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

Canadian Intellectual Property Office
An agency of Industry Canada

CA 2599784 A1 2006/09/08

(21) 2 599 784

(12) DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION

(13) A1

(86) Date de dépôt PCT/PCT Filing Date: 2006/03/06
(87) Date publication PCT/PCT Publication Date: 2006/09/08
(85) Entrée phase nationale/National Entry: 2007/08/31
(86) N° demande PCT/PCT Application No.: AU 2006/000278
(87) N° publication PCT/PCT Publication No.: 2006/092023
(30) Priorité/Priority: 2005/03/04 (AU2005901040)

(51) Cl.Int./Int.Cl. *C12N 5/10* (2006.01), *C12N 5/16* (2006.01), *C12N 5/22* (2006.01)

(71) Demandeur/Applicant:
AUSTRALIAN STEM CELL CENTRE LIMITED, AU

(72) Inventeurs/Inventors:
COSTA, MAGDALINE, AU;
DOTTORI, MIRELLA, AU;
STANLEY, EDOUARD, AU;
ELEFANTY, ANDREW, AU;
PERA, MARTIN, AU

(74) Agent: BERESKIN & PARR

(54) Titre : CELLULES SOUCHES HUMAINES PLURIPOTENTES MODIFIEES GENETIQUEMENT PORTANT UN GENE RAPPORTEUR INSERE DANS LE CHROMOSOME 12 A UN LOCUS NOUVELLEMENT IDENTIFIE, ENVY

(54) Title: PLURIPOTENT GENETICALLY MODIFIED HUMAN STEM CELLS CARRYING REPORTER GENE INSERTED IN CHROMOSOME 12 AT NEWLY IDENTIFIED LOCUS, ENVY

(57) Abrégé/Abstract:
The present invention provides a genetically modified cell comprising a polynucleotide sequence integrated into the ENVY locus from chromosome 12. The ENVY locus advantageously provides a conduit for the ubiquitous or cell type specific expression of heterologous genes in human stem cells.

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
8 September 2006 (08.09.2006)

PCT

(10) International Publication Number
WO 2006/092023 A1

(51) International Patent Classification:
C12N 5/10 (2006.01) *C12N 5/22* (2006.01)
C12N 5/16 (2006.01)

(21) International Application Number:
PCT/AU2006/000278

(22) International Filing Date: 6 March 2006 (06.03.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
2005901040 4 March 2005 (04.03.2005) AU

(71) Applicant *(for all designated States except US)*: **AUSTRALIAN STEM CELL CENTRE LIMITED** [AU/AU]; Ground Floor Building 75 (strip), Monash Univeristy, Clayton, Victoria 3800 (AU).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **COSTA, Magdaline** [AU/AU]; 69 Pine Hill Drive, Doncaster East, Victoria 3109 (AU). **DOTTORI, Mirella** [AU/AU]; 2/12 Austin Avenue, Elwood, Victoria 3184 (AU). **STANLEY, Edouard** [AU/AU]; 20 Mascoma Street, Ascot Vale, Victoria 3800 (AU). **ELEFANTY, Andrew** [AU/AU]; 8 Arundel Crescent, Surrey Hills, Victoria 3127 (AU). **PERA, Martin** [AU/AU]; 53 Highbury Grove, Prahran, Victoria 3181 (AU).

(74) Agent: **BLAKE DAWSON WALDRON PATENT SERVICES**; Level 39, 101 Collins Street, Melbourne, Victoria 3000 (AU).

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

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

Published:
— *with international search report*
— *with amended claims*

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.

WO 2006/092023 A1

(54) Title: PLURIPOTENT GENETICALLY MODIFIED HUMAN STEM CELLS CARRYING REPORTER GENE INSERTED IN CHROMOSOME 12 AT NEWLY IDENTIFIED LOCUS, ENVY

(57) Abstract: The present invention provides a genetically modified cell comprising a polynucleotide sequence integrated into the ENVY locus from chromosome 12. The ENVY locus advantageously provides a conduit for the ubiquitous or cell type specific expression of heterologous genes in human stem cells.

DEMANDE OU BREVET VOLUMINEUX

LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVET COMPREND PLUS D'UN TOME.

CECI EST LE TOME 1 DE 2
CONTENANT LES PAGES 1 À 12

NOTE : Pour les tomes additionels, veuillez contacter le Bureau canadien des brevets

JUMBO APPLICATIONS/PATENTS

THIS SECTION OF THE APPLICATION/PATENT CONTAINS MORE THAN ONE VOLUME

THIS IS VOLUME 1 OF 2
CONTAINING PAGES 1 TO 12

NOTE: For additional volumes, please contact the Canadian Patent Office

NOM DU FICHIER / FILE NAME :

NOTE POUR LE TOME / VOLUME NOTE:

Pluripotent Genetically Modified Human Stem Cells carrying reporter gene inserted in chromosome 12 at newly identified locus, ENVY

FIELD OF THE INVENTION

The present invention relates to the identification of a genetic locus present on chromosome 12 and its use as a site for insertion and expression of exogenous polynucleotide sequences in stem cells and their progeny.

BACKGROUND OF THE INVENTION

Stem cells, particularly human embryonic stem cells (hESCs) are a potential source of cells for use in cell replacement therapies. The ability to readily identify stem cells, such as hESCs and their differentiated progeny in transplantation experiments and recombinant tissue grafting experiments would be useful to facilitate the analysis of hESC potential and function *in vivo*.

Existing methods of identifying human ES-derived cells in animal grafting experiments has relied on the use of species-specific antibodies to ubiquitously express intracellular markers or *in situ* hybridisation, both procedures that are cumbersome and require fixation of the cells. The Rosa26 locus has been previously used in mouse ES cells to express exogenous genes (3;4). However, an effective marker system for hES cells is required,

We have now generated genetically modified cell comprising a polynucleotide sequence inserted into the ENVY gene locus from chromosome 12 that can be used as effective marker system for human embryonic stem cells. The ENVY locus advantageously provides a conduit for the ubiquitous expression of exogenous genes in human stem cells. Although other groups have generated GFP^+ HESCs (5;6), we have generated a hESC line, designated "*Envy*", that advantageously expresses sustained and high levels of a exogenous polynucleotide sequences, such as, green fluorescent protein (GFP) in stem cells and all differentiated progeny in the absence of ongoing selection. This uniformity and intensity of exogenous polynucleotide expression allows the analysis and enables the viable recovery of HESC-derived material from animal transplantation experiments (7) and recombinant tissue grafting experiments (8).

SUMMARY OF THE INVENTION

The present inventors have successfully identified a genetic locus called ENVY which is present on human chromosome 12 that has been found to be an effective conduit for the ubiquitous expression of exogenous genes in human stem cells and their differentiated progeny. Thus the locus can be targeted for the integration and expression of exogenous DNA in human stem cells, particularly embryonic stem cells. Expression of exogenous DNA can therefore be carried through to the differentiated progeny which derive from that embryonic stem cell.

One aspect of the present invention provides a genetically modified comprising a polynucleotide sequence inserted in a gene locus positioned between Thrombopoietin (Tpo) gene and Solute Carrier 25 (SC25) gene.

In a preferred embodiment of the invention the gene locus has at least 95% nucleotide identity to nucleotides sequence of SEQ ID NO:1 excluding nucleotides 2037 to 6095 of SEQ ID NO:1. Preferably, the polynucleotide sequence comprises a promoter sequence, more preferably the promoter sequence is a beta actin promoter sequence. In a preferred embodiment the polynucleotide sequence encodes a reporter.

A second aspect of the present invention provides a genetically modified cell comprising the nucleotide sequence of SEQ ID NO:1, wherein nucleotides 2036 to 6096 of SEQ ID NO:1 is replaced with a second polynucleotide sequence.

A third aspect of the present invention provides a genetically modified cell comprising the nucleotide sequence of SEQ ID NO:1, wherein nucleotides 4740 to 6096 of SEQ ID NO:1 is replaced with a second polynucleotide sequence.

In a fourth aspect of the present invention there is provided a cell line comprising a genetically modified cell as hereinbefore described.

A fifth aspect of the present invention provides a mixture of cells comprising at least two types of cells, wherein at least one cell type comprises a polynucleotide sequence encoding a first reporter inserted in a gene locus positioned between Thrombopoietin (Tpo) gene and Solute Carrier 25 (SC25) gene; and at least a second cell type comprises a polynucleotide sequence encoding a second reporter inserted in a gene locus positioned between Thrombopoietin (Tpo) gene and Solute Carrier 25 (SC25) gene; wherein the first reporter is distinguishable from the second reporter.

A sixth aspect of the present invention provides the use of a vector comprising a polynucleotide sequence homologous with the nucleotide sequence of SEQ ID NO:1 excluding nucleotides 2037 to 6095 of SEQ ID NO:1, and a promoter for insertion into a gene locus positioned between Thrombopoietin (Tpo) gene and Solute Carrier 25 (SC25) gene.

In still further aspects the present invention provides a nucleotide sequence with at least 95% identity to sequence ID NO:1.

In yet another aspect the present invention provides a derivative formed by the insertion of a polynucleotide sequence into any part of sequence ID NO:1.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 shows the nucleotide sequence including variant nucleotide positions of the envy locus from chromosome 12 including the integrated sequence of human beta actin promoter, Green fluorescent protein (GFP); and internal ribosome entry site (IRES). This nucleotide sequence is referred to as SEQ ID NO:1.

Figure 2 shows the nucleotide sequence of the envy locus from chromosome 12, together with the complementary sequence underneath. Location of restriction sites are indicated. The location of the human beta actin promoter, Green fluorescent protein (GFP); and internal ribosome entry site (IRES) are indicated. This nucleotide sequence is referred to as SEQ ID NO:2. This annotated sequence is in genbank as sequence entry AY952326.

Figure 3 shows genetic and phenotypic analysis of *Envy* cells. (A) Structure of the *Envy* locus showing the vector, locus and chromosomal location relative to neighbouring genes, Thrombopoietin (Tpo) and Solute Carrier 25 (SC25). Cm, centromere; Tm, telomere. *Envy* cells grown in bulk culture displayed uniform high level GFP expression clearly visible by fluorescence microscopy (B,C,D) and were easily distinguishable from unmanipulated hESCs by flow cytometry (E). Flow cytometric experiments indicated that *Envy* cells retained robust expression of the stem cell markers GCTM-2 (F) and E-Cadherin (G) while immunofluorescence microscopy showed uniform expression of Oct4 (H-J). Xenotransplanted *Envy* cells formed teratomas containing a variety of different tissues including those reminiscent of neural rosettes (K), muscle and cartilage (L) and endoderm derived secretory epithelia (M). Methods: *Envy* cells were grown as described (9) and the β-Actin promoter GFP expression vector introduced by electroporation essentially as for mouse ES cells (10). For FACS analysis, ES colonies were dissociated to single cells using trypsin and stained with antibodies recognising GCTM-2 (11) and E-Cadherin

(Zymed). ES cells were stained for Oct4 expression as previously described (12). Generation and histological analysis of teratomas was performed as described by Reubinoff and colleagues (1).

Figure 4 shows that differentiated progeny of *Envy* cells retain robust levels of GFP expression. GFP expression was maintained during noggin-induced neural differentiation (A,B), neurosphere formation (C) and in differentiated neurons and astrocytes marked by expression of β-tubulin (Tub) (D-F) and GFAP (G-I) respectively. In suspension culture, *Envy* cells formed GFP^+ embryoid bodies (J-L) that differentiated further to form GFP^+ blood cells (M-O). FACS analysis indicated that all of the cell types within these cultures were GFP^+, including a proportion of cells that expressed the pan haematopoietic marker CD45 (Q) and the myeloid marker CD116 (GM-CSF receptor) (R). In cultures that favoured the formation of endoderm, GFP expression was retained in cells expressing albumin (S-U). *Methods* Differentiation and antibody staining of neural cells was performed as described by (7; 13). For mesoderm (blood) formation, ES colonies were digested to give small clumps using collagenase type IV (14) and transferred to low attachment dishes in serum free medium (15) supplemented with 5 ng/ml Bone Morphogenetic Protein 4, 5ng/ml Vascular endothelial growth factor (VEGF), 5ng/ml Flt3 ligand, 10 ng/ml Stem cell factor (SCF), 5ng/ml Interleukin-6 (IL-6), 5ng/ml Insulin like growth factor-2 (IGF-2). Following 11 days of suspension culture, EBs were transferred to 24 well plates containing media supplemented with VEGF, Flt3 ligand, 10 ng/ml SCF, 5 units/ml Erythropoietin, 5 ng/ml Interleukin-3, 5ng/ml Thrombopoietin and grown for 9 more days. For FACS analysis, blood cells were stained with either mouse IgG (control), or antibodies directed to CD45 or CD116 (BD Biosciences). Endodermal cells were identified in day 6 to 8 week old cultures of ES cells differentiated under reduced serum conditions. Albumin expression was visualized by immuno-staining ethanol fixed cells with a rabbit anti-human albumin antibody (Dako) followed by anti-rabbit Alexa 568 (Molecular Probes).

DETAILED DESCRIPTION OF THE INVENTION

The present invention is related to the identification of a genetic locus which is present on chromosome 12 in human cells. This locus is referred to as the ENVY locus. The locus has been found to be a conduit for the ubiquitous expression of genes in human stem cells and their differentiated progeny. Thus the locus can be used as a target site for the integration and expression of exogenous polynucleotides in human stem cells, particularly embryonic stem cells. Expression of the exogenous polynucleotides will therefore be carried through to the differentiated progeny which derive from that embryonic stem cell.

Accordingly, in a first aspect, the present invention provides invention provides a genetically modified comprising a polynucleotide sequence inserted in a gene locus positioned between Thrombopoietin (Tpo) gene and Solute Carrier 25 (SC25) gene.

In a preferred embodiment of the invention the gene locus has at least 95%, 96%, 97%, 98% or 99% nucleotide identity to nucleotides sequence of SEQ ID NO:1 excluding nucleotides 2037 to 6095 of SEQ ID NO:1. Nucleotides 2037 to 6095 of SEQ ID NO:1 corresponds to inserted polynucleotide sequence encoding human beta actin promoter, Green fluorescent protein (GFP) and an internal ribosome entry site (IRES).

In still further aspects the present invention provides a nucleotide sequence with at least 95%, 96%, 97%, 98% or 99% identity to sequence ID NO:1.

In yet another aspect the present invention provides a derivative formed by the insertion of a polynucleotide sequence into any part of sequence ID NO:1.

For the purpose of this invention, the "nucleotide identity" or "sequence identity" of two related nucleotide or amino acid sequences, expressed as a percentage, refers to the number of positions in the two optimally aligned sequences which have identical residues (x100) divided by the number of positions compared. A gap, i.e., a position in an alignment where a residue is present in one sequence but not in the other is regarded as a position with non-identical residues. The alignment of the two sequences is performed by the Needleman and Wunsch algorithm (Needleman and Wunsch (1970) "A general method applicable to the search for similarities in the amino acid sequence of two proteins" J Mol Biol 48: 443-453). The computer-assisted sequence alignment above can be conveniently performed using standard software program such as GAP, which is part of the Wisconsin Package Version 10.1 (Genetics Computer Group, Madison, Wisconsin, USA) using the default scoring matrix with a gap creation penalty of 50 and a gap extension penalty of 3. It is clear that when RNA sequences are to be essentially similar or have a certain degree of sequence identity with DNA sequences, thymine (T) in the DNA sequence is considered equal to uracil (U) in the RNA sequence.

In a preferred embodiment, the genetically modified cell of the present invention comprises a polynucleotide sequence inserted in a gene locus having the nucleotides sequence of SEQ ID NO:1 excluding nucleotides 2037 to 6095 of SEQ ID NO:1.

It is preferred that the polynucleotide sequence inserted in the gene locus of the present invention comprises a promoter sequence. The promoter sequence preferably is useful for expression of exogenous polypeptides in stem cells. Suitable promoters would be well

known by those skilled in the art. Preferably, the promoter confers constitutive expression of polynucleotide sequence when operably joined to the sequence and is effective in a variety of different cell types. Most preferably, the promoter sequence is, but not limited to, a beta actin promoter sequence.

Suitable polynucleotide sequence that can be inserted in the gene locus of the present invention, includes but is not limited to, DNA encoding a reporter, gene regulatory elements, such as promoter sequences, gene elements involved in homologous recombination of DNA, such as internal ribosome binding sites and/ or genes that regulates the survival, proliferation and/or differentiation of stem cells. The polynucleotide sequence inserted in the gene locus of the present invention preferably comprises an internal ribosome binding site (IRES). Most preferably, the genetically modified cells of the present invention comprises the nucleotide sequence of SEQ ID NO:1 or the sequence as shown in Figure 2 or SEQ ID NO:2. (ie sequence that corresponds to the ENVY locus including inserted polynucleotide sequence encoding human beta actin promoter, GFP and an IRES).

In a preferred embodiment of the invention there is provided a genetically modified cell as hereinbefore described, wherein the polynucleotide sequence encodes a polypeptide that regulates the survival, proliferation and/or differentiation of stem cells. For example, the polynucleotide sequence can encode a polypeptide involved in the suicide or apoptosis of stem cells, a polypeptide involved in promoting proliferation of stem cells or their differentiated progeny, a polypeptide that is a drug target, a polypeptide or RNA that confers resistance to bacterial or viral infection, or a polypeptide that that drives differentiation of stem cells towards a particular lineage.

The inserted sequences will typically confer on the stem cell or its differentiated progeny a new phenotypic characteristic.

Therefore, the present invention contemplates a number of different types of polynucleotides that may be inserted into the envy locus. For example, this may include genes which regulate the survival, proliferation and/or differentiation of stem cells, such as suicide/apoptosis or survival genes for example from the *bcl-2* gene family or proliferation promoting genes such as *myc or raf*; genes whose expression products represent appropriate drug targets, for example mutated Ras proteins found in human tumours; dominant negative p53 protein, TAT protein from HIV, an oncogenic form of the adenomatous polyposis coli protein, genes whose expression product effect a particular condition, such as an isomerase that alters the conformation of amyloid proteins in Alzheimer's disease; genes that encode proteins or RNA's that confer resistance to bacterial

or viral infection; and genes that drive differentiation of stem cells towards a particular lineage for use as a therapeutic agent. Examples of appropriate genes that may be used would be familiar to persons skilled in the art. Reference to the term "gene" is understood as referring to either the genomic sequence or the cDNA sequence. The gene may be an endogenous or exogenous gene. The gene may be human derived or derived from another species.

The genetically modified cell of the present invention preferably comprises polynucleotide sequence that encodes a reporter. The reporter is most preferably a fluorescent protein, most preferably the fluorescent protein is green fluorescent protein (GFP). Other suitable reporter can include biochemical, enzymatic reporters or other markers known to those skilled in the art, or combinations thereof.

In this embodiment, an appropriate reporter, such as fluorescent or biochemical marker may be inserted directly into the ENVY locus so that the progeny of a 'marked' stem cell may be tracked. This may be useful where it is, for instance, intended to be used in harvested stem cells for therapy following chemotherapy so that it can be determined whether the cells home to the bone marrow upon re-infusion.

In an alternative embodiment, the polynucleotide sequence which is the complement of the sequence inserted into the ENVY locus may be labelled with an appropriate label such as a fluorescent or biochemical tag so that sequences which hybridise to the labelled sequence can be isolated.

A second aspect of the present invention provides a genetically modified cell comprising the nucleotide sequence of SEQ ID NO:1, wherein nucleotides 2036 to 6096 of SEQ ID NO:1 is replaced with a second polynucleotide sequence. Nucleotides 2036 to 6096 of SEQ ID NO:1 corresponds to polynucleotide sequence encoding human beta actin promoter, Green fluorescent protein (GFP) and an internal ribosome entry site (IRES).

A third aspect of the present invention provides a genetically modified cell comprising the nucleotide sequence of SEQ ID NO:1, wherein nucleotides 4740 to 6096 of SEQ ID NO:1 is replaced with a second polynucleotide sequence. Nucleotides 4740 to 6096 of SEQ ID NO:1 corresponds to polynucleotide sequence encoding Green fluorescent protein (GFP) and an internal ribosome entry site (IRES). When the second polynucleotide sequence is substituted into the GPF or GPF and IRES sites, this brings it under the control of the beta actin promoter. Accordingly, it is not necessary to include additional promoter sequences upstream of the inserted polynucleotide sequence.

In a fourth aspect of the present invention there is provided a cell line comprising a genetically modified cell as hereinbefore described.

The genetically modified cell of the present invention is preferably a stem cell. The stem cell may be a totipotent cell which has the potential to become any type of cell in the adult body or any cell of the extraembryonic membranes (e.g., placenta). Typically, the only totipotent cells are the fertilized egg and the first 4 or so cells produced by its cleavage. The stem cell may be a pluriopotent stem cell having the potential to make any differentiated cell in the body, but cannot contribute to making the extraembryonic membranes (which are derived from the trophoblast). Types of pluriopotent stem cells include embryonic stem (ES) cells which can be isolated from the inner cell mass (ICM) of a blastocyst - the stage of embryonic development when implantation occurs. For humans, excess embryos produced during *in vitro* fertilization (IVF) procedures are used. Embryonic germ (EG) cells are also pluriopotent cells that can be isolated from the precursor to the gonads in aborted fetuses. Embryonic carcinoma (EC) cells are also pluriopotent cells that can be isolated from teratocarcinomas, a tumor that occasionally occurs in a gonad of a fetus. Pluriopotent stem cells can only be isolated from embryonic or fetal tissue. The stem cell may be a multipotent stem cell that can only differentiate into a limited number of cell types. For example, the bone marrow contains multipotent stem cells that give rise to all the cells of the blood but not to other types of cells. Multipotent stem cells are found in adult animals; including most organs in the body (e.g., brain, liver) contain them where they can replace dead or damaged cells. Most stem cells can develop into any of the three major tissue types: endoderm (interior gut lining), mesoderm (muscle, bone, blood), and ectoderm (epidermal tissues and nervous system). Pluriopotent stem cells can eventually specialize in any bodily tissue, but they cannot themselves develop into a human being.

Most preferably, the genetically modified cell of the present invention is a human embryonic stem cell.

A fifth aspect of the present invention provides a mixture of cells comprising at least two types of cells, wherein at least one cell type comprises a polynucleotide sequence encoding a first reporter inserted in a gene locus positioned between Thrombopoietin (Tpo) gene and Solute Carrier 25 (SC25) gene; and at least a second cell type comprises a polynucleotide sequence encoding a second reporter inserted in a gene locus positioned between Thrombopoietin (Tpo) gene and Solute Carrier 25 (SC25) gene; wherein the first reporter is distinguishable from the second reporter.

Preferably the cells are human embryonic stem cells and the first and second reporter is most preferably a fluorescent protein. Accordingly, the present invention can be utilised to compare different stem cell lineages and observe the differentiation of various stem cells to specific cell types. In addition, the effects of exogenous factors on cultured stem cells can be examined by the reporter system of the present invention.

A sixth aspect of the present invention provides the use of a vector comprising a polynucleotide sequence homologous with the nucleotide sequence of SEQ ID NO:1 excluding nucleotides 2037 to 6095 of SEQ ID NO:1, and a promoter for insertion into a gene locus positioned between Thrombopoietin (Tpo) gene and Solute Carrier 25 (SC25) gene.

In a preferred embodiment, the vector comprises a promoter that is a beta actin promoter. Preferably, the vector comprises polynucleotide sequence encoding a reporter, most preferably the reporter is a fluorescent protein, such as green fluorescent protein (GFP). The vector preferably comprises a polynucleotide sequence encoding an internal ribosome binding site (IRES). Most preferably, the vector comprises nucleotides 2037 to 6095 of SEQ ID NO:1.

The vector sequence of the present invention can also be used as a transgene when inserted into a site outside human chromosome 12.

Throughout this specification the word "comprise", or variations such as "comprises" or "comprising", will be understood to imply the inclusion of a stated element, integer or step, or group of elements, integers or steps, but not the exclusion of any other element, integer or step, or group of elements, integers or steps.

All publications mentioned in this specification are herein incorporated by reference. Any discussion of documents, acts, materials, devices, articles or the like which has been included in the present specification is solely for the purpose of providing a context for the present invention. It is not to be taken as an admission that any or all of these matters form part of the prior art base or were common general knowledge in the field relevant to the present invention as it existed in Australia or elsewhere before the priority date of each claim of this application.

It will be appreciated by persons skilled in the art that numerous variations and/or modifications may be made to the invention as shown in the specific embodiments without departing from the spirit or scope of the invention as broadly described. The present

embodiments are, therefore, to be considered in all respects as illustrative and not restrictive.

EXAMPLE 1

1. *Envy* Human Embryonic Stem Cell line That Expresses High Levels Of Green Fluorescent Protein In All Differentiated Progeny

The National Institutes of Health Human Embryonic Stem Cell Registry listed hES cell line, hES 3 (1) (http://www.escellinternational.com) was electroporated with a vector in which GFP was expressed under the control of the human β–Actin promoter (Figure 3A). A GFP positive hES clone identified three days after electroporation was isolated under epifluorescence illumination and expanded. Southern blot and PCR analysis suggested this clone harboured a single copy of the vector that had undergone significant rearrangement upon integration into the genome (data not shown). 3' RACE experiments revealed the vector had integrated into chromosome 12 q23.1 (Figure 3A), an assignment confirmed by PCR using primers predicted to lie 5' and 3' of the proposed integration site (data not shown). Examination of this genomic region showed the GFP transgene was positioned between the genes encoding Thrombopoietin (Tpo) and Solute Carrier 25 (SC25) and had not disrupted any known genes (Figure 3A). *Envy* cells could be maintained in bulk culture and expressed uniform, high levels of GFP making them easily distinguishable from wild type ES cells (Figure 3B-E). *Envy* cells expressed the stem cell markers GCTM-2, E-Cadherin and Oct4 in their undifferentiated state and had a normal karyotype (Figure 3F-J and data not shown). When injected into the testes capsule of SCID mice, *Envy* cells formed teratomas containing a spectrum of different cell types including derivatives of ectoderm (neural rosettes), mesoderm (cartilage) and endoderm (glandular epithelia) confirming their pluripotentiality (Figure 3K-M)).

Envy cells retained uniform and robust GFP expression following *in vitro* differentiation into cells representative of the three embryonic germ layers. Differentiation of *Envy* cells towards the neuronal lineage showed that GFP expression was present at high levels during neurosphere formation and in differentiated cells expressing the neural and glial markers β-tubulin and GFAP respectively (Figure 4A-I).*Envy* embryoid bodies formed in suspension culture showed robust expression of GFP that was maintained in all of the differentiated progeny (Figure 4J-O). Flow cytometric analysis indicated that in cultures of differentiated *Envy* cells containing blood cells, a mesoderm derivative, all of the cells expressed uniform high levels of GFP (Figure 4P-R). When *Envy* cells differentiated into cells of the endoderm lineage, as determined by expression of alphafetoprotein, the hepatic

stem cell marker GCTM5 ((2); data not shown) or albumin (Figure 4S-U), GFP expression was also retained at high level.

REFERENCES

1 Reubinoff BE, Pera MF, Fong CY, Trounson A, Bongso A: *Nat Biotechnol* 18:399-404, 2000

2 Stamp L, Crosby HA, Hawes SM, Strain AJ, Pera MF: *Stem Cells* 23:in press, 2005

3 Soriano P: *Nat Genet* 21:70-71., 1999

4 Mao X, Fujiwara Y, Chapdelaine A, Yang H, Orkin SH: *Blood* 97:324-326., 2001

5 Ma Y, Ramezani A, Lewis R, Hawley RG, Thomson JA: *Stem Cells* 21:111-117, 2003

6 Gropp M, Itsykson P, Singer O, Ben-Hur T, Reinhartz E, Galun E, Reubinoff BE: *Mol Ther* 7:281-287, 2003

7 Reubinoff BE, Itsykson P, Turetsky T, Pera MF, Reinhartz E, Itzik A, Ben-Hur T: *Nat Biotechnol* 19:1134-1140, 2001

8 Goldstein RS, Drukker M, Reubinoff BE, Benvenisty N: *Dev Dyn* 225:80-86, 2002

9 Tian X, Kaufman DS: *Methods Mol Med* 105:425-436., 2004

10 Barnett LD, Kontgen F: *Methods Mol Biol* 158:65-82, 2001

11 Cooper S, Pera MF, Bennett W, Finch JT: *Biochem J* 286 (Pt 3):959-966, 1992

12 Pera MF, Filipczyk AA, Hawes SM, Laslett AL: *Methods Enzymol* 365:429-446., 2003

13 Pera MF, Andrade J, Houssami S, Reubinoff B, Trounson A, Stanley EG, Ward-van Oostwaard D, Mummery C: *J Cell Sci* 117:1269-1280, 2004

14 Thomson JA, Itskovitz-Eldor J, Shapiro SS, Waknitz MA, Swiergiel JJ, Marshall VS, Jones JM: *Science* 282:1145-1147, 1998

15 Wiles MV, Johansson BM: *Leukemia* 11 Suppl 3:454-456, 1997

DEMANDE OU BREVET VOLUMINEUX

LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVET COMPREND PLUS D'UN TOME.

CECI EST LE TOME 1 DE 2
CONTENANT LES PAGES 1 À 12

NOTE : Pour les tomes additionels, veuillez contacter le Bureau canadien des brevets

JUMBO APPLICATIONS/PATENTS

THIS SECTION OF THE APPLICATION/PATENT CONTAINS MORE THAN ONE VOLUME

THIS IS VOLUME 1 OF 2
CONTAINING PAGES 1 TO 12

NOTE: For additional volumes, please contact the Canadian Patent Office

NOM DU FICHIER / FILE NAME :

NOTE POUR LE TOME / VOLUME NOTE:

AMENDED CLAIMS
[received by the International Bureau on 27 June 2006 (27.06.06); original claims 1-31 replaced by new claims 1-31 (3 pages)]

Claims

1. A genetically modified cell comprising a polynucleotide sequence inserted in a gene locus positioned between Thymopoietin (Tmpo) gene and Solute Carrier 25 (SC25) gene.

2. A genetically modified cell according to claim 1, wherein the gene locus has at least 95% nucleotide identity to nucleotides sequence of SEQ ID NO:1 excluding nucleotides 2037 to 6095 of SEQ ID NO:1.

3. A genetically modified cell according to claim 1 or 2, wherein the gene locus has the nucleotides sequence of SEQ ID NO:1 excluding nucleotides 2037 to 6095 of SEQ ID NO:1.

4. A genetically modified cell according to any one of claims 1 to 3, wherein the polynucleotide sequence comprises a promoter sequence.

5. A genetically modified cell according to claim 4, wherein the promoter sequence is a beta actin promoter sequence.

6. A genetically modified cell according to any one of claims 1 to 5, wherein the polynucleotide sequence encodes a reporter.

7. A genetically modified cell according to claim 6, wherein the reporter is a fluorescent protein.

8. A genetically modified cell according to claim 7, wherein the fluorescent protein is green fluorescent protein (GFP).

9. A genetically modified cell according to any one of claims 1 to 8, wherein the polynucleotide sequence encodes an internal ribosome binding site (IRES).

10. A genetically modified cell according to any one of claims 1 to 9, wherein the cell comprises the nucleotide sequence of SEQ ID NO:1.

11. A genetically modified cell according to any one of claims 1 to 5, wherein the polynucleotide sequence encodes a polypeptide that regulates the survival, proliferation and/or differentiation of stem cells or their differentiated progeny.

12. A genetically modified cell according to claim 6, wherein the polynucleotide sequence encodes a polypeptide involved in the suicide or apoptosis of stem cells or their differentiated progeny, a polypeptide involved in promoting proliferation of stem cells or

their differentiated progeny, a polypeptide that is a drug target, a polypeptide or RNA that confers resistance to bacterial or viral infection, or a polypeptide that that drives differentiation of stem cells towards a particular lineage.

13. A genetically modified cell comprising the nucleotide sequence of SEQ ID NO:1, wherein nucleotides 2036 to 6096 of SEQ ID NO:1 is replaced with a second polynucleotide sequence.

14. A genetically modified cell comprising the nucleotide sequence of SEQ ID NO:1, wherein nucleotides 4740 to 6096 of SEQ ID NO:1 is replaced with a second polynucleotide sequence.

15. A genetically modified cell according to any one of claims 1 to 13, wherein the polynucleotide sequence is inserted into the gene locus by homologous recombination or site specific recombination.

16. A genetically modified cell according to any one of claims 1 to 15, wherein the cell is a human embryonic stem cell or a derivative of a human embryonic stem cell.

17. A genetically modified cell according to claims 13 or 14, wherein the second polynucleotide sequence encodes a polypeptide that regulates the survival, proliferation and/or differentiation of stem cells.

18. A genetically modified cell according to claim 17, wherein the second polynucleotide sequence encodes a polypeptide involved in the suicide or apoptosis of stem cells or their differentiated progeny, a polypeptide involved in promoting proliferation of stem cells or their differentiated progeny, a polypeptide that is a drug target, a polypeptide or RNA that confers resistance to bacterial or viral infection, or a polypeptide that that drives differentiation of stem cells towards a particular lineage.

19. A cell line comprising a genetically modified cell according to any one of claims 1 to 18.

20. A mixture of cells comprising at least two types of cells, wherein at least one cell type comprises a polynucleotide sequence encoding a first reporter inserted in a gene locus positioned between Thymopoietin (Tmpo) gene and Solute Carrier 25 (SC25) gene; and at least a second cell type comprises a polynucleotide sequence encoding a second reporter inserted in a gene locus positioned between Thymopoietin (Tmpo) gene and Solute Carrier 25 (SC25) gene; wherein the first reporter is distinguishable from the second reporter.

21. A mixture of cells according to claim 20, wherein the cells are human embryonic stem cells.

22. A mixture of cells according to claims 20 and 21, wherein the first and second reporters are fluorescent proteins.

23. Use of a vector comprising a polynucleotide sequence homologous with the nucleotide sequence of SEQ ID NO:1 excluding nucleotides 2037 to 6095 of SEQ ID NO:1, and a promoter for insertion into a gene locus positioned between Thymopoietin (Tmpo) gene and Solute Carrier 25 (SC25) gene.

24 Use of a vector according to claim 23, wherein the promoter is a beta actin promoter or other promoter sequences.

25. Use of a vector according to claim 23 or 24, wherein the polynucleotide sequence encodes a reporter.

26. Use of a vector according to claim 25, wherein the reporter is a fluorescent protein.

27. Use of a vector according to claim 26, wherein the fluorescent protein is green fluorescent protein (GFP).

28. Use of a vector according to any one of claims 23 to 27, wherein the polynucleotide sequence encodes an internal ribosome binding site (IRES).

29. Use of a vector according to any one of claims 23 to 28, wherein the polynucleotide sequence comprises nucleotides 2037 to 6095 of SEQ ID NO:1.

30. A nucleotide sequence with at least 95% identity to sequence ID NO:1.

31. A derivative formed by the insertion of a polynucleotide sequence into any part of sequence ID NO:1.

FIGURE 1

LOCUS envy 8081 bp DNA

AGATCGCACCACTGCACTCCAGCCTGGGAGACAGAGAGAGACTCTATCTCAAAAAAAA
AAGATAAAAGATAAAAAAATACTACTAATAAAACAAAATCAGAAAGTGTGAAACCTC
CAACTTTGCTCTTCTTCAAGATTGTTTTTGCTATTCAGCTAAGACACAACTCACTTTTTT
TTGTTGTTGTTTTGGAGACAGGGTCTCACTCTTGCCTAGGCTGGAGTGAGGTGGTGCGA
TCTCAGCTCACTGCACTGCTGGGTTCAAGTGATCCTTCAACCTCAGCCTCTCAAGTAGC
TGGGACTACAGGTGTGTGCCACCATGCCCAGCTAATTAAGTATACAATTTCGTGACCAT
GAAGCAACACCTATTTAACSMCCWYCCTGATCAAGAAACACAACACTCCCCAGTTCCT
TCCCTTTCTCCTGCATCCTTTCCTGATTACAATTCCCTCCTTTCCACAGGTTATCAGCAG
TCTGATTTTTTTGGTAATTATTTTCTTCTCTTTATAGTTTTAACACTTAATTATGCATCCT
CAAACAATATAATTTTACCTATTTTTCAACTTGAATTTAAATCATATCATGTAATACTTT
TAACTTGCATATATTACTGAACATCATGAAATTCATCCATGCTGTGTAGCAGTAATTCA
CACATTACATATATTATATTTCTGGCTTTTGGTAACTATGAACAGTACTGCTATAAATTT
TCTTTACATGTGTTCTGGTCTGCAGGTGTAAGAGTTTACCTAGAGGCCATGCGCGGTGG
CTCATGCCTGTAATCCCAGCACTTCGGGAGGCCGAGGCAGGCCGATCACTTGAGTCCA
GGAATTCGAGACTAGCCTGGCCAACATGGCGAAACCCTGTCTCTACTAAAAATACAAA
AAATTAGCTAGGTGTGGTGGTGGTGTGCCTGTAGTCTCAGCTATTTGAGAGGGTGAGG
TARGGAGAATTGCTTGAACCTGGGAAGTGGAGGTTGCAATAAGCCGAAATGGCACCM
HCTGCWCTCMAWGTCTGGSCAAAAGARCAAGASTGTTWAAAAAAAAAAAACAAAATTT
AMCTAKKGTGTATATCTAAYYATGGWATWMYTCAATKGTKGGATMCWMKMTGCWG
AATAGAYATGTCAAACCGCTTTTCAGAGTTGTTCCAATTTATACTTCCACCAGTGCACC
TGGGTATCCCCCAAGGATCACTTGAGCCCAAAAGTTTGAGATCAGCCTGGGCAACATA
CTGAGACCCCCATCTGTATTTAAAAAAAAAAAAAATTCCAACAAATAAATGAAGAGAC
AATATTTTGCCAAAAAAGGACAAAAGTCTTTAACAAGCAGTTCACACATAGCTAATAA
ACATTAAAAAGTGTATGGCATGACTGGTAAGTAAACCACAGAGAGATATCATTACACA
CCCATCAGCCGGCTAAAAATTAAATCATGCAATAGCAAAAGAGTATGTGGGGCCAGGA
ACGGTGGCTCATGCCTGCAATTGCTTAATTGCTTATTATTATTATTCCATTATTCCACAT
GCTCTTTTGTTTTTTTGTTTGTTTATTGTTTTTTGAGATGGAGTCTCACTTTGTGGCCCAT
GCTGGAGTGCAGTGGCATGATCTCGGCTCACTGCAACCTCCACCTCCCAGGTTCAAGA
ATTCTCCTGCCTCAGTAGCTGGGATTACAGGTGCGTGCCACGACACCAAGCTAATTTTT
TTTTTTTTTTTTTTTTTTTTTTTTGAGATGGAGTCTCGCTCTTTTGCCCAGGCTGGAGTGC

AGTGGTGTGATCTTGGCTCACTRCAATCTCCACCTCCCAGGTTCAAGTGATTCTCCTGC

CTCAGCCTCCCGAGTAGCTGGAACTACAGGCACCTGCCACAGCATCCGGCTAATTTTTT

GGGTATTTTTAGTACAGGTGGGGTTTCAGCATGTTGGCCATGCTGGTCTCGAACTCCTG

ACCTCAAATGATCCACCCGCCTCAGCTGCCCAAAGTACTGGGATTACAGGTGTGAGCC

ACTGTTCCTGGCCCCACATACTCTTTTGCTATTGCATGATTGTAGCGTTGGCAGGTCCTG

AGGCAGCTGGCAAGACGCCTGCAGCTGAAAGATACAAGGCCAGGGACAGGACAGTCC

CATCCCCAGGAGGCAGGGAGTATACAGGCTGGGGAAGTTTGCCCTTGCGTGGGGTGGT

GATGGAGGAGGCTCAGCAAGTCTTCTGGACTGTGAACCTGTGTCTGCCACTGTGTGCTG

GGTGGTGGTCATCTTTCCCACCAGGCTGTGGCCTCTGCAACCTTCAAGGGAGGAGCAG

GTCCCATTGGCTGAGCACAGCCTTGTACCGTGAACTGGAACAAGCAGCCTCCTTCCTGG

CCACAGGTTCCATGTCCTTATATGGACTCATCTTTGCCTATTGCGACACACACTCAATG

AACACCTACTACGCGCTGCAAAGAGCCCCGCAGGCCTGAGGTGCCCCCACCTCACCAC

TCTTCCTATTTTTGTGTAAAAATCCAGCTTCTTGTCACCACCTCCAAGGAGGGGGAGGA

GGAGGAAGGCAGGTTCCTCTAGGCTGAGCCGAATGCCCCTCTGTGGTCCCACGCCACT

GATCGCTGCATGCCCACCACCTGGGTACACACAGTCTGTGATTCCCGGAGCAGAACGG

ACCCTGCCCACCCGGTCTTGTGTGCTACTCAGTGGACAGACCCAAGGCAAGAAAGGGT

GACAAGGACAGGGTCTTCCCAGGCTGGCTTTGAGTTCCTAGCACCGCCCCGCCCCCAA

TCCTCTGTGGCACATGGAGTCTTGGTCCCCAGAGTCCCCCAGCGGCCTCCAGATGGTCT

GGGAGGGCAGTTCAGCTGTGGCTGCGCATAGCAGACATACAACGGACGGTGGGCCCA

GACCCAGGCTGTGTAGACCCAGCCCCCCCGCCCCGCAGTGCCTAGGTCACCCACTAAC

GCCCCAGGCCTGGTCTTGGCTGGGCGTGACTGTTACCCTCAAAAGCAGGCAGCTCCAG

GGTAAAAGGTGCCCTGCCCTGTAGAGCCCACCTTCCTTCCCAGGGCTGCGGCTGGGTA

GGTTTGTAGCCTTCATCACGGGCCACCTCCAGCCACTGGACCGCTGGCCCCTGCCCTGT

CCTGGGGAGTGTGGTCCTGCGACTTCTAAGTGGCCGCAAGCCACCTGACTCCCCCAAC

ACCACACTCTACCTCTCAAGCCCAGGTCTCTCCCTAGTGACCCACCCAGCACATTTAGC

TAGCTGAGCCCCACAGCCAGAGGTCCTCAGGCCCTGCTTTCAGGGCAGTTGCTCTGAA

GTCGGCAAGGGGGAGTGACTGCCTGGCCACTCCATGCCCTCCAAGAGCTCCTTCTGCA

GGAGCGTACAGAACCCAGGGCCCTGGCACCCGTGCAGACCCTGGCCCACCCCACCTGG

GCGCTCAGTGCCCAAGAGATGTCCACACCTAGGATGTCCCGCGGTGGGTGGGGGGCCC

GAGAGACGGGCAGGCCGGGGGGCAGGCCTGGCCATGCGGGGCCGAACCGGGCACTGCC

CAGCGTGGGGCGCGGGGGCCACGGCGCGCGCCCCCAGCCCCCGGGCCCAGCACCCCA

AGGCGGCCAACGCCAAAACTCTCCCTCCTCCTCTTCCTCAATCTCGCTCTCGCTCTTTTT

TTTTTTCGCAAAAGGAGGGGAGAGGGGGTAAAAAAATGCTGCACTGTGCGGCGAAGC

CGGTGAGTGAGCGGCGCGGGGCCAATCAGCGTGCGCCGTTCCGAAAGTTGCCTTTTAT

GGCTCGAGCGGCCGCGGCGGCGCCCTATAAAACCCAGCGGCGCGACGCGCCACCACC

GCCGAGACCGCGTCCGCCCCGCGAGCACAGAGCCTCGCCTTTGCCGATCCGCCGCCCG

TCCACACCCGCCGCCAGGTAAGCCCGGCCAGCCGACCGGGGCATGCGGCCGCGGCCCC

TTCGCCCGTGCAGAGCCGCCGTCTGGGCCGCAGCGGGGGGCGCATGGGGGGGGAACC

GGACCGCCGTGGGGGGCGCGGGAGAAGCCCCTGGGCCTCCGGAGATGGGGGACACCC

CACGCCAGTTCGGAGGCGCGAGGCCGCGCTCGGGAGGCGCGCTCCGGGGGTGCCGCTC

TCGGGGCGGGGGCAACCGGCGGGGTCTTTGTCTGAGCCGGGCTCTTGCCAATGGGGAT

CGCAGGGTGGGCGCGGCGKAGCCCCCGCCAGGCCCGGTGGGGGCTGGGGCGCCATTG

CGCGTGCGCGCTGGTCCTTTGGGCGCTAACTGCGTGCGCGCTGGGAATTGGCGCTAATT

GCGCGTGCGCGCTGGGACTCAAGGCGCTAAYTGCGCGTGCGTTCTGGGGCCCGGGGTG

CCGCGGCCTGGGCTGGGGCGAAGGCGGGCTCGGCCGGAAGGGGTGGGGTCGCCGCGG

CTCCCGGGCGCTTGCGCGCACTTCCTGCCCGAGCCGCTGGCCGCCCGAGGGTGTGGCC

GCTGCGTGCGCGCGCGCCGACCCGGCGCTGTTTGAACCGGGCGGAGGCGGGGCTGGCG

CCCGGTTGGGAGGGGGTTGGGGCCTGGCTTCCTGCCGCGCGCCGCGGGGACGCCTCCG

ACCAGTGTTTGCCTTTTATGGTAATAACGCGGCCGGCCCGGCTTCCTTTGTCCCCAATC

TGGGCGCGCGCCGGCGCCCCCTGGCGGCCTAAGGACTCGGCGCGCCGGAAGTGGCCAG

GGCGGGGGCGACCTCGGCTCACAGCGCGCCCGGCTATTCTCGCAGCTCACSTCGAYGG

TYWACGAATTCCCCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTG

CCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCTGGCG

AGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGG

CAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCT

TCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGA

AGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGC

GCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATC

GACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGC

CACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGCGAACTTCAAG

ATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACA

CCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTC

CGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTG

ACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAATGAATTAATTAAG

AATTATCACCGCTTCTATTCAGCCAGTAAGGCCTGTCTTTAATGGCCTCCGGCATGAGA

CACTTTCTAGAATTCGGTACCCCCCCCCCCCAACGTTACTGGCCGAAGCCGCTTGGAATA

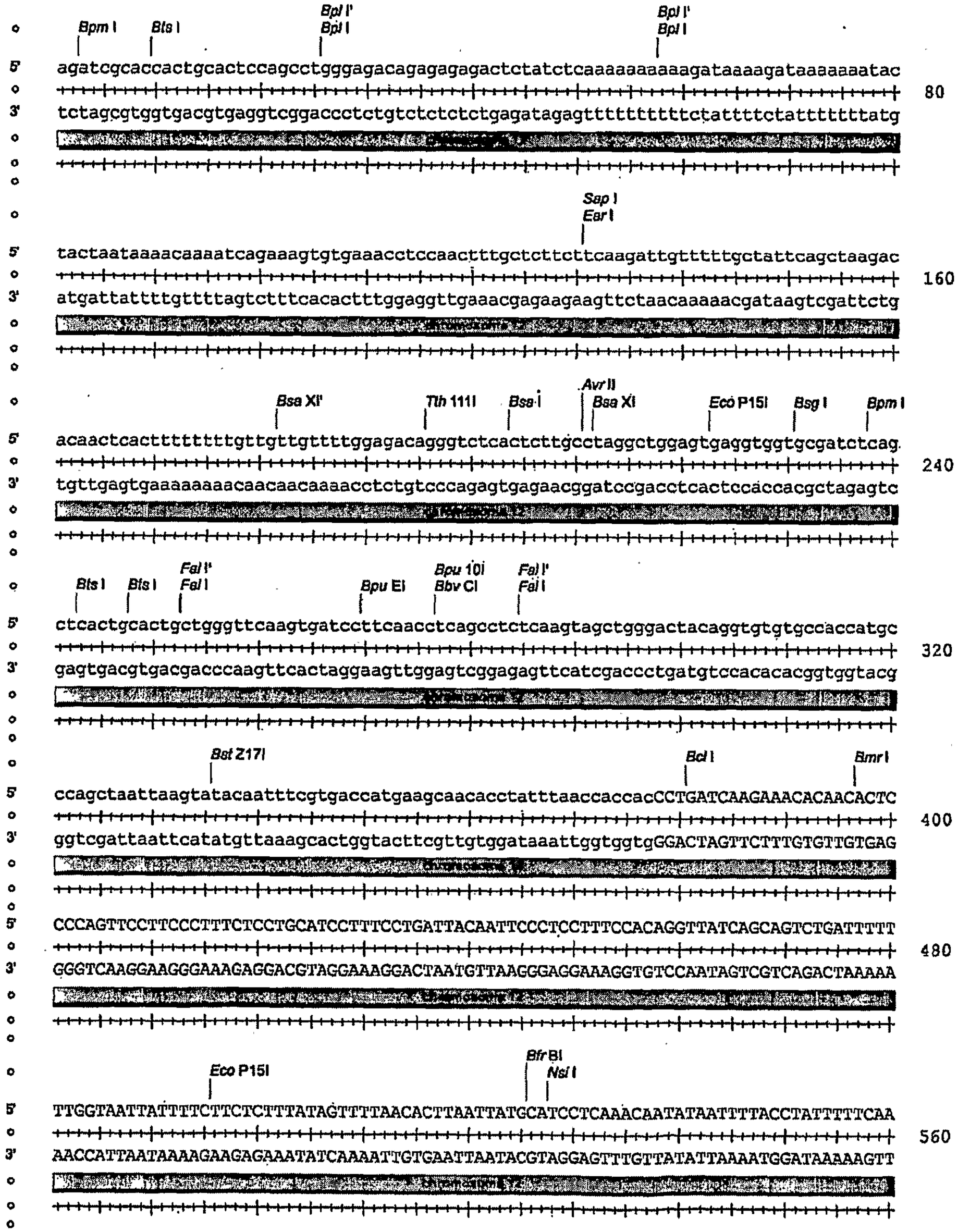
AGGCCGGTGTGCGTTTGTCTATATGTTATTTTCCACCATATTGCCGTCTTTTGGCAATGT
GAGGGCCCGGAAACCTGGCCCTGTCTTCTTGACGAGCATTCCTAGGGGTCTTTCCCCTC
TCGCCAAAGGAATGCAAGGTCTGTTGAATGTCGTGAAGGAAGCAGTTCCTCTGGAAGC
TTCTTGAAGACAAACAACGTCTGTAGCGACCCTTTGCAGGCAGCGGAACCCCCCACCT
GGCGACAGGTGCCTCTGCGGCCAAAAGCCACGTGTATAAGATACACCTGCAAAGGCGG
CACAACCCCAGTGCCACGTTGTGAGTTGGATAGTTGTGGAAAGAGTCAAATGGCTCTC
CTCAAGCGTATTCAACAAGGGGCTGAAGGATGCCCAGAAGGTACCCCATTGTATGGGA
TCTGATCTGGGGCCTCGGTGCACATGCTTTACATGTGTTTAGTCGAGGTTAAAAAAACG
TCTAGGCCCCCCGAACCACGGGGACGTGGTTTCCTAATTTTTAGCAGGCTGATGGGTGT
GTAATGATATCTCTCTGTGGTTAATTACCAGTCATGCCATACGCTTTTTAATGTTTATTA
GCTATGTGTGAACTGCTTGTTAAAGACTTTTGCCCATTGTTGGCAAAAGGTTGTCTCTT
CGTTTATTTGTTGGAATTTTTTGTTTGTTTGTTTGTTTTTAAATACAGGTGGGGGTCTCA
GTATGTTACCCAGGCTGGTCTCAAACTTCTGGGCTCAAGTGATCCTCCTGCCTCTGCCT
TCTCAAAGAACTGGGATTACAGGCATGAGCCATTATGCCCAGCCTATTGGTAGAAGTT
CTTTATATGTTTTGAATATTTATCGCACATTTAGGCATCCAGTTTTGTGAAAGCCCTCAG
TTTTGGATCACTTTTATATTAGTTTTTTTTTTAGATCAGTTTGTAGAAGTTCTTTATATAT
TTTTGATATGACTGCTTTATCAGACCCATGTACAGCAAATATTTTCTCCCACTCTTTGGT
TTGCCTTTCCACTCTTTCGGTGATGTCTTTTGGTGAACAGAAGTTTTTAATTTTACTACA
GTCCCATGTATGAATTGTTTTAATTATGATTACTGCCTTTCAAAAACATTTAAGAAAGT
TTTACCTACTCAAAGTTTATGAAAATATTCTTCTCCTATGCTCTCTTCTAAAATATGTAT
TGCTTTACCTTTCCAATTTTAAGGATAATACAACTGAAACTAACTTTTTTTAAAAGTAA
CTTTTAGGCTGGGCGTGGTGGCTCACGCCTGTAATCCCAGTACTTCAGGAGGCCGACAT
TGGGGATATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGGTGAAACTC
CATATCTACTAAAAACAAAACTTAGCCAGGTGTGGTGGCAGACACCTGTAATCCCAGC
TACTCGGGAGGCTGAGGCAGGAGAATCACTTGAACCCAGGAGGCGGAGGTTGCAGTG
AGCCGAGATTATGCCATTGCACTCCAGCCTGGGTGACAGAGCAAGACTCCATCTCAAA
AAGAAATAAATAAATAAAAAATAAAAGTAACTTTTAAATGATTGTGAGGTAGGGATCA
AGTTTAATGTTTTTCCATAGGGATATTCAATTGGCCCATCCCAAAAAGACCGGCCTTTC
TCCACTACACAGCAGTGTTGACTTTGTCATCTATCAGGTGACCCTTCAGCCTGTCTGTG
GACTCTTTGTTGTGTTCCATTGGCCTGTTTGTCTATTGTTGGAACAAAACCACACTGTCT
GAATTGCCACAGCTTTATAATAGGGCATGCTATCAAGCAGTCAGTGTCATTTTACAGGT
TTGTTCTTCTTTAGGTTTGCCTTAGCTATTCTTTGGTCTTTGCATTGCCATCTAAATTTTA
GAATCAACTAAGTCAAAATGTTGTATCACTTTCCACACCAAAAATCTGCTGGAATTTTG

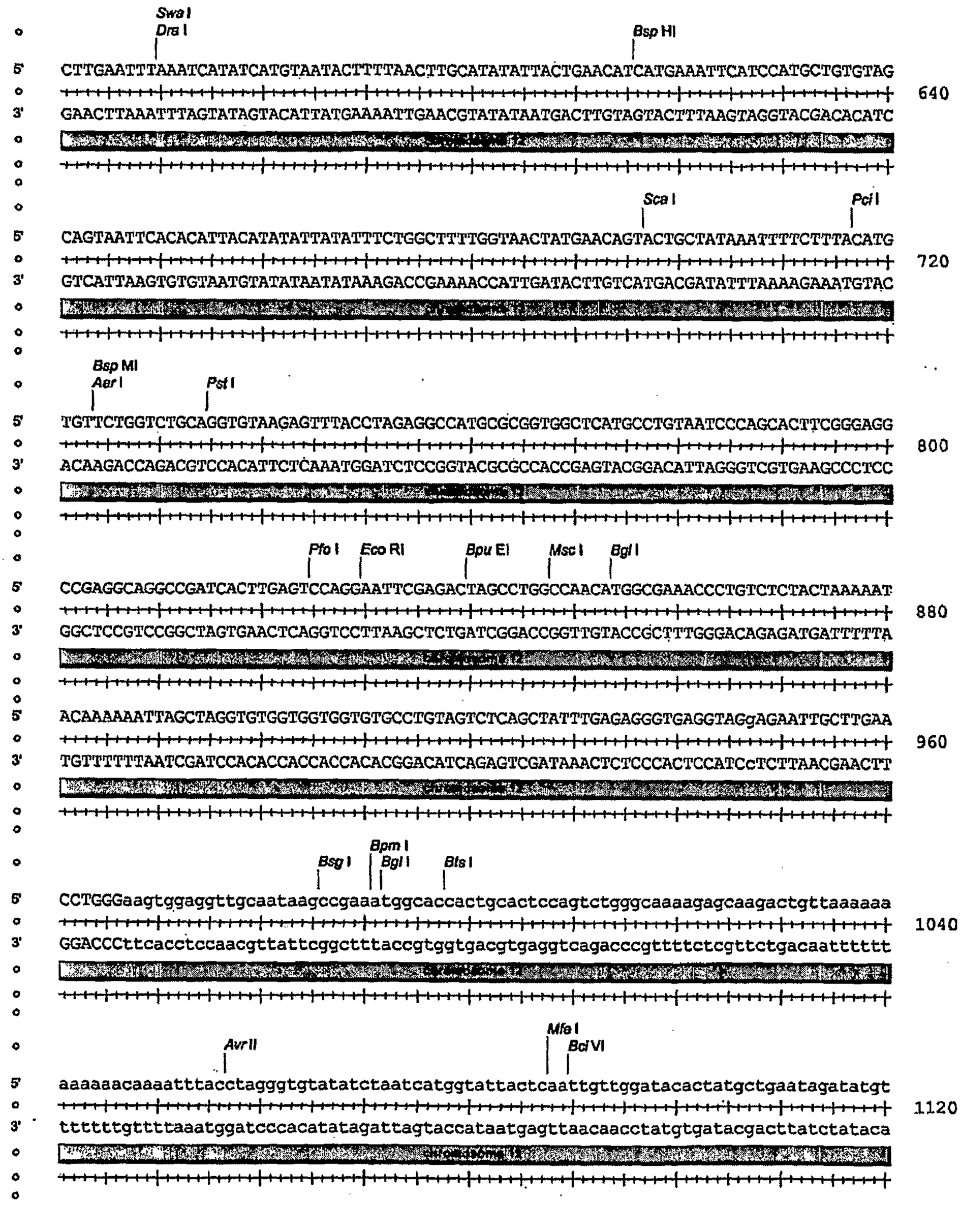
ATTGGGATTATATTAAATTTGTAGATCAATTTGGGAAGAAATAAGATCTTTACAACATT
AACTCTTCCAATCTATAAATGTGGTATTTATTTTTTCTTTATTTTTTTCTCAGTGGTTTGT
AGTTTCTAGTGCAAAGGTCTTTGACATCGTTAGATTTAATGTTTTTTGATATATTTTGGT
GTTATAAGTGATACTTTTCAAAATTTCATTTTCTATTTGTTTCTTGTTGACACAGCAACA
CAACTGGTTTTAGTGTAATAATCTTATATTTAGAGCCCTTGCTAATTAAACCTATTAATT
ATAATAGTTCTTCTATAGATTCTTTCAGCTTTTTTGCATATACGATCATGTCATCTGTGA
GTAATGACATGAACCCAGGAGGCGTTCTTTCCAGGCTTTCTTTTTATAGTCATGTACTG
CTTAACGACGTTTTTGTCGATGATGAACTGCATATACAACAGTGGTCCCATAAGATTAT
AAAGAGGCTGAAAAATTCTTACTGCCTAGTGACACTGTAGCCATCATAATGTAGCACA
ATTCAGTAC

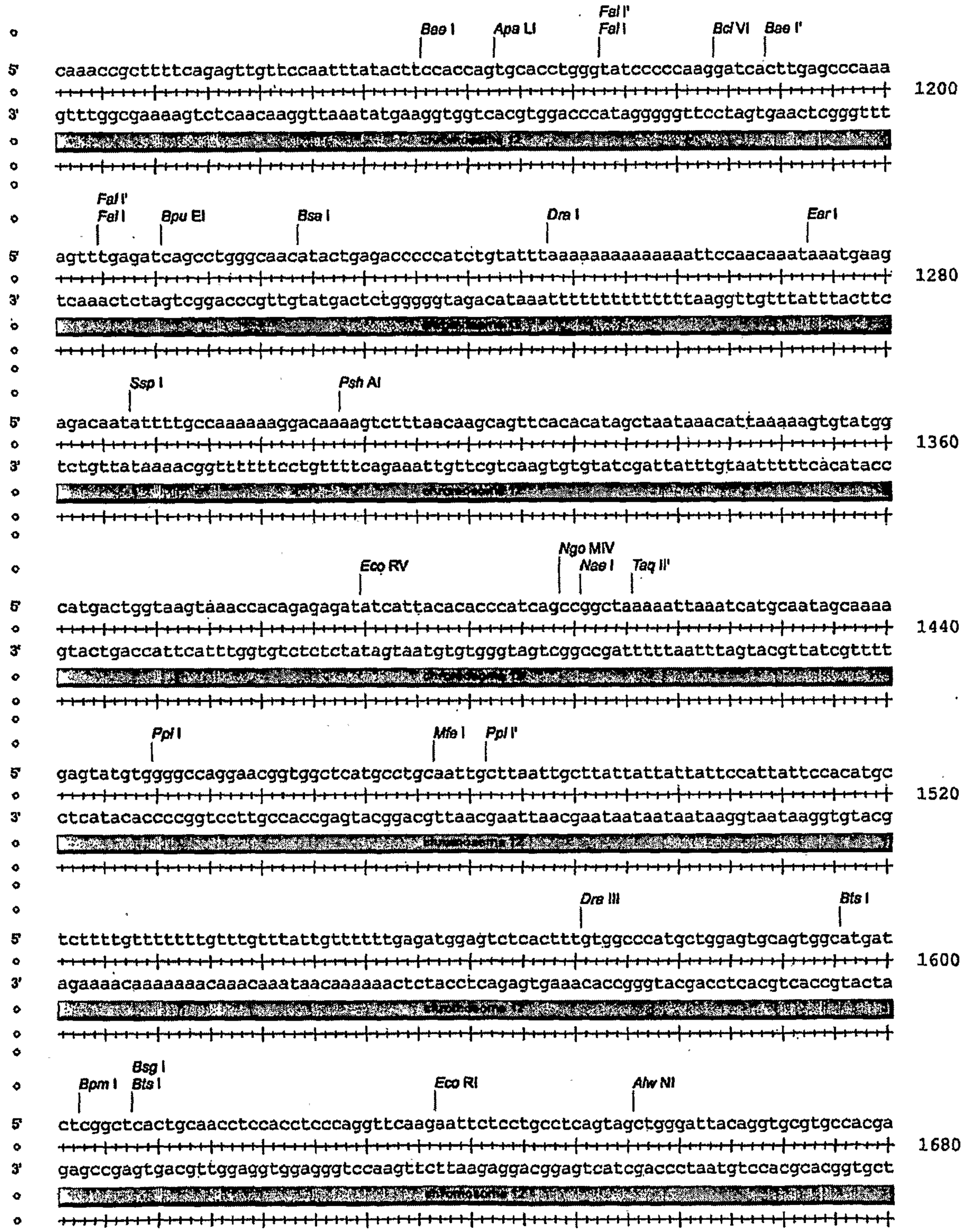
Legend to sequence:

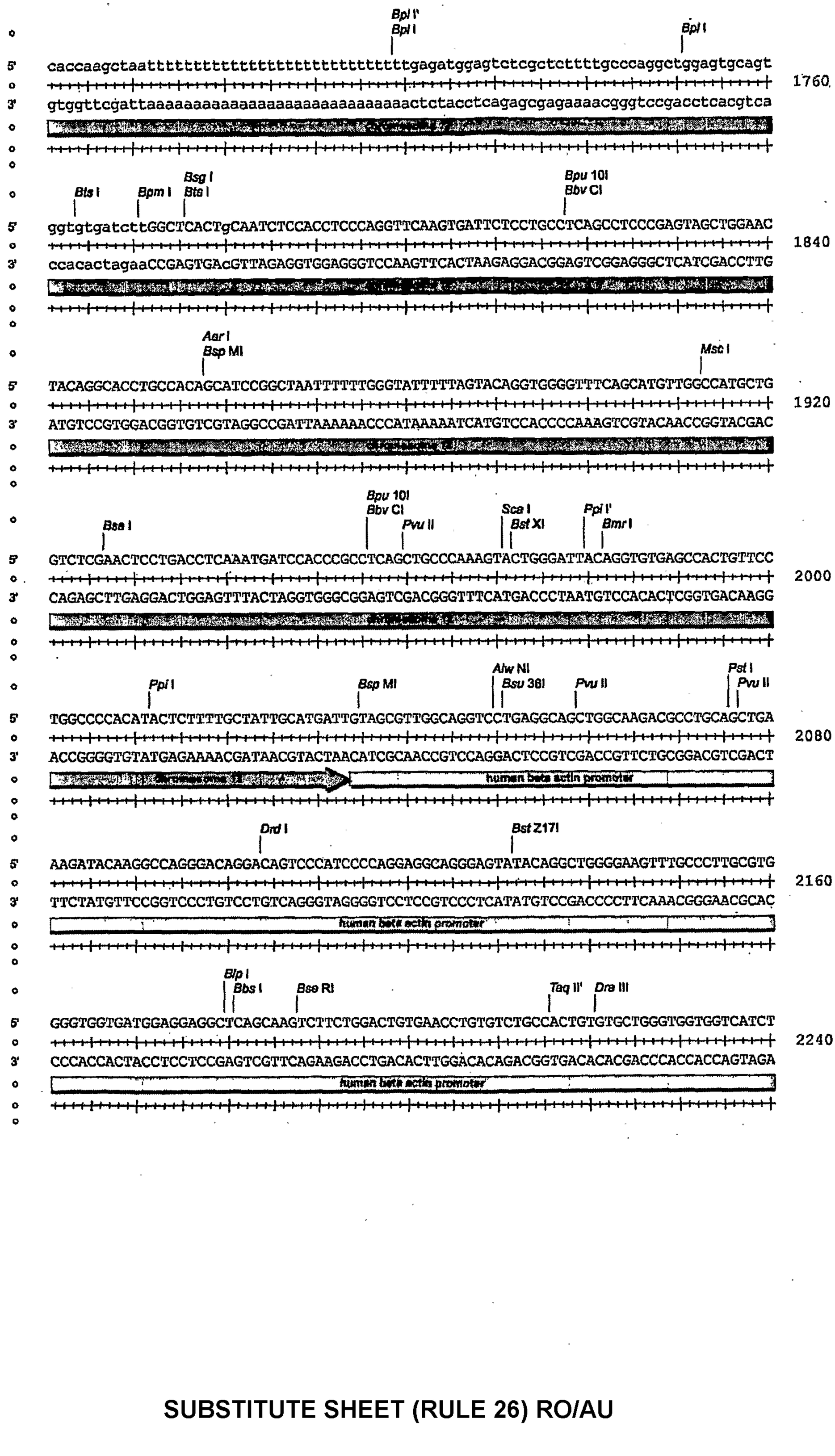
Symbol	Meaning
G	G
A	A
T	T
C	C
R	G or A
Y	T or C
M	A or C
K	G or T
S	G or C
W	A or T
H	A or C or T
B	G or T or C
V	G or C or A
D	G or A or T
N	G or A or T or C

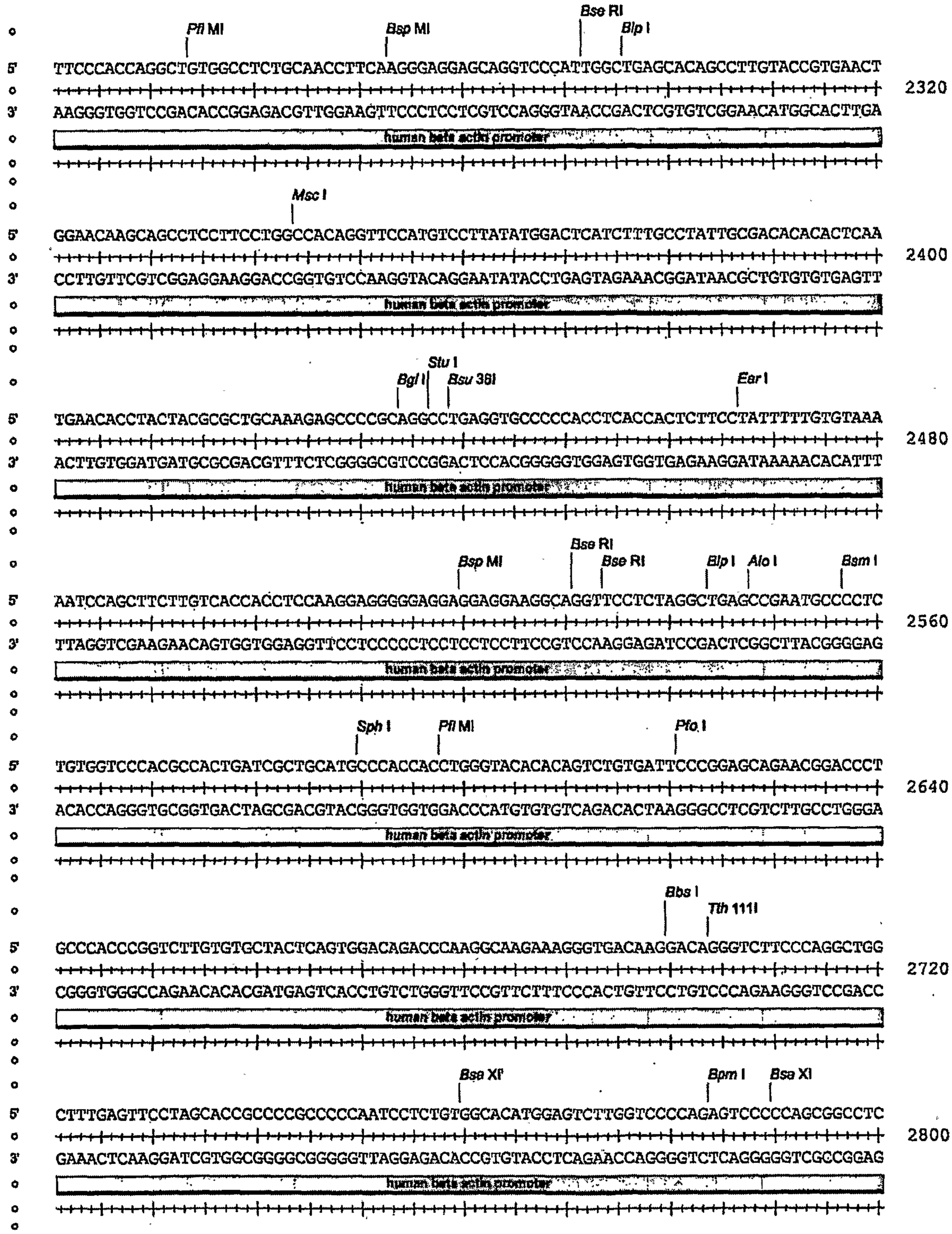
FIGURE 2

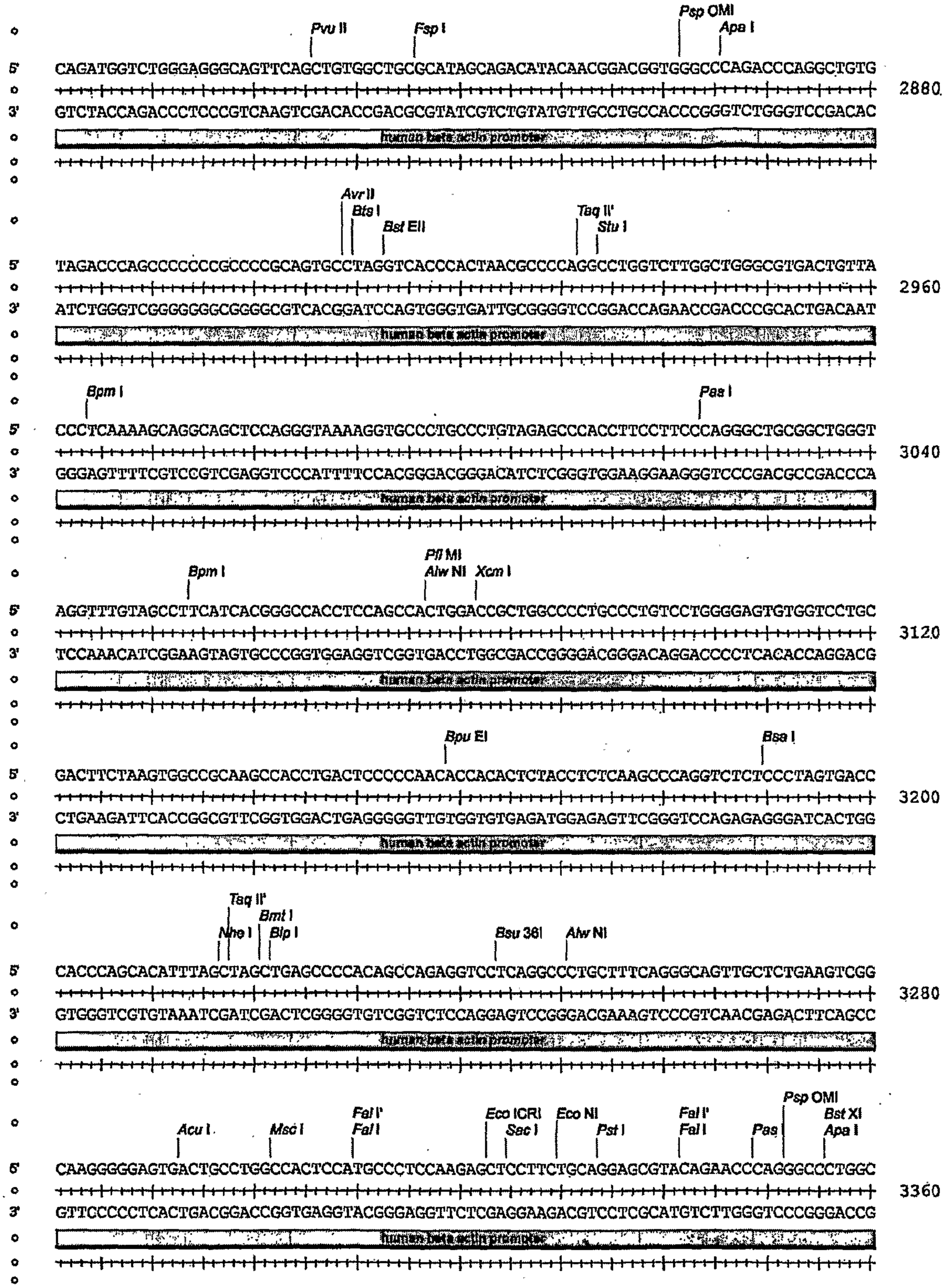

Bpm I
Bts I
Bpl I'
Bpl I
Bpl I'
Bpl I
5' agatcgcaccactgcactccagcctgggagacagagagactctatctcaaaaaaaaaagataaaagataaaaaaatac
3' tctagcgtggtgacgtgaggtcggaccctctgtctctctctgagatagagttttttttttctattttctatttttttatg
80
Sap I
Ear I
5' tactaataaaacaaaatcagaaagtgtgaaacctccaactttgctcttcttcaagattgtttttgctattcagctaagac
3' atgattattttgttttagtctttcacactttggaggttgaaacgagaagaagttctaacaaaaacgataagtcgattctg
160
Avr II
Bsa XI'
Tth 111I
Bsa I
Bsa XI
Eco P15I
Bsg I
Bpm I
5' acaactcacttttttttgttgttgttttggagacagggtctcactcttgcctaggctggagtgaggtggtgcgatctcag
3' tgttgagtgaaaaaaaacaacaacaaaacctctgtcccagagtgagaacggatccgacctcactccaccacgctagagtc
240
Fal I'
Fal I
Bts I
Bts I
Bpu EI
Bpu 10I
Bbv CI
Fal I'
Fal I
5' ctcactgcactgctgggttcaagtgatccttcaacctcagcctctcaagtagctgggactacaggtgtgtgccaccatgc
3' gagtgacgtgacgacccaagttcactaggaagttggagtcggagagttcatcgaccctgatgtccacacacggtggtacg
320
Bst Z17I
Bcl I
Bmr I
5' ccagctaattaagtatacaatttcgtgaccatgaagcaacacctatttaaccaccacCCTGATCAAGAAACACAACACTC
3' ggtcgattaattcatatgttaaagcactggtacttcgttgtggataaattggtggtgGGACTAGTTCTTTGTGTTGTGAG
400
5' CCCAGTTCCTTCCCTTTCTCCTGCATCCTTTCCTGATTACAATTCCCTCCTTTCCACAGGTTATCAGCAGTCTGATTTTT
3' GGGTCAAGGAAGGGAAAGAGGACGTAGGAAAGGACTAATGTTAAGGGAGGAAAGGTGTCCAATAGTCGTCAGACTAAAAA
480
Bfr BI
Nsi I
Eco P15I
5' TTGGTAATTATTTTCTTCTCTTTATAGTTTTAACACTTAATTATGCATCCTCAAACAATATAATTTTACCTATTTTTCAA
3' AACCATTAATAAAAGAAGAGAAATATCAAAATTGTGAATTAATACGTAGGAGTTTGTTATATTAAAATGGATAAAAAGTT
560

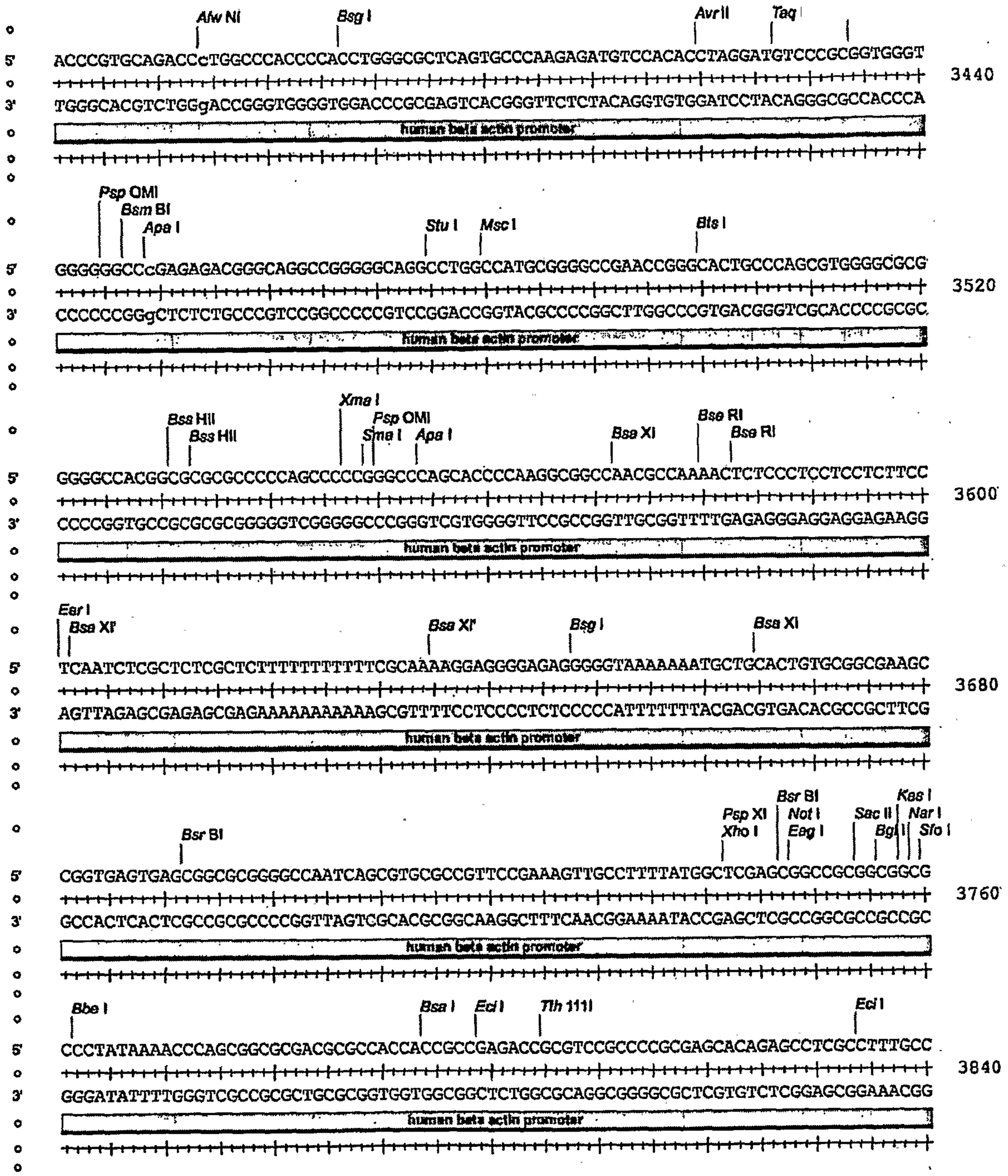

SwaI
DraI
BspHI
5' CTTGAATTTAAATCATATCATGTAATACTTTTAACTTGCATATATTACTGAACATCATGAAATTCATCCATGCTGTGTAG
640
3' GAACTTAAATTTAGTATAGTACATTATGAAAATTGAACGTATATAATGACTTGTAGTACTTTAAGTAGGTACGACACATC
ScaI
PciI
5' CAGTAATTCACACATTACATATATTATATTTCTGGCTTTTGGTAACTATGAACAGTACTGCTATAAATTTTCTTTACATG
720
3' GTCATTAAGTGTGTAATGTATATAATATAAAGACCGAAAACCATTGATACTTGTCATGACGATATTTAAAAGAAATGTAC
BspMI
AarI
PstI
5' TGTTCTGGTCTGCAGGTGTAAGAGTTTACCTAGAGGCCATGCGCGGTGGCTCATGCCTGTAATCCCAGCACTTCGGGAGG
800
3' ACAAGACCAGACGTCCACATTCTCAAATGGATCTCCGGTACGCGCCACCGAGTACGGACATTAGGGTCGTGAAGCCCTCC
PfoI
EcoRI
BpuEI
MscI
BglI
5' CCGAGGCAGGCCGATCACTTGAGTCCAGGAATTCGAGACTAGCCTGGCCAACATGGCGAAACCCTGTCTCTACTAAAAAT
880
3' GGCTCCGTCCGGCTAGTGAACTCAGGTCCTTAAGCTCTGATCGGACCGGTTGTACCGCTTTGGGACAGAGATGATTTTTA
5' ACAAAAAATTAGCTAGGTGTGGTGGTGGTGTGCCTGTAGTCTCAGCTATTTGAGAGGGTGAGGTAGgAGAATTGCTTGAA
960
3' TGTTTTTTAATCGATCCACACCACCACCACACGGACATCAGAGTCGATAAACTCTCCCACTCCATCcTCTTAACGAACTT
BpmI
BsgI
BglI
BfsI
5' CCTGGGaagtggaggttgcaataagccgaaatggcaccactgcactccagtctgggcaaaagagcaagactgttaaaaaa
1040
3' GGACCCttcacctccaacgttattcggctttaccgtggtgacgtgaggtcagacccgttttctcgttctgacaatttttt
MfeI
AvrII
BclVI
5' aaaaaacaaaatttacctagggtgtatatctaatcatggtattactcaattgttggatacactatgctgaatagatatgt
1120
3' ttttttgttttaaatggatcccacatatagattagtaccataatgagttaacaacctatgtgatacgacttatctataca

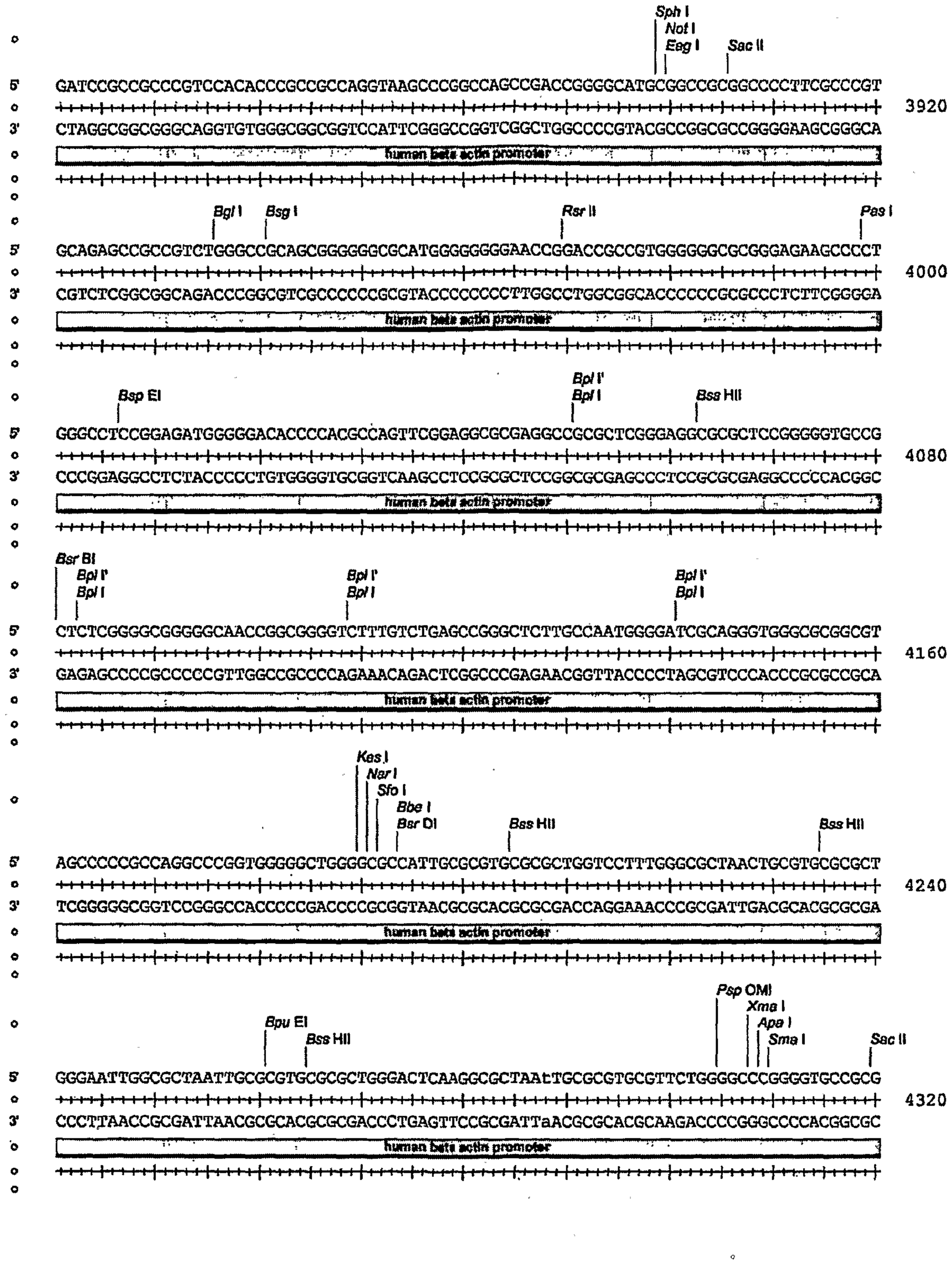

Bae I
Apa LI
Fal I'
Fal I
Bcl VI
Bae I'
5' caaaccgcttttcagagttgttccaatttatacttccaccagtgcacctgggtatcccccaaggatcacttgagcccaaa
1200
3' gtttggcgaaaagtctcaacaaggttaaatatgaaggtggtcacgtggacccatagggggttcctagtgaactcgggttt
Fal I'
Fal I
Bpu EI
Bsa I
Dra I
Ear I
5' agtttgagatcagcctgggcaacatactgagacccccatctgtatttaaaaaaaaaaaaaattccaacaaataaatgaag
1280
3' tcaaactctagtcggacccgttgtatgactctgggggtagacataaatttttttttttttttaaggttgtttatttacttc
Ssp I
Psh AI
5' agacaatattttgccaaaaaaggacaaaagtctttaacaagcagttcacacatagctaataaacattaaaaaagtgtatgg
1360
3' tctgttataaaacggtttttttcctgttttcagaaattgttcgtcaagtgtgtatcgattatttgtaatttttcacatacc
Eco RV
Ngo MIV
Nae I
Taq II'
5' catgactggtaagtaaaccacagagagatatcattacacacccatcagccggctaaaaattaaatcatgcaatagcaaaa
1440
3' gtactgaccattcatttggtgtctctctatagtaatgtgtgggtagtcggccgatttttaatttagtacgttatcgtttt
Ppl I
Mfe I
Ppl I'
5' gagtatgtggggccaggaacggtggctcatgcctgcaattgcttaattgcttattattattattccattattccacatgc
1520
3' ctcatacaccccggtccttgccaccgagtacggacgttaacgaattaacgaataataataataaggtaataaggtgtacg
Dra III
Bts I
5' tcttttgtttttttgtttgtttattgtttttttgagatggagtctcactttgtggcccatgctggagtgcagtggcatgat
1600
3' agaaaacaaaaaaacaaacaaataacaaaaaactctacctcagagtgaaacaccgggtacgacctcacgtcaccgtacta
Bpm I
Bsg I
Bts I
Eco RI
Alw NI
5' ctcggctcactgcaacctccacctcccaggttcaagaattctcctgcctcagtagctgggattacaggtgcgtgccacga
1680
3' gagccgagtgacgttggaggtggagggtccaagttcttaagaggacggagtcatcgaccctaatgtccacgcacggtgct

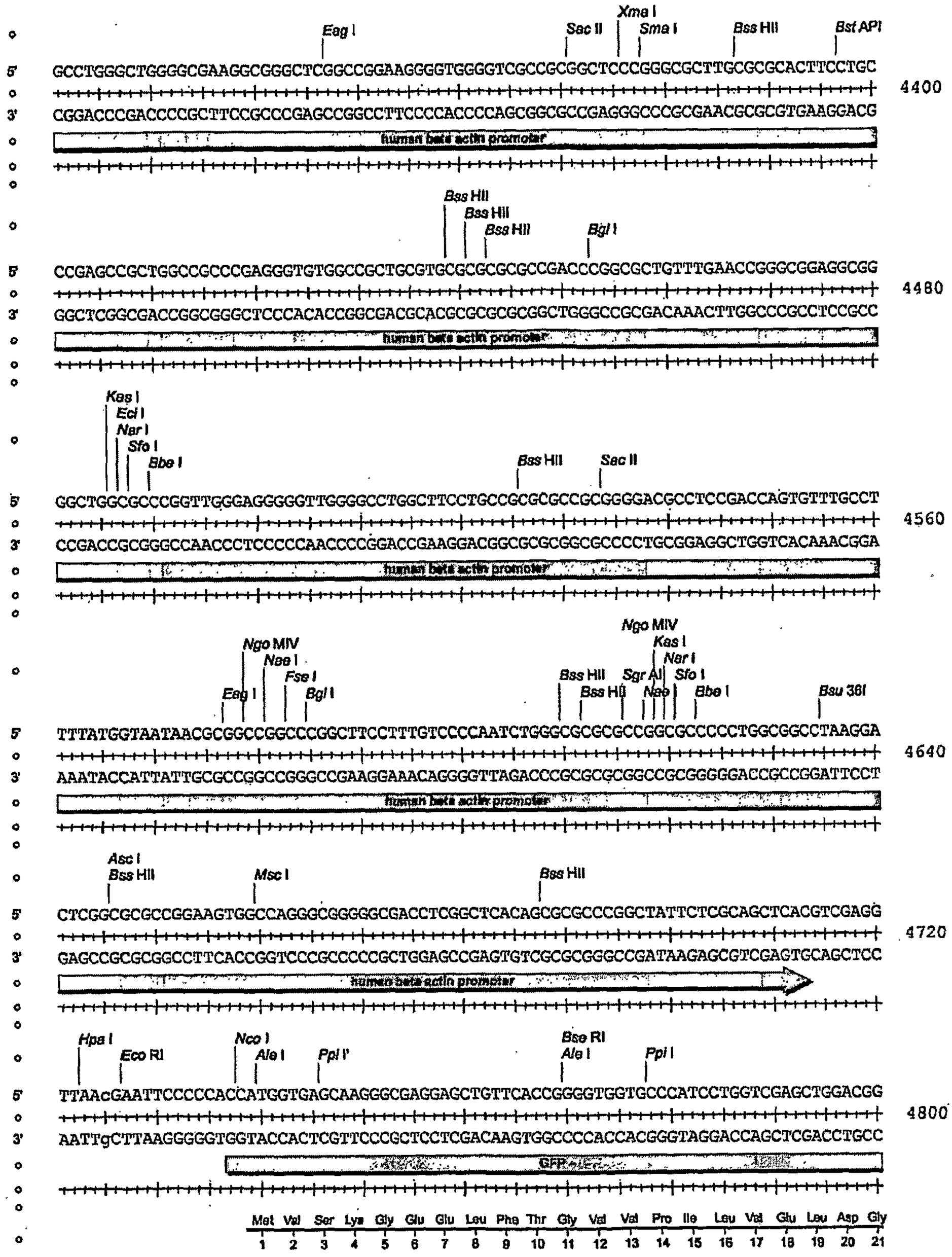

BplI'
BplI
BplI
5' caccaagctaattttttttttttttttttttttttttttgagatggagtctcgctcttttgcccaggctggagtgcagt
1760
3' gtggttcgattaaaaaaaaaaaaaaaaaaaaaaaaaaaactctacctcagagcgagaaaacgggtccgacctcacgtca
BtsI
BpmI
BsgI
BtsI
Bpu10I
BbvCI
5' ggtgtgatcttGGCTCACTgCAATCTCCACCTCCCAGGTTCAAGTGATTCTCCTGCCTCAGCCTCCCGAGTAGCTGGAAC
1840
3' ccacactagaaCCGAGTGAcGTTAGAGGTGGAGGGTCCAAGTTCACTAAGAGGACGGAGTCGGAGGGCTCATCGACCTTG
AsrI
BspMI
MscI
5' TACAGGCACCTGCCACAGCATCCGGCTAATTTTTTGGGTATTTTTAGTACAGGTGGGGTTTCAGCATGTTGGCCATGCTG
1920
3' ATGTCCGTGGACGGTGTCGTAGGCCGATTAAAAAACCCATAAAAATCATGTCCACCCCAAAGTCGTACAACCGGTACGAC
Bpu10I
BbvCI
ScaI
PpiI'
BseI
PvuII
BstXI
BmrI
5' GTCTCGAACTCCTGACCTCAAATGATCCACCCGCCTCAGCTGCCCAAAGTACTGGGATTACAGGTGTGAGCCACTGTTCC
2000
3' CAGAGCTTGAGGACTGGAGTTTACTAGGTGGGCGGAGTCGACGGGTTTCATGACCCTAATGTCCACACTCGGTGACAAGG
AlwNI
PstI
PpiI
BspMI
BsuMI
PvuII
PvuII
5' TGGCCCCACATACTCTTTTGCTATTGCATGATTGTAGCGTTGGCAGGTCCTGAGGCAGCTGGCAAGACGCCTGCAGCTGA
2080
3' ACCGGGGTGTATGAGAAAACGATAACGTACTAACATCGCAACCGTCCAGGACTCCGTCGACCGTTCTGCGGACGTCGACT
human beta actin promoter
DrdI
BstZ17I
5' AAGATACAAGGCCAGGGACAGGACAGTCCCATCCCCAGGAGGCAGGGAGTATACAGGCTGGGGAAGTTTGCCCTTGCGTG
2160
3' TTCTATGTTCCGGTCCCTGTCCTGTCAGGGTAGGGGTCCTCCGTCCCTCATATGTCCGACCCCTTCAAACGGGAACGCAC
human beta actin promoter
BlpI
BbsI
BseRI
TaqII'
DraIII
5' GGGTGGTGATGGAGGAGGCTCAGCAAGTCTTCTGGACTGTGAACCTGTGTCTGCCACTGTGTGCTGGGTGGTGGTCATCT
2240
3' CCCACCACTACCTCCTCCGAGTCGTTCAGAAGACCTGACACTTGGACACAGACGGTGACACACGACCCACCACCAGTAGA
human beta actin promoter

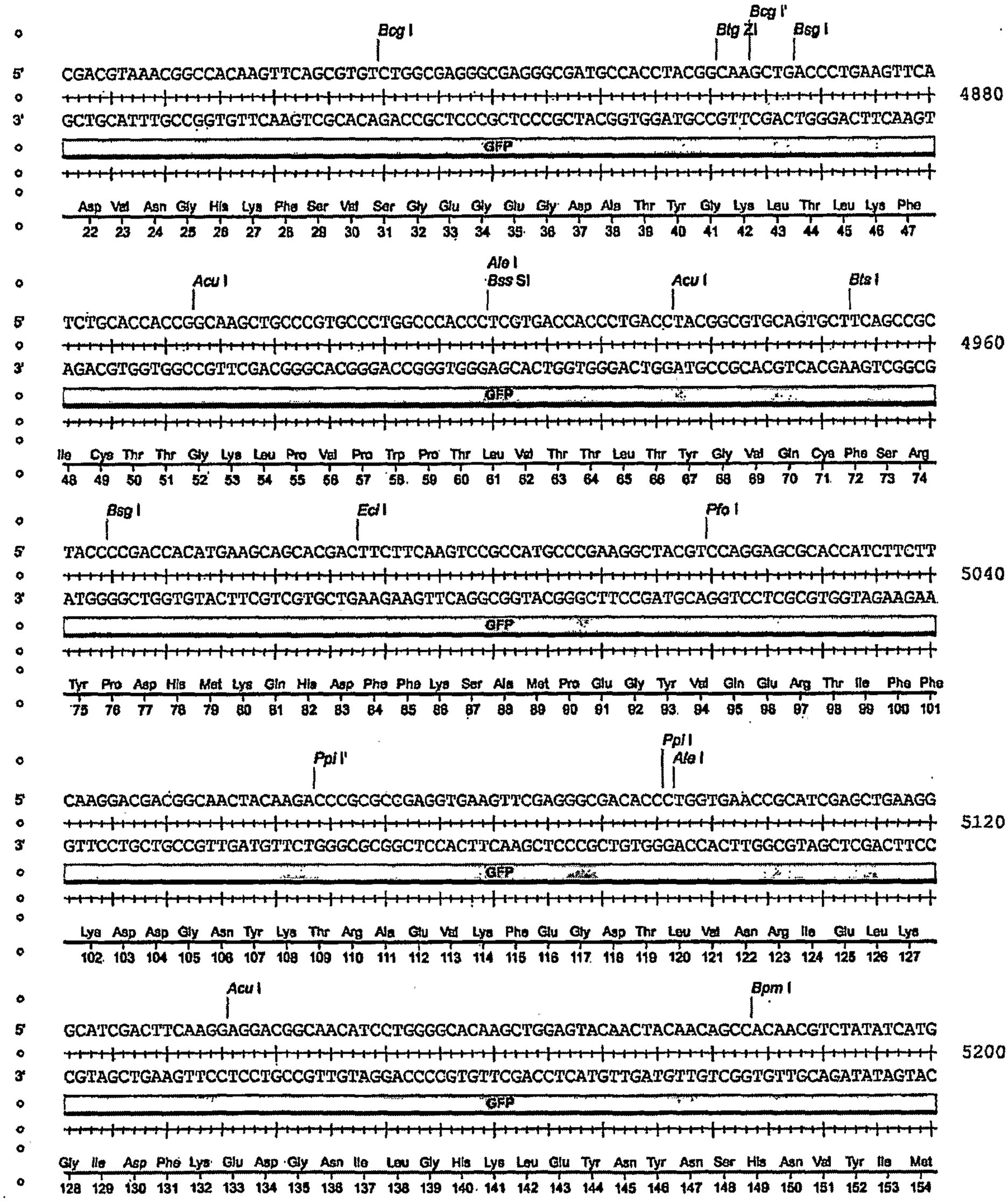

Pfl MI
Bsp MI
Bse RI
Blp I
5' TTCCCACCAGGCTGTGGCCTCTGCAACCTTCAAGGGAGGAGCAGGTCCCATTGGCTGAGCACAGCCTTGTACCGTGAACT
2320
3' AAGGGTGGTCCGACACCGGAGACGTTGGAAGTTCCCTCCTCGTCCAGGGTAACCGACTCGTGTCGGAACATGGCACTTGA
human beta actin promoter
Msc I
5' GGAACAAGCAGCCTCCTTCCTGGCCACAGGTTCCATGTCCTTATATGGACTCATCTTTGCCTATTGCGACACACACTCAA
2400
3' CCTTGTTCGTCGGAGGAAGGACCGGTGTCCAAGGTACAGGAATATACCTGAGTAGAAACGGATAACGCTGTGTGTGAGTT
human beta actin promoter
Stu I
Bgl I
Bsu 36I
Ear I
5' TGAACACCTACTACGCGCTGCAAAGAGCCCCGCAGGCCTGAGGTGCCCCCACCTCACCACTCTTCCTATTTTTGTGTAAA
2480
3' ACTTGTGGATGATGCGCGACGTTTCTCGGGGCGTCCGGACTCCACGGGGGTGGAGTGGTGAGAAGGATAAAAACACATTT
human beta actin promoter
Bsp MI
Bse RI
Bse RI
Blp I
Alo I
Bsm I
5' AATCCAGCTTCTTGTCACCACCTCCAAGGAGGGGGAGGAGGAGGAAGGCAGGTTCCTCTAGGCTGAGCCGAATGCCCCTC
2560
3' TTAGGTCGAAGAACAGTGGTGGAGGTTCCTCCCCCTCCTCCTCCTTCCGTCCAAGGAGATCCGACTCGGCTTACGGGGAG
human beta actin promoter
Sph I
Pfl MI
Pfo I
5' TGTGGTCCCACGCCACTGATCGCTGCATGCCCACCACCTGGGTACACACAGTCTGTGATTCCCGGAGCAGAACGGACCCT
2640
3' ACACCAGGGTGCGGTGACTAGCGACGTACGGGTGGTGGACCCATGTGTGTCAGACACTAAGGGCCTCGTCTTGCCTGGGA
human beta actin promoter
Bbs I
Tth 111I
5' GCCCACCCGGTCTTGTGTGCTACTCAGTGGACAGACCCAAGGCAAGAAAGGGTGACAAGGACAGGGTCTTCCCAGGCTGG
2720
3' CGGGTGGGCCAGAACACACGATGAGTCACCTGTCTGGGTTCCGTTCTTTCCCACTGTTCCTGTCCCAGAAGGGTCCGACC
human beta actin promoter
Bsa XI
Bpm I
Bsa XI
5' CTTTGAGTTCCTAGCACCGCCCCGCCCCCAATCCTCTGTGGCACATGGAGTCTTGGTCCCCAGAGTCCCCCAGCGGCCTC
2800
3' GAAACTCAAGGATCGTGGCGGGGCGGGGGTTAGGAGACACCGTGTACCTCAGAACCAGGGGTCTCAGGGGGTCGCCGGAG
human beta actin promoter

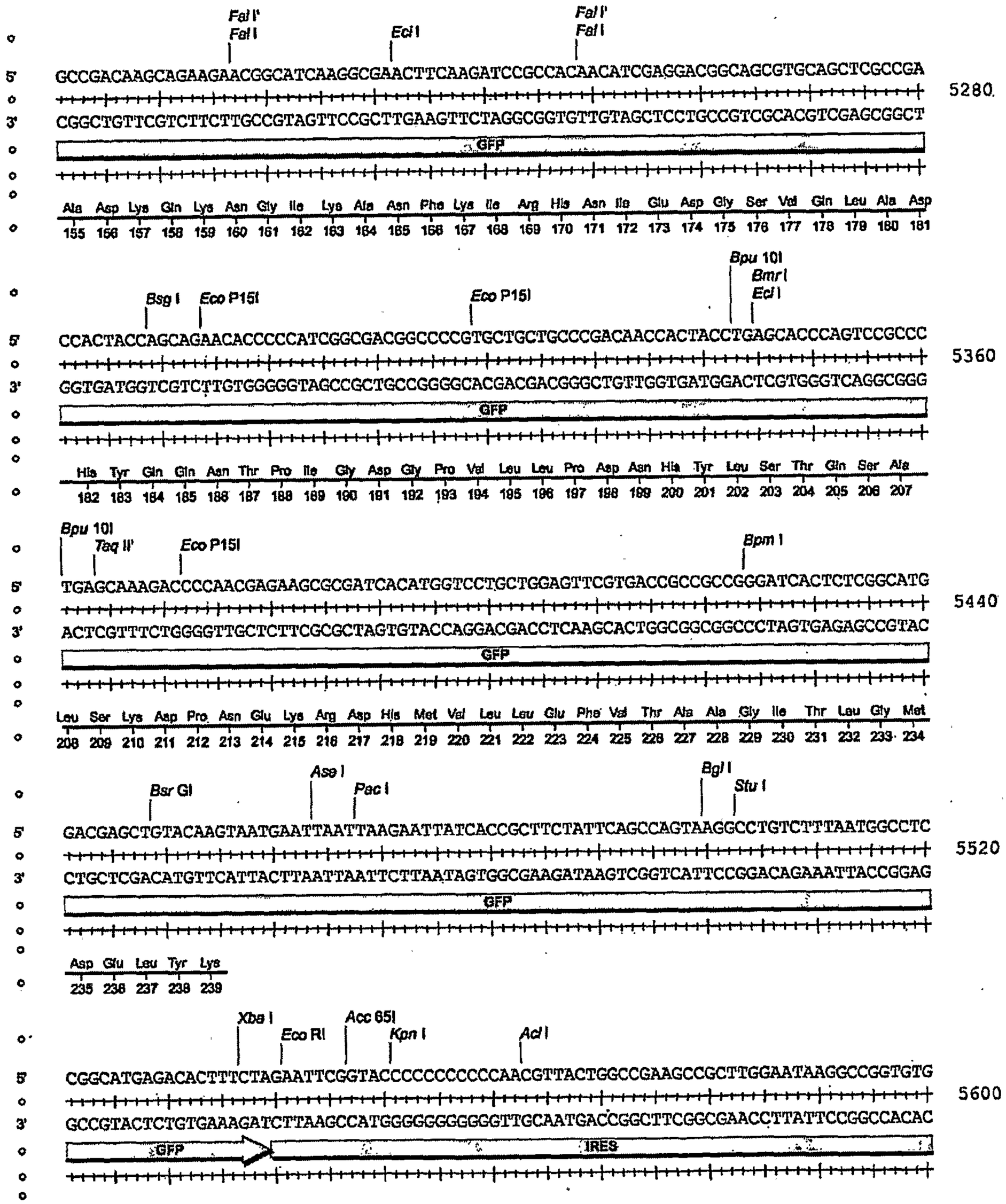

Psp OMI
Pvu II
Fsp I
Apa I
5' CAGATGGTCTGGGAGGGCAGTTCAGCTGTGGCTGCGCATAGCAGACATACAACGGACGGTGGGCCCAGACCCAGGCTGTG
2880
3' GTCTACCAGACCCTCCCGTCAAGTCGACACCGACGCGTATCGTCTGTATGTTGCCTGCCACCCGGGTCTGGGTCCGACAC
human beta actin promoter
Avr II
Bfa I
Bst EII
Taq II'
Stu I
5' TAGACCCAGCCCCCCCGCCCCGCAGTGCCTAGGTCACCCACTAACGCCCCAGGCCTGGTCTTGGCTGGGCGTGACTGTTA
2960
3' ATCTGGGTCGGGGGGGCGGGGCGTCACGGATCCAGTGGGTGATTGCGGGGTCCGGACCAGAACCGACCCGCACTGACAAT
human beta actin promoter
Bpm I
Pas I
5' CCCTCAAAAGCAGGCAGCTCCAGGGTAAAAGGTGCCCTGCCCTGTAGAGCCCACCTTCCTTCCCAGGGCTGCGGCTGGGT
3040
3' GGGAGTTTTCGTCCGTCGAGGTCCCATTTTCCACGGGACGGGACATCTCGGGTGGAAGGAAGGGTCCCGACGCCGACCCA
human beta actin promoter
Pfl MI
Alw NI
Bpm I
Xcm I
5' AGGTTTGTAGCCTTCATCACGGGCCACCTCCAGCCACTGGACCGCTGGCCCCTGCCCTGTCCTGGGGAGTGTGGTCCTGC
3120
3' TCCAAACATCGGAAGTAGTGCCCGGTGGAGGTCGGTGACCTGGCGACCGGGGACGGGACAGGACCCCTCACACCAGGACG
human beta actin promoter
Bpu EI
Bsa I
5' GACTTCTAAGTGGCCGCAAGCCACCTGACTCCCCCAACACCACACTCTACCTCTCAAGCCCAGGTCTCTCCCTAGTGACC
3200
3' CTGAAGATTCACCGGCGTTCGGTGGACTGAGGGGGTTGTGGTGTGAGATGGAGAGTTCGGGTCCAGAGAGGGATCACTGG
human beta actin promoter
Taq II'
Bmt I
Nhe I
Blp I
Bsu 36I
Alw NI
5' CACCCAGCACATTTAGCTAGCTGAGCCCCACAGCCAGAGGTCCTCAGGCCCTGCTTTCAGGGCAGTTGCTCTGAAGTCGG
3280
3' GTGGGTCGTGTAAATCGATCGACTCGGGGTGTCGGTCTCCAGGAGTCCGGGACGAAAGTCCCGTCAACGAGACTTCAGCC
human beta actin promoter
Psp OMI
Fal I'
Fal I
Eco ICRI
Sac I
Eco NI
Pst I
Fal I'
Fal I
Pas
Bst XI
Apa I
Acu I
Msc I
5' CAAGGGGGAGTGACTGCCTGGCCACTCCATGCCCTCCAAGAGCTCCTTCTGCAGGAGCGTACAGAACCCAGGGCCCTGGC
3360
3' GTTCCCCCTCACTGACGGACCGGTGAGGTACGGGAGGTTCTCGAGGAAGACGTCCTCGCATGTCTTGGGTCCCGGGACCG
human beta actin promoter

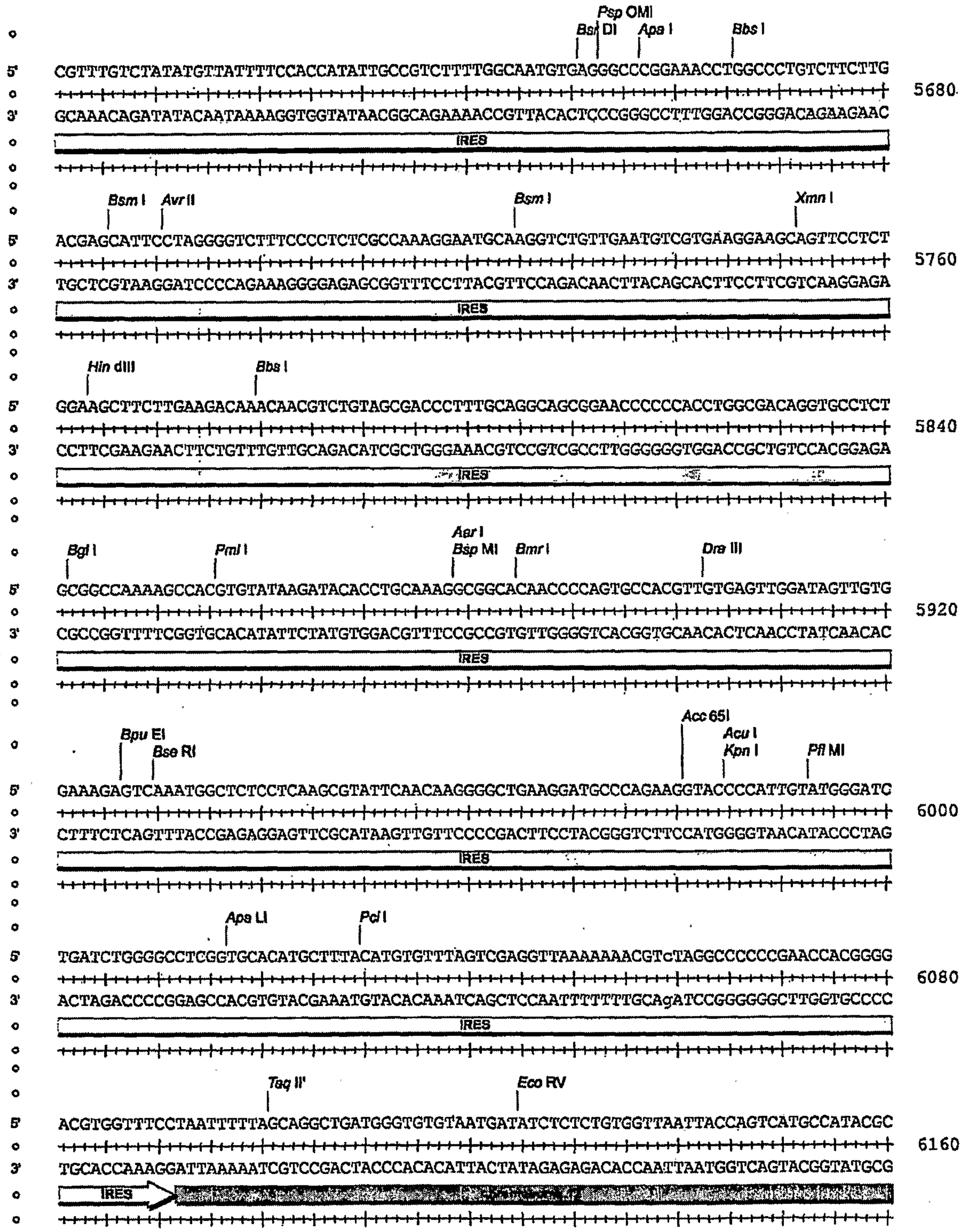

AfIw NI
Bsg I
Avr II
Taq I
5' ACCCGTGCAGACCcTGGCCCACCCCACCTGGGCGCTCAGTGCCCAAGAGATGTCCACACCTAGGATGTCCCGCGGTGGGT 3440
3' TGGGCACGTCTGGgACCGGGTGGGGTGGACCCGCGAGTCACGGGTTCTCTACAGGTGTGGATCCTACAGGGCGCCACCCA
human beta actin promoter
Psp OMI
Bsm BI
Apa I
Stu I
Msc I
Bts I
5' GGGGGGCCCcGAGAGACGGGCAGGCCGGGGGCAGGCCTGGCCATGCGGGGCCGAACCGGGCACTGCCCAGCGTGGGGCGCG 3520
3' CCCCCCGGgCTCTCTGCCCGTCCGGCCCCCGTCCGGACCGGTACGCCCCGGCTTGGCCCGTGACGGGTCGCACCCCGCGC
human beta actin promoter
Xma I
Bsa HII
Bss HII
Psp OMI
Sma I
Apa I
Bse XI
Bse RI
Bse RI
5' GGGGCCACGGCGCGCGCCCCCAGCCCCCGGGCCCAGCACCCCAAGGCGGCCAACGCCAAAACTCTCCCTCCTCCTCTTCC 3600
3' CCCCGGTGCCGCGCGCGGGGGTCGGGGGCCCGGGTCGTGGGGTTCCGCCGGTTGCGGTTTTGAGAGGGAGGAGGAGAAGG
human beta actin promoter
Ear I
Bsa XI
Bsa XI
Bsg I
Bsa XI
5' TCAATCTCGCTCTCGCTCTTTTTTTTTTCGCAAAAGGAGGGGAGAGGGGGTAAAAAAATGCTGCACTGTGCGGCGAAGC 3680
3' AGTTAGAGCGAGAGCGAGAAAAAAAAAAAGCGTTTTCCTCCCCTCTCCCCCATTTTTTTACGACGTGACACGCCGCTTCG
human beta actin promoter
Bsr BI
Psp XI
Xho I
Bsr BI
Not I
Eag I
Sac II
Kas I
Bgl II
Nar I
Sfo I
5' CGGTGAGTGAGCGGCGCGGGGCCAATCAGCGTGCGCCGTTCCGAAAGTTGCCTTTTATGGCTCGAGCGGCCGCGGCGGCG 3760
3' GCCACTCACTCGCCGCGCCCCGGTTAGTCGCACGCGGCAAGGCTTTCAACGGAAAATACCGAGCTCGCCGGCGCCGCCGC
human beta actin promoter
Bbe I
Bsa I
Ecl I
Tth 111I
Ecl I
5' CCCTATAAAACCCAGCGGCGCGACGCGCCACCACCGCCGAGACCGCGTCCGCCCCGCGAGCACAGAGCCTCGCCTTTGCC 3840
3' GGGATATTTTGGGTCGCCGCGCTGCGCGGTGGTGGCGGCTCTGGCGCAGGCGGGGCGCTCGTGTCTCGGAGCGGAAACGG
human beta actin promoter

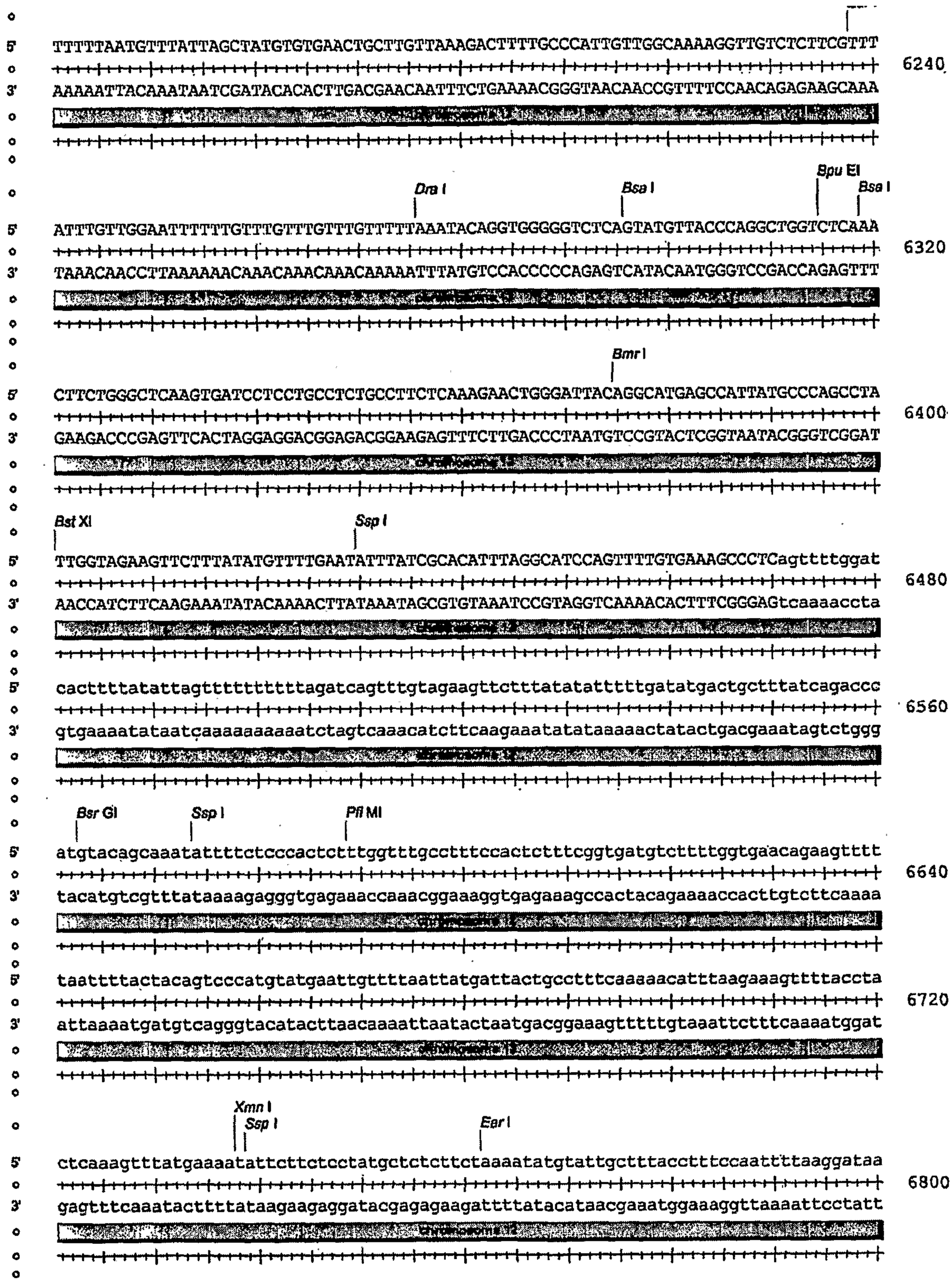

Sph I
Not I
Eag I
Sac II
5' GATCCGCCGCCCCGTCCACACCCGCCGCCAGGTAAGCCCGGCCAGCCGACCGGGGCATGCGGCCGCGGCCCCTTCGCCCGT
3920
3' CTAGGCGGCGGGGCAGGTGTGGGCGGCGGTCCATTCGGGCCGGTCGGCTGGCCCCGTACGCCGGCGCCGGGGAAGCGGGCA
human beta actin promoter
Bgl I
Bsg I
Rsr II
Pas I
5' GCAGAGCCGCCGTCTGGGCCGCAGCGGGGGGCGCATGGGGGGGGAACCGGACCGCCGTGGGGGGCGCGGGAGAAGCCCCT
4000
3' CGTCTCGGCGGCAGACCCGGCGTCGCCCCCCGCGTACCCCCCCCTTGGCCTGGCGGCACCCCCCGCGCCCTCTTCGGGGA
human beta actin promoter
Bsp EI
Bpl I'
Bpl I
Bsa HII
5' GGGCCTCCGGAGATGGGGGACACCCCACGCCAGTTCGGAGGCGCGAGGCCGCGCTCGGGAGGCGCGCTCCGGGGGTGCCG
4080
3' CCCGGAGGCCTCTACCCCCTGTGGGGTGCGGTCAAGCCTCCGCGCTCCGGCGCGAGCCCTCCGCGCGAGGCCCCCACGGC
human beta actin promoter
Bsr BI
Bpl I'
Bpl I
Bpl I'
Bpl I
Bpl I'
Bpl I
5' CTCTCGGGGCGGGGGCAACCGGCGGGGTCTTTGTCTGAGCCGGGCTCTTGCCAATGGGGATCGCAGGGTGGGCGCGGCGT
4160
3' GAGAGCCCCGCCCCCGTTGGCCGCCCCAGAAACAGACTCGGCCCGAGAACGGTTACCCCTAGCGTCCCACCCGCGCCGCA
human beta actin promoter
Kes I
Nar I
Sfo I
Bbe I
Bsr DI
Bss HII
Bss HII
5' AGCCCCCGCCAGGCCCGGTGGGGGCTGGGGCGCCATTGCGCGTGCGCGCTGGTCCTTTGGGCGCTAACTGCGTGCGCGCT
4240
3' TCGGGGGCGGTCCGGGCCACCCCCGACCCCGCGGTAACGCGCACGCGCGACCAGGAAACCCGCGATTGACGCACGCGCGA
human beta actin promoter
Psp OMI
Xma I
Ape I
Sma I
Bpu EI
Bss HII
Sac II
5' GGGAATTGGCGCTAATTGCGCGTGCGCGCTGGGACTCAAGGCGCTAAtTGCGCGTGCGTTCTGGGGCCCGGGGGTGCCGCG
4320
3' CCCTTAACCGCGATTAACGCGCACGCGCGACCCTGAGTTCCGCGATTaACGCGCACGCAAGACCCCGGGCCCCACGGCGC
human beta actin promoter

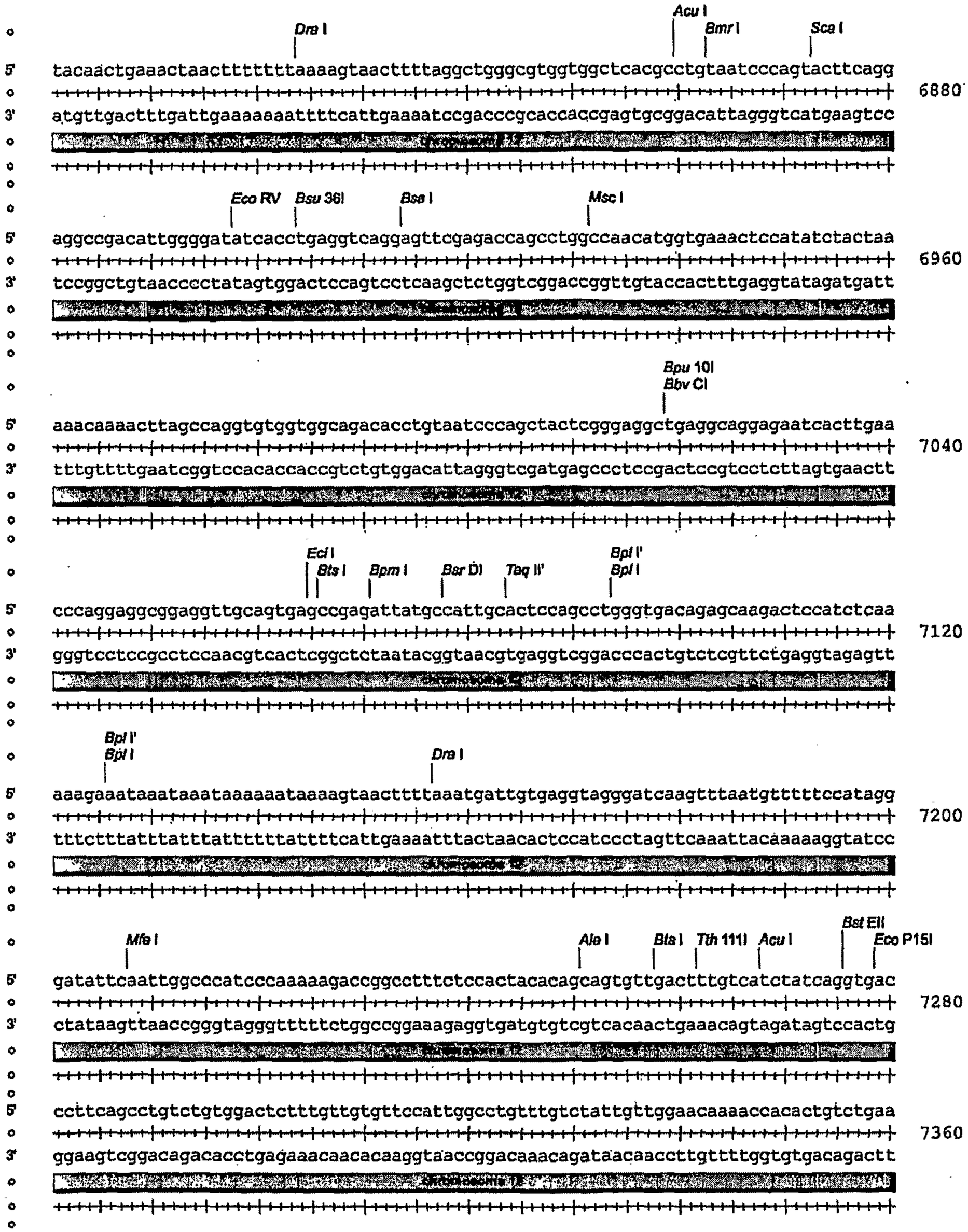

Xma I
Eag I
Sac II
Sma I
Bss HII
Bst API
5' GCCTGGGCTGGGGCGAAGGCGGGCTCGGCCGGAAGGGGTGGGGTCGCCGCGGCTCCCGGGCGCTTGCGCGCACTTCCTGC
4400
3' CGGACCCGACCCCGCTTCCGCCCGAGCCGGCCTTCCCCACCCCAGCGGCGCCGAGGGCCCGCGAACGCGCGTGAAGGACG
human beta actin promoter
Bss HII
Bss HII
Bss HII
Bgl I
5' CCGAGCCGCTGGCCGCCCGAGGGTGTGGCCGCTGCGTGCGCGCGCGCCGACCCGGCGCTGTTTGAACCGGGCGGAGGCGG
4480
3' GGCTCGGCGACCGGCGGGCTCCCACACCGGCGACGCACGCGCGCGCGGCTGGGCCGCGACAAACTTGGCCCGCCTCCGCC
human beta actin promoter
Kas I
Ecl I
Nar I
Sfo I
Bbe I
Bss HII
Sac II
5' GGCTGGCGCCCGGTTGGGAGGGGGTTGGGGCCTGGCTTCCTGCCGCGCGCCGCGGGGACGCCTCCGACCAGTGTTTGCCT
4560
3' CCGACCGCGGGCCAACCCTCCCCCAACCCCGGACCGAAGGACGGCGCGCGGCGCCCCTGCGGAGGCTGGTCACAAACGGA
human beta actin promoter
Ngo MIV
Kas I
Ngo MIV
Nae I
Nar I
Fse I
Bss HII
Sgr AI
Sfo I
Eag I
Bgl I
Bss HII
Nae I
Bbe I
Bsu 36I
5' TTTATGGTAATAACGCGGCCGGCCCGGCTTCCTTTGTCCCCAATCTGGGCGCGCGCCGGCGCCCCCTGGCGGCCTAAGGA
4640
3' AAATACCATTATTGCGCCGGCCGGGCCGAAGGAAACAGGGGTTAGACCCGCGCGCGGCCGCGGGGGACCGCCGGATTCCT
human beta actin promoter
Asc I
Bss HII
Msc I
Bss HII
5' CTCGGCGCGCCGGAAGTGGCCAGGGCGGGGGCGACCTCGGCTCACAGCGCGCCCGGCTATTCTCGCAGCTCACGTCGAGG
4720
3' GAGCCGCGCGGCCTTCACCGGTCCCGCCCCCGCTGGAGCCGAGTGTCGCGCGGGCCGATAAGAGCGTCGAGTGCAGCTCC
human beta actin promoter
Hpa I
Eco RI
Nco I
Ala I
Ppl I'
Bse RI
Ala I
Ppl I
5' TTAAcGAATTCCCCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGG
4800
3' AATTgCTTAAGGGGGTGGTACCACTCGTTCCCGCTCCTCGACAAGTGGCCCCACCACGGGTAGGACCAGCTCGACCTGCC
GFP
Met Val Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu Val Glu Leu Asp Gly
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21

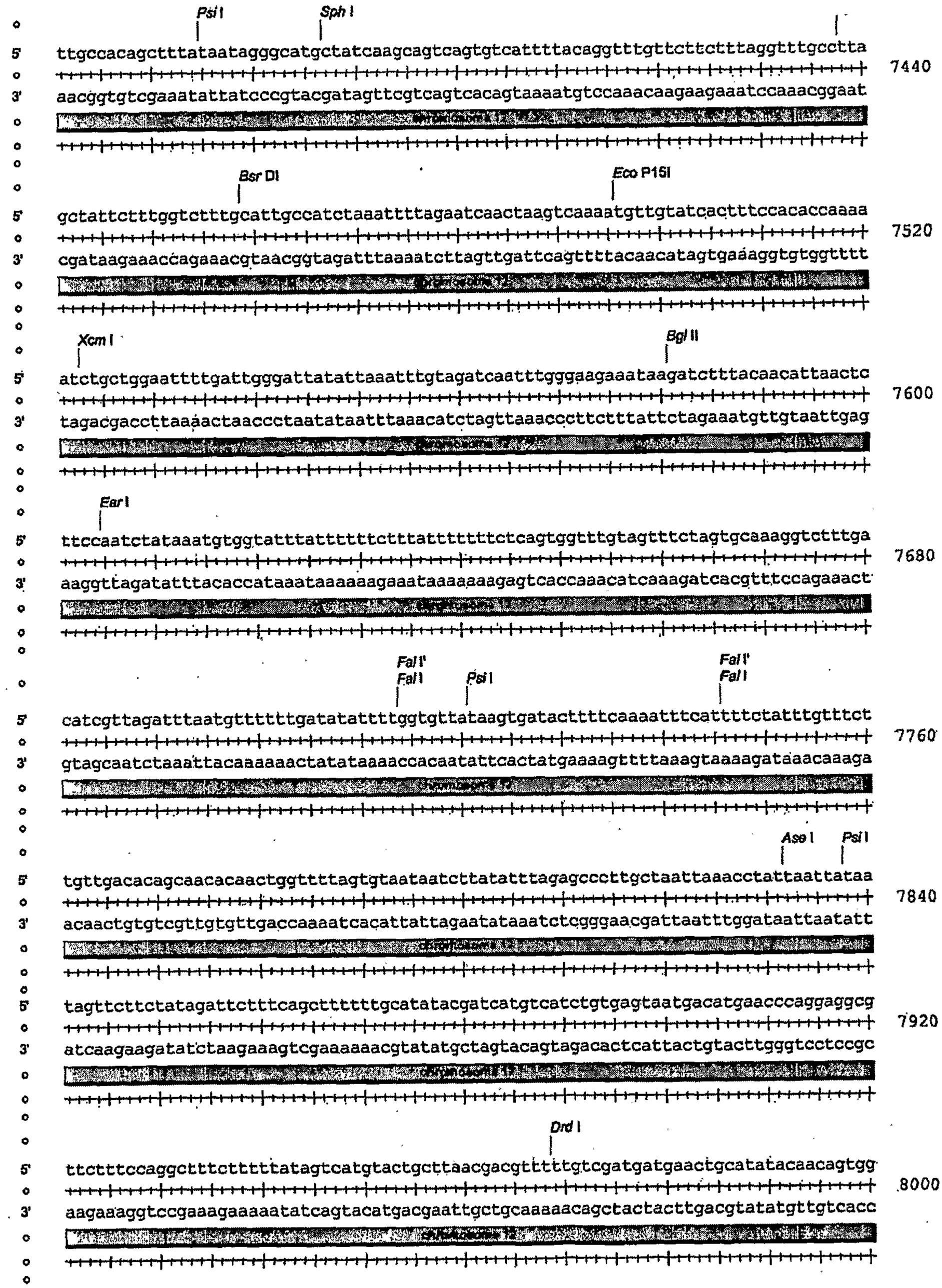

Bcg I
Bcg I'
Btg ZI
Bsg I
5' CGACGTAAACGGCCACAAGTTCAGCGTGTCTGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCA 4880
3' GCTGCATTTGCCGGTGTTCAAGTCGCACAGACCGCTCCCGCTCCCGCTACGGTGGATGCCGTTCGACTGGGACTTCAAGT
GFP
Asp Val Asn Gly His Lys Phe Ser Val Ser Gly Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe
22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47
Ale I
Acu I
Bss SI
Acu I
Bts I
5' TCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGC 4960
3' AGACGTGGTGGCCGTTCGACGGGCACGGGACCGGGTGGGAGCACTGGTGGGACTGGATGCCGCACGTCACGAAGTCGGCG
GFP
Ile Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr Leu Thr Tyr Gly Val Gln Cys Phe Ser Arg
48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74
Bsg I
Ecl I
Pfo I
5' TACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTT 5040
3' ATGGGGCTGGTGTACTTCGTCGTGCTGAAGAAGTTCAGGCGGTACGGGCTTCCGATGCAGGTCCTCGCGTGGTAGAAGAA
GFP
Tyr Pro Asp His Met Lys Gln His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu Arg Thr Ile Phe Phe
75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101
Ppi I'
Ppi I
Ale I
5' CAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGG 5120
3' GTTCCTGCTGCCGTTGATGTTCTGGGCGCGGCTCCACTTCAAGCTCCCGCTGTGGGACCACTTGGCGTAGCTCGACTTCC
GFP
Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys
102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127
Acu I
Bpm I
5' GCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATG 5200
3' CGTAGCTGAAGTTCCTCCTGCCGTTGTAGGACCCCGTGTTCGACCTCATGTTGATGTTGTCGGTGTTGCAGATATAGTAC
GFP
Gly Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr Asn Tyr Asn Ser His Asn Val Tyr Ile Met
128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154


FatI'
FatI
EclI
FatI'
FatI
5' GCCGACAAGCAGAAGAACGGCATCAAGGCGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGA 5280
3' CGGCTGTTCGTCTTCTTGCCGTAGTTCCGCTTGAAGTTCTAGGCGGTGTTGTAGCTCCTGCCGTCGCACGTCGAGCGGCT
GFP
Ala Asp Lys Gln Lys Asn Gly Ile Lys Ala Asn Phe Lys Ile Arg His Asn Ile Glu Asp Gly Ser Val Gln Leu Ala Asp
155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181
Bpu 10I
BmrI
EclI
BsgI
Eco P15I
Eco P15I
5' CCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCC 5360
3' GGTGATGGTCGTCTTGTGGGGGTAGCCGCTGCCGGGGCACGACGACGGGCTGTTGGTGATGGACTCGTGGGTCAGGCGGG
GFP
His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly Pro Val Leu Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Ala
182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207
Bpu 10I
TaqII'
Eco P15I
BpmI
5' TGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATG 5440
3' ACTCGTTTCTGGGGTTGCTCTTCGCGCTAGTGTACCAGGACGACCTCAAGCACTGGCGGCGGCCCTAGTGAGAGCCGTAC
GFP
Leu Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe Val Thr Ala Ala Gly Ile Thr Leu Gly Met
208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234
AseI
BglI
BsrGI
PacI
StuI
5' GACGAGCTGTACAAGTAATGAATTAATTAAGAATTATCACCGCTTCTATTCAGCCAGTAAGGCCTGTCTTTAATGGCCTC 5520
3' CTGCTCGACATGTTCATTACTTAATTAATTCTTAATAGTGGCGAAGATAAGTCGGTCATTCCGGACAGAAATTACCGGAG
GFP
Asp Glu Leu Tyr Lys
235 236 237 238 239
XbaI
AccI65I
EcoRI
KpnI
AclI
5' CGGCATGAGACACTTTCTAGAATTCGGTACCCCCCCCCCCAACGTTACTGGCCGAAGCCGCTTGGAATAAGGCCGGTGTG 5600
3' GCCGTACTCTGTGAAAGATCTTAAGCCATGGGGGGGGGGGTTGCAATGACCGGCTTCGGCGAACCTTATTCCGGCCACAC
GFP
IRES


Psp OMI
BsiWI
Apa I
Bbs I
5'
CGTTTGTCTATATGTTATTTTCCACCATATTGCCGTCTTTTGGCAATGTGAGGGCCCGGAAACCTGGCCCTGTCTTCTTG
5680
3'
GCAAACAGATATACAATAAAAGGTGGTATAACGGCAGAAAACCGTTACACTCCCGGGCCTTTGGACCGGGACAGAAGAAC
IRES
Bsm I
Avr II
Bsm I
Xmn I
5'
ACGAGCATTCCTAGGGGTCTTTCCCCTCTCGCCAAAGGAATGCAAGGTCTGTTGAATGTCGTGAAGGAAGCAGTTCCTCT
5760
3'
TGCTCGTAAGGATCCCCAGAAAGGGGAGAGCGGTTTCCTTACGTTCCAGACAACTTACAGCACTTCCTTCGTCAAGGAGA
IRES
Hin dIII
Bbs I
5'
GGAAGCTTCTTGAAGACAAACAACGTCTGTAGCGACCCTTTGCAGGCAGCGGAACCCCCCACCTGGCGACAGGTGCCTCT
5840
3'
CCTTCGAAGAACTTCTGTTTGTTGCAGACATCGCTGGGAAACGTCCGTCGCCTTGGGGGGTGGACCGCTGTCCACGGAGA
IRES
Bgl I
Pml I
Aar I
Bsp MI
Bmr I
Dra III
5'
GCGGCCAAAAGCCACGTGTATAAGATACACCTGCAAAGGCGGCACAACCCCAGTGCCACGTTGTGAGTTGGATAGTTGTG
5920
3'
CGCCGGTTTTCGGTGCACATATTCTATGTGGACGTTTCCGCCGTGTTGGGGTCACGGTGCAACACTCAACCTATCAACAC
IRES
Acc 65I
Bpu EI
Bse RI
Acu I
Kpn I
Pfl MI
5'
GAAAGAGTCAAATGGCTCTCCTCAAGCGTATTCAACAAGGGGCTGAAGGATGCCCAGAAGGTACCCCATTGTATGGGATC
6000
3'
CTTTCTCAGTTTACCGAGAGGAGTTCGCATAAGTTGTTCCCCGACTTCCTACGGGTCTTCCATGGGGTAACATACCCTAG
IRES
Apa LI
Pcl I
5'
TGATCTGGGGCCTCGGTGCACATGCTTTACATGTGTTTAGTCGAGGTTAAAAAAACGTcTAGGCCCCCCGAACCACGGGG
6080
3'
ACTAGACCCCGGAGCCACGTGTACGAAATGTACACAAATCAGCTCCAATTTTTTTGCAgATCCGGGGGGCTTGGTGCCCC
IRES
Taq II'
Eco RV
5'
ACGTGGTTTCCTAATTTTTAGCAGGCTGATGGGTGTGTAATGATATCTCTCTGTGGTTAATTACCAGTCATGCCATACGC
6160
3'
TGCACCAAAGGATTAAAAATCGTCCGACTACCCACACATTACTATAGAGAGACACCAATTAATGGTCAGTACGGTATGCG
IRES


5' TTTTTAATGTTTATTAGCTATGTGTGAACTGCTTGTTAAAGACTTTTGCCCATTGTTGGCAAAAGGTTGTCTCTTCGTTT
6240
3' AAAAATTACAAATAATCGATACACACTTGACGAACAATTTCTGAAAACGGGTAACAACCGTTTTCCAACAGAGAAGCAAA
Dra I
Bsa I
Bpu EI
Bsa I
5' ATTTGTTGGAATTTTTTGTTTGTTTGTTTGTTTTTAAATACAGGTGGGGGTCTCAGTATGTTACCCAGGCTGGTCTCAAA
6320
3' TAAACAACCTTAAAAAACAAACAAACAAACAAAAATTTATGTCCACCCCCAGAGTCATACAATGGGTCCGACCAGAGTTT
Bmr I
5' CTTCTGGGCTCAAGTGATCCTCCTGCCTCTGCCTTCTCAAAGAACTGGGATTACAGGCATGAGCCATTATGCCCAGCCTA
6400
3' GAAGACCCGAGTTCACTAGGAGGACGGAGACGGAAGAGTTTCTTGACCCTAATGTCCGTACTCGGTAATACGGGTCGGAT
Bst XI
Ssp I
5' TTGGTAGAAGTTCTTTATATGTTTTGAATATTTATCGCACATTTAGGCATCCAGTTTTGTGAAAGCCCTCagttttggat
6480
3' AACCATCTTCAAGAAATATACAAAACTTATAAATAGCGTGTAAATCCGTAGGTCAAAACACTTTCGGGAGtcaaaaccta
5' cacttttatattagttttttttttagatcagtttgtagaagttctttatatattttttgatatgactgctttatcagaccc
6560
3' gtgaaaatataatcaaaaaaaaaatctagtcaaacatcttcaagaaatatataaaaactatactgacgaaatagtctggg
Bsr GI
Ssp I
Pfl MI
5' atgtacagcaaatattttctcccactctttggtttgcctttccactctttcggtgatgtcttttggtgaacagaagtttt
6640
3' tacatgtcgtttataaaagagggtgagaaaccaaacggaaaggtgagaaagccactacagaaaaccacttgtcttcaaaa
5' taattttactacagtcccatgtatgaattgttttaattatgattactgcctttcaaaaacatttaagaaagttttaccta
6720
3' attaaaatgatgtcagggtacatacttaacaaaattaatactaatgacggaaagtttttgtaaattctttcaaaatggat
Xmn I
Ssp I
Ear I
5' ctcaaagtttatgaaaatattcttctcctatgctctcttctaaaatatgtattgctttacctttccaattttaaggataa
6800
3' gagtttcaaatacttttataagaagaggatacgagagaagattttatacataacgaaatggaaaggttaaaattcctatt


Acu I
Dra I
Bmr I
Sca I
5' tacaactgaaactaacttttttttaaaagtaacttttaggctgggcgtggtggctcacgcctgtaatcccagtacttcagg
6880
3' atgttgactttgattgaaaaaaattttcattgaaaatccgacccgcaccaccgagtgcggacattagggtcatgaagtcc
Eco RV
Bsu 36I
Bse I
Msc I
5' aggccgacattggggatatcacctgaggtcaggagttcgagaccagcctggccaacatggtgaaactccatatctactaa
6960
3' tccggctgtaacccctatagtggactccagtcctcaagctctggtcggaccggttgtaccactttgaggtatagatgatt
Bpu 10I
Bbv CI
5' aaacaaaacttagccaggtgtggtggcagacacctgtaatcccagctactcgggaggctgaggcaggagaatcacttgaa
7040
3' tttgttttgaatcggtccacaccaccgtctgtggacattagggtcgatgagccctccgactccgtcctcttagtgaactt
Ecl I
Bts I
Bpm I
Bsr DI
Taq II'
Bpl I'
Bpl I
5' cccaggaggcggaggttgcagtgagccgagattatgccattgcactccagcctgggtgacagagcaagactccatctcaa
7120
3' gggtcctccgcctccaacgtcactcggctctaatacggtaacgtgaggtcggacccactgtctcgttctgaggtagagtt
Bpl I'
Bpl I
Dra I
5' aaagaaataaataaataaaaaataaaagtaacttttaaatgattgtgaggtagggatcaagtttaatgtttttccatagg
7200
3' tttctttatttatttatttttttatttttcattgaaaatttactaacactccatccctagttcaaattacaaaaaggtatcc
Bst EII
Mfe I
Ale I
Bts I
Tth 111I
Acu I
Eco P15I
5' gatattcaattggcccatcccaaaaagaccggcctttctccactacacagcagtgttgactttgtcatctatcaggtgac
7280
3' ctataagttaaccgggtagggttttttctggccggaaagaggtgatgtgtcgtcacaactgaaacagtagatagtccactg
5' ccttcagcctgtctgtggactctttgttgtgttccattggcctgtttgtctattgttggaacaaaaccacactgtctgaa
7360
3' ggaagtcggacagacacctgagaaacaacacaaggtaaccggacaaacagataacaaccttgttttggtgtgacagactt


Psi I
Sph I
5' ttgccacagctttataatagggcatgctatcaagcagtcagtgtcattttacaggtttgttcttctttaggtttgcctta
7440
3' aacggtgtcgaaatattatcccgtacgatagttcgtcagtcacagtaaaatgtccaaacaagaagaaatccaaacggaat
Bsr DI
Eco P15I
5' gctattctttggtctttgcattgccatctaaattttagaatcaactaagtcaaaatgttgtatcactttccacaccaaaa
7520
3' cgataagaaaccagaaacgtaacggtagatttaaaatcttagttgattcagttttacaacatagtgaaaaggtgtggtttt
Xcm I
Bgl II
5' atctgctggaattttgattgggattatattaaatttgtagatcaatttgggaagaaataagatctttacaacattaactc
7600
3' tagacgaccttaaaactaaccctaatataatttaaacatctagttaaacccttctttattctagaaatgttgtaattgag
Ear I
5' ttccaatctataaatgtggtatttatttttttctttatttttttctcagtggtttgtagtttctagtgcaaaggtctttga
7680
3' aaggttagatatttacaccataaataaaaaagaaataaaaaaagagtcaccaaacatcaaagatcacgtttccagaaact
Fal I'
Fal I
Psi I
Fal I'
Fal I
5' catcgttagatttaatgtttttttgatatattttggtgttataagtgatacttttcaaaatttcattttctatttgtttct
7760
3' gtagcaatctaaattacaaaaaactatataaaaccacaatattcactatgaaaagttttaaagtaaaagataaacaaaga
Ase I
Psi I
5' tgttgacacagcaacacaactggttttagtgtaataatcttatatttagagcccttgctaattaaacctattaattataa
7840
3' acaactgtgtcgttgtgttgaccaaaatcacattattagaatataaatctcgggaacgattaatttggataattaatatt
5' tagttcttctatagattctttcagctttttgcatatacgatcatgtcatctgtgagtaatgacatgaacccaggaggcg
7920
3' atcaagaagatatctaagaaagtcgaaaaaacgtatatgctagtacagtagacactcattactgtacttgggtcctccgc
Dra I
5' ttctttccaggctttctttttatagtcatgtactgcttaacgacgtttttgtcgatgatgaactgcatatacaacagtgg
8000
3' aagaaaggtccgaaagaaaaatatcagtacatgacgaattgctgcaaaaacagctactacttgacgtatatgttgtcacc

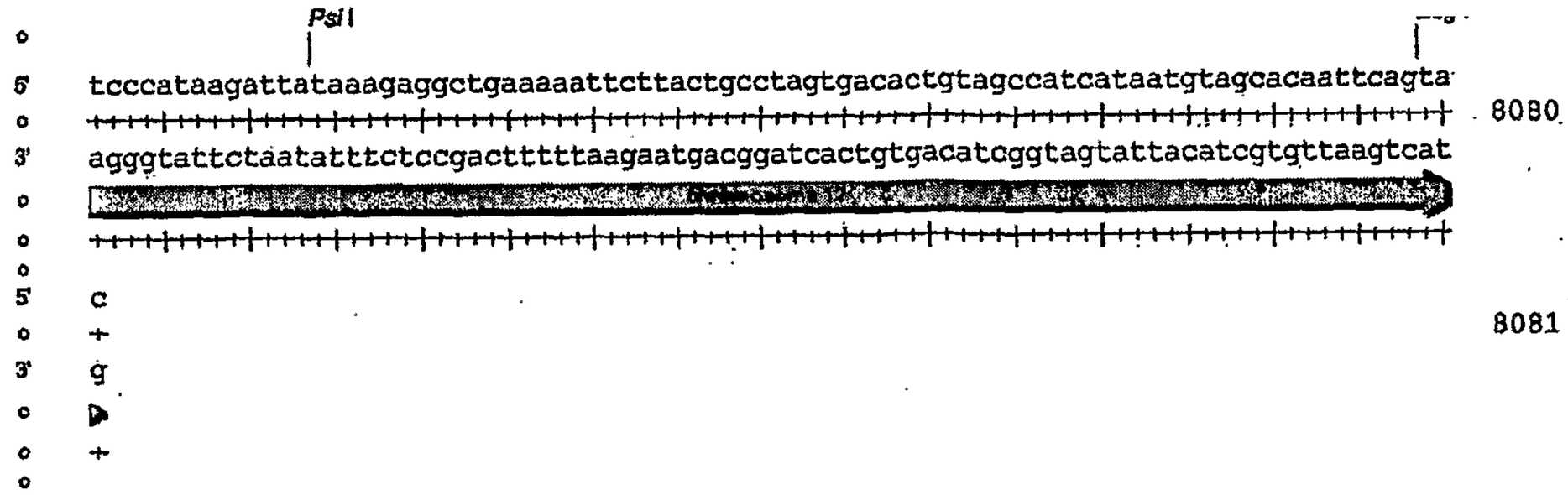

PstI
5' tcccataagattataaagaggctgaaaaattcttactgcctagtgacactgtagccatcataatgtagcacaattcagta
8080
3' agggtattctaatatttctccgacttttaagaatgacggatcactgtgacatcggtagtattacatcgtgttaagtcat
5' c
8081
3' g

Figure 3

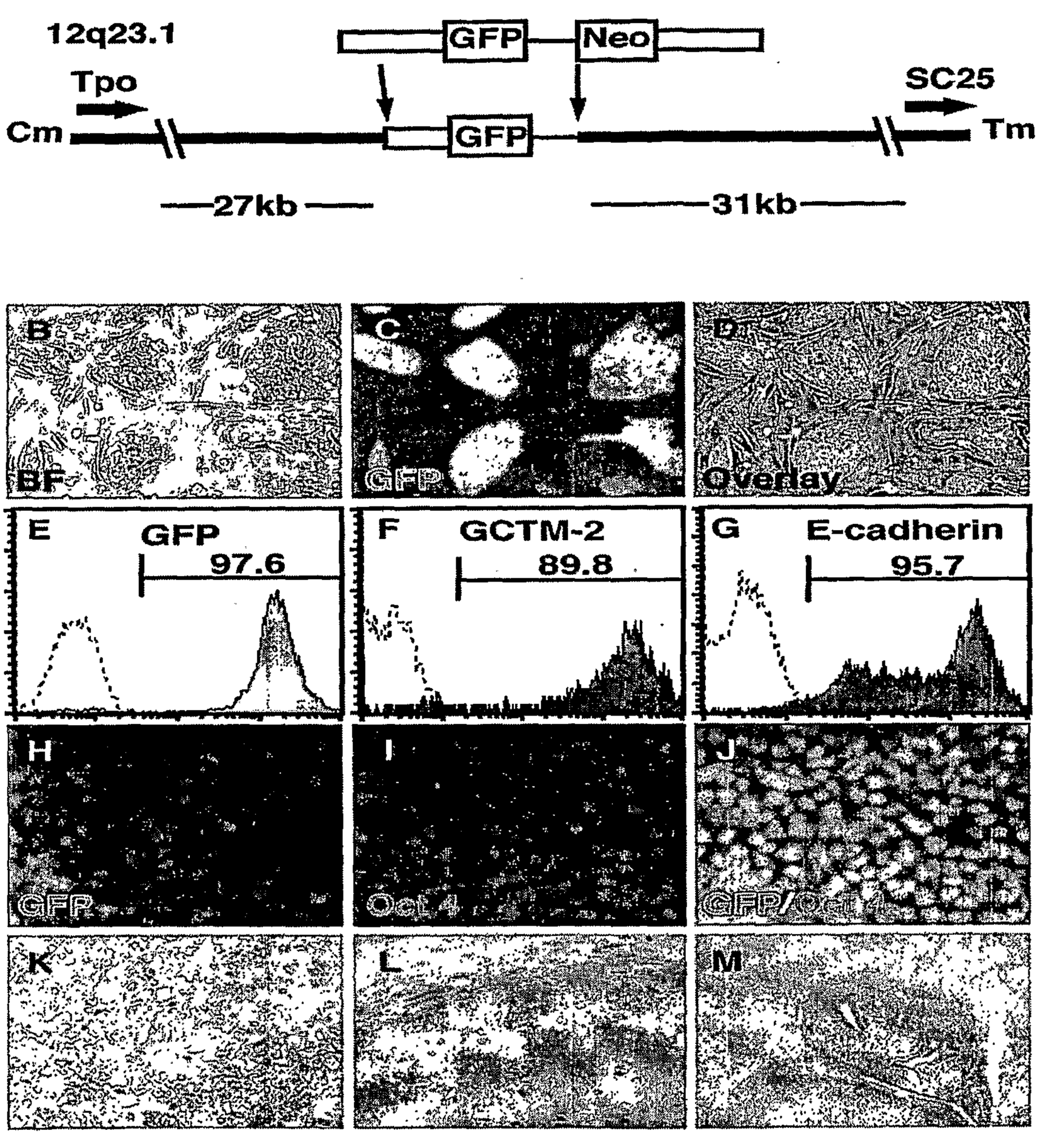
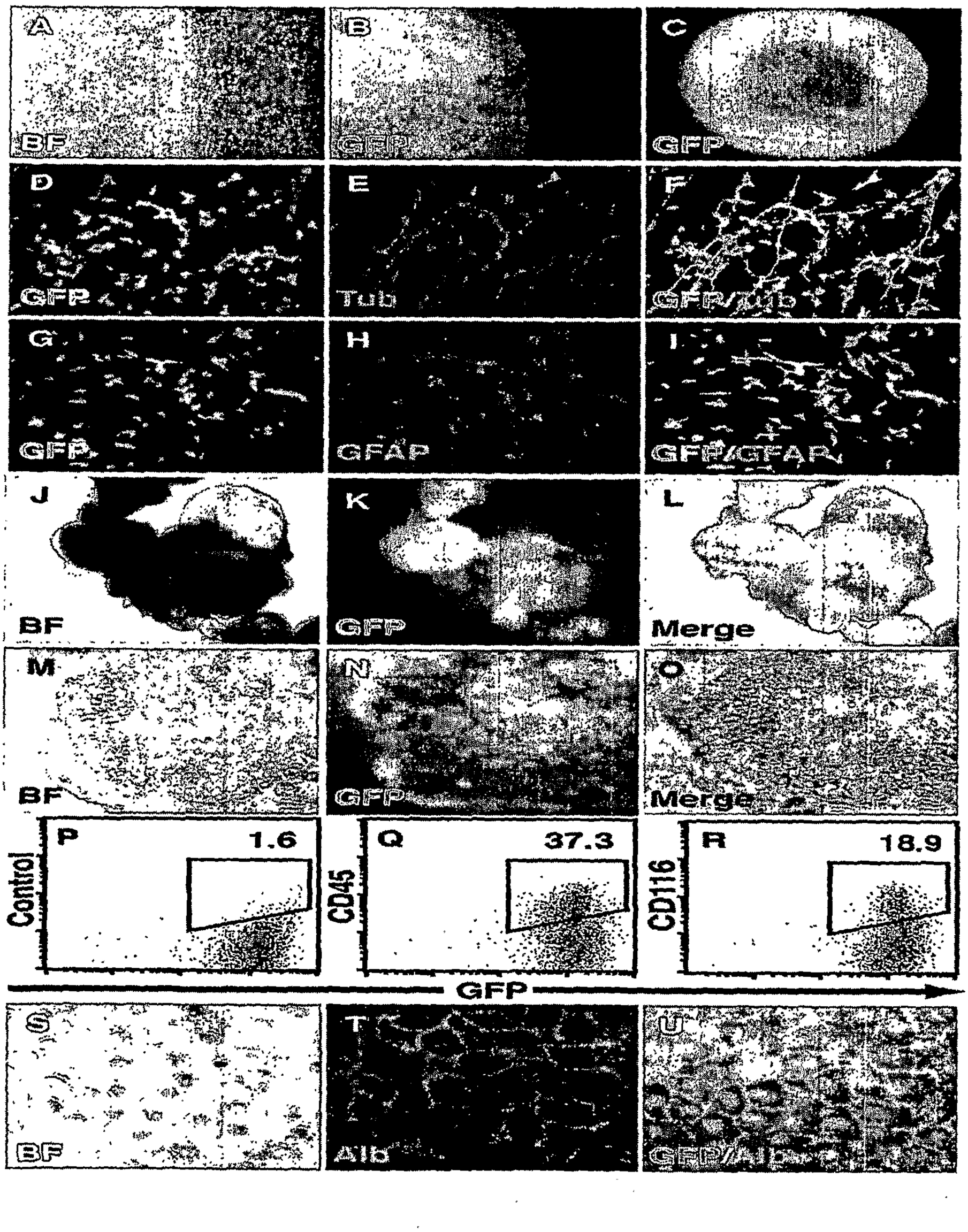

A
12q23.1
Tpo
Cm
GFP
Neo
GFP
SC25
Tm
27kb
31kb
B
BF
C
GFP
D
Overlay
E
GFP
97.6
F
GCTM-2
89.8
G
E-cadherin
95.7
H
GFP
I
Oct 4
J
GFP/Oct 4
K
L
M

Figure 4


A
BF
B
GFP
C
GFP
D
GFP
E
Tub
F
GFP/Tub
G
GFP
H
GFAP
I
GFP/GFAP
J
BF
K
GFP
L
Merge
M
BF
N
GFP
O
Merge
Control
P
1.6
CD45
Q
37.3
CD116
R
18.9
GFP
S
BF
T
Alb
U
GFP/Alb