS
MET
loxP
Chloramphenicol Resistance Cm^R
loxP
5'-GAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAG-3'
3'-CTCGCGTTGCGTTAATTACACTCAATCGAGTGAGTAATC-5'
5'-ATT.ACG.GAT.TCA.CTG.GCC.GTC.GTT.TTA.CAA.CGT.CGT.GAC-3'
3'-TAA.TGC.CTA.AGT.GAC.CGG.CAG.CAA.AAT.GTT.GCA.GCA.CTG-5'
5'-GGATGCATATGGCGGCCGCA-3'
3'-CCTACGTATACCGCCGGCGT-5
5'-TTTCACACAGGAACAGACC.ATG-3'
3'-AAAGTGTGTCCTTTGTCTGG.TAC-5

FIG._7

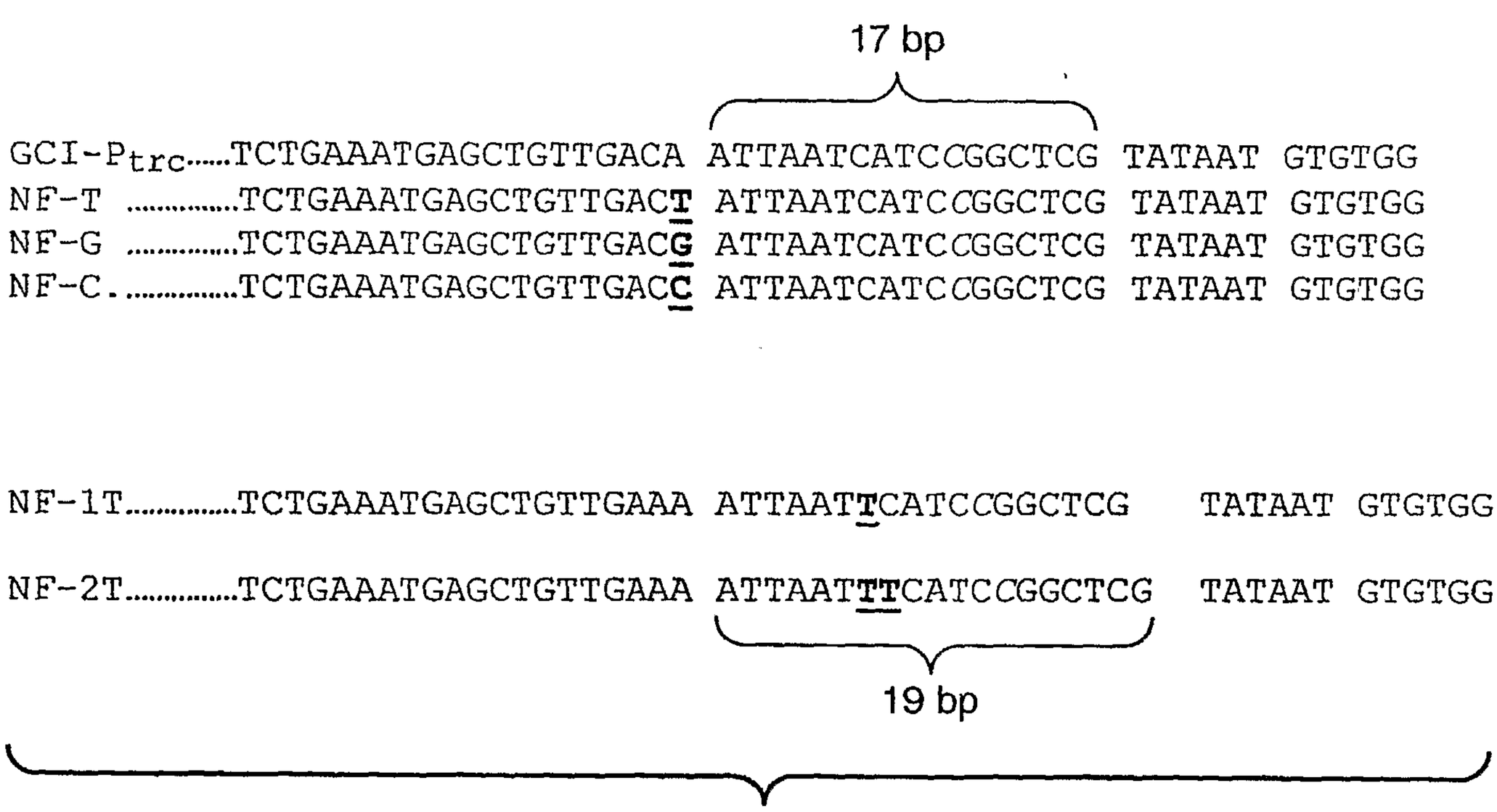
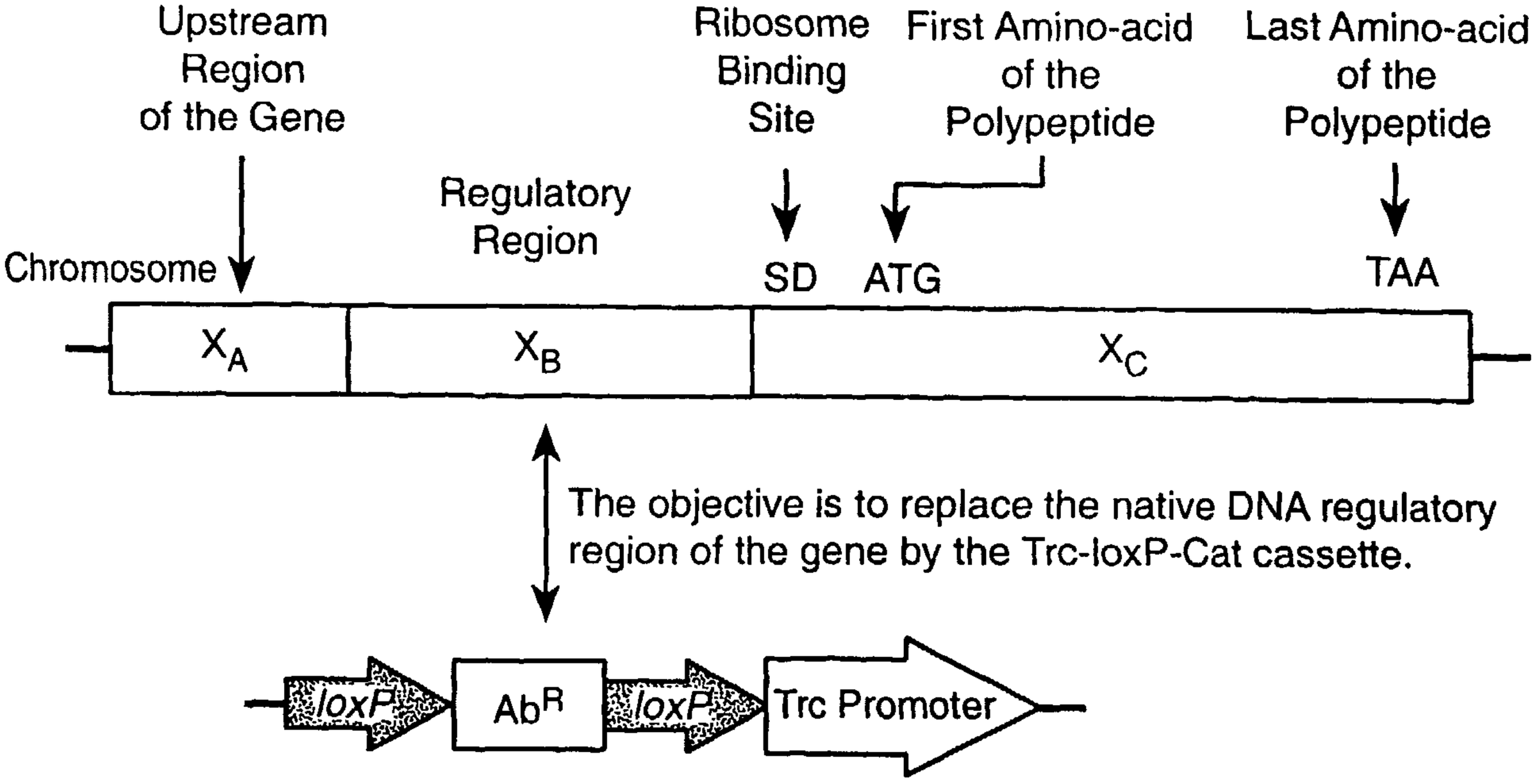

Plac .....Chromosomal Sequence.....
17 bp
GCI-Ptrc.....TCTGAAATGAGCTGTTGACA ATTAATCATCCGGCTCG TATAAT GTGTGG
NF-T .....TCTGAAATGAGCTGTTGACT ATTAATCATCCGGCTCG TATAAT GTGTGG
NF-G .....TCTGAAATGAGCTGTTGACG ATTAATCATCCGGCTCG TATAAT GTGTGG
NF-C.....TCTGAAATGAGCTGTTGACC ATTAATCATCCGGCTCG TATAAT GTGTGG
NF-1T.....TCTGAAATGAGCTGTTGAAA ATTAATTCATCCGGCTCG TATAAT GTGTGG
NF-2T.....TCTGAAATGAGCTGTTGAAA ATTAATTTCATCCGGCTCG TATAAT GTGTGG
19 bp

FIG._8


Upstream Region of the Gene
Ribosome Binding Site
First Amino-acid of the Polypeptide
Last Amino-acid of the Polypeptide
Regulatory Region
Chromosome
SD
ATG
TAA
X_A
X_B
X_C
The objective is to replace the native DNA regulatory region of the gene by the Trc-loxP-Cat cassette.
loxP
Ab^R
loxP
Trc Promoter