

PCTWORLD INTELLECTUAL
PROPERTY ORGANIZATION

INTERNATIONAL APPLICATION PUBLISHED



WO 9604391A1

(51) International Patent Classification⁶ :

C12N 15/64, 15/67, 15/85, 9/72, 5/10

A1

(43) International Publication Date: 15 February 1996 (15.02.96)

(21) International Application Number: PCT/US95/09576

(22) International Filing Date: 28 July 1995 (28.07.95)

(30) Priority Data:
08/286,740 5 August 1994 (05.08.94) US

(71) Applicant: GENENTECH, INC. [US/US]; 460 Point San Bruno Boulevard, South San Francisco, CA 94080-4990 (US).

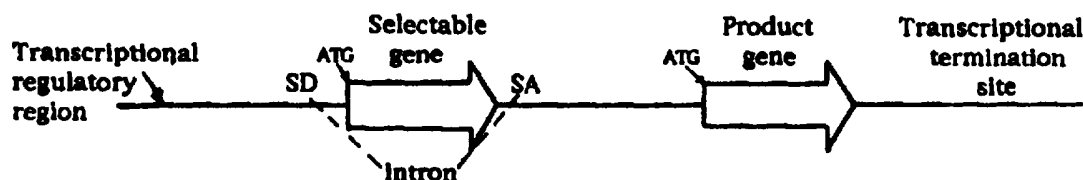
(72) Inventor: CROWLEY, Craig, W.; 151 Durazno Way, Portola Valley, CA 94028 (US).

(74) Agents: LEE, Wendy, M. et al.; Genentech, Inc., 460 Point San Bruno Boulevard, South San Francisco, CA 94080-4990 (US).

(81) Designated States: AU, CA, JP, MX, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published*With international search report.**Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.*

(54) Title: METHOD FOR SELECTING HIGH-EXPRESSING HOST CELLS

**(57) Abstract**

A method for selecting recombinant host cells expressing high levels of a desired protein is described. This method utilizes eukaryotic host cells harboring a DNA construct comprising a selectable gene (preferably an amplifiable gene) and a product gene provided 3' to the selectable gene. The selectable gene is positioned within an intron defined by a splice donor site and a splice acceptor site and the selectable gene and product gene are under the transcriptional control of a single transcriptional regulatory region. The splice donor site is generally an efficient splice donor site and thereby regulates expression of the product gene using the transcriptional regulatory region. The transfected cells are cultured so as to express the gene encoding the product in a selective medium comprising an amplifying agent for sufficient time to allow amplification to occur, whereupon either the desired product is recovered or cells having multiple copies of the product gene are identified.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgyzstan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LU	Luxembourg	TD	Chad
CS	Czechoslovakia	LV	Latvia	TG	Togo
CZ	Czech Republic	MC	Monaco	TJ	Tajikistan
DE	Germany	MD	Republic of Moldova	TT	Trinidad and Tobago
DK	Denmark	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	US	United States of America
FI	Finland	MN	Mongolia	UZ	Uzbekistan
FR	France			VN	Viet Nam
GA	Gabon				

METHOD FOR SELECTING HIGH-EXPRESSING HOST CELLSBACKGROUND OF THE INVENTIONField of the Invention

This invention relates to a method of selecting for high-expressing host cells, a method of producing a protein of interest in high yields and a method of producing eukaryotic cells having multiple copies of a sequence encoding a protein of interest.

Description of Background and Related Art

The discovery of methods for introducing DNA into living host cells in a functional form has provided the key to understanding many fundamental biological processes, and has made possible the production of important proteins and other molecules in commercially useful quantities.

Despite the general success of such gene transfer methods, several common problems exist that may limit the efficiency with which a gene encoding a desired protein can be introduced into and expressed in a host cell. One problem is knowing when the gene has been successfully transferred into recipient cells. A second problem is distinguishing between those cells that contain the gene and those that have survived the transfer procedures but do not contain the gene. A third problem is identifying and isolating those cells that contain the gene and that are expressing high levels of the protein encoded by the gene.

In general, the known methods for introducing genes into eukaryotic cells tend to be highly inefficient. Of the cells in a given culture, only a small proportion take up and express exogenously added DNA, and an even smaller proportion stably maintain that DNA.

Identification of those cells that have incorporated a product gene encoding a desired protein typically is achieved by introducing into the same cells another gene, commonly referred to as a selectable gene, that encodes a selectable marker. A selectable marker is a protein that is necessary for the growth or survival of a host cell under the particular culture conditions chosen, such as an enzyme that confers resistance to an antibiotic or other drug, or an enzyme that compensates for a metabolic or catabolic defect in the host cell. For example, selectable genes commonly used with eukaryotic cells include the genes for aminoglycoside phosphotransferase (APH), hygromycin phosphotransferase (hyg), dihydrofolate reductase (DHFR), thymidine kinase (tk), neomycin, puromycin, glutamine synthetase, and asparagine synthetase.

The method of identifying a host cell that has incorporated one gene on the basis of expression by the host cell of a second incorporated gene encoding a selectable marker is referred to as cotransfection (or cotransfection). In that method, a gene encoding a desired polypeptide and a selection gene typically are introduced into the host cell simultaneously, although they may be introduced sequentially. In the case of simultaneous cotransfection, the gene encoding the desired polypeptide

and the selectable gene may be present on a single DNA molecule or on separate DNA molecules prior to being introduced into the host cells. Wigler et al., Cell, 16:777 (1979). Cells that have incorporated the gene encoding the desired polypeptide then are identified or isolated by
5 culturing the cells under conditions that preferentially allow for the growth or survival of those cells that synthesize the selectable marker encoded by the selectable gene.

The level of expression of a gene introduced into a eukaryotic host cell depends on multiple factors, including gene copy number, efficiency
10 of transcription, messenger RNA (mRNA) processing, stability, and translation efficiency. Accordingly, high level expression of a desired polypeptide typically will involve optimizing one or more of those factors.

For example, the level of protein production may be increased by covalently joining the coding sequence of the gene to a "strong" promoter
15 or enhancer that will give high levels of transcription. Promoters and enhancers are nucleotide sequences that interact specifically with proteins in a host cell that are involved in transcription. Kriegler, Meth. Enzymol., 185:512 (1990); Maniatis et al., Science, 236:1237 (1987). Promoters are located upstream of the coding sequence of a gene and
20 facilitate transcription of the gene by RNA polymerase. Among the eukaryotic promoters that have been identified as strong promoters for high-level expression are the SV40 early promoter, adenovirus major late promoter, mouse metallothionein-I promoter, Rous sarcoma virus long terminal repeat, and human cytomegalovirus immediate early promoter (CMV).

Enhancers stimulate transcription from a linked promoter. Unlike
25 promoters, enhancers are active when placed downstream from the transcription initiation site or at considerable distances from the promoter, although in practice enhancers may overlap physically and functionally with promoters. For example, all of the strong promoters
30 listed above also contain strong enhancers. Bendig, Genetic Engineering, 7:91 (Academic Press, 1988).

The level of protein production also may be increased by increasing the gene copy number in the host cell. One method for obtaining high gene copy number is to directly introduce into the host cell multiple copies of
35 the gene, for example, by using a large molar excess of the product gene relative to the selectable gene during cotransfection. Kaufman, Meth. Enzymol., 185:537 (1990). With this method, however, only a small proportion of the cotransfected cells will contain the product gene at high copy number. Furthermore, because no generally applicable, convenient
40 method exists for distinguishing such cells from the majority of cells that contain fewer copies of the product gene, laborious and time-consuming screening methods typically are required to identify the desired high-copy number transfectants.

Another method for obtaining high gene copy number involves cloning
45 the gene in a vector that is capable of replicating autonomously in the host cell. Examples of such vectors include mammalian expression vectors

derived from Epstein-Barr virus or bovine papilloma virus, and yeast 2-micron plasmid vectors. Stephens & Hentschel, Biochem. J., 248:1 (1987); Yates et al., Nature, 313:812 (1985); Beggs, Genetic Engineering, 2:175 (Academic Press, 1981).

5 Yet another method for obtaining high gene copy number involves gene amplification in the host cell. Gene amplification occurs naturally in eukaryotic cells at a relatively low frequency. Schimke, J. Biol. Chem., 263:5989 (1988). However, gene amplification also may be induced, or at least selected for, by exposing host cells to appropriate selective
10 pressure. For example, in many cases it is possible to introduce a product gene together with an amplifiable gene into a host cell and subsequently select for amplification of the marker gene by exposing the cotransfected cells to sequentially increasing concentrations of a selective agent. Typically the product gene will be coamplified with the marker gene under
15 such conditions.

The most widely used amplifiable gene for that purpose is a DHFR gene, which encodes a dihydrofolate reductase enzyme. The selection agent used in conjunction with a DHFR gene is methotrexate (Mtx). A host cell is cotransfected with a product gene encoding a desired protein and a DHFR
20 gene, and transfectants are identified by first culturing the cells in culture medium that contains Mtx. A suitable host cell when a wild-type DHFR gene is used is the Chinese Hamster Ovary (CHO) cell line deficient in DHFR activity, prepared and propagated as described by Urlaub & Chasin, Proc. Nat. Acad. Sci. USA, 77:4216 (1980). The transfected cells then are
25 exposed to successively higher amounts of Mtx. This leads to the synthesis of multiple copies of the DHFR gene, and concomitantly, multiple copies of the product gene. Schimke, J. Biol. Chem., 263:5989 (1988); Axel et al., U.S. Patent No. 4,399,216; Axel et al., U.S. Patent No. 4,634,665. Other references directed to co-transfection of a gene together with a genetic
30 marker that allows for selection and subsequent amplification include Kaufman in Genetic Engineering, ed. J. Setlow (Plenum Press, New York), Vol. 9 (1987); Kaufman and Sharp, J. Mol. Biol., 159:601 (1982); Ringold et al., J. Mol. Appl. Genet., 1:165-175 (1981); Kaufman et al., Mol. Cell Biol., 5:1750-1759 (1985); Kaetzel and Nilson, J. Biol. Chem., 263:6244-
35 6251 (1988); Hung et al., Proc. Natl. Acad. Sci. USA, 83:261-264 (1986); Kaufman et al., EMBO J., 6:87-93 (1987); Johnston and Kucey, Science, 242:1551-1554 (1988); Urlaub et al., Cell, 33:405-412 (1983).

To extend the DHFR amplification method to other cell types, a mutant DHFR gene that encodes a protein with reduced sensitivity to methotrexate
40 may be used in conjunction with host cells that contain normal numbers of an endogenous wild-type DHFR gene. Simonsen and Levinson, Proc. Natl. Acad. Sci. USA, 80:2495 (1983); Wigler et al., Proc. Natl. Acad. Sci. USA, 77:3567-3570 (1980); Haber and Schimke, Somatic Cell Genetics, 8:499-508 (1982).

45 Alternatively, host cells may be co-transfected with the product gene, a DHFR gene, and a dominant selectable gene, such as a neo^r gene. Kim

and Wold, Cell, 42:129 (1985); Capon et al., U.S. Pat. No. 4,965,199. Transfectants are identified by first culturing the cells in culture medium containing neomycin (or the related drug G418), and the transfectants so identified then are selected for amplification of the DHFR gene and the product gene by exposure to successively increasing amounts of Mtx.

As will be appreciated from this discussion, the selection of recombinant host cells that express high levels of a desired protein generally is a multi-step process. In the first step, initial transfectants are selected that have incorporated the product gene and the selectable gene. In subsequent steps, the initial transfectants are subject to further selection for high-level expression of the selectable gene and then random screening for high-level expression of the product gene. To identify cells expressing high levels of the desired protein, typically one must screen large numbers of transfectants. The majority of transfectants produce less than maximal levels of the desired protein. Further, Mtx resistance in DHFR transformants is at least partially conferred by varying degrees of gene amplification. Schimke, Cell, 37:705-713 (1984). The inadequacies of co-expression of the non-selected gene have been reported by Wold et al., Proc. Natl. Acad. Sci. USA, 76:5684-5688 (1979). Instability of the amplified DNA is reported by Kaufman and Schimke, Mol. Cell Biol., 1:1069-1076 (1981); Haber and Schimke, Cell, 26:355-362 (1981); and Fedespiel et al., J. Biol. Chem., 259:9127-9140 (1984).

Several methods have been described for directly selecting such recombinant host cells in a single step. One strategy involves co-transfecting host cells with a product gene and a DHFR gene, and selecting those cells that express high levels of DHFR by directly culturing in medium containing a high concentration of Mtx. Many of the cells selected in that manner also express the co-transfected product gene at high levels. Page and Sydenham, Bio/Technology, 9:64 (1991). This method for single-step selection suffers from certain drawbacks that limit its usefulness. High-expressing cells obtained by direct culturing in medium containing a high level of a selection agent may have poor growth and stability characteristics, thus limiting their usefulness for long-term production processes. Page and Snyderman, Bio/Technology, 9:64 (1991). Single-step selection for high-level resistance to Mtx may produce cells with an altered, Mtx-resistant DHFR enzyme, or cells that have altered Mtx transport properties, rather than cells containing amplified genes. Haber et al., J. Biol. Chem., 256:9501 (1981); Assaraf and Schimke, Proc. Natl. Acad. Sci. USA, 84:7154 (1987).

Another method involves the use of polycistronic mRNA expression vectors containing a product gene at the 5' end of the transcribed region and a selectable gene at the 3' end. Because translation of the selectable gene at the 3' end of the polycistronic mRNA is inefficient, such vectors exhibit preferential translation of the product gene and require high levels of polycistronic mRNA to survive selection. Kaufman, Meth.

Enzymol., 185:487 (1990); Kaufman, Meth. Enzymol., 185:537 (1990); Kaufman et al., EMBO J., 6:187 (1987). Accordingly, cells expressing high levels of the desired protein product may be obtained in a single step by culturing the initial transfectants in medium containing a selection agent appropriate for use with the particular selectable gene. However, the utility of these vectors is variable because of the unpredictable influence of the upstream product reading frame on selectable marker translation and because the upstream reading frame sometimes becomes deleted during methotrexate amplification (Kaufman et al., J. Mol. Biol., 159:601-621 [1982]; Levinson, Methods in Enzymology, San Diego: Academic Press, Inc. [1990]). Later vectors incorporated an internal translation initiation site derived from members of the picornavirus family which is positioned between the product gene and the selectable gene (Pelletier et al., Nature, 334:320 [1988]; Jang et al., J. Virol., 63:1651 [1989]).

15 A third method for single-step selection involves use of a DNA construct with a selectable gene containing an intron within which is located a gene encoding the protein of interest. See U.S. Patent No. 5,043,270 and Abrams et al., J. Biol. Chem., 264(24): 14016-14021 (1989). In yet another single-step selection method, host cells are co-transfected with an intron-modified selectable gene and a gene encoding the protein of interest. See WO 92/17566, published October 15, 1992. The intron-modified gene is prepared by inserting into the transcribed region of a selectable gene an intron of such length that the intron is correctly spliced from the corresponding mRNA precursor at low efficiency, so that the amount of selectable marker produced from the intron-modified selectable gene is substantially less than that produced from the starting selectable gene. These vectors help to insure the integrity of the integrated DNA construct, but transcriptional linkage is not achieved as selectable gene and the protein gene are driven by separate promoters.

30 Other mammalian expression vectors that have single transcription units have been described. Retroviral vectors have been constructed (Cepko et al., Cell, 37:1053-1062 [1984]) in which a cDNA is inserted between the endogenous Moloney murine leukemia virus (M-MuLV) splice donor and splice acceptor sites which are followed by a neomycin resistance gene. This vector has been used to express a variety of gene products following retroviral infection of several cell types.

With the above drawbacks in mind, it is one object of the present invention to increase the level of homogeneity with regard to expression levels of stable clones transfected with a product gene of interest, by expressing a selectable marker (DHFR) and the protein of interest from a single promoter.

It is another object to provide a method for selecting stable, recombinant host cells that express high levels of a desired protein product, which method is rapid and convenient to perform, and reduces the numbers of transfected cells which need to be screened. Furthermore, it is

an object to allow high levels of single and two unit polypeptides to be rapidly generated from clones or pools of stable host cell transfectants.

It is an additional object to provide expression vectors which bias for active integration events (i.e. have an increased tendency to generate transformants wherein the DNA construct is inserted into a region of the genome of the host cell which results in high level expression of the product gene) and can accommodate a variety of product genes without the need for modification.

10

SUMMARY OF THE INVENTION

Accordingly, the present invention is directed to a DNA construct (DNA molecule) alternative terminology comprising a 5' transcriptional initiation site and a 3' transcriptional termination site, a selectable gene (preferably an amplifiable gene) and a product gene provided 3' to the selectable gene, a transcriptional regulatory region regulating transcription of both the selectable gene and the product gene, the selectable gene positioned within an intron defined by a splice donor site and a splice acceptor site. The splice donor site preferably comprises an effective splice donor sequence as herein defined and thereby regulates expression of the product gene using the transcriptional regulatory region.

In another embodiment, the invention provides a method for producing a product of interest comprising culturing a eukaryotic cell which has been transfected with the DNA construct described above, so as to express the product gene and recovering the product.

In a further embodiment, the invention provides a method for producing eukaryotic cells having multiple copies of the product gene comprising transfecting eukaryotic cells with the DNA construct described above (where the selectable gene is an amplifiable gene), growing the cells in a selective medium comprising an amplifying agent for a sufficient time for amplification to occur, and selecting cells having multiple copies of the product gene. Preferably transfection of the cells is achieved using electroporation.

After transfection of the host cells, most of the transfectants fail to exhibit the selectable phenotype characteristic of the protein encoded by the selectable gene, but surprisingly a small proportion of the transfectants do exhibit the selectable phenotype, and among those transfectants, the majority are found to express high levels of the desired product encoded by the product gene. Thus, the invention provides an improved method for the selection of recombinant host cells expressing high levels of a desired product, which method is useful with a wide variety of eukaryotic host cells and avoids the problems inherent in existing cell selection technology.

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1A-1D illustrate schematically various DNA constructs encompassed by the instant invention. The large arrows represent the selectable gene and the product gene, the V formed by the dashed lines shows the region of the precursor RNA internal to the 5' splice donor site (SD) and 3' splice acceptor site (SA) that is excised from vectors that contain a functional SD. The transcriptional regulatory region, selectable gene, product gene and transcriptional termination site are depicted in Figure 1A. Figure 1B depicts the DNA constructs of Example 1. The various splice donor sequences are depicted, i.e., wild type ras splice donor sequence (WT ras), mutant ras splice donor sequence (MUTANT ras) and non-functional splice donor sequence (Δ GT). The probes used for Northern blot analysis in Example 1 are shown in Figure 1B. Figure 1C depicts the DNA constructs of Example 2 and Figure 1D depicts the DNA construct of Example 3 used for expression of anti-IgE V_H.

Figure 2 depicts schematically the control DNA construct used in Example 1.

Figures 3A-Q depict the nucleotide sequence (SEQ ID NO: 1) of the DHFR/intron-(WT ras SD)-tPA expression vector of Example 1.

Figure 4 is a bar graph which shows the number of colonies that form in selective medium after electroporation of linearized duplicate miniprep DNA's prepared in parallel from the three vectors shown in Figure 1B (i.e. with wild type ras splice donor sequence [WT ras], mutant ras splice donor sequence [MUTANT ras] and non-functional splice donor sequence [Δ GT]) and from the control vector that has DHFR under control of SV40 promoter and tPA under control of CMV promoter (see Figure 2). Cells were selected in nucleoside free medium and counted with an automated colony counter.

Figures 5A-C are bar graphs depicting expression of tPA from stable pools and clones generated from the vectors shown in Figure 1B. In Figure 5A greater than 100 clones from each vector transfection were mixed, plated in 24 well plates, and assayed by tPA ELISA at "saturation". In Figure 5B, twenty clones chosen at random derived from each of the vectors were assayed by tPA ELISA at "saturation". In Figure 5C, the pools mentioned in Figure 5A (except the Δ GT pool) were exposed to 200nM Mtx to select for DHFR amplification and then pooled and assayed for tPA expression.

Figures 6A-P depict the nucleotide sequence (SEQ ID NO: 2) of the DHFR/intron-(WT ras SD)-TNFr-IgG expression vector of Example 2.

Figures 7A-B are bar graphs depicting expression of TNFr-IgG using dicistronic or control vectors (see Example 2). Vectors containing TNFr-IgG (but otherwise identical to those described for tPA expression in Example 1) were constructed (see Figure 1C), introduced into dp12.CHO cells by electroporation, pooled, and assayed for product expression before (Figure 7A) and after (Figure 7B) being subjected to amplification in 200nM Mtx.

Figure 8 depicts schematically the DNA construct used for expression of the V_L of anti-IgE in Example 3.

Figures 9A-O depict the nucleotide sequence (SEQ ID NO: 3) of the anti-IgE V_H expression vector of Example 3.

Figures 10A-Q depict the nucleotide sequence (SEQ ID NO: 4) of the anti-IgE V_L expression vector of Example 3.

5 Figure 11 is a bar graph depicting anti-IgE expression in Example 3. Heavy (V_H) and light (V_L) chain expression vectors were constructed, co-electroporated into CHO cells, clones were selected and assayed for antibody expression. Additionally, pools were established and assessed with regard to expression before and after Mtx selection at 200nM and 1μM.

10 DESCRIPTION OF THE PREFERRED EMBODIMENTS

Definitions:

The "DNA construct" disclosed herein comprises a non-naturally occurring DNA molecule which can either be provided as an isolate or integrated in another DNA molecule e.g. in an expression vector or the
15 chromosome of an eukaryotic host cell.

The term "selectable gene" as used herein refers to a DNA that encodes a selectable marker necessary for the growth or survival of a host cell under the particular cell culture conditions chosen. Accordingly, a host cell that is transformed with a selectable gene will be capable of
20 growth or survival under certain cell culture conditions wherein a non-transfected host cell is not capable of growth or survival. Typically, a selectable gene will confer resistance to a drug or compensate for a metabolic or catabolic defect in the host cell. Examples of selectable genes are provided in the following table. See also Kaufman, Methods in
25 Enzymology, 185: 537-566 (1990), for a review of these.

TABLE 1
Selectable Genes and their Selection Agents

Selection Agent	Selectable Gene
Methotrexate	Dihydrofolate reductase
30 Cadmium	Metallothionein
PALA	CAD
Xyl-A-or adenosine and 2'-deoxycoformycin	Adenosine deaminase
35 Adenine, azaserine, and coformycin	Adenylate deaminase
6-Azauridine, pyrazofuran	UMP Synthetase
Mycophenolic acid	IMP 5'-dehydrogenase

	Mycophenolic acid with limiting xanthine	Xanthine-guanine phosphoribosyltransferase
	Hypoxanthine, aminopterin, and thymidine (HAT)	Mutant HGPRTase or mutant thymidine kinase
5	5-Fluorodeoxyuridine	Thymidylate synthetase
	Multiple drugs e.g. adriamycin, vincristine or colchicine	P-glycoprotein 170
	Aphidicolin	Ribonucleotide reductase
10	Methionine sulfoximine	Glutamine synthetase
	β -Aspartyl hydroxamate or Albizziin	Asparagine synthetase
	Canavanine	Arginosuccinate synthetase
	α -Difluoromethylornithine	Ornithine decarboxylase
15	Compactin	HMG-CoA reductase
	Tunicamycin	N-Acetylglucosaminyl transferase
	Borrelidin	Threonyl-tRNA synthetase
	Ouabain	Na ⁺ K ⁺ -ATPase

The preferred selectable gene is an amplifiable gene. As used herein, the term "amplifiable gene" refers to a gene which is amplified (i.e. additional copies of the gene are generated which survive in intrachromosomal or extrachromosomal form) under certain conditions. The amplifiable gene usually encodes an enzyme (i.e. an amplifiable marker) which is required for growth of eukaryotic cells under those conditions. For example, the gene may encode DHFR which is amplified when a host cell transformed therewith is grown in Mtx. According to Kaufman, the selectable genes in Table 1 above can also be considered amplifiable genes. An example of a selectable gene which is generally not considered to be an amplifiable gene is the neomycin resistance gene (Cepko et al., *supra*).

As used herein, "selective medium" refers to nutrient solution used for growing eukaryotic cells which have the selectable gene and therefore includes a "selection agent". Commercially available media such as Ham's F10 (Sigma), Minimal Essential Medium ([MEM], Sigma), RPMI-1640 (Sigma), and Dulbecco's Modified Eagle's Medium ([DMEM], Sigma) are exemplary nutrient solutions. In addition, any of the media described in Ham and Wallace, Meth. Enz., 58:44 (1979), Barnes and Sato, Anal. Biochem., 102:255

(1980), U.S. Patent Nos. 4,767,704; 4,657,866; 4,927,762; or 4,560,655; WO 90/03430; WO 87/00195; U.S. Patent Re. 30,985; or U.S. Patent No. 5,122,469, the disclosures of all of which are incorporated herein by reference, may be used as culture media. Any of these media may be
5 supplemented as necessary with hormones and/or other growth factors (such as insulin, transferrin, or epidermal growth factor), salts (such as sodium chloride, calcium, magnesium, and phosphate), buffers (such as HEPES), nucleosides (such as adenosine and thymidine), antibiotics (such as Gentamycin™ drug), trace elements (defined as inorganic compounds usually
10 present at final concentrations in the micromolar range), and glucose or an equivalent energy source. Any other necessary supplements may also be included at appropriate concentrations that would be known to those skilled in the art. The preferred nutrient solution comprises fetal bovine serum.

The term "selection agent" refers to a substance that interferes with
15 the growth or survival of a host cell that is deficient in a particular selectable gene. Examples of selection agents are presented in Table 1 above. The selection agent preferably comprises an "amplifying agent" which is defined for purposes herein as an agent for amplifying copies of the amplifiable gene, such as Mtx if the amplifiable gene is DHFR. See Table
20 1 for examples of amplifying agents.

As used herein, the term "transcriptional initiation site" refers to the nucleic acid in the DNA construct corresponding to the first nucleic acid incorporated into the primary transcript, i.e., the mRNA precursor, which site is generally provided at, or adjacent to, the 5' end of the DNA
25 construct.

The term "transcriptional termination site" refers to a sequence of DNA, normally represented at the 3' end of the DNA construct, that causes RNA polymerase to terminate transcription.

As used herein, "transcriptional regulatory region" refers to a
30 region of the DNA construct that regulates transcription of the selectable gene and the product gene. The transcriptional regulatory region normally refers to a promoter sequence (i.e. a region of DNA involved in binding of RNA polymerase to initiate transcription) which can be constitutive or inducible and, optionally, an enhancer (i.e. a cis-acting DNA element,
35 usually from about 10-300 bp, that acts on a promoter to increase its transcription).

As used herein, "product gene" refers to DNA that encodes a desired protein or polypeptide product. Any product gene that is capable of expression in a host cell may be used, although the methods of the
40 invention are particularly suited for obtaining high-level expression of a product gene that is not also a selectable or amplifiable gene. Accordingly, the protein or polypeptide encoded by a product gene typically will be one that is not necessary for the growth or survival of a host cell under the particular cell culture conditions chosen. For example, product
45 genes suitably encode a peptide, or may encode a polypeptide sequence of

amino acids for which the chain length is sufficient to produce higher levels of tertiary and/or quaternary structure.

Examples of bacterial polypeptides or proteins include, e.g., alkaline phosphatase and β -lactamase. Examples of mammalian polypeptides or proteins include molecules such as renin; a growth hormone, including human growth hormone, and bovine growth hormone; growth hormone releasing factor; parathyroid hormone; thyroid stimulating hormone; lipoproteins; alpha-1-antitrypsin; insulin A-chain; insulin B-chain; proinsulin; follicle stimulating hormone; calcitonin; luteinizing hormone; glucagon; clotting factors such as factor VIIIC, factor IX, tissue factor, and von Willebrands factor; anti-clotting factors such as Protein C; atrial natriuretic factor; lung surfactant; a plasminogen activator, such as urokinase or human urine or tissue-type plasminogen activator (t-PA); bombesin; thrombin; hemopoietic growth factor; tumor necrosis factor-alpha and -beta; enkephalinase; RANTES (regulated on activation normally T-cell expressed and secreted); human macrophage inflammatory protein (MIP-1-alpha); a serum albumin such as human serum albumin; mullerian-inhibiting substance; relaxin A-chain; relaxin B-chain; prorelaxin; mouse gonadotropin-associated peptide; a microbial protein, such as beta-lactamase; DNase; inhibin; activin; vascular endothelial growth factor (VEGF); receptors for hormones or growth factors; integrin; protein A or D; rheumatoid factors; a neurotrophic factor such as bone-derived neurotrophic factor (BDNF), neurotrophin-3, -4, -5, or -6 (NT-3, NT-4, NT-5, or NT-6), or a nerve growth factor such as NGF- β ; platelet-derived growth factor (PDGF); fibroblast growth factor such as aFGF and bFGF; epidermal growth factor (EGF); transforming growth factor (TGF) such as TGF-alpha and TGF-beta, including TGF- β 1, TGF- β 2, TGF- β 3, TGF- β 4, or TGF- β 5; insulin-like growth factor-I and -II (IGF-I and IGF-II); des(1-3)-IGF-I (brain IGF-I), insulin-like growth factor binding proteins; CD proteins such as CD-3, CD-4, CD-8, and CD-19; erythropoietin; osteoinductive factors; immunotoxins; a bone morphogenetic protein (BMP); an interferon such as interferon-alpha, -beta, and -gamma; colony stimulating factors (CSFs), e.g., M-CSF, GM-CSF, and G-CSF; interleukins (ILs), e.g., IL-1 to IL-10; superoxide dismutase; T-cell receptors; surface membrane proteins; decay accelerating factor; viral antigen such as, for example, a portion of the AIDS envelope; transport proteins; homing receptors; addressins; regulatory proteins; antibodies; chimeric proteins such as immunoadhesins and fragments of any of the above-listed polypeptides.

The product gene preferably does not consist of an anti-sense sequence for inhibiting the expression of a gene present in the host. Preferred proteins herein are therapeutic proteins such as TGF- β , TGF- α , PDGF, EGF, FGF, IGF-I, DNase, plasminogen activators such as t-PA, clotting factors such as tissue factor and factor VIII, hormones such as relaxin and insulin, cytokines such as IFN- γ , chimeric proteins such as TNF receptor IgG immunoadhesin (TNFr-IgG) or antibodies such as anti-IgE.

The term "intron" as used herein refers to a nucleotide sequence present within the transcribed region of a gene or within a messenger RNA precursor, which nucleotide sequence is capable of being excised, or spliced, from the messenger RNA precursor by a host cell prior to translation. Introns suitable for use in the present invention are suitably prepared by any of several methods that are well known in the art, such as purification from a naturally occurring nucleic acid or de novo synthesis. The introns present in many naturally occurring eukaryotic genes have been identified and characterized. Mount, Nuc. Acids Res., 10:459 (1982). Artificial introns comprising functional splice sites also have been described. Winey et al., Mol. Cell Biol., 9:329 (1989); Gattermann et al., Mol. Cell Biol., 9:1526 (1989). Introns may be obtained from naturally occurring nucleic acids, for example, by digestion of a naturally occurring nucleic acid with a suitable restriction endonuclease, or by PCR cloning using primers complementary to sequences at the 5' and 3' ends of the intron. Alternatively, introns of defined sequence and length may be prepared synthetically using various methods in organic chemistry. Narang et al., Meth. Enzymol., 68:90 (1979); Caruthers et al., Meth. Enzymol., 154:287 (1985); Froehler et al., Nuc. Acids Res., 14:5399 (1986).

As used herein "splice donor site" or "SD" refers to the DNA sequence immediately surrounding the exon-intron boundary at the 5' end of the intron, where the "exon" comprises the nucleic acid 5' to the intron. Many splice donor sites have been characterized and Ohshima et al., J. Mol. Biol., 195:247-259 (1987) provides a review of these. An "efficient splice donor sequence" refers to a nucleic acid sequence encoding a splice donor site wherein the efficiency of splicing of messenger RNA precursors having the splice donor sequence is between about 80 to 99% and preferably 90 to 95% as determined by quantitative PCR. Examples of efficient splice donor sequences include the wild type (WT) ras splice donor sequence and the GAC:GTAAGT sequence of Example 3. Other efficient splice donor sequences can be readily selected using the techniques for measuring the efficiency of splicing disclosed herein.

The terms "PCR" and "polymerase chain reaction" as used herein refer to the in vitro amplification method described in US Patent No. 4,683,195 (issued July 28, 1987). In general, the PCR method involves repeated cycles of primer extension synthesis, using two DNA primers capable of hybridizing preferentially to a template nucleic acid comprising the nucleotide sequence to be amplified. The PCR method can be used to clone specific DNA sequences from total genomic DNA, cDNA transcribed from cellular RNA, viral or plasmid DNAs. Wang & Mark, in PCR Protocols, pp. 70-75 (Academic Press, 1990); Scharf, in PCR Protocols, pp. 84-98; Kawasaki & Wang, in PCR Technology, pp. 89-97 (Stockton Press, 1989). Reverse transcription-polymerase chain reaction (RT-PCR) can be used to analyze RNA samples containing mixtures of spliced and unspliced mRNA transcripts. Fluorescently tagged primers designed to span the intron are used to

amplify both spliced and unspliced targets. The resultant amplification products are then separated by gel electrophoresis and quantitated by measuring the fluorescent emission of the appropriate band(s). A comparison is made to determine the amount of spliced and unspliced transcripts present in the RNA sample.

One preferred splice donor sequence is a "consensus splice donor sequence". The nucleotide sequences surrounding intron splice sites, which sequences are evolutionarily highly conserved, are referred to as "consensus splice donor sequences". In the mRNAs of higher eukaryotes, the 5' splice site occurs within the consensus sequence AG:GUAAGU (wherein the colon denotes the site of cleavage and ligation). In the mRNAs of yeast, the 5' splice site is bounded by the consensus sequence :GUAUGU. Padgett, et al., Ann. Rev. Biochem., 55:1119 (1986).

The expression "splice acceptor site" or "SA" refers to the sequence immediately surrounding the intron-exon boundary at the 3' end of the intron, where the "exon" comprises the nucleic acid 3' to the intron. Many splice acceptor sites have been characterized and Ohshima et al., J. Mol. Biol., 195:247-259 (1987) provides a review of these. The preferred splice acceptor site is an efficient splice acceptor site which refers to a nucleic acid sequence encoding a splice acceptor site wherein the efficiency of splicing of messenger RNA precursors having the splice acceptor site is between about 80 to 99% and preferably 90 to 95% as determined by quantitative PCR. The splice acceptor site may comprise a consensus sequence. In the mRNAs of higher eukaryotes, the 3' splice acceptor site occurs within the consensus sequence (U/C)₁₁NCAG:G. In the mRNAs of yeast, the 3' acceptor splice site is bounded by the consensus sequence (C/U)AG:. Padgett, et al., *supra*.

As used herein "culturing for sufficient time to allow amplification to occur" refers to the act of physically culturing the eukaryotic host cells which have been transformed with the DNA construct in cell culture media containing the amplifying agent, until the copy number of the amplifiable gene (and preferably also the copy number of the product gene) in the host cells has increased relative to the transformed cells prior to this culturing.

The term "expression" as used herein refers to transcription or translation occurring within a host cell. The level of expression of a product gene in a host cell may be determined on the basis of either the amount of corresponding mRNA that is present in the cell or the amount of the protein encoded by the product gene that is produced by the cell. For example, mRNA transcribed from a product gene is desirably quantitated by northern hybridization. Sambrook, et al., Molecular Cloning: A Laboratory Manual, pp. 7.3-7.57 (Cold Spring Harbor Laboratory Press, 1989). Protein encoded by a product gene can be quantitated either by assaying for the biological activity of the protein or by employing assays that are independent of such activity, such as western blotting or radioimmunoassay using antibodies that are capable of reacting with the protein. Sambrook,

et al., Molecular Cloning: A Laboratory Manual, pp. 18.1-18.88 (Cold Spring Harbor Laboratory Press, 1989).

Modes for Carrying Out the Invention

5 Methods and compositions are provided for enhancing the stability and/or copy number of a transcribed sequence in order to allow for elevated levels of a RNA sequence of interest. In general, the methods of the present invention involve transfecting a eukaryotic host cell with an expression vector comprising both a product gene encoding a desired polypeptide and a selectable gene (preferably an amplifiable gene).

10 Selectable genes and product genes may be obtained from genomic DNA, cDNA transcribed from cellular RNA, or by *in vitro* synthesis. For example, libraries are screened with probes (such as antibodies or oligonucleotides of about 20-80 bases) designed to identify the selectable gene or the product gene (or the protein(s) encoded thereby). Screening the cDNA or
15 genomic library with the selected probe may be conducted using standard procedures as described in chapters 10-12 of Sambrook et al., Molecular Cloning: A Laboratory Manual (New York: Cold Spring Harbor Laboratory Press, 1989). An alternative means to isolate the selectable gene or product gene is to use PCR methodology as described in section 14 of
20 Sambrook et al., *supra*.

 A preferred method of practicing this invention is to use carefully selected oligonucleotide sequences to screen cDNA libraries from various tissues known to contain the selectable gene or product gene. The oligonucleotide sequences selected as probes should be of sufficient length
25 and sufficiently unambiguous that false positives are minimized.

 The oligonucleotide generally is labeled such that it can be detected upon hybridization to DNA in the library being screened. The preferred method of labeling is to use ³²P- labeled ATP with polynucleotide kinase, as is well known in the art, to radiolabel the oligonucleotide. However,
30 other methods may be used to label the oligonucleotide, including, but not limited to, biotinylation or enzyme labeling.

 Sometimes, the DNA encoding the selectable gene and product gene is preceded by DNA encoding a signal sequence having a specific cleavage site at the N-terminus of the mature protein or polypeptide. In general, the
35 signal sequence may be a component of the expression vector, or it may be a part of the selectable gene or product gene that is inserted into the expression vector. If a heterologous signal sequence is used, it preferably is one that is recognized and processed (*i.e.*, cleaved by a signal peptidase) by the host cell. For yeast secretion the native signal
40 sequence may be substituted by, *e.g.*, the yeast invertase leader, alpha factor leader (including *Saccharomyces* and *Kluyveromyces* α -factor leaders, the latter described in U.S. Pat. No. 5,010,182 issued 23 April 1991), or acid phosphatase leader, the *C. albicans* glucoamylase leader (EP 362,179 published 4 April 1990), or the signal described in WO 90/13646 published
45 15 November 1990. In mammalian cell expression the native signal sequence

of the protein of interest is satisfactory, although other mammalian signal sequences may be suitable, such as signal sequences from secreted polypeptides of the same or related species, as well as viral secretory leaders, for example, the herpes simplex gD signal. The DNA for such precursor region is ligated in reading frame to the selectable gene or product gene.

As shown in Figure 1A, the selectable gene is generally provided at the 5' end of the DNA construct and this selectable gene is followed by the product gene. Therefore, the full length (non-spliced) message will contain DHFR as the first open reading frame and will therefore generate DHFR protein to allow selection of stable transfectants. The full length message is not expected to generate appreciable amounts of the protein of interest as the second AUG in a dicistronic message is an inefficient initiator of translation in mammalian cells (Kozak, J. Cell Biol., 115: 887-903 [1991]).

The selectable gene is positioned within an intron. Introns are noncoding nucleotide sequences, normally present within many eukaryotic genes, which are removed from newly transcribed mRNA precursors in a multiple-step process collectively referred to as splicing.

A single mechanism is thought to be responsible for the splicing of mRNA precursors in mammalian, plant, and yeast cells. In general, the process of splicing requires that the 5' and 3' ends of the intron be correctly cleaved and the resulting ends of the mRNA be accurately joined, such that a mature mRNA having the proper reading frame for protein synthesis is produced. Analysis of a variety of naturally occurring and synthetically constructed mutant genes has shown that nucleotide changes at many of the positions within the consensus sequences at the 5' and 3' splice sites have the effect of reducing or abolishing the synthesis of mature mRNA. Sharp, Science, 235:766 (1987); Padgett, et al., Ann. Rev. Biochem., 55:1119 (1986); Green, Ann. Rev. Genet., 20:671 (1986). Mutational studies also have shown that RNA secondary structures involving splicing sites can affect the efficiency of splicing. Solnick, Cell, 43:667 (1985); Konarska, et al., Cell, 42:165 (1985).

The length of the intron may also affect the efficiency of splicing. By making deletion mutations of different sizes within the large intron of the rabbit beta-globin gene, Wieringa, et al. determined that the minimum intron length necessary for correct splicing is about 69 nucleotides. Cell, 37:915 (1984). Similar studies of the intron of the adenovirus E1A region have shown that an intron length of about 78 nucleotides allows correct splicing to occur, but at reduced efficiency. Increasing the length of the intron to 91 nucleotides restores normal splicing efficiency, whereas truncating the intron to 63 nucleotides abolishes correct splicing. Ulfendahl, et al., Nuc. Acids Res., 13:6299 (1985).

To be useful in the invention, the intron must have a length such that splicing of the intron from the mRNA is efficient. The preparation of introns of differing lengths is a routine matter, involving methods well known in the art, such as *de novo* synthesis or *in vitro* deletion

mutagenesis of an existing intron. Typically, the intron will have a length of at least about 150 nucleotides, since introns which are shorter than this tend to be spliced less efficiently. The upper limit for the length of the intron can be up to 30 kB or more. However, as a general
5 proposition, the intron is generally less than about 10 kB in length.

The intron is modified to contain the selectable gene not normally present within the intron using any of the various known methods for modifying a nucleic acid *in vitro*. Typically, a selectable gene will be introduced into an intron by first cleaving the intron with a restriction
10 endonuclease, and then covalently joining the resulting restriction fragments to the selectable gene in the correct orientation for host cell expression, for example by ligation with a DNA ligase enzyme.

The DNA construct is dicistronic, i.e. the selectable gene and product gene are both under the transcriptional control of a single
15 transcriptional regulatory region. As mentioned above, the transcriptional regulatory region comprises a promoter. Suitable promoting sequences for use with yeast hosts include the promoters for 3-phosphoglycerate kinase (Hitzeman et al., J. Biol. Chem., 255:2073 [1980]) or other glycolytic enzymes (Hess et al., J. Adv. Enzyme Req., 7:149 [1968]; and Holland,
20 Biochemistry, 17:4900 [1978]), such as enolase, glyceraldehyde-3-phosphate dehydrogenase, hexokinase, pyruvate decarboxylase, phosphofructokinase, glucose-6-phosphate isomerase, 3-phosphoglycerate mutase, pyruvate kinase, triosephosphate isomerase, phosphoglucose isomerase, and glucokinase.

Other yeast promoters, which are inducible promoters having the
25 additional advantage of transcription controlled by growth conditions, are the promoter regions for alcohol dehydrogenase 2, isocytochrome C, acid phosphatase, degradative enzymes associated with nitrogen metabolism, metallothionein, glyceraldehyde-3-phosphate dehydrogenase, and enzymes responsible for maltose and galactose utilization. Suitable vectors and
30 promoters for use in yeast expression are further described in Hitzeman et al., EP 73,657A. Yeast enhancers also are advantageously used with yeast promoters.

Expression control sequences are known for eukaryotes. Virtually all eukaryotic genes have an AT-rich region located approximately 25 to 30
35 bases upstream from the site where transcription is initiated. Another sequence found 70 to 80 bases upstream from the start of transcription of many genes is a CXCAAT region where X may be any nucleotide.

Product gene transcription from vectors in mammalian host cells is controlled by promoters obtained from the genomes of viruses such as
40 polyoma virus, fowlpox virus (UK 2,211,504 published 5 July 1989), adenovirus (such as Adenovirus 2), bovine papilloma virus, avian sarcoma virus, cytomegalovirus, a retrovirus, hepatitis-B virus and most preferably Simian Virus 40 (SV40), from heterologous mammalian promoters, e.g. the actin promoter or an immunoglobulin promoter, from heat-shock promoters,
45 and from the promoter normally associated with the product gene, provided such promoters are compatible with the host cell systems.

The early and late promoters of the SV40 virus are conveniently obtained as an SV40 restriction fragment that also contains the SV40 viral origin of replication. Fiers et al., Nature, 273:113 (1978); Mulligan and Berg, Science, 209:1422-1427 (1980); Pavlakis et al., Proc. Natl. Acad. Sci. USA, 78:7398-7402 (1981). The immediate early promoter of the human cytomegalovirus (CMV) is conveniently obtained as a HindIII E restriction fragment. Greenaway et al., Gene, 18:355-360 (1982). A system for expressing DNA in mammalian hosts using the bovine papilloma virus as a vector is disclosed in U.S. 4,419,446. A modification of this system is described in U.S. 4,601,978. See also Gray et al., Nature, 295:503-508 (1982) on expressing cDNA encoding immune interferon in monkey cells; , Reyes et al., Nature, 297:598-601 (1982) on expression of human β -interferon cDNA in mouse cells under the control of a thymidine kinase promoter from herpes simplex virus, Canaani and Berg, Proc. Natl. Acad. Sci. USA, 79:5166-5170 (1982) on expression of the human interferon β 1 gene in cultured mouse and rabbit cells, and Gorman et al., Proc. Natl. Acad. Sci. USA, 79:6777-6781 (1982) on expression of bacterial CAT sequences in CV-1 monkey kidney cells, chicken embryo fibroblasts, Chinese hamster ovary cells, HeLa cells, and mouse NIH-3T3 cells using the Rous sarcoma virus long terminal repeat as a promoter.

Preferably the transcriptional regulatory region in higher eukaryotes comprises an enhancer sequence. Enhancers are relatively orientation and position independent having been found 5' (Lainins et al., Proc. Natl. Acad. Sci. USA, 78:993 [1981]) and 3' (Lusky et al., Mol. Cell Bio., 3:1108 [1983]) to the transcription unit, within an intron (Banerji et al., Cell, 33:729 [1983]) as well as within the coding sequence itself (Osborne et al., Mol. Cell Bio., 4:1293 [1984]). Many enhancer sequences are now known from mammalian genes (globin, elastase, albumin, α -fetoprotein and insulin). Typically, however, one will use an enhancer from a eukaryotic cell virus. Examples include the SV40 enhancer on the late side of the replication origin (bp 100-270), the cytomegalovirus early promoter enhancer (CMV), the polyoma enhancer on the late side of the replication origin, and adenovirus enhancers. See also Yaniv, Nature, 297:17-18 (1982) on enhancing elements for activation of eukaryotic promoters. The enhancer may be spliced into the vector at a position 5' or 3' to the product gene, but is preferably located at a site 5' from the promoter.

The DNA construct has a transcriptional initiation site following the transcriptional regulatory region and a transcriptional termination region following the product gene (see Figure 1A). These sequences are provided in the DNA construct using techniques which are well known in the art.

The DNA construct normally forms part of an expression vector which may have other components such as an origin of replication (i.e., a nucleic acid sequence that enables the vector to replicate in one or more selected host cells) and, if desired, one or more additional selectable gene(s). Construction of suitable vectors containing the desired coding and control sequences employs standard ligation techniques. Isolated plasmids or DNA

fragments are cleaved, tailored, and religated in the form desired to generate the plasmids required.

Generally, in cloning vectors the origin of replication is one that enables the vector to replicate independently of the host chromosomal DNA, and includes origins of replication or autonomously replicating sequences. Such sequences are well known. The 2 μ plasmid origin of replication is suitable for yeast, and various viral origins (SV40, polyoma, adenovirus, VSV or BPV) are useful for cloning vectors in mammalian cells. Generally, the origin of replication component is not needed for mammalian expression vectors (the SV40 origin may typically be used only because it contains the early promoter).

Most expression vectors are "shuttle" vectors, i.e., they are capable of replication in at least one class of organisms but can be transfected into another organism for expression. For example, a vector is cloned in *E. coli* and then the same vector is transfected into yeast or mammalian cells for expression even though it is not capable of replicating independently of the host cell chromosome.

For analysis to confirm correct sequences in plasmids constructed, plasmids from the transformants are prepared, analyzed by restriction, and/or sequenced by the method of Messing et al., Nucleic Acids Res., 9:309 (1981) or by the method of Maxam et al., Methods in Enzymology, 65:499 (1980).

The expression vector having the DNA construct prepared as discussed above is transformed into a eukaryotic host cell. Suitable host cells for cloning or expressing the vectors herein are yeast or higher eukaryote cells.

Eukaryotic microbes such as filamentous fungi or yeast are suitable hosts for vectors containing the product gene. *Saccharomyces cerevisiae*, or common baker's yeast, is the most commonly used among lower eukaryotic host microorganisms. However, a number of other genera, species, and strains are commonly available and useful herein, such as *S. pombe* [Beach and Nurse, Nature, 290:140 (1981)], *Kluyveromyces lactis* [Louvencourt et al., J. Bacteriol., 737 (1983)], *Yarrowia* [EP 402,226], *Pichia pastoris* [EP 183,070], *Trichoderma reesia* [EP 244,234], *Neurospora crassa* [Case et al., Proc. Natl. Acad. Sci. USA, 76:5259-5263 (1979)], and *Aspergillus* hosts such as *A. nidulans* [Ballance et al., Biochem. Biophys. Res. Commun., 112:284-289 (1983); Tilburn et al., Gene, 26:205-221 (1983); Yelton et al., Proc. Natl. Acad. Sci. USA, 81:1470-1474 (1984)] and *A. niger* [Kelly and Hynes, EMBO J., 4:475-479 (1985)].

Suitable host cells for the expression of the product gene are derived from multicellular organisms. Such host cells are capable of complex processing and glycosylation activities. In principle, any higher eukaryotic cell culture is workable, whether from vertebrate or invertebrate culture. Examples of invertebrate cells include plant and insect cells. Numerous baculoviral strains and variants and corresponding permissive insect host cells from hosts such as *Spodoptera frugiperda*

(caterpillar), *Aedes aegypti* (mosquito), *Aedes albopictus* (mosquito), *Drosophila melanogaster* (fruitfly), and *Bombyx mori* host cells have been identified. See, e.g., Luckow et al., Bio/Technology, 6:47-55 (1988); Miller et al., in Genetic Engineering, Setlow, J.K. et al., eds., Vol. 8 (Plenum Publishing, 1986), pp. 277-279; and Maeda et al., Nature, 315:592-594 (1985). A variety of such viral strains are publicly available, e.g., the L-1 variant of *Autographa californica* NPV and the Bm-5 strain of *Bombyx mori* NPV, and such viruses may be used as the virus herein according to the present invention, particularly for transfection of *Spodoptera frugiperda* cells.

Plant cell cultures of cotton, corn, potato, soybean, petunia, tomato, and tobacco can be utilized as hosts. Typically, plant cells are transfected by incubation with certain strains of the bacterium *Agrobacterium tumefaciens*, which has been previously manipulated to contain the product gene. During incubation of the plant cell culture with *A. tumefaciens*, the product gene is transferred to the plant cell host such that it is transfected, and will, under appropriate conditions, express the product gene. In addition, regulatory and signal sequences compatible with plant cells are available, such as the nopaline synthase promoter and polyadenylation signal sequences. Depicker et al., J. Mol. Appl. Gen., 1:561 (1982). In addition, DNA segments isolated from the upstream region of the T-DNA 780 gene are capable of activating or increasing transcription levels of plant-expressible genes in recombinant DNA-containing plant tissue. EP 321,196 published 21 June 1989.

However, interest has been greatest in vertebrate cells, and propagation of vertebrate cells in culture (tissue culture) has become a routine procedure in recent years [Tissue Culture, Academic Press, Kruse and Patterson, editors (1973)]. Examples of useful mammalian host cell lines are monkey kidney CV1 line transformed by SV40 (COS-7, ATCC CRL 1651); human embryonic kidney line (293 or 293 cells subcloned for growth in suspension culture, Graham et al., J. Gen. Virol., 36:59 [1977]); baby hamster kidney cells (BHK, ATCC CCL 10); Chinese hamster ovary cells/-DHFR (CHO, Urlaub and Chasin, Proc. Natl. Acad. Sci. USA, 77:4216 [1980]); dp12.CHO cells (EP 307,247 published 15 March 1989); mouse sertoli cells (TM4, Mather, Biol. Reprod., 23:243-251 [1980]); monkey kidney cells (CV1 ATCC CCL 70); African green monkey kidney cells (VERO-76, ATCC CRL-1587); human cervical carcinoma cells (HELA, ATCC CCL 2); canine kidney cells (MDCK, ATCC CCL 34); buffalo rat liver cells (BRL 3A, ATCC CRL 1442); human lung cells (W138, ATCC CCL 75); human liver cells (Hep G2, HB 8065); mouse mammary tumor (MMT 060562, ATCC CCL51); TRI cells (Mather et al., Annals N.Y. Acad. Sci., 383:44-68 [1982]); MRC 5 cells; FS4 cells; and a human hepatoma line (Hep G2).

Host cells are transformed with the above-described expression or cloning vectors of this invention and cultured in conventional nutrient media modified as appropriate for inducing promoters, selecting transformants, or amplifying the genes encoding the desired sequences.

Infection with *Agrobacterium tumefaciens* is used for transformation of certain plant cells, as described by Shaw et al., Gene, 23:315 (1983) and WO 89/05859 published 29 June 1989. For mammalian cells without such cell walls, the calcium phosphate precipitation method of Graham and van der Eb, Virology, 52:456-457 (1978) may be used. General aspects of mammalian cell host system transformations have been described by Axel in U.S. 4,399,216 issued 16 August 1983. Transformations into yeast are typically carried out according to the method of Van Solingen et al., J. Bact., 130:946 (1977) and Hsiao et al., Proc. Natl. Acad. Sci. (USA), 76:3829 (1979). However, other methods for introducing DNA into cells such as by nuclear injection or by protoplast fusion may also be used.

In the preferred embodiment the DNA is introduced into the host cells using electroporation. See Andreason, J. Tiss. Cult. Meth., 15:56-62 (1993), for a review of electroporation techniques useful for practicing the instantly claimed invention. It was discovered that electroporation techniques for introducing the DNA construct into the host cells were preferable over calcium phosphate precipitation techniques insofar as the latter could cause the DNA to break up and forming concatemers.

The mammalian host cells used to express the product gene herein may be cultured in a variety of media as discussed in the definitions section above. The media contains the selection agent used for selecting transformed host cells which have taken up the DNA construct (either as an intra- or extra-chromosomal element). To achieve selection of the transformed eukaryotic cells, the host cells may be grown in cell culture plates and individual colonies expressing the selectable gene (and thus the product gene) can be isolated and grown in growth medium until the nutrients are depleted. The host cells are then analyzed for transcription and/or transformation as discussed below. The culture conditions, such as temperature, pH, and the like, are those previously used with the host cell selected for expression, and will be apparent to the ordinarily skilled artisan.

Gene amplification and/or expression may be measured in a sample directly, for example, by conventional Southern blotting, Northern blotting to quantitate the transcription of mRNA (Thomas, Proc. Natl. Acad. Sci. USA, 77:5201-5205 [1980]), dot blotting (DNA analysis), or in situ hybridization, using an appropriately labeled probe, based on the sequences provided herein. Various labels may be employed, most commonly radioisotopes, particularly ³²P. However, other techniques may also be employed, such as using biotin-modified nucleotides for introduction into a polynucleotide. The biotin then serves as the site for binding to avidin or antibodies, which may be labeled with a wide variety of labels, such as radionuclides, fluorescents, enzymes, or the like. Alternatively, antibodies may be employed that can recognize specific duplexes, including DNA duplexes, RNA duplexes, and DNA-RNA hybrid duplexes or DNA-protein duplexes. The antibodies in turn may be labeled and the assay may be carried out where the duplex is bound to a surface, so that upon the

formation of duplex on the surface, the presence of antibody bound to the duplex can be detected.

Gene expression, alternatively, may be measured by immunological methods, such as immunohistochemical staining of tissue sections and assay
5 of cell culture or body fluids, to quantitate directly the expression of gene product. With immunohistochemical staining techniques, a cell sample is prepared, typically by dehydration and fixation, followed by reaction with labeled antibodies specific for the gene product coupled, where the labels are usually visually detectable, such as enzymatic labels,
10 fluorescent labels, luminescent labels, and the like. A particularly sensitive staining technique suitable for use in the present invention is described by Hsu et al., Am. J. Clin. Path., 75:734-738 (1980).

In the preferred embodiment, the mRNA is analyzed by quantitative PCR (to determine the efficiency of splicing) and protein expression is
15 measured using ELISA as described in Example 1 herein.

The product of interest preferably is recovered from the culture medium as a secreted polypeptide, although it also may be recovered from host cell lysates when directly expressed without a secretory signal. When the product gene is expressed in a recombinant cell other than one of human
20 origin, the product of interest is completely free of proteins or polypeptides of human origin. However, it is necessary to purify the product of interest from recombinant cell proteins or polypeptides to obtain preparations that are substantially homogeneous as to the product of interest. As a first step, the culture medium or lysate is centrifuged
25 to remove particulate cell debris. The product of interest thereafter is purified from contaminant soluble proteins and polypeptides, for example, by fractionation on immunoaffinity or ion-exchange columns; ethanol precipitation; reverse phase HPLC; chromatography on silica or on a cation exchange resin such as DEAE; chromatofocusing; SDS-PAGE; ammonium sulfate
30 precipitation; gel electrophoresis using, for example, Sephadex G-75; chromatography on plasminogen columns to bind the product of interest and protein A Sepharose columns to remove contaminants such as IgG.

The following examples are offered by way of illustration only and are not intended to limit the invention in any manner. All patent and
35 literature references cited herein are expressly incorporated by reference.

EXAMPLE 1

tPA production using the dicistronic expression vectors

It was sought to increase the level of homogeneity with regard to expression levels of stable clones by expressing a selectable marker (such
40 as DHFR) and the protein of interest from a single promoter. These vectors divert most of the transcript to product expression while linking it at a fixed ratio to DHFR expression via differential splicing.

Vectors were constructed which were derived from the vector pRK (Suva et al., Science, 237:893-896 [1987]) which contains an intron between the
45 cytomegalovirus immediate early promoter (CMV) and the cDNA that encodes

the polypeptide of interest. The intron of pRK is 139 nucleotides in length, has a splice donor site derived from cytomegalovirus immediate early gene (CMVIE), and a splice acceptor site from an IgG heavy chain variable region (V_H) gene (Eaton et al., Biochem., 25:8343 [1986]).

5 DHFR/intron vectors were constructed by inserting an EcoRV linker into the BSTX1 site present in the intron of pRK7. An 830 base-pair fragment containing a mouse DHFR coding fragment was inserted to obtain DHFR intron expression vectors which differ only in the sequence that comprises the splice donor site. Those sequences were altered by
10 overlapping PCR mutagenesis to obtain sequences that match splice donor sites found between exons 3 and 4 of normal and mutant Ras genes. PCR was also used to destroy the splice donor site.

A mouse DHFR cDNA fragment (Simonsen et al., Proc. Natl. Acad. Sci. USA, 80:2495-2499 [1983]) was inserted into the intron of this vector 59
15 nucleotides downstream of the splice donor site. The splice donor site of this vector was altered by mutagenesis to change the ratio of spliced to non-spliced message in transfected cells. It has previously been shown that a single nucleotide change (G to A) converted a relatively efficient splice donor site found in the normal ras gene into an inefficient splice
20 site (Cohen et al., Nature, 334:119-124 [1988]). This effect has been demonstrated in the context of the ras gene and confirmed when these sequences were transferred to human growth hormone constructs (Cohen et al., Cell, 58:461-472 [1989]). Additionally, a non functional 5' splice site (GT to CA) was constructed as a control (Δ GT). A polylinker was
25 inserted 35 nucleotides downstream of the 3' splice site to accept the cDNA of interest. A vector containing tPA (Pennica et al., Nature, 301:214-221 [1983]) was linearized downstream of the polyadenylation site before it was introduced into CHO cells (Potter et al., Proc. Natl. Acad. Sci. USA, 81:7161 [1984]).

30 Plasmid DNA's that contained DHFR/intron, tPA and (a) wild type ras (WT ras), i.e. Figure 3 (SEQ ID NO: 1), (b) mutant ras, or (c) non-functional splice donor site (Δ GT) were introduced into CHO DHFR minus cells by electroporation. The intron vectors were each linearized downstream of the polyadenylation site by restriction endonuclease
35 treatment. The control vector was linearized downstream of the second polyadenylation site. The DNA's were ethanol precipitated after phenol/chloroform extraction and were resuspended in 20 μ l 1/10 Tris EDTA. Then, 10 μ g of DNA was incubated with 10⁷ CHO.dp12 cells (EP 307,247 published 15 March 1989) in 1 ml of PBS on ice for 10 min. before
40 electroporation at 400 volts and 330 μ f using a BRL Cell Porator.

Cells were returned to ice for 10 min. before being plated into non-selective medium. After 24 hours cells were fed nucleoside-free medium to select for stable DHFR+ clones which were pooled. The pooled DHFR+ clones were lysed and mRNA's were prepared.

45 To prepare the mRNA, RNA was extracted from 5 x 10⁷ cells which were grown from pools of more than 200 clones derived from the stable

transfection of the three vectors, the essential construction of which is shown in Figure 1B and from non-transfected CHO cells. RNA was purified over oligo-DT cellulase (Collaborative Biomedical Products). 10 μ g of mRNA was then subjected to Northern blotting which involved running the mRNA on a 1.2% agarose, 6.6% formaldehyde gel, and transferring it to a nylon filter (Stratagene Duralon-UV membrane), prehybridized, probed and washed according to the manufacturer's instructions.

The filter was probed sequentially using probes (shown in Figure 1B) that would detect (a) the full length message, (b) both full length and spliced message, or (c) beta actin. Probing with the long probe showed that the vector that contains the efficient splice donor site (*i.e.* WT ras) generates predominately a mRNA of the size predicted for the spliced product while the other two vectors gave rise primarily to a mRNA that corresponds in size to non-spliced message. The DHFR probe detected only full length message and demonstrated that the WT ras splice donor derived vector generates very little full length message with which to confer a DHFR positive phenotype.

Figure 4 shows the number of DHFR positive colonies obtained after duplicate electroporations with the three intron vectors described above and from a conventional vector that has a CMV promoter driving tPA and a SV40 promoter driving DHFR (see Figure 2). The increase in colony number parallels the increase in full length message that accumulates with the modification of the splice donor sites. The conventional vector efficiently generates colonies and does not vary significantly from the Δ GT construct.

The level of tPA expression was determined by seeding cells in 1 ml of F12:DMEM (50:50, with 5% FBS) in 24 well dishes to near confluency. Growth of the cells continued until the media was exhausted. Media was then assayed by ELISA for tPA production. Briefly, anti-tPA antibody was coated onto the wells of an ELISA microtiter plate, media samples were added to the wells followed by washing. Binding of the antigen (tPA) was then quantified using horse radish peroxidase (HRPO) labelled anti-tPA antibody.

Figure 5A depicts the titers of secreted tPA protein after pooling the clones of each group shown in Figure 4. While the number of colonies increased with a weakening of splice donor function, the inverse was seen with respect to tPA expression. The expression levels are consistent with the RNA products that are observed; as more of the dicistronic message is spliced an increased amount of message will contain tPA as the first open reading frame resulting in increased tPA expression. A mutation of GT to CA in the splice donor site results in an abundance of DHFR positive colonies which express undetectable levels of tPA, possibly resulting from inefficient utilization of the second AUG. Importantly, Figure 5A also shows that expression levels obtained from one of the dicistronic vectors (with WT ras SD) was about threefold higher than that obtained with the control vector containing a CMV promoter/enhancer driving tPA, SV40

promoter/enhancer controlling DHFR and SV40 polyadenylation signals controlling the expression of tPA and DHFR.

Additionally, the homogeneity of expression in the pools was investigated. Figure 5B shows that all 20 clones generated by the WT ras splice donor site derived dicistronic vectors express detectable levels of tPA while only 4 of 20 clones generated by the control vector express tPA. None of the clones transfected with the non-splicing (Δ GT) vector expressed tPA levels detectable by ELISA. This finding is consistent with previous observations that relatively few clones generated by conventional vectors make useful levels of protein.

Expression of tPA was increased following methotrexate amplification of pools. Figure 5C shows that 2 of the dicistronic vector derived pools (i.e. with WT ras and MUTANT ras SD sites) increased in expression markedly (8.4 and 7.7 fold), while the pool generated by the conventional vector increased only slightly (2.8 fold) when each was subjected to 200 nM Mtx. An overall increase of 9 fold was obtained using the best dicistronic (WT ras SD) versus the conventional vector following amplification. Growth of the highest expressing amplified pool in nutrient rich production medium yielded titers of 4.2 μ g/ml tPA.

It was shown that manipulation of the splice donor sequence alters the ratio of spliced to full length message and the number of colonies that form in selective medium. It was also shown that dicistronic expression vectors generate clones that express high levels of recombinant proteins. Surprisingly, it was possible to isolate high expressors which had the efficient WT ras splice donor site by selection for DHFR^r cells despite the efficiency with which the DHFR gene was spliced from the RNA precursors formed in these cells.

EXAMPLE 2

TNFr-IgG production using the dicistronic expression vectors

To prove the general applicability of this approach, a second product was evaluated in the dicistronic vector system containing, as the DNA of interest, an immunoadhesin (TNFr-IgG) capable of binding tumor necrosis factor (TNF) (Ashkenazi et al., Proc. Natl. Acad. Sci. USA, **88**:10535-10539 [1991]). The experiments described in Example 1 above were essentially repeated except that the product gene encoded the immunoadhesin TNFr-IgG. Plasmid DNA's that contained a TNFr-IgG cDNA and (a) WT ras, i.e. Figure 6 (SEQ ID NO: 2), (b) mutant ras or (c) nonfunctional splice donor site (Δ GT) were introduced into the dpl2.CHO cells as discussed for Example 1. See Figure 1C for an illustration of the DNA constructs.

It was discovered that the number of DHFR positive colonies generated by three of these vectors was similar to that seen with the tPA constructs. Expression of TNFr-IgG also paralleled that seen with the tPA constructs (Figure 7A). Amplification of pools from two of the constructs showed a marked increase in expression of immunoadhesin (9.6 and 6.8 fold) (Figure

7B). The best of these amplified pools expressed 9.5 µg/ml when grown in nutrient rich production medium.

Thus, it was again shown that dicistronic expression vectors generate clones that express high levels of recombinant proteins. Furthermore, 5 contrary to expectations, it was discovered that isolation of high product expressing host DHFR⁺ cells was possible using an efficient splice donor site (i.e. the WT ras splice donor site).

EXAMPLE 3

Antibody production using a dicistronic expression vector

10 The usefulness of this system for antibody expression was evaluated by testing production of an antibody directed against IgE (Presta et al., Journal of Immunology, 151:2623-2632 [1993]). Further, the flexibility of the system with regard to transcription initiation was tested by replacing the CMV promoter/enhancer present in the previous vectors with the 15 promoter/ enhancer derived from the early region of SV40 virus (Griffin, B., Structure and Genomic Organization of SV40 and Polyoma Virus, In J. Tooze [Ed] DNA Tumor Viruses, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York). The heavy chain of the antibody was inserted downstream of DHFR as described in the earlier tPA and TNFr-IgG constructs. 20 Additionally, a new splice donor site sequence (GAC:GTAAGT) was engineered into the vector which matches the consensus splice donor site more closely than did the splice donor sites present in the vectors tested in Examples 1 and 2. The resultant expression vector is shown in Figures 1D and 9.

It was discovered that this vector produced fewer colonies than the 25 vectors previously tested, and produced predominantly a spliced RNA product. A second vector was constructed to have the light chain of the antibody under control of the SV40 promoter/enhancer and poly-A and the hygromycin B resistance gene under control of the CMV promoter/enhancer and SV40 poly-A. These vectors were linearized at unique HpaI sites downstream 30 of the poly-A signal, mixed at a ratio of light chain vector to heavy chain vector of 10:3 and electroporated into CHO cells using an optimized protocol (as discussed in Examples 1 and 2).

Figure 11 shows the levels of antibody expressed by clones and pools after selection in hygromycin B followed by selection for DHFR expression. 35 All 20 of the clones analyzed expressed high levels of antibody when grown in rich medium and varied from one another by only a factor of four. A pool of antibody producing clones was generated and assayed shortly after it was established. That pool was grown continuously for 6 weeks without a significant decrease in productivity demonstrating that its stability was 40 sufficient to generate gram quantities of protein from its large scale culture.

The pool was subjected to methotrexate amplification at 200nM and 1µM and achieved a greater than 2 fold increase in antibody titer. The 1µM Mtx resistant pool achieved a titer of 41 mg/L when grown under optimal 45 conditions in suspension culture.

The structure of the expressed antibody was examined. Proteins expressed by the 200nM methotrexate resistant pool and by a well characterized expression clone generated by conventional vectors (Presta et al. [1993], supra) were metabolically labeled with S³⁵ cysteine and methionine. In particular, confluent 35mm plates of cells were metabolically labeled with 50μCi each S-35 methionine and S-35 cysteine (Amersham) in serum free cysteine and methionine free F12:DMEM. After one hour, nutrient rich production media was added and labeled proteins were allowed to "chase" into the medium for six more hours. Proteins were run on a 12% SDS/PAGE gel (NOVEX) non-reduced or following reduction with B-mercaptoethanol. Dried gels were exposed to film for 16 hours. CHO control cells were also labeled.

The majority of the antibody protein is secreted with a molecular weight of about 155 kilodaltons, consistent with a properly disulfide-linked antibody molecule with 2 light and 2 heavy chains. Upon reduction the molecular weight shifts to 2 approximately equally abundant proteins of 22.5 and 55 kilodaltons. The protein generated from the pool is indistinguishable from the antibody produced by the well characterized expression clone, with no apparent increase of free heavy or light chain expressed by the pool.

CONCLUSION

The efficient expression system described herein utilizes vectors consisting of promoter/enhancer elements followed by an intron containing the selectable marker coding sequence, followed by the cDNA of interest and a polyadenylation signal.

Several splice donor site sequences were tested for their effect on colony number and expression of the cDNA of interest. A non-functional splice donor site, splice donor sites found in an intron between exons 3 and 4 of mutant (mutant ras) and normal (WT ras) forms of the Harvey Ras gene and another efficient SD site (see Example 3) were used. The vectors were designed to direct expression of dicistronic primary transcripts. Within a transfected cell some of the transcripts remain full length while the remainder are spliced to excise the DHFR coding sequence. When the splice donor site is weakened or destroyed an increase in colony number is observed.

Expression levels show the inverse pattern, with the most efficient splice donor sites generating the highest levels of tPA, TNF α immunoadhesin or anti-IgE V α .

The homogeneity of expression of clones generated by the ras splice donor site intron DHFR vectors was compared to clones generated from a conventional vector with a separate promoter/enhancer and polyadenylation signal for each DHFR and tPA. The DHFR intron vector gives rise to colonies that are much more homogeneous with regard to expression than those generated by the conventional vector. Non-expressing clones derived from the conventional vector may be the result of breaks in the tPA or

TNFr-IgG domain of the plasmid during integration into the genome or the result of methylation of promoter elements (Busslinger et al., Cell, 34:197-206 [1983]; Watt et al., Genes and Development, 2:1136-1143 [1988]) driving tPA or TNFr-IgG expression. Promoter silencing by methylation or breaks in the DHFR-intron vectors would very likely render them incapable of conferring a DHFR positive phenotype.

It was found that pools generated by the DHFR-intron vectors could be amplified in methotrexate and would increase in expression by a factor of 8.4 (tPA), or 9.8 (TNFr-IgG). Pools from conventional vectors increased by only 2.8 and 3.0 fold for tPA and TNFr-IgG when amplified similarly. Amplified pools resulted in 9 fold higher tPA levels and 15 fold higher TNFr-IgG levels when compared to the conventional vector amplified pools.

Without being limited to any theory, the increase in expression of methotrexate resistant pools derived from the dicistronic vectors is likely due to the transcriptional linkage of DHFR and the product; when cells are selected for increased DHFR expression they consistently over-express product. Conventional approaches lack selectable marker and cDNA expression linkage and therefore methotrexate amplification often generates DHFR overexpression without the concomitant increase in product expression.

A further increase of 4 and 6.3 fold in expression were obtained when amplified tPA and TNFr-IgG pools were transferred from the media used for the selections and amplifications to a nutrient rich production medium.

In Example 3, the expression vector had a splice donor site that more closely matches the consensus splice donor sequence and had the heavy chain of a humanized anti-IgE antibody inserted downstream. This vector was linearized and co-electroporated with a second linearized vector that expresses the hygromycin resistance gene and the light chain of the antibody each under the control of its own promoter/enhancer and poly-A signals. An excess of light chain expression vector over the heavy chain dicistronic expression vector was used to bias in favor of light chain expression. Clones and a pool were generated after hygromycin B and DHFR selections. The clones were found to express relatively consistent, high levels of antibody, as did the pool. The 1 μ M pool achieved a titer of 41mg/L when grown under optimal conditions in suspension culture.

The anti-IgE antibody was assessed by metabolic labeling followed by SDS/PAGE under reducing and non reducing conditions and found to be indistinguishable from the protein expressed by a highly characterized clonal cell line. Of particular importance is the finding that no free light chain is observed in the pool relative to the clone.

A stable expression system for CHO cells has been developed that produces high levels of recombinant proteins rapidly and with less effort than that required by other expression systems. The vector system generates stable clones that express consistently high levels thereby reducing the number of clones that must be screened to obtain a highly productive clonal line. Alternatively, pools have been used to conveniently generate moderate to high levels of protein. This approach

may be particularly useful when a number of related proteins are to be expressed and compared.

Without being limited to this theory, it is possible the vectors that have very efficient splice donor sites generate very productive clones because so little transcript remains non spliced that only integration events that lead to the generation of high levels of RNA produce enough DHFR protein to give rise to colonies in selective medium. The high level of spliced message from such clones is then translated into abundant amounts of the protein of interest. Pools of clones made concurrently by introducing conventional vectors expressed lower levels of protein, and were unstable with regard to long term expression, and expression could not be appreciably increased when the cells were subjected to methotrexate amplification.

The system developed herein is versatile in that it allows high levels of single and multiple subunit polypeptides to be rapidly generated from clones or pools of stable transfectants. This expression system combines the advantages of transient expression systems (rapid and labor non intensive generation of research amounts of protein) with the concurrent development of highly productive stable production cell lines.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- 5 (i) APPLICANT: GENENTECH, INC.
- (ii) TITLE OF INVENTION: METHOD FOR SELECTING HIGH-EXPRESSING HOST CELLS
- 10 (iii) NUMBER OF SEQUENCES: 4
- (iv) CORRESPONDENCE ADDRESS:
- (A) ADDRESSEE: Genentech, Inc.
- (B) STREET: 460 Point San Bruno Blvd
- 15 (C) CITY: South San Francisco
- (D) STATE: California
- (E) COUNTRY: USA
- (F) ZIP: 94080
- 20 (v) COMPUTER READABLE FORM:
- (A) MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
- (B) COMPUTER: IBM PC compatible
- (C) OPERATING SYSTEM: PC-DOS/MS-DOS
- (D) SOFTWARE: patin (Genentech)
- 25 (vi) CURRENT APPLICATION DATA:
- (A) APPLICATION NUMBER:
- (B) FILING DATE:
- (C) CLASSIFICATION:
- 30 (vii) PRIOR APPLICATION DATA:
- (A) APPLICATION NUMBER: 08/286740
- (B) FILING DATE: 05-AUG-1994
- (viii) ATTORNEY/AGENT INFORMATION:
- 35 (A) NAME: Lee, Wendy M.
- (B) REGISTRATION NUMBER: 00,000
- (C) REFERENCE/DOCKET NUMBER: 798PCT
- (ix) TELECOMMUNICATION INFORMATION:
- 40 (A) TELEPHONE: 415/225-1994
- (B) TELEFAX: 415/952-9881
- (C) TELEX: 910/371-7168

(2) INFORMATION FOR SEQ ID NO:1:

- 45 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 7360 bases
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- 50 (D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:
- 55 TTCGAGCTCG CCCGACATTG ATTATTGACT AGTTATTAAT AGTAATCAAT 50
- TACGGGGTCA TTAGTTCATA GCCCATATAT GGAGTTCCGC GTTACATAAC 100
- 60 TTACGGTAAA TGGCCCGCCT GGCTGACCGC CCAACGACCC CCGCCCATTG 150
- ACGTCAATAA TGACGTATGT TCCCATAGTA ACGCCAATAG GGACTTTCCA 200
- 65 TTGACGTCAA TGGGTGGAGT ATTTACGGTA AACTGCCCAC TTGGCAGTAC 250

ATCAAGTGTA TCATATGCCA AGTACGCCCC CTATTGACGT CAATGACGGT 300

5 AAATGGCCCG CCTGGCATTG TGGCCAGTAC ATGACCTTAT GGGACTTTCC 350

TACTTGGCAG TACATCTACG TATTAGTCAT CGCTATTACC ATGGTGATGC 400

10 GGTTTTGGCA GTACATCAAT GGGCGTGGAT AGCGGTTTGA CTCACGGGGA 450

TTTCCAAGTC TCCACCCCAT TGACGTCAAT GGGAGTTTGT TTTGGCACCA 500

15 AAATCAACGG GACTTTCCAA AATGTCGTAA CAACTCCGCC CCATTGACGC 550

AAATGGGCGG TAGGCGTGTA CGGTGGGAGG TCTATATAAG CAGAGCTCGT 600

20 TTAGTGAACC GTCAGATCGC CTGGAGACGC CATCCACGCT GTTTTGACCT 650

25 CCATAGAAGA CACCGGGACC GATCCAGCCT CCGCGGCCCG GAACGGTGCA 700

TTGGAACGCG GATTCCCCGT GCCAAGAGTG CTGTAAGTAC CGCCTATAGA 750

30 GCGATAAGAG GATTTTATCC CCGCTGCCAT CATGGTTCGA CCATTGAACT 800

35 GCATCGTCGC CGTGTCCCAA AATATGGGGA TTGGCAAGAA CGGAGACCTA 850

CCCTGCCCTC CGCTCAGGAA CGCGTTCAAG TACTTCCAAA GAATGACCAC 900

40 AACCTCTTCA GTGGAAGGTA AACAGAATCT GGTGATTATG GGTAGGAAAA 950

CCTGGTTCTC CATTCTGAG AAGAATCGAC CTTTAAAGGA CAGAATTAAT 1000

45 ATAGTTCTCA GTAGAGAACT CAAAGAACCA CCACGAGGAG CTCATTTTCT 1050

TGCCAAAAGT TTGGATGATG CCTTAAGACT TATTGAACAA CCGGAATTGG 1100

50 CAAGTAAAGT AGACATGGTT TGGATAGTCG GAGGCAGTTC TGTTTACCAG 1150

55 GAAGCCATGA ATCAACCAGG CCACCTTAGA CTCTTTGTGA CAAGGATCAT 1200

GCAGGAATTT GAAAGTGACA CGTTTTTCCC AGAAATTGAT TTGGGGAAAT 1250

60 ATAAACCTCT CCCAGAATAC CCAGGCGTCC TCTCTGAGGT CCAGGAGGAA 1300

AAAGGCATCA AGTATAAGTT TGAAGTCTAC GAGAAGAAAG ACTAACAGGA 1350

65 AGATGCTTTC AAGTTCTCTG CTCCCCTCCT AAAGCTATGC ATTTTTATAA 1400

GACCATGGGA CTTTTGCTGG CTTTAGACCC CCTTGGCTTC GTTAGAACGC 1450

5 GGCTACAATT AATACATAAC CTTATGTATC ATACACATAG ATTTAGGTGA 1500

CACTATAGAA TAACATCCAC TTTGCCTTTC TCTCCACAGG TGTCACTCCA 1550

10 GGTCAACTGC ACCTCGGTTC TAAGCTTGGG CTGCAGGTCG CCGTGAATTT 1600

AAGGGACGCT GTGAAGCAAT CATGGATGCA ATGAAGAGAG GGCTCTGCTG 1650

15 TGTGCTGCTG CTGTGTGGAG CAGTCTTCGT TTCGCCCAGC CAGGAAATCC 1700

ATGCCCCGATT CAGAAGAGGA GCCAGATCTT ACCAAGTGAT CTGCAGAGAT 1750

20 GAAAAAACGC AGATGATATA CCAGCAACAT CAGTCATGGC TCGGCCCTGT 1800

25 GCTCAGAAGC AACC GG GTGG AATATTGCTG GTGCAACAGT GGCAGGGCAC 1850

AGTGCCACTC AGTGCCTGTC AAAAGTTGCA GCGAGCCAAG GTGTTTCAAC 1900

30 GGGGGCACCT GCCAGCAGGC CCTGTACTTC TCAGATTTCG TGTGCCAGTG 1950

CCCCGAAGGA TTTGCTGGGA AGTGCTGTGA AATAGATACC AGGGCCACGT 2000

35 GCTACGAGGA CCAGGGCATC AGCTACAGGG GCACGTGGAG CACAGCGGAG 2050

40 AGTGGCGCCG AGTGCACCAA CTGGAACAGC AGCGCGTTGG CCCAGAAGCC 2100

CTACAGCGGG CGGAGGCCAG ACGCCATCAG GCTGGGCCTG GGAACCACA 2150

45 ACTACTGCAG AAACCCAGAT CGAGACTCAA AGCCCTGGTG CTACGTCTTT 2200

AAGGCGGGGA AGTACAGCTC AGAGTTCTGC AGCACCCTG CCTGCTCTGA 2250

50 GGGAAACAGT GACTGCTACT TTGGGAATGG GTCAGCCTAC CGTGGCACGC 2300

55 ACAGCCTCAC CGAGTCGGGT GCCTCCTGCC TCCCGTGGAA TTCCATGATC 2350

CTGATAGGCA AGGTTTACAC AGCACAGAAC CCCAGTGCCC AGGCACTGGG 2400

60 CCTGGGCAAA CATAATTACT GCCGGAATCC TGATGGGGAT GCCAAGCCCT 2450

GGTGCCACGT GCTGAAGAAC CGCAGGCTGA CGTGGGAGTA CTGTGATGTG 2500

65 CCCTCCTGCT CCACCTGCGG CCTGAGACAG TACAGCCAGC CTCAGTTTCG 2550

CATCAAAGGA GGGCTCTTCG CCGACATCGC CTCCCACCCC TGGCAGGCTG 2600

5 CCATCTTTGC CAAGCACAGG AGGTCGCCCC GAGAGCGGTT CCTGTGCGGG 2650

GGCATACTCA TCAGCTCCTG CTGGATTCTC TCTGCCGCCC ACTGCTTCCA 2700

10 GGAGAGGTTT CCGCCCCACC ACCTGACGGT GATCTTGGGC AGAACATACC 2750

GGGTGGTCCC TGGCGAGGAG GAGCAGAAAT TTGAAGTCGA AAAATACATT 2800

15 GTCCATAAGG AATTCGATGA TGACACTTAC GACAATGACA TTGCGTGCT 2850

GCAGCTGAAA TCGGATTCGT CCCGCTGTGC CCAGGAGAGC AGCGTGGTCC 2900

20 GCACTGTGTG CCTTCCCCCG GCGGACCTGC AGCTGCCGGA CTGGACGGAG 2950

25 TGTGAGCTCT CCGGCTACGG CAAGCATGAG GCCTTGCTCT CTTTCTATTC 3000

GGAGCGGCTG AAGGAGGCTC ATGTCAGACT GTACCCATCC AGCCGCTGCA 3050

30 CATCACAACA TTTACTTAAC AGAACAGTCA CCGACAACAT GCTGTGTGCT 3100

GGAGAACTC GGAGCGGCGG GCCCCAGGCA AACTTGCACG ACGCCTGCCA 3150

35 GGGCGATTCT GGAGGCCCCC TGGTGTGTCT GAACGATGGC CGCATGACTT 3200

40 TGGTGGGCAT CATCAGCTGG GGCCTGGGCT GTGGACAGAA GGATGTCCCG 3250

GGTGTGTACA CCAAGGTTAC CAACTACCTA GACTGGATTG GTGACAACAT 3300

45 GCGACCGTGA CCAGGAACAC CCGACTCCTC AAAAGCAAAT GAGATCCCGC 3350

CTCTTCTTCT TCAGAAGACA CTGCAAAGGC GCAGTGCTTC TCTACAGACT 3400

50 TCTCCAGACC CACCACACCG CAGAAGCGGG ACGAGACCCT ACAGGAGAGG 3450

55 GAAGAGTGCA TTTTCCCAGA TACTTCCCAT TTTGGAAGTT TTCAGGACTT 3500

GGTCTGATTT CAGGATACTC TGTCAGATGG GAAGACATGA ATGCACACTA 3550

60 GCCTCTCCAG GAATGCCTCC TCCCTGGGCA GAAGTGGGGG GAATTCAATC 3600

GATGGCCGCC ATGGCCCAAC TTGTTTATTG CAGCTTATAA TGGTTACAAA 3650

65 TAAAGCAATA GCATCACAAA TTTCACAAAT AAAGCATTTT TTTCAGTCA 3700

TTCTAGTTGT GGT TTGTCCA AACTCATCAA TGTATCTTAT CATGTCTGGA 3750

TCGATCGGGA ATTAATTCGG CGCAGCACCA TGGCCTGAAA TAACCTCTGA 3800

5 AAGAGGAACT TGGTTAGGTA CCTTCTGAGG CGGAAAGAAC CAGCTGTGGA 3850

10 ATGTGTGTCA GTTAGGGTGT GGAAAGTCCC CAGGCTCCCC AGCAGGCAGA 3900

AGTATGCAAA GCATGCATCT CAATTAGTCA GCAACCAGGT GTGGAAAGTC 3950

15 CCCAGGCTCC CCAGCAGGCA GAAGTATGCA AAGCATGCAT CTCAATTAGT 4000

CAGCAACCAT AGTCCCGCCC CTA ACTCCGC CCATCCCGCC CCTAACTCCG 4050

20 CCCAGTTCCG CCCATTCTCC GCCCCATGGC TGACTAATTT TTTTATTTA 4100

25 TGCAGAGGCC GAGGCCGCCT CGGCCTCTGA GCTATTCCAG AAGTAGTGAG 4150

GAGGCTTTTT TGGAGGCTTA GGCTTTTGCA AAAAGCTGTT AACAGCTTGG 4200

30 CACTGGCCGT CGTTTTACAA CGTCGTGACT GGGAAAACCC TGGCGTTACC 4250

CAACTTAATC GCCTTGCAGC ACATCCCCC TTCGCCAGCT GCGTAATAG 4300

35 CGAAGAGGCC CGCACCGATC GCCCTTCCCA ACAGTTGCGT AGCCTGAATG 4350

40 GCGAATGGCG CCTGATGCGG TATTTTCTCC TTACGCATCT GTGCGGTATT 4400

TCACACCGCA TACGTCAAAG CAACCATAGT ACGCGCCCTG TAGCGGCGCA 4450

45 TTAAGCGCGG CGGGTGTGGT GGT TACGCGC AGCGTGACCG CTACACTTGC 4500

CAGCGCCCTA GCGCCCGCTC CTTTCGCTTT CTTCCCTTCC TTTCTCGCCA 4550

50 CGTTGCGCCG CTTTCCCCGT CAAGCTCTAA ATCGGGGGCT CCCTTTAGGG 4600

55 TTCCGATTTA GTGCTTTACG GCACCTCGAC CCCAAAAAC TTGATTTGGG 4650

TGATGGTTCA CGTAGTGGG CATCGCCCTG ATAGACGGTT TTTGCCCCCTT 4700

60 TGACGTTGGA GTCCACGTTT TTTAATAGTG GACTCTTGTT CCAA ACTGGA 4750

ACAACACTCA ACCCTATCTC GGGCTATTCT TTTGATTTAT AAGGGATTTT 4800

65 GCCGATTTG GCCTATTGGT TAAAAAATGA GCTGATTTAA CAAAAATTTA 4850

ACGCGAATTT TAACAAAATA TTAACGTTTA CAATTTTATG GTGCACTCTC 4900

5 AGTACAATCT GCTCTGATGC CGCATAGTTA AGCCAACTCC GCTATCGCTA 4950

CGTGACTGGG TCATGGCTGC GCCCCGACAC CCGCCAACAC CCGCTGACGC 5000

10 GCCCTGACGG GCTTGTCTGC TCCCGGCATC CGCTTACAGA CAAGCTGTGA 5050

CCGTCTCCGG GAGCTGCATG TGTCAGAGGT TTTCACCGTC ATCACCGAAA 5100

15 CGCGCGAGGC AGTATTCTTG AAGACGAAAG GGCTCTGTGA TACGCCTATT 5150

TTTATAGGTT AATGTCATGA TAATAATGGT TTCTTAGACG TCAGGTGGCA 5200

20 CTTTTCGGGG AAATGTGCGC GGAACCCCTA TTTGTTTATT TTTCTAAATA 5250

CATTCAAATA TGTATCCGCT CATGAGACAA TAACCCTGAT AAATGCTTCA 5300

ATAATATTGA AAAAGGAAGA GTATGAGTAT TCAACATTTC CGTGTCGCCC 5350

30 TTATTCCCTT TTTTGC GGCA TTTTGCCTTC CTGTTTTTGC TCACCCAGAA 5400

ACGCTGGTGA AAGTAAAAGA TGCTGAAGAT CAGTTGGGTG CACGAGTGGG 5450

35 TTACATCGAA CTGGATCTCA ACAGCGGTAA GATCCTTGAG AGTTTTCGCC 5500

CCGAAGAACG TTTTCCAATG ATGAGCACTT TTAAAGTTCT GCTATGTGGC 5550

GCGGTATTAT CCCGTGATGA CGCCGGGCAA GAGCAACTCG GTCGCCGCAT 5600

45 ACACTATTCT CAGAATGACT TGTTGAGTA CTCACCAGTC ACAGAAAAGC 5650

ATCTTACGGA TGGCATGACA GTAAGAGAAT TATGCAGTGC TGCCATAACC 5700

50 ATGAGTGATA ACACTGCGGC CAACTTACTT CTGACAACGA TCGGAGGACC 5750

GAAGGAGCTA ACCGCTTTTT TGCACAACAT GGGGGATCAT GTAACGCGC 5800

TTGATCGTTG GGAACCGGAG CTGAATGAAG CCATACCAA CGACGAGCGT 5850

60 GACACCACGA TGCCAGCAGC AATGGCAACA ACGTTGCGCA AACTATTAAC 5900

TGGCGAACTA CTTACTCTAG CTTCCCGGCA ACAATTAATA GACTGGATGG 5950

65 AGGCGGATAA AGTTGCAGGA CCACTTCTGC GCTCGGCCCT TCCGGCTGGC 6000

TGGTTTATTG CTGATAAATC TGGAGCCGGT GAGCGTGGGT CTCGCGGTAT 6050

5 CATTGCAGCA CTGGGGCCAG ATGGTAAGCC CTCCCGTATC GTAGTTATCT 6100

ACACGACGGG GAGTCAGGCA ACTATGGATG AACGAAATAG ACAGATCGCT 6150

10 GAGATAGGTG CCTCACTGAT TAAGCATTGG TAACTGTCAG ACCAAGTTTA 6200

CTCATATATA CTTTAGATTG ATTTAAAACT TCATTTTTAA TTTAAAAGGA 6250

15 TCTAGGTGAA GATCCTTTTT GATAATCTCA TGACCAAAAT CCCTTAACGT 6300

GAGTTTTCGT TCCACTGAGC GTCAGACCCC GTAGAAAAGA TCAAAGGATC 6350

20 TTCTTGAGAT CCTTTTTTTC TGCGCGTAAT CTGCTGCTTG CAAACAAAAA 6400

25 AACCACCGCT ACCAGCGGTG GTTTGTTTGC CGGATCAAGA GCTACCAACT 6450

CTTTTTCCGA AGGTAACTGG CTTCAGCAGA GCGCAGATAC CAAATACTGT 6500

30 CCTTCTAGTG TAGCCGTAGT TAGGCCACCA CTTCAAGAAC TCTGTAGCAC 6550

CGCCTACATA CCTCGCTCTG CTAATCCTGT TACCAGTGGC TGCTGCCAGT 6600

35 GGCGATAAGT CGTGTCTTAC CGGGTTGGAC TCAAGACGAT AGTTACCGGA 6650

40 TAAGGCGCAG CGGTCGGGCT GAACGGGGGG TTCGTGCACA CAGCCCAGCT 6700

TGGAGCGAAC GACCTACACC GAACTGAGAT ACCTACAGCG TGAGCATTGA 6750

45 GAAAGCGCCA CGCTTCCCGA AGGGAGAAAG GCGGACAGGT ATCCGGTAAG 6800

CGGCAGGGTC GGAACAGGAG AGCGCACGAG GGAGCTTCCA GGGGGAAACG 6850

50 CCTGGTATCT TTATAGTCCT GTCGGGTTTC GCCACCTCTG ACTTGAGCGT 6900

55 CGATTTTTGT GATGCTCGTC AGGGGGGCGG AGCCTATGGA AAAACGCCAG 6950

CAACGCGGCC TTTTACGGT TCCTGGCCTT TTGCTGGCCT TTTGCTCACA 7000

60 TGTTCTTTCC TCGTTATCC CCTGATTCTG TGGATAACCG TATTACCGCC 7050

TTTGAGTGAG CTGATACCGC TCGCCGACG CGAACGACCG AGCGCAGCGA 7100

65 GTCAGTGAGC GAGGAAGCGG AAGAGCGCCC AATACGCAA CCGCCTCTCC 7150

CCGCGCGTTG GCCGATTCAT TAATCCAGCT GGCACGACAG GTTTCCTCGAC 7200
5 TGGAAAGCGG GCAGTGAGCG CAACGCAATT AATGTGAGTT ACCTCACTCA 7250
TTAGGCACCC CAGGCTTTAC ACTTTATGCT TCCGGCTCGT ATGTTGTGTG 7300
10 GAATTGTGAG CGGATAACAA TTTCACACAG GAAACAGCTA TGACCATGAT 7350
TACGAATTAA 7360

15

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
20 (A) LENGTH: 6889 bases
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

- 25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

TTCGAGCTCG CCCGACATTG ATTATTGACT AGTTATTAAT AGTAATCAAT 50
30 TACGGGGTCA TTAGTTCATA GCCCATATAT GGAGTTCCGC GTTACATAAC 100
TTACGGTAAA TGGCCCGCCT GGCTGACCGC CCAACGACCC CCGCCCATTG 150
35 ACGTCAATAA TGACGTATGT TCCCATAGTA ACGCCAATAG GGACTTTCCA 200
TTGACGTCAA TGGGTGGAGT ATTTACGGTA AACTGCCCCAC TTGGCAGTAC 250
40 ATCAAGTGTA TCATATGCCA AGTACGCCCC CTATTGACGT CAATGACGGT 300
45 AAATGGCCCG CCTGGCATTG TGCCAGTAC ATGACCTTAT GGGACTTTCC 350
TACTTGGCAG TACATCTACG TATTAGTCAT CGCTATTACC ATGGTGATGC 400
50 GGTTTTGGCA GTACATCAAT GGGCGTGGAT AGCGGTTTGA CTCACGGGGA 450
55 TTTCCAAGTC TCCACCCCAT TGACGTCAAT GGGAGTTTGT TTTGGCACCA 500
AAATCAACGG GACTTTCCAA AATGTCGTAA CAACTCCGCC CCATTGACGC 550
60 AAATGGGCGG TAGGCGTGTA CGGTGGGAGG TCTATATAAG CAGAGCTCGT 600
TTAGTGAACC GTCAGATCGC CTGGAGACGC CATCCACGCT GTTTTGACCT 650
65 CCATAGAAGA CACCGGGACC GATCCAGCCT CCGCGGCCGG GAACGGTGCA 700

TTGGAACGCG GATTCCCCGT GCCAAGAGTG CTGTAAGTAC CGCCTATAGA 750
GCGATAAGAG GATTTTATCC CCGCTGCCAT CATGGTTCGA CCATTGAACT 800
5 GCATCGTCGC CGTGTCCTCAA AATATGGGGA TTGGCAAGAA CGGAGACCTA 850
CCCTGCCCTC CGCTCAGGAA CGCGTTCAAG TACTTCCAAA GAATGACCAC 900
AACCTCTTCA GTGGAAGGTA AACAGAATCT GGTGATTATG GGTAGGAAAA 950
15 CCTGGTTCTC CATTCCTGAG AAGAATCGAC CTTTAAAGGA CAGAATTAAT 1000
ATAGTTCTCA GTAGAGAACT CAAAGAACCA CCACGAGGAG CTCATTTTCT 1050
20 TGCCAAAAGT TTGGATGATG CCTTAAGACT TATTGAACAA CCGGAATTGG 1100
CAAGTAAAGT AGACATGGTT TGGATAGTCG GAGGCAGTTC TGTTTACCAG 1150
GAAGCCATGA ATCAACCAGG CCACCTTAGA CTCTTTGTGA CAAGGATCAT 1200
30 GCAGGAATTT GAAAGTGACA CGTTTTTCCC AGAAATTGAT TTGGGGAAAT 1250
ATAAACCTCT CCCAGAATAC CCAGGCGTCC TCTCTGAGGT CCAGGAGGAA 1300
35 AAAGGCATCA AGTATAAGTT TGAAGTCTAC GAGAAGAAAG ACTAACAGGA 1350
AGATGCTTTC AAGTTCTCTG CTCCCCTCCT AAAGCTATGC ATTTTTATAA 1400
GACCATGGGA CTTTTGCTGG CTTTAGACCC CCTTGGCTTC GTTAGAACGC 1450
45 GGCTACAATT AATACATAAC CTTATGTATC ATACACATAG ATTTAGGTGA 1500
CACTATAGAA TAACATCCAC TTTGCCTTTC TCTCCACAGG TGTCACTCCA 1550
50 GGTCAACTGC ACCTCGGTTC TATCGATTGA ATTCCCCGGC CATAGCTGTC 1600
TGGCATGGGC CTCTCCACCG TGCCTGACCT GCTGCTGCCG CTGGTGCTCC 1650
TGGAGCTGTT GGTGGGAATA TACCCCTCAG GGGTTATTGG ACTGGTCCCT 1700
60 CACCTAGGGG ACAGGGAGAA GAGAGATAGT GTGTGTCCCC AAGGAAAATA 1750
TATCCACCCT CAAAATAATT CGATTGCTG TACCAAGTGC CACAAAGGAA 1800
65 CCTACTTGTA CAATGACTGT CCAGGCCCGG GGCAGGATAC GGAAGTGCAGG 1850

GAGTGTGAGA GCGGCTCCTT CACCGCTTCA GAAAACCACC TCAGACACTG 1900

5 CCTCAGCTGC TCCAAATGCC GAAAGGAAAT GGGTCAGGTG GAGATCTCTT 1950

CTTGACACAGT GGACCGGGAC ACCGTGTGTG GCTGCAGGAA GAACCAGTAC 2000

10 CGGCATTATT GGAGTGAAAA CCTTTTCCAG TGCTTCAATT GCAGCCTCTG 2050

CCTCAATGGG ACCGTGCACC TCTCCTGCCA GGAGAAACAG AACACCGTGT 2100

15 GCACCTGCCA TGCAGGTTTC TTTCTAAGAG AAAACGAGTG TGTCTCCTGT 2150

AGTAACTGTA AGAAAAGCCT GGAGTGCACG AAGTTGTGCC TACCCCAGAT 2200

20 TGAGAATGTT AAGGGCACTG AGGACTCAGG CACCACAGAC AAGAGAGTTG 2250

AGCTCAAAAC CCCACTTGGT GACACAACTC ACACATGCCC ACGGTGCCCA 2300

GAGCCCAAAT CTTGTGACAC ACCTCCCCCG TGCCCACGGT GCCCAGAGCC 2350

30 CAAATCTTGT GACACACCTC CCCCATGCCC ACGGTGCCCA GAGCCCAAAT 2400

CTTGTGACAC ACCTCCCCCA TGCCCACGGT GCCCAGCACC TGAACCTCTG 2450

35 GGAGGACCGT CAGTCTTCCT CTTCCCCCA AAACCAAGG ATACCCTTAT 2500

GATTTCCCGG ACCCCTGAGG TCACGTGCGT GGTGGTGGAC GTGAGCCACG 2550

AAGACCCCGA GGTCCAGTTC AAGTGGTACG TGGACGGCGT GGAGGTGCAT 2600

45 AATGCCAAGA CAAAGCCGCG GGAGGAGCAG TTCAACAGCA CGTTCCGTGT 2650

GGTCAGCGTC CTCACCGTCC TGCACCAGGA CTGGCTGAAC GGCAAGGAGT 2700

50 ACAAGTGCAA GGTCTCCAAC AAAGCCCTCC CAGCCCCCAT CGAGAAAACC 2750

ATCTCCAAAA CCAAGGACA GCCCCGAGAA CCACAGGTGT ACACCCTGCC 2800

CCCATCCCGG GAGGAGATGA CCAAGAACCA GGTCAGCCTG ACCTGCCTGG 2850

60 TCAAAGGCTT CTACCCACAGC GACATCGCCG TGGAGTGGGA GAGCAGCGGG 2900

CAGCCGGAGA ACAACTACAA CACCACGCCT CCCATGCTGG ACTCCGACGG 2950

65 CTCCTTCTTC CTCTACAGCA AGCTCACCGT GGACAAGAGC AGGTGGCAGC 3000

AGGGGAACAT CTTCTCATGC TCCGTGATGC ATGAGGCTCT GCACAACCGC 3050

5 TTCACGCAGA AGAGCCTCTC CCTGTCTCCG GGTAAATGAG TGCACGGCC 3100

GGGGATCCTC TAGAGTCGAC CTGCAGAAGC TTGGCCGCCA TGGCCCAACT 3150

10 TGTATTATGC AGCTTATAAT GGTTACAAAT AAAGCAATAG CATCACAAAT 3200

TTCACAAATA AAGCATTTTT TCACTGTCAT TCTAGTTGTG GTTTGTCCAA 3250

15 ACTCATCAAT GTATCTTATC ATGTCTGGAT CGATCGGGAA TTAATTCGGC 3300

GCAGCACCAT GGCCTGAAAT AACCTCTGAA AGAGGAACTT GGTTAGGTAC 3350

20 CTTCTGAGGC GGAAAGAACC AGCTGTGGAA TGTGTGTCAG TTAGGGTGTG 3400

25 GAAAGTCCCC AGGCTCCCCA GCAGGCAGAA GTATGCAAAG CATGCATCTC 3450

AATTAGTCAG CAACCAGGTG TGGAAAGTCC CCAGGCTCCC CAGCAGGCAG 3500

30 AAGTATGCAA AGCATGCATC TCAATTAGTC AGCAACCATA GTCCCGCCCC 3550

TAATCCGCC CATCCCGCCC CTAATCCGC CCAGTTCCGC CCATTCTCCG 3600

35 CCCCATGGCT GACTAATTTT TTTTATTTAT GCAGAGGCCG AGGCCGCCTC 3650

40 GGCCTCTGAG CTATTCCAGA AGTAGTGAGG AGGCTTTTTT GGAGGCCTAG 3700

GCTTTTGCAA AAAGCTGTTA ACAGCTTGGC ACTGGCCGTC GTTTTACAAC 3750

45 GTCGTGACTG GGAAAACCTT GCGGTTACCC AACTTAATCG CCTTGCAGCA 3800

CATCCCCCCT TCGCCAGCTG GCGTAATAGC GAAGAGGCCG GCACCGATCG 3850

50 CCCTTCCCAA CAGTTGCGTA GCCTGAATGG CGAATGGCGC CTGATGCGGT 3900

55 ATTTTCTCCT TACGCATCTG TGCGGTATTT CACACCGCAT ACGTCAAAGC 3950

AACCATAGTA CGCGCCCTGT AGCGGCGCAT TAAGCGCGGC GGGTGTGGTG 4000

60 GTTACGCGCA GCGTGACCGC TACACTTGCC AGCGCCCTAG CGCCCGCTCC 4050

TTTCGCTTTC TTCCCTTCCT TTCTCGCCAC GTTCGCGGC TTTCCCGTC 4100

65 AAGCTCTAAA TCGGGGGCTC CCTTTAGGGT TCCGATTTAG TGCTTTACGG 4150

CACCTCGACC CCAAAAACT TGATTTGGGT GATGGTTCAC GTAGTGGGCC 4200

5 ATCGCCCTGA TAGACGGTTT TTCGCCCTTT GACGTTGGAG TCCACGTTCT 4250

TTAATAGTGG ACTCTGTTC CAACTGGAA CAACACTCAA CCCTATCTCG 4300

10 GGCTATTCTT TTGATTTATA AGGGATTTTG CCGATTTCGG CCTATTGGTT 4350

AAAAATGAG CTGATTTAAC AAAAATTTAA CGCGAATTTT AACAAAATAT 4400

15 TAACGTTTAC AATTTTATGG TGCACCTCTCA GTACAATCTG CTCTGATGCC 4450

GCATAGTTAA GCCAACTCCG CTATCGCTAC GTGACTGGGT CATGGCTGCG 4500

20 CCCCACACC CGCCAACACC CGCTGACGCG CCCTGACGGG CTTGTCTGCT 4550

25 CCCGGCATCC GCTTACAGAC AAGCTGTGAC CGTCTCCGGG AGCTGCATGT 4600

GTCAGAGGTT TTCACCGTCA TCACCGAAAC GCGCGAGGCA GTATTCTTGA 4650

30 AGACGAAAGG GCCTCGTGAT ACGCCTATTT TTATAGGTTA ATGTCATGAT 4700

AATAATGGTT TCTTAGACGT CAGGTGGCAC TTTTCGGGGA AATGTGCGCG 4750

35 GAACCCCTAT TTGTTTATTT TTCTAAATAC ATTCAAATAT GTATCCGCTC 4800

40 ATGAGACAAT AACCTGATA AATGCTTCAA TAATATTGAA AAAGGAAGAG 4850

TATGAGTATT CAACATTTCC GTGTCGCCCT TATTCCTTTT TTTGCGGCAT 4900

45 TTTGCCTTCC TGTTTTTGCT CACCCAGAAA CGCTGGTGAA AGTAAAAGAT 4950

GCTGAAGATC AGTTGGGTGC ACGAGTGGGT TACATCGAAC TGGATCTCAA 5000

50 CAGCGGTAAG ATCCTTGAGA GTTTTCGCCC CGAAGAACGT TTTCCAATGA 5050

55 TGAGCACTTT TAAAGTTCTG CTATGTGGCG CGGTATTATC CCGTGATGAC 5100

GCCGGGCAAG AGCAACTCGG TCGCCGCATA CACTATTCTC AGAATGACTT 5150

60 GGTGAGTAC TCACCAGTCA CAGAAAAGCA TCTTACGGAT GGCATGACAG 5200

TAAGAGAATT ATGCAGTGCT GCCATAACCA TGAGTGATAA CACTGCGGCC 5250

65 AACTTACTTC TGACAACGAT CGGAGGACCG AAGGAGCTAA CCGCTTTTTT 5300

GCACAACATG GGGGATCATG TAACTCGCCT TGATCGTTGG GAACCGGAGC 5350

5 TGAATGAAGC CATACCAAAC GACGAGCGTG ACACCACGAT GCCAGCAGCA 5400

ATGGCAACAA CGTTGCGCAA ACTATTA ACT GGCGAACTAC TTA CTCTAGC 5450

10 TTCCCGGCAA CAATTAATAG ACTGGATGGA GGCGGATAAA GTTGCAGGAC 5500

CACTTCTGCG CTCGGCCCTT CCGGCTGGCT GGTTTATTGC TGATAAATCT 5550

15 GGAGCCGGTG AGCGTGGGTC TCGCGGTATC ATTGCAGCAC TGGGGCCAGA 5600

TGGTAAGCCC TCCCGTATCG TAGTTATCTA CACGACGGGG AGTCAGGCAA 5650

20 CTATGGATGA ACGAAATAGA CAGATCGCTG AGATAGGTGC CTCACTGATT 5700

25 AAGCATTGGT AACTGTCAGA CCAAGTTTAC TCATATATAC TTTAGATTGA 5750

TTTAAACTT CATTTTAAAT TTAAAAGGAT CTAGGTGAAG ATCCTTTTGT 5800

30 ATAATCTCAT GACCAAATC CCTTAACGTG AGTTTTCGTT CCACTGAGCG 5850

TCAGACCCCG TAGAAAAGAT CAAAGGATCT TCTTGAGATC CTTTTTTTCT 5900

35 GCGCGTAATC TGCTGCTTGC AAACAAAAAA ACCACCGCTA CCAGCGGTGG 5950

40 TTTGTTTGCC GGATCAAGAG CTACCAACTC TTTTCCGAA GGTAAGTGGC 6000

TTCAGCAGAG CGCAGATACC AAATACTGTC CTTCTAGTGT AGCCGTAGTT 6050

45 AGGCCACCAC TTCAAGAACT CTGTAGCACC GCCTACATAC CTCGCTCTGC 6100

TAATCCTGTT ACCAGTGGCT GCTGCCAGTG GCGATAAGTC GTGTCTTACC 6150

50 GGGTTGGACT CAAGACGATA GTTACCGGAT AAGGCGCAGC GGTCGGGCTG 6200

55 AACGGGGGGT TCGTGACAC AGCCCAGCTT GGAGCGAACG ACCTACACCG 6250

AACTGAGATA CCTACAGCGT GAGCATTGAG AAAGCGCCAC GCTTCCCGAA 6300

60 GGGAGAAAGG CGGACAGGTA TCCGGTAAGC GGCAGGGTCG GAACAGGAGA 6350

GCGCACGAGG GAGCTTCCAG GGGGAAACGC CTGGTATCTT TATAGTCCTG 6400

65 TCGGGTTTCG CCACCTCTGA CTTGAGCGTC GATTTTGTG ATGCTCGTCA 6450

GGGGGGCGGA GCCTATGGAA AAACGCCAGC AACGCGGCCT TTTTACGGTT 6500
5 CCTGGCCTTT TGCTGGCCTT TTGCTCACAT GTTCTTTCCT GCGTTATCCC 6550
CTGATTCTGT GGATAACCGT ATTACCGCCT TTGAGTGAGC TGATACCGCT 6600
10 CGCCGCAGCC GAACGACCGA GCGCAGCGAG TCAGTGAGCG AGGAAGCGGA 6650
AGAGCGCCCA ATACGCAAAC CGCCTCTCCC CGCGCGTTGG CCGATTCAAT 6700
15 AATCCAGCTG GCACGACAGG TTTCCCGACT GGAAAGCGGG CAGTGAGCGC 6750
AACGCAATTA ATGTGAGTTA CCTCACTCAT TAGGCACCCC AGGCTTTACA 6800
20 CTTTATGCTT CCGGCTCGTA TGTGTGTGG AATTGTGAGC GGATAACAAT 6850
25 TTCACACAGG AAACAGCTAT GACCATGATT ACGAATTAA 6889

(2) INFORMATION FOR SEQ ID NO:3:

30

(i) SEQUENCE CHARACTERISTICS:

35

- (A) LENGTH: 6557 bases
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

40 TTCGAGCTCG CCCGACATTG ATTATTGACT AGAGTCGATC GACAGCTGTG 50
GAATGTGTGT CAGTTAGGGT GTGGAAAGTC CCCAGGCTCC CCAGCAGGCA 100
45 GAAGTATGCA AAGCATGCAT CTCAATTAGT CAGCAACCAG GTGTGGAAAG 150
TCCCCAGGCT CCCCAGCAGG CAGAAGTATG CAAAGCATGC ATCTCAATTA 200
50 GTCAGCAACC ATAGTCCCGC CCCTAACTCC GCCCATCCCG CCCCTAACTC 250
55 CGCCAGTTC CGCCCATCTT CCGCCCCATG GCTGACTAAT TTTTTTTATT 300
TATGCAGAGG CCGAGGCCGC CTCGGCCTCT GAGCTATTCC AGAAGTAGTG 350
60 AGGAGGCTTT TTTGGAGGCC TAGGCTTTTG CAAAAGCTA GCTTATCCGG 400
CCGGGAACGG TGCATTGGAA CGCGGATTCC CCGTGCCAAG AGTGACGTAA 450
65 GTACCGCCTA TAGAGCGATA AGAGGATTTT ATCCCCGCTG CCATCATGGT 500

TCGACCATTG AACTGCATCG TCGCCGTGTC CCAAATATG GGGATTGGCA 550
5 AGAACGGAGA CCTACCCTGG CCTCCGCTCA GGAACGAGTT CAAGTACTTC 600
CAAAGAATGA CCACAACCTC TTCAGTGGAA GGTAACAGA ATCTGGTGAT 650
10 TATGGGTAGG AAAACCTGGT TCTCCATTCC TGAGAAGAAT CGACCTTTAA 700
AGGACAGAAT TAATATAGTT CTCAGTAGAG AACTCAAAGA ACCACCACGA 750
15 GGAGCTCATT TTCTTGCCAA AAGTTTGGAT GATGCCTTAA GACTTATTGA 800
ACAACCGGAA TTGGCAAGTA AAGTAGACAT GGTTTGGATA GTCGGAGGCA 850
20 GTTCTGTTTA CCAGGAAGCC ATGAATCAAC CAGGCCACCT TAGACTCTTT 900
GTGACAAGGA TCATGCAGGA ATTTGAAAGT GACACGTTTT TCCCAGAAAT 950
TGATTTGGGG AAATATAAAC CTCTCCCAGA ATACCCAGGC GTCCTCTCTG 1000
30 AGGTCCAGGA GGAAAAGGC ATCAAGTATA AGTTTGAAGT CTACGAGAAG 1050
AAAGACTAAC AGGAAGATGC TTTCAAGTTC TCTGCTCCCC TCCTAAAGCT 1100
35 ATGCATTTTT ATAAGACCAT GGGACTTTTG CTGGCTTTAG ATCCCCTTGG 1150
CTTCGTTAGA ACGCAGCTAC AATTAATACA TAACCTTATG TATCATACAC 1200
ATACGATTTA GGTGACACTA TAGATAACAT CCACTTTGCC TTTCTCTCCA 1250
45 CAGGTGTCCA CTCCCAGGTC CAACTGCACC TCGGTTCTAT CGATTGAATT 1300
CCACCATGGG ATGGTCATGT ATCATCCTTT TTCTAGTAGC AACTGCAACT 1350
50 GGAGTACATT CAGAAGTTCA GCTGGTGGAG TCTGGCGGTG GCCTGGTGCA 1400
GCCAGGGGGC TCACTCCGTT TGTCTGTGC AGTTTCTGGC TACTCCATCA 1450
CCTCCGGATA TAGCTGGAAC TGGATCCGTC AGGCCCCGGG TAAGGGCCTG 1500
60 GAATGGGTTG CATCGATTAC GTATGCCGGA TCGACTAACT ATAACCCTAG 1550
CGTCAAGGGC CGTATCACTA TAAGTCGCGA CGATTCCAAA AACACATTCT 1600
65 ACCTGCAGAT GAACAGCCTG CGTGCTGAGG AACTGCCGT CTATTATTGT 1650

GCTCGAGGCA GCCACTATTT CGGCGCCTGG CACTTCGCCG TGTGGGGTCA 1700

5 AGGAACCCTG GTCACCGTCT CCTCGGCCTC CACCAAGGGC CCATCGGTCT 1750

TCCCCCTGGC ACCCTCCTCC AAGAGCACCT CTGGGGGCAC AGCGGCCCTG 1800

10 GGCTGCCTGG TCAAGGACTA CTTCCCCGAA CCGGTGACGG TGTCGTGGAA 1850

CTCAGGCGCC CTGACCAGCG GCGTGACAC CTTCCCGGCT GTCCTACAGT 1900

15 CCTCAGGACT CTACTCCCTC AGCAGCGTGG TGACTGTGCC CTCTAGCAGC 1950

TTGGGCACCC AGACCTACAT CTGCAACGTG AATCACAAGC CCAGCAACAC 2000

20 CAAGGTGGAC AAGAAAAGTTG AGCCCAAATC TTGTGACAAA ACTCACACAT 2050

25 GCCCACCGTG CCCAGCACCT GAACTCCTGG GGGGACCGTC AGTCTTCCTC 2100

TTCCCCCAA AACC CAAGGA CACCCTCATG ATCTCCCGGA CCCCTGAGGT 2150

30 CACATGCGTG GTGGTGGACG TGAGCCACGA AGACCCTGAG GTCAAGTTCA 2200

ACTGGTACGT GGACGGCGTG GAGGTGCATA ATGCCAAGAC AAAGCCGCGG 2250

35 GAGGAGCAGT ACAACAGCAC GTACCGTGTG GTCAGCGTCC TCACCGTCCT 2300

GCACCAGGAC TGGCTGAATG GCAAGGAGTA CAAGTGCAAG GTCTCCAACA 2350

40 AAGCCCTCCC AGCCCCCATC GAGAAAACCA TCTCCAAAGC CAAAGGGCAG 2400

45 CCCC GAGAAC CACAGGTGTA CACCCTGCCC CCATCCCGGG AAGAGATGAC 2450

CAAGAACCAG GTCAGCCTGA CCTGCCTGGT CAAAGGCTTC TATCCCAGCG 2500

50 ACATCGCCGT GGAGTGGGAG AGCAATGGGC AGCCGGAGAA CAACTACAAG 2550

55 ACCACGCCTC CCGTGCTGGA CTCCGACGGC TCCTTCTTCC TCTACAGCAA 2600

GCTCACCCTG GACAAGAGCA GGTGGCAGCA GGGGAACGTC TTCTCATGCT 2650

60 CCGTGATGCA TGAGGCTCTG CACAACCACT ACACGCAGAA GAGCCTCTCC 2700

CTGTCTCCGG GTAAATGAGT GCGACGGCCC TAGAGTCGAC CTGCAGAAGC 2750

65 TTGGCCGCCA TGGCCCAACT TGTTTATTGC AGCTTATAAT GGTACAAAT 2800

AAAGCAATAG CATCACAAAT TTCACAAATA AAGCATTTTT TTCACTGCAT 2850
TCTAGTTGTG GTTTGTCCAA ACTCATCAAT GTATCTTATC ATGTCTGGAT 2900
5 CGATCGGGAA TTAATTCGGC GCAGCACCAT GGCCTGAAAT AACCTCTGAA 2950
AGAGGAACTT GGTTAGGTAC CTTCTGAGGC GGAAAGAACC AGCTGTGGAA 3000
TGTGTGTCAG TTAGGGTGTG GAAAGTCCCC AGGCTCCCCA GCAGGCAGAA 3050
15 GTATGCAAAG CATGCATCTC AATTAGTCAG CAACCAGGTG TGGAAAGTCC 3100
CCAGGCTCCC CAGCAGGCAG AAGTATGCAA AGCATGCATC TCAATTAGTC 3150
20 AGCAACCATA GTCCCGCCCC TAACTCCGCC CATCCCGCCC CTAATCCGC 3200
CCAGTTCCGC CCATTCTCCG CCCCATGGCT GACTAATTTT TTTTATTTAT 3250
GCAGAGGCCG AGGCCGCCTC GGCCTCTGAG CTATTCCAGA AGTAGTGAGG 3300
30 AGGCTTTTTT GGAGGCCTAG GCTTTTGCAA AAAGCTGTTA CCTCGAGCGG 3350
CCGCTTAATT AAGGCGCGCC ATTTAAATCC TGCAGGTAAC AGCTTGGCAC 3400
35 TGGCCGTCGT TTTACAACGT CGTGA CTGGG AAAACCCTGG CGTTACCCAA 3450
CTTAATCGCC TTGCAGCACA TCCCCCTTC GCCAGCTGGC GTAATAGCGA 3500
AGAGGCCCCG ACCGATCGCC CTTCCCAACA GTTGCGTAGC CTGAATGGCG 3550
45 AATGGCGCCT GATGCGGTAT TTTCTCCTTA CGCATCTGTG CGGTATTTCA 3600
CACCGCATAC GTCAAAGCAA CCATAGTACG CGCCCTGTAG CGGCGCATTA 3650
50 AGCGCGGCGG GTGTGGTGGT TACGCGCAGC GTGACCGCTA CACTTGCCAG 3700
CGCCCTAGCG CCCGCTCCTT TCGCTTTCTT CCCTTCCTTT CTCGCCACGT 3750
TCGCCGGCTT TCCCGTCAA GCTCTAAATC GGGGGCTCCC TTTAGGGTTC 3800
60 CGATTTAGTG CTTTACGGCA CCTCGACCCC AAAAACTTG ATTTGGGTGA 3850
TGGTTCACGT AGTGGGCCAT CGCCCTGATA GACGGTTTTT CGCCCTTTGA 3900
65 CGTTGGAGTC CACGTTCTTT AATAGTGGAC TCTTGTTCCA AACTGGAACA 3950

ACACTCAACC CTATCTCGGG CTATTCTTTT GATTTATAAG GGATTTTGCC 4000
GATTTTCGGCC TATTGGTTAA AAAATGAGCT GATTTAACAA AAATTTAACG 4050
5 CGAATTTTAA CAAAATATTA ACGTTTACAA TTTTATGGTG CACTCTCAGT 4100
10 ACAATCTGCT CTGATGCCGC ATAGTTAAGC CAACTCCGCT ATCGCTACGT 4150
GACTGGGTCA TGGCTGCGCC CCGACACCCG CCAACACCCG CTGACGCGCC 4200
15 CTGACGGGCT TGTCTGCTCC CGGCATCCGC TTACAGACAA GCTGTGACCG 4250
TCTCCGGGAG CTGCATGTGT CAGAGGTTTT CACCGTCATC ACCGAAACGC 4300
20 GCGAGGCAGT ATTCTTGAAG ACGAAAGGGC CTCGTGATAC GCCTATTTTT 4350
25 ATAGGTAAAT GTCATGATAA TAATGGTTTC TTAGACGTCA GGTGGCACTT 4400
TTCGGGGAAA TGTGCGCGGA ACCCCTATTT GTTTATTTTT CTAAATACAT 4450
30 TCAAATATGT ATCCGCTCAT GAGACAATAA CCCTGATAAA TGCTTCAATA 4500
ATATTGAAAA AGGAAGAGTA TGAGTATTCA ACATTTCGGT GTCGCCCTTA 4550
35 TTCCCTTTTT TCGGCGATTT TGCCTTCCTG TTTTGCTCA CCCAGAAACG 4600
40 CTGGTGAAAG TAAAAGATGC TGAAGATCAG TTGGGTGCAC GAGTGGGTTA 4650
CATCGAACTG GATCTCAACA GCGGTAAGAT CCTTGAGAGT TTTCGCCCCG 4700
45 AAGAACGTTT TCCAATGATG AGCACTTTTA AAGTTCTGCT ATGTGGCGCG 4750
GTATTATCCC GTGATGACGC CGGGCAAGAG CAACTCGGTC GCCGCATACA 4800
50 CTATTCTCAG AATGACTTGG TTGAGTACTC ACCAGTCACA GAAAAGCATC 4850
55 TTACGGATGG CATGACAGTA AGAGAATTAT GCAGTGCTGC CATAACCATG 4900
AGTGATAACA CTGCGGCCAA CTTACTTCTG ACAACGATCG GAGGACCGAA 4950
60 GGAGCTAACC GCTTTTTTGC ACAACATGGG GGATCATGTA ACTCGCCTTG 5000
ATCGTTGGGA ACCGGAGCTG AATGAAGCCA TACCAAACGA CGAGCGTGAC 5050
65 ACCACGATGC CAGCAGCAAT GGCAACAACG TTGCGCAAAC TATTAACCTGG 5100

CGAACTACTT ACTCTAGCTT CCCGGCAACA ATTAATAGAC TGGATGGAGG 5150

5 CGGATAAAGT TGCAGGACCA CTTCTGCGCT CGGCCCTTCC GGCTGGCTGG 5200

TTTATTGCTG ATAAATCTGG AGCCGGTGAG CGTGGGTCTC GCGGTATCAT 5250

10 TGCAGCACTG GGGCCAGATG GTAAGCCCTC CCGTATCGTA GTTATCTACA 5300

CGACGGGGAG TCAGGCAACT ATGGATGAAC GAAATAGACA GATCGCTGAG 5350

15 ATAGGTGCCT CACTGATTAA GCATTGGTAA CTGTCAGACC AAGTTTACTC 5400

ATATATACTT TAGATTGATT TAAAACTTCA TTTTAAATTT AAAAGGATCT 5450

20 AGGTGAAGAT CCTTTTTGAT AATCTCATGA CCAAATCCC TTAACGTGAG 5500

25 TTTTCGTTCC ACTGAGCGTC AGACCCCGTA GAAAAGATCA AAGGATCTTC 5550

TTGAGATCCT TTTTTTCTGC GCGTAATCTG CTGCTTGCAA ACAAAAAAAC 5600

30 CACCGCTACC AGCGGTGGTT TGTTTGCCGG ATCAAGAGCT ACCAACTCTT 5650

TTTCCGAAGG TAACTGGCTT CAGCAGAGCG CAGATACCAA ATACTGTCCT 5700

35 TCTAGTGTAG CCGTAGTTAG GCCACCACTT CAAGAACTCT GTAGCACCGC 5750

40 CTACATACCT CGCTCTGCTA ATCCTGTTAC CAGTGGCTGC TGCCAGTGGC 5800

GATAAGTCGT GTCTTACCGG GTTGGACTCA AGACGATAGT TACCGGATAA 5850

45 GGCGCAGCGG TCGGGCTGAA CGGGGGGTTT GTGCACACAG CCCAGCTTGG 5900

AGCGAACGAC CTACACCGAA CTGAGATACC TACAGCGTGA GCATTGAGAA 5950

50 AGCGCCACGC TTCCCGAAGG GAGAAAGGCG GACAGGTATC CGGTAAGCGG 6000

55 CAGGGTCGGA ACAGGAGAGC GCACGAGGGA GCTTCCAGGG GGAAACGCCT 6050

GGTATCTTTA TAGTCCTGTC GGGTTTCGCC ACCTCTGACT TGAGCGTCGA 6100

60 TTTTGTGAT GCTCGTCAGG GGGGCGGAGC CTATGGAAAA ACGCCAGCAA 6150

CGCGGCCTTT TTACGGTTCC TGGCCTTTTG CTGGCCTTTT GCTCACATGT 6200

65 TCTTTCCTGC GTTATCCCCT GATTCTGTGG ATAACCGTAT TACCGCCTTT 6250

GAGTGAGCTG ATACCGCTCG CCGCAGCCGA ACGACCGAGC GCAGCGAGTC 6300
AGTGAGCGAG GAAGCGGAAG AGCGCCCAAT ACGCAAACCG CCTCTCCCCG 6350
5 CGCGTTGGCC GATTCATTAA TCCAGCTGGC ACGACAGGTT TCCCGACTGG 6400
AAAGCGGGCA GTGAGCGCAA CGCAATTAAT GTGAGTTACC TCACTCATTA 6450
10 GGCACCCAG GCTTTACACT TTATGCTTCC GGCTCGTATG TTGTGTGGAA 6500
15 TTGTGAGCGG ATAACAATTT CACACAGGAA ACAGCTATGA CCATGATTAC 6550
GAATTAA 6557
20

(2) INFORMATION FOR SEQ ID NO:4:

25 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7305 bases
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

TTCGAGCTCG CCCGACATTG ATTATTGACT AGTTATTAAT AGTAATCAAT 50
35 TACGGGGTCA TTAGTTCATA GCCCATATAT GGAGTTCCGC GTTACATAAC 100
TTACGGTAAA TGGCCCGCCT GGCTGACCGC CCAACGACCC CCGCCCATTG 150
40 ACGTCAATAA TGACGTATGT TCCCATAGTA ACGCCAATAG GGACTTTCCA 200
45 TTGACGTCAA TGGGTGGAGT ATTTACGGTA AACTGCCCAC TTGGCAGTAC 250
ATCAAGTGTA TCATATGCCA AGTACGCCCC CTATTGACGT CAATGACGGT 300
50 AAATGGCCCG CCTGGCATTG TGCCAGTAC ATGACCTTAT GGGACTTTCC 350
TACTTGGCAG TACATCTACG TATTAGTCAT CGCTATTACC ATGGTGATGC 400
GGTTTTGGCA GTACATCAAT GGGCGTGGAT AGCGGTTTGA CTCACGGGGA 450
60 TTTCCAAGTC TCCACCCCAT TGACGTCAAT GGGAGTTTGT TTTGGCACCA 500
AAATCAACGG GACTTTCCAA AATGTCGTAA CAACTCCGCC CCATTGACGC 550
65 AAATGGGCGG TAGGCGTGTA CGGTGGGAGG TCTATATAAG CAGAGCTCGT 600

TTAGTGAACC GTCAGATCGC CTGGAGACGC CATCCACGCT GTTTTGACCT 650

CCATAGAAGA CACCGGGACC GATCCAGCCT CCGCGGCCGG GAACGGTGCA 700

5 TTGGAACGCG GATTCCCCGT GCCAAGAGTG ACGTAAGTAC CGCCTATAGA 750

GTCTATAGGC CCACCCCTT GGCTTCGTTA GAACGCGGCT ACAATTAATA 800

CATAACCTTA TGTATCATAC ACATACGATT TAGGTGACAC TATAGAATAA 850

15 CATCCACTTT GCCTTTCTCT CCACAGGTGT CCACTCCCAG GTCCAACTGC 900

ACCTCGGTTC TAAGCTTATC GATATGAAAA AGCCTGAACT CACCGCGACG 950

20 TCTGTCGAGA AGTTTCTGAT CGAAAAGTTC GACAGCGTCT CCGACCTGAT 1000

GCAGCTCTCG GAGGGCGAAG AATCTCGTGC TTTCAGCTTC GATGTAGGAG 1050

GGCGTGATA TGTCTGCGG GTAAATAGCT GCGCCGATGG TTTCTACAAA 1100

30 GATCGTTATG TTTATCGGCA CTTTGCATCG GCCGCGCTCC CGATTCCGGA 1150

AGTGCTTGAC ATTGGGGAAT TCAGCGAGAG CCTGACCTAT TGCATCTCCC 1200

35 GCCGTGCACA GGGTGTACG TTGCAACACC TGCCTGAAAC CGAACTGCCC 1250

GCTGTTCTGC AGCCGGTCGC GGAGGCCATG GATGCGATCG CTGCGGCCGA 1300

TCTTAGCCAG ACGAGCGGGT TCGGCCCATT CGGACCGCAA GGAATCGGTC 1350

45 AATACACTAC ATGGCGTGAT TTCATATGCG CGATTGCTGA TCCCCATGTG 1400

TATCACTGGC AACTGTGAT GGACGACACC GTCAGTGCGT CCGTCGCGCA 1450

50 GGCTCTCGAT GAGCTGATGC TTTGGGCCGA GGAAGTCCGC 1500

ACCTCGTGCA CGCGGATTTC GGCTCCAACA ATGTCCTGAC GGACAATGGC 1550

CGCATAACAG CGGTCATTGA CTGGAGCGAG GCGATGTTTCG GGGATTCCCA 1600

60 ATACGAGGTC GCCAACATCT TCTTCTGGAG GCCGTGGTTG GCTTGATATG 1650

AGCAGCAGAC GACTTCGAG CGGAGGCATC CGGAGCTTGC AGGATCGCCG 1700

65 CGGCTCCGGG CGTATATGCT CCGCATTGGT CTTGACCAAC TCTATCAGAG 1750

CTTGTTGAC GGCAATTTTCG ATGATGCAGC TTGGGCGCAG GGTTCGATGCG 1800

ACGCAATCGT CCGATCCGGA GCCGGGACTG TCGGGCGTAC ACAAATCGCC 1850

5 CGCAGAAGCG CGGCCGTCTG GACCGATGGC TGTGTAGAAG TACTCGCCGA 1900

10 TAGTGAAAC CGACGCCCCA GCACTCGTCC GAGGGCAAAG GAATAGAGTA 1950

GATGCCGACC GAAGGATCCC CGGGGAATTC AATCGATGGC CGCCATGGCC 2000

15 CAACTTGTTT ATTGCAGCTT ATAATGGTTA CAAATAAAGC AATAGCATCA 2050

CAAATTTTAC AAATAAAGCA TTTTTTTCAC TGCATTCTAG TTGTGGTTTG 2100

20 TCCAACTCA TCAATGTATC TTATCATGTC TGGATCGATC GGGAATTAAT 2150

25 TCGGCGCAGC ACCATGGCCT GAAATAACCT CTGAAAGAGG AACTTGGTTA 2200

GGTACCTTCT GAGGCGGAAA GAACCAGCTG TGAATGTGT GTCAGTTAGG 2250

30 GTGTGGAAAG TCCCCAGGCT CCCCAGCAGG CAGAAGTATG CAAAGCATGC 2300

ATCTCAATTA GTCAGCAACC AGGTGTGGAA AGTCCCCAGG CTCCCCAGCA 2350

35 GGCAGAAGTA TGCAAAGCAT GCATCTCAAT TAGTCAGCAA CCATAGTCCC 2400

40 GCCCCTAACT CCGCCCATCC CGCCCCTAAC TCCGCCCAGT TCCGCCCAT 2450

CTCCGCCCCA TGGCTGACTA ATTTTTTTTA TTTATGCAGA GGCCGAGGCC 2500

45 GCCTCGGCCT CTGAGCTATT CCAGAAGTAG TGAGGAGGCT TTTTGGAGG 2550

CCTAGGCTTT TGCAAAAAGC TAGCTTATCC GGCCGGGAAC GGTGCATTGG 2600

50 AACGCGGATT CCCCCTGCCA AGAGTCAGGT AAGTACCGCC TATAGAGTCT 2650

55 ATAGGCCAC CCCCTTGGCT TCGTTAGAAC GCGGCTACAA TTAATACATA 2700

ACCTTTTGGA TCGATCCTAC TGACACTGAC ATCCACTTTT TCTTTTCTC 2750

60 CACAGGTGTC CACTCCCAGG TCCAACGCA CCTCGGTTTC CGAAGCTAGC 2800

TTGGGCTGCA TCGATTGAAT TCCACCATGG GATGGTCATG TATCATCCTT 2850

65 TTTCTAGTAG CAACTGCAAC TGGAGTACAT TCAGATATCC AGCTGACCCA 2900

GTCCCCGAGC TCCCTGTCCG CCTCTGTGGG CGATAGGGTC ACCATCACCT 2950

5 GCCGTGCCAG TCAGAGCGTC GATTACGATG GTGATAGCTA CATGAACTGG 3000

TATCAACAGA AACCAGGAAA AGCTCCGAAA CTACTGATTT ACGCGGCCTC 3050

10 GTACCTGGAG TCTGGAGTCC CTTCTCGCTT CTCTGGATCC GGTTCCTGGGA 3100

CGGATTTTAC TCTGACCATC AGCAGTCTGC AGCCGGAAGA CTTGCAACT 3150

15 TATTACTGTC AGCAAAGTCA CGAGGATCCG TACACATTTG GACAGGGTAC 3200

CAAGGTGGAG ATCAAACGAA CTGTGGCTGC ACCATCTGTC TTCATCTTCC 3250

20 CGCCATCTGA TGAGCAGTTG AAATCTGGAA CTGCCTCTGT TGTGTGCCTG 3300

CTGAATAACT TCTATCCCAG AGAGGCCAAA GTACAGTGGA AGGTGGATAA 3350

CGCCCTCCAA TCGGGTAACT CCCAGGAGAG TGTCACAGAG CAGGACAGCA 3400

30 AGGACAGCAC CTACAGCCTC AGCAGCACCC TGACGCTGAG CAAAGCAGAC 3450

TACGAGAAAC ACAAAGTCTA CGCCTGCGAA GTCACCCATC AGGGCCTGAG 3500

35 CTCGCCCCGTC ACAAAGAGCT TCAACAGGGG AGAGTGTTAA GCTTCGATGG 3550

CCGCCATGGC CCAACTTGTT TATTGCAGCT TATAATGGTT ACAAATAAAG 3600

CAATAGCATC ACAAATTTCA CAAATAAAGC ATTTTTTTTCA CTGCATTCTA 3650

45 GTTGTGGTTT GTCCAAACTC ATCAATGTAT CTTATCATGT CTGGATCGAT 3700

CGGGAATTAA TTCGGCGCAG CACCATGGCC TGAAATAACC TCTGAAAGAG 3750

50 GAACTTGTTT AGGTACCTTC TGAGGCGGAA AGAACCAGCT GTGGAATGTG 3800

TGTCAGTTAG GGTGTGGAAA GTCCCCAGGC TCCCCAGCAG GCAGAAGTAT 3850

GCAAAGCATG CATCTCAATT AGTCAGCAAC CAGGTGTGGA AAGTCCCCAG 3900

60 GCTCCCCAGC AGGCAGAAAGT ATGCAAAGCA TGCATCTCAA TTAGTCAGCA 3950

ACCATAGTCC CGCCCCTAAC TCCGCCCATC CCGCCCCTAA CTCCGCCAG 4000

65 TTCCGCCCAT TCTCCGCCCC ATGGCTGACT AATTTTTTTT ATTTATGCAG 4050

AGGCCGAGGC CGCCTCGGCC TCTGAGCTAT TCCAGAAGTA GTGAGGAGGC 4100

5 TTTTTTGGAG GCCTAGGCTT TTGCAAAAAG CTGTTAACAG CTTGGCACTG 4150

GCCGTCGTTT TACAACGTCG TGA CTGGGAA AACCTGGCG TTACCCAACT 4200

10 TAATCGCCTT GCAGCACATC CCCCCTTCGC CAGCTGGCGT AATAGCGAAG 4250

AGGCCCGCAC CGATCGCCCT TCCCAACAGT TGC GTAGCCT GAATGGCGAA 4300

15 TGGCGCCTGA TGC GG TATTT TCTCCTTACG CATCTGTGCG GTATTT CACA 4350

CCGCATACGT CAAAGCAACC ATAGTACGCG CCCTGTAGCG GCGCATTAAG 4400

20 CGCGGCGGGT GTGGTGGTTA CGCGCAGCGT GACCGCTACA CTTGCCAGCG 4450

25 CCCTAGCGCC CGCTCCTTTC GCTTTCTTCC CTTCTTTTCT CGCCACGTTT 4500

GCCGGCTTTC CCCGTCAAGC TCTAAATCGG GGGCTCCCTT TAGGGTTCCG 4550

30 ATTTAGTGCT TTACGGCACC TCGACCCCAA AAAACTTGAT TTGGGTGATG 4600

G TTCACGTAG TGGGCCATCG CCCTGATAGA CGGTTTTTCG CCCTTTGACG 4650

35 TTGGAGTCCA CGTTCTTTAA TAGTGGACTC TTGTTCCAAA CTGGAACAAC 4700

40 ACTCAACCCT ATCTCGGGCT ATTCTTTTGA TTTATAAGGG ATTTTGCCGA 4750

TTTCGGCCTA TTGGTTAAAA AATGAGCTGA TTTAACAAAA ATTTAACGCG 4800

45 AATTTTAACA AAATATTAAC GTTTACAATT TTATGGTGCA CTCTCAGTAC 4850

AATCTGCTCT GATGCCGCAT AGTTAAGCCA ACTCCGCTAT CGCTACGTGA 4900

50 CTGGGTCATG GCTGCGCCCC GACACCCGCC AACACCCGCT GACGCGCCCT 4950

55 GACGGGCTTG TCTGCTCCCG GCATCCGCTT ACAGACAAGC TGTGACCGTC 5000

TCCGGGAGCT GCATGTGTCA GAGGTTTTCA CCGTCATCAC CGAAACGCGC 5050

60 GAGGCAGTAT TCTTGAAGAC GAAAGGGCCT CGTGATACGC CTATTTTTTAT 5100

AGGTTAATGT CATGATAATA ATGGTTTCTT AGACGTCAGG TGGCACTTTT 5150

65 CGGGGAAATG TGCGCGGAAC CCCTATTTGT TTATTTTTTCT AAATACATTC 5200

AAATATGTAT CCGCTCATGA GACAATAACC CTGATAAATG CTTCAATAAT 5250

ATTGAAAAAG GAAGAGTATG AGTATTCAAC ATTTCCGTGT CGCCCTTATT 5300

5 CCCTTTTTTG CGGCATTTTG CCTTCCTGTT TTTGCTCACC CAGAAACGCT 5350

GGTGAAAGTA AAAGATGCTG AAGATCAGTT GGGTGCACGA GTGGGTTACA 5400

10 TCGAACTGGA TCTCAACAGC GGTAAGATCC TTGAGAGTTT TCGCCCCGAA 5450

15 GAACGTTTTT CAATGATGAG CACTTTTAAA GTTCTGCTAT GTGGCGCGGT 5500

ATTATCCCGT GATGACGCCG GGCAAGAGCA ACTCGGTCGC CGCATACACT 5550

20 ATTCTCAGAA TGACTTGGTT GAGTACTCAC CAGTCACAGA AAAGCATCTT 5600

25 ACGGATGGCA TGACAGTAAG AGAATTATGC AGTGCTGCCA TAACCATGAG 5650

TGATAACACT GCGGCCAACT TACTTCTGAC AACGATCGGA GGACCGAAGG 5700

30 AGCTAACCGC TTTTTTGCAC AACATGGGGG ATCATGTAAC TCGCCTTGAT 5750

CGTTGGGAAC CGGAGCTGAA TGAAGCCATA CCAAACGACG AGCGTGACAC 5800

35 CACGATGCCA GCAGCAATGG CAACAACGTT GCGCAAATA TTAAGTGGCG 5850

AACTACTTAC TCTAGCTTCC CGGCAACAAT TAATAGACTG GATGGAGGCG 5900

GATAAAGTTG CAGGACCACT TCTGCGCTCG GCCCTTCCGG CTGGCTGGTT 5950

45 TATTGCTGAT AAATCTGGAG CCGGTGAGCG TGGGTCTCGC GGTATCATTG 6000

CAGCACTGGG GCCAGATGGT AAGCCCTCCC GTATCGTAGT TATCTACACG 6050

50 ACGGGGAGTC AGGCAACTAT GGATGAACGA AATAGACAGA TCGCTGAGAT 6100

AGGTGCCTCA CTGATTAAGC ATTGGTAACT GTCAGACCAA GTTTACTCAT 6150

ATATACTTTA GATTGATTTA AAACCTCATT TTTAATTTAA AAGGATCTAG 6200

60 GTGAAGATCC TTTTTGATAA TCTCATGACC AAAATCCCTT AACGTGAGTT 6250

TTCGTTCCAC TGAGCGTCAG ACCCCGTAGA AAAGATCAAA GGATCTTCTT 6300

65 GAGATCCTTT TTTTCTGCGC GTAATCTGCT GCTTGCAAAC AAAAAACCA 6350

CCGCTACCAG CGGTGGTTTG TTTGCCGGAT CAAGAGCTAC CAACTCTTTT 6400

5 TCCGAAGGTA ACTGGCTTCA GCAGAGCGCA GATACCAAAT ACTGTCCTTC 6450

TAGTGTAGCC GTAGTTAGGC CACCACTTCA AGAACTCTGT AGCACCGCCT 6500

10 ACATACCTCG CTCTGCTAAT CCTGTTACCA GTGGCTGCTG CCAGTGGCGA 6550

TAAGTCGTGT CTTACCGGGT TGGACTCAAG ACGATAGTTA CCGGATAAGG 6600

15 CGCAGCGGTC GGGCTGAACG GGGGGTTCGT GCACACAGCC CAGCTTGGAG 6650

CGAACGACCT ACACCGAACT GAGATACCTA CAGCGTGAGC ATTGAGAAAAG 6700

20 CGCCACGCTT CCCGAAGGGA GAAAGGCGGA CAGGTATCCG GTAAGCGGCA 6750

GGGTCGGAAC AGGAGAGCGC ACGAGGGAGC TTCCAGGGGG AAACGCCTGG 6800

25 TATCTTTATA GTCCTGTCGG GTTTCGCCAC CTCTGACTTG AGCGTCGATT 6850

30 TTTGTGATGC TCGTCAGGGG GCGGAGCCT ATGGAAAAAC GCCAGCAACG 6900

CGGCCTTTTT ACGGTTCTTG GCCTTTTGCT GGCCTTTTGC TCACATGTTC 6950

35 TTTCTGCGT TATCCCCTGA TTCTGTGGAT AACC GTATTA CCGCCTTTGA 7000

40 GTGAGCTGAT ACCGCTCGCC GCAGCCGAAC GACCGAGCGC AGCGAGTCAG 7050

TGAGCGAGGA AGCGGAAGAG CGCCCAATAC GCAAACCGCC TCTCCCCGCG 7100

45 CGTTGGCCGA TTCATTAATC CAGCTGGCAC GACAGGTTTC CCGACTGGAA 7150

AGCGGGCAGT GAGCGCAACG CAATTAATGT GAGTTACCTC ACTCATTAGG 7200

50 CACCCCAGGC TTTACACTTT ATGCTTCCGG CTCGTATGTT GTGTGGAATT 7250

55 GTGAGCGGAT AACAATTTCA CACAGGAAAC AGCTATGACC ATGATTACGA 7300

ATTAA 7305

60

CLAIMS

1. A DNA construct comprising a transcriptional initiation site, a transcriptional termination site, a selectable gene, a product gene
5 provided 3' to the selectable gene, a transcriptional regulatory region regulating transcription of both the selectable gene and the product gene, the selectable gene being positioned within an intron having a splice donor site 5' of the intron, which splice donor site regulates expression of the product gene using the transcriptional
10 regulatory region.
2. The DNA construct of claim 1 wherein the splice donor site comprises an efficient splice donor sequence.
- 15 3. The DNA construct of claim 2 wherein the splice donor site comprises a consensus splice donor sequence.
4. The DNA construct of claim 2 wherein the splice donor site comprises the sequence GACGTAAGT.
20
5. The DNA construct of claim 1 wherein the selectable gene is an amplifiable gene.
6. The DNA construct of claim 5 wherein the amplifiable gene is DHFR.
25
7. The DNA construct of claim 1 wherein the transcriptional regulatory region comprises a promoter and an enhancer.
8. A vector comprising the DNA construct of claim 1.
30
9. The vector of claim 8 wherein the selectable gene of the DNA construct is an amplifiable gene.
10. The vector of claim 8 that is capable of replication in a eukaryotic
35 host.
11. A eukaryotic host cell comprising the vector of claim 10.
12. A eukaryotic host cell comprising the DNA construct of claim 5.
40
13. The host cell of claim 11 wherein the vector is introduced into the host cell by electroporation.
14. A eukaryotic host cell comprising the DNA construct of claim 1
45 integrated into a chromosome of the host cell.

15. The host cell of claim 14 that is a mammalian cell.
16. A method for producing a product of interest comprising culturing the host cell of claim 11 so as to express the product gene and recovering the product from the host cell culture.
5
17. The method of claim 16 further comprising recovering the product from the culture medium.
- 10 18. The method of claim 16 wherein the selectable gene is an amplifiable gene and the splice donor site comprises an efficient splice donor sequence.
- 15 19. A method for producing a product of interest comprising culturing the host cell of claim 12 so as to express the product gene in a selective medium comprising an amplifying agent for sufficient time to allow amplification to occur, and recovering the product.
- 20 20. A method for producing eukaryotic cells having multiple copies of a product gene comprising transforming eukaryotic cells with the DNA construct of claim 5, growing the cells in a selective medium comprising an amplifying agent for a sufficient time for amplification to occur, and selecting cells having multiple copies of the product gene.
25
21. The method of claim 20 further comprising recovering from the selected cells the product of interest.

1 / 81

FIG. 1A

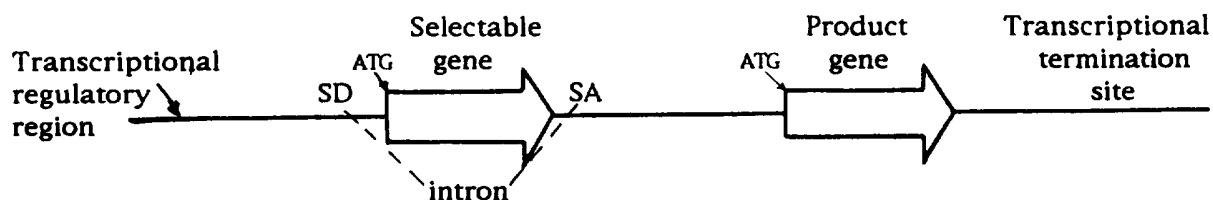


FIG. 1B

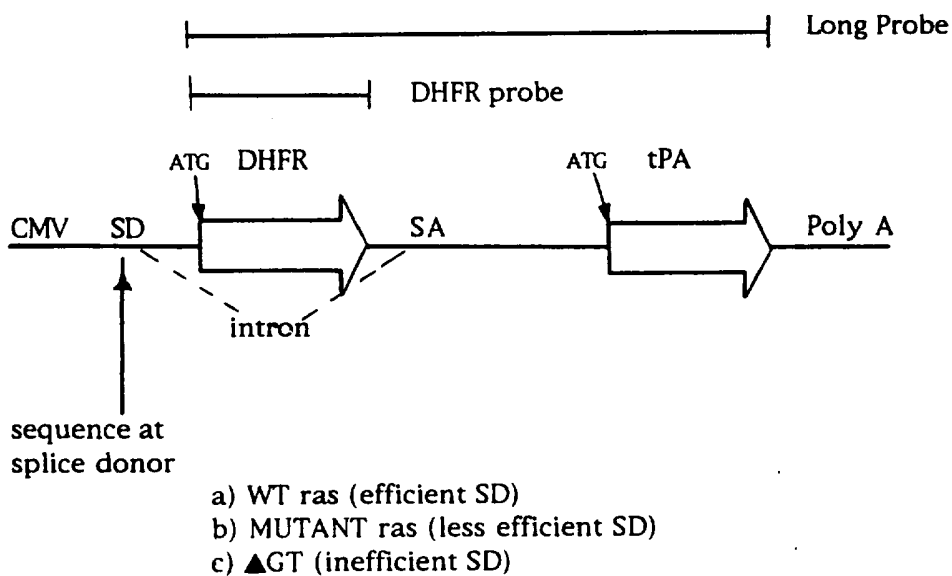
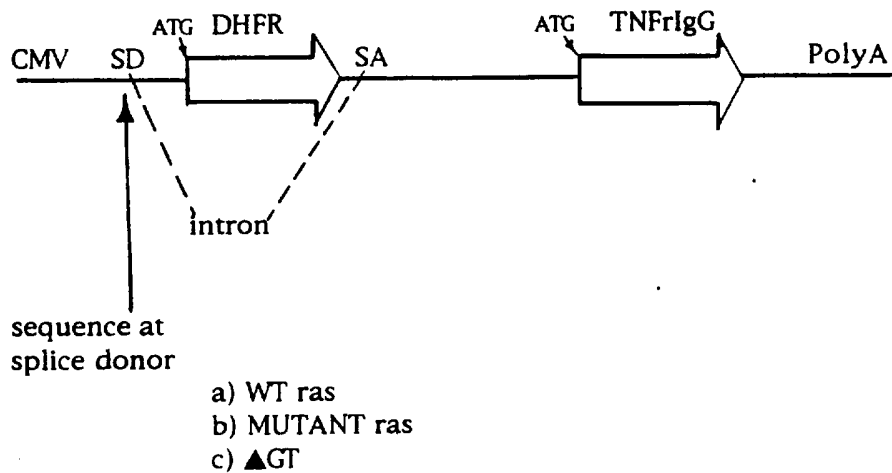


FIG. 1C



2 / 81

FIG. 1D

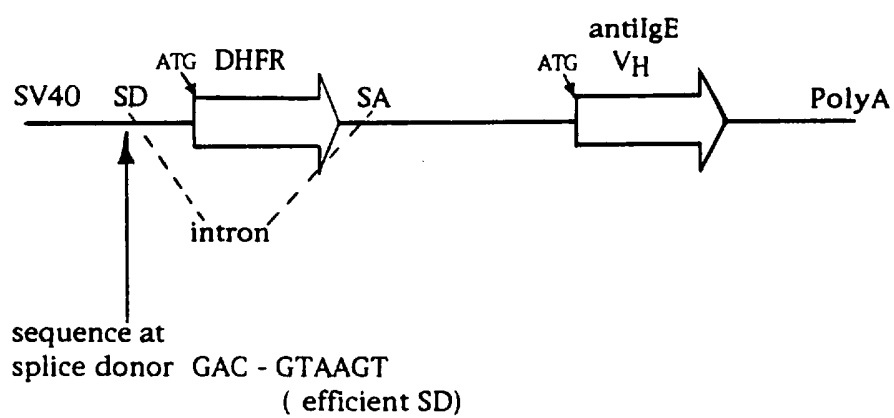


FIG. 2

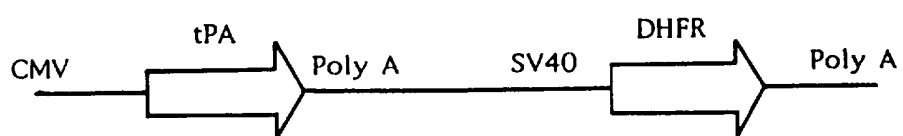


FIG. 3A

FIG. 3A

```
aluI
sstI
sacI
hgiJII
hgiAI/aspHI
ecII36II
bsp1286
bsiHKA1
bmyI
banII
taqI
1 TTCGAGCTCG CCCGACATTG ATTATTGACT AGTTATTAAAT AGTAATCAAT TACGGGGTCA TTAGTTTCATA GCCCATATAT GGAGTTCCGC GTTACATAAC
AAGCTCGAGC GGGCTGTAAC TAATAACTGA TCAATAATTA TCATTAGTTA ATGCCCCAGT AATCAAGTAT CCGGTATATA CCTCAAGGCG CAATGTATTG
thaI
fnuDII/mvni
bstUI
bsh1236I
aciI maeIII
bsII
scrFI
mvaI
ecoRII
dsav
aciI
bglI bstNI
sau96I
haeIII/palI aciI
asuI apyI[dcM+]
101 TTACGGTAA TGGCCGCCT GGCTGACCGC CCAACGACCC CGCCCATTTG AGTCAATAA TGACGTATGT TCCCATAGTA ACGCCAATAG GGACTTTCCA
AATGCCATT ACCGGCGGA CCGACTGGCG GGTGCTGGG GCGGGTAAAC TGCAGTTATT ACTGCATACA AGGGTATCAT TCGGGTTATC CCTGAAAGGT
maeII
hinII/acyI
ahaII/bsaHI
aatII
bglI
rsal
csp6I
ndeI
maeII
hinII/acyI
ahaII/bsaHI
aatII
201 TTGACGTCAA TGGGTGGAGT ATTTACGGTA AACTGCCAC TTGGCAGTAC ATCAAGTGA TCATATGCCA AGTACGCCCC CTATTGACGT CAATGACGGT
AACTGCAGTT ACCACCTCA TAAATGCCAT TTGACGGGTG AACCCTCATG TAGTTCACAT AGTATACGGT TCATGCGGGG GATAACTGCA GTTACTGCCA
nlaIII
scrFI
mvaI
ecoRII
aciI
bglI dsav
sau96I bstNI
haeIII/palI
asuI apyI[dcM+] bsrI nlaIII
301 AAATGGCCG CCTGGCATTG TGCCAGTAC ATGACCTTAT GGGACTTTCC TACTTGGCAG TACATCTACG TATTAGTCAT CGCTATTACC ATGGTGATGC
TTTACCGGC GGACCGTAAT ACGGCTCATG TACTGGAATA CCCTGAAAGG ATGAACCGTC ATGTAGATGC ATAATCAGTA GCGATAATGG TACCACCTAGC
ncoI
dsal hphI acil
bsaJI sfaNI
```

4 / 81

FIG. 3B

maeII
hinII/acyI
ahaII/bsaHI
aatII
bsmAI
pleI
hinFI
aciI
acII
rsal
csp6I
GATCATCAAT GGGCGTGGAT AGCGTTTGA CTCACGGGGA TTTCCAAAGTC TCCACCCCAT TGACGTCAAT GGGAGTTTGT TTTGGCACCA
CCAAAACCGT CATGTAGTTA CCGGCACCTA TCGCCAAACT GAGTGCCCT AAAGTTTCAAG AGGTGGGTA ACTGCAGTTA CCCTCAAACA AAACCGTGGT

alul
sstI
sacI
hgiJII
hgiAI/aspHI
ecII36II
bsp1286
bsiHKA
bmyI
banII

rsal
csp6I
mnII
nmlI
rsal
csp6I
aciI
hgaI
maeIII
aciI
acII
ccATTGACGC AAATGGCGG TAGCGGTGA CGGTGGGAGG TCTATATAAG CAGAGCTCGT
TTTAGTTGCC CTGAAAGGTT TTACAGCAAT GTTGAGGCGG GGTAACCTGCG TTACCCCGCC ATCCGCACAT GCCACCCCTC AGATATATTC GTCTCGAGCA

haeIII/palI
mcrI
eagI/xmaIII/ecI XI
eaeI
cfrI
fnu4HI
aciI
thaI
fnuDII/mvnI
sacII/sstII
nspBII
kspI scrFI
dsal nciI
bglI bslI mspI
sau3AI mnlI bstUI
mboI/ndeII[dam-] hpaII
dpnI[dam+] bsaJI dsav
dpnII[dam-] bsh1236I
alwI[dam-] aciI cauII

esp3I
scrFI
mvaI bsmAI
ecoRII
dsav
bstNI hinII/acyI
apyI[dcM+]
sau3AI gsuI/bpmI
mboI/ndeII[dam-]
dpnI[dam+] hgaI foki
dpnII[dam-] ahaII/bsaHI
gTCAGATCGC CTGGAGACGC CATCCACGCT GTTTTGACCT CCATAGAAGA CACCGGACC GATCCAGCCT CCGCGGCCGG GAACGTGCA
AATCACTGG CAGTCTAGCG GACCTCTGCG GTAGGTGCGA CAAAACCTGGA GGTATCTTCT GTGGCCCTGG CTAGGTGGA GCGCGCGGCC CTTGCCACGT

FIG. 3C

701 TTGGAACGCG GATTCCCGT GCCAAGAGTG CTGTAAAGTAC CGCCTATAGA GCGATAAGAG GATTTTATCC CGCTGCCAT CATGTTCCGA CCATTGAACT
 AACCTTGCGC CTAAGGGGCA CGGTTCTCAC GACATTTCATG GCGGATATCT CGCTATTCTC CTAATAATAGG GCGACGGTA GTACCAAGCT GGTAACTTGA
 801 GCATCGTCG CGTGTCCTCAA AATATGGGGA TTGGCAAGAA CGGAGACCTA CCTGCCCCC CGCTCAGGAA CGCGTTCAAG TACTTCCAAA GAATGACCAC
 CGTAGCAGCG GCACAGGGTT TTATACCCCT AACCGTTCTT GCCTCTGGAT GGGACGGGAG GCGAGTCCTT GCGCAAGTTC ATGAAGGTTT CTTACTGGTG
 901 AACCTCTTCA GTGGAAGGTA AACAGATCT GGTGATTATG GGTAGGAAA CCTGGTTCTC CATTCCTGAG AAGAATCGAC CTTTAAAGGA CAGAATTAAAT
 TTGGAGAGAGT CACCTTCCAT TTGTCTTAGA CCACTAATAC CCATCCTTTT GGACCAAGAG GTAAGGACTC TTCTTAGCTG GAAATTTCCT GTCTTAATTA
 1001 ATAGTTCTCA GTAGAGAACT CAAAGAACCA CCACGAGGAG CTCATTTTCT TGCCAAAAGT TTGGATGATG CCTTAAGACT TATTGAACAA CCGGAATTGG
 TATCAAGAGT CATCTCTTGA GTTCTTGGT GGTGCTCCTC GAGTAAAGA ACGGTTTTCA AACCTACTAC GGAATTCTGA ATAACCTTGT GGCCTTAACC

tfil
 acil
 thal
 fnuDII/mvnI
 bstUI
 bsh1236I
 fnu4HI
 bbvI
 nspBII
 acil
 nlaIII
 taqI
 mnlI
 rsal
 csp6I
 scfI
 bsmAI
 bsaI
 pflMI
 bslI
 bsrBI
 acil
 mnlI
 ddeI
 asp700
 scaI
 aflIII
 xmnI
 mluI
 thal
 fnuDII/mvnI
 bstUI
 bsh1236I
 scrFI
 mvaI
 ecoRII
 dsav
 bstNI
 apyI[dcn+]
 sexAI
 tfil
 hinfI
 alwNI
 hphI
 mboII
 earI/ksp632I
 mnlI
 tfil
 hinfI
 alwNI
 hphI
 mboII
 taqI
 msel
 ahaIII/draI
 aseI/asnI/vspI
 tru9I
 msel
 aluI
 sstI
 sacI
 hgiJII
 hgiAI/aspHI
 eclI36II
 bsp1286
 bsiHKAI
 bmyI
 banII
 mnlI
 bslI
 bstXI
 foki
 sfanI
 msel
 tru9I
 mspI
 hpaII
 bsaWI

6 / 81

FIG. 3D

```

1101 CAAGTAAAGT AGACATGGTT TGGATAGTCG GAGGCAGTTC TGTTTACCAG GAAGCCATGA ATCAACCAGG CCACCTTAGA CTCTTGTGA CTAAGGATCAT
GTTCAATTCA TCTGTACCAA ACCTATCAGC CTCCGTCAAG ACAATGGTC CTTCGGTACT TAGTTGGTCC GGTGGAATCT GAGAAACACT GTTCCTAGTA

accI nlaIII mnlI haeIII/palI haeI
scrFI mvaI mvaI mvaI mvaI
ecoRII dsav dsav pleI
bstNI nlaIII bstNI ddeI
apyI[dcM+] hinFI apyI[dcM+] hinFI maeIII alwI[dam-]
dpnII[dam-]
nlaIII mboI/ndeII[dam-]
dpnI[dam+] sau3AI
nlaIII

1201 GCAGGAATTT GAAAGTGACA CGTTTTTCCC AGAAATTGAT TTGGGGAAT ATAAACCTCT CCCAGATATC CCAGGCGTCC TCTCTGAGGT CCAGGAGGAA
CGTCCCTTAA CTTCCTACTGT GCAAAAAGGG TCTTTAACTA AACCCCTTA TATTTGGAGA GGTCTTTATG GGTCCGCGAGG AGAGACTCCA GGTCCCTCCTT

maeII mnlI
aflIII
apoI maeIII
bsaJI hgaI ddeI apyI[dcM+]
bstNI bslI asuI mnlI
dsav dsav dsav dsav
ecoRII sau96I
mvaI ecoNI
scrFI ecorII
ahaiI/bsaHI mvaI
hinII/acyI scrFI
mnlI

1301 AAAGGCATCA AGTATAAGTT TGAAGTCTAC GAGAAGAAAG ACTAACAGGA AGATGCTTTC AAGTTCTCTG CTCCCCTCCT AAAGCTATGC ATTTTATATA
TTTCCGCTAGT TCATATTCAA ACTTCAGATG CTCTTCTTTC TGATTGTCTT TCTACGAAAG TTCAAGAGAC GAGGGGAGGA TTTGATACG TAAAAATATT

sfaNI mboII accI mboII
scrFI mboII
nsiI/avaII
ppu10I

1401 GACCATGGA CTTTGTGCTGG CTTTAGACCC CTTGGCTTC GTTAGAACGC GGCTACAATT AATACATAAC bsh1236I aseI/asnI/vspI
CTGTGACCCT GAAAACGACC GAAATCTGGG GGAACCGAAG CAATCTTGG CCGATGTATA TTAGGTATTG GAATACATAG TATGTGTATC TAAATCCACT
hphI maeIII

```

FIG. 3E

FIG. 3E

1501 CACTATAGAA TAACATCCAC TTTGCCCTTTC TCTCCACAGG TGTCACCTCCA GGTCAACTGC ACCTCGGTTT TAAGCTTGGG CTGCAGGTGC CCGTGAATTT
 GTGATATCTT ATTGTAGGTG AAACGGAAAG AGAGGTGTCC ACAGTGAGGT CCAGTTGACG TGGAGCCAAG ATTCGAACCC GACGTCCAGC GGCACCTTAA

1601 AAGGACGCT GTGAAGCAAT CATGGATGCA ATGAAGAGAG GGCTCTGCTG TGTGCTGCTG CTGTGTGGAG CAGTCTTCGT TTCGCCAGC CAGGAAATCC
 TTCCCTGCGA CACTTCGTTA GTACCTACGT TACTTCTCTC CCGAGACGAC ACACGACGAC GACACACCTC GTCAGAAGCA AAGCGGCTCG GTCTTTTAGG

1701 ATGCCCCGATT CAGAAGAGGA GCCAGATCTT ACCAAGTGAT CTCAGAGAT GAAAAAAGC AGATGATATA CCAGCAACAT CAGTCATGGC TGCGCCCTGT
 TACGGGCTAA GTCTTCTCCT CGGTCTAGAA TGGTCTACTA GACGTCTCTA CTTTCTTGGG TCTACTATAT GGTCGTTGTA GTCAGTACCG ACGCGGACA

1801 GCTCAGAAGC AACCGGCTG .AATATTGCTG GTGCAACAGT GGCAGGCGAC AGTGCCACTC AGTGCCTGTC AAAAGTTGCA CGCAGCCAAG GTGTTTCAAC
 CGAGTCTTCG TTGGCCCCACC TTATAACGAC CACGTTGTCA CCGTCCCGTG TCACGGTGAG TCACGGACAG TTTTCAACGT CGCTCGGTTT CACAAAAGTTG

Restriction enzymes used for cloning and mapping are indicated above the sequences. The sequences are presented in 5' to 3' orientation. The sequences are presented in 5' to 3' orientation. The sequences are presented in 5' to 3' orientation.

FIG. 3F

[illegible]

FIG. 3G

fnu4HI
 bbvI
 scfI
 pstI
 bsgI
 aluI
 rsaI
 csp6I ddeI
 aciI
 2201 AAGCGGGGA AGTACAGCTC AGAGTTCTGC AGCACCCCTG CCTGCTCTGA GGGAAACAGT GACTGCTACT TTGGGAATGG GTCAGCCTAC CGTGGCAGCG
 TTCCGCCCT. TCATGTCGAG TCTCAAGACG TCGTGGGGAC GGACGAGACT CCTTTGTCA CTGACGATGA AACCTTACC CAGTCGGATG GCACCGTGG
 scrFI
 pflMI
 mvaI
 ecorII
 dsav
 bstNI
 bsli
 apyI[dcm+] haeIII/palI
 bsp1286 sau96I
 bmyI alwNI asuI
 bsrI bsaJI bsrI
 2301 ACAGCCTCAC CGAGTCGGGT GCCTCCTGCC TCCCGTGGAA TTCCATGATC CTGATAGGCA AGTTTACAC AGCACAGAAC CCCAGTGCCC AGGCACCTGGG
 TGTCGGAGTG GCTCAGCCCCA CGGAGGACGG AGGCACCTT AAGGTACTAG GACTATCCGT TCCAAATGTG TCGTGTCTTG GGGTCAAGGG TCCGTGACCC
 nlaIV
 hgiCI
 banI maeII
 scrFI
 mvaI
 pmlI
 ecorII
 dsav
 bstNI
 pflMI
 bsalI
 tfiI
 hinfI
 mspI
 hpaII
 apyI[dcm+] CATAATTACT GCCGGAATCC TGATGGGGAT GCCAGCCCT GGTGCCACGT GCTGAAGAAC CGCAGGCTGA CGTGGGAGTA CTGTGATGTG
 2401 CCTGGGCAAA GGACCCGTTT GTATTATGA CGGCCTTAGG ACTACCCCTA CGGTTCCGGA CCACGGTGCA CGACTTCTTG GCGTCCGACT GCACCTCAT GACACTACAC
 scrFI
 mboII
 earI/ksp632I
 hgiJII
 bsp1286
 bmyI
 banII
 mnlI sapi
 sfanI
 ddeI
 mnlI
 csp6I
 rsaI
 fnu4HI bsmAI
 aciI ddeI
 bspMI haeIII/palI
 2501 CCCTCCTGCT CCACCTGCGG CCTGAGACAG TACAGCCAGC CTCAGTTTCG CATCAAAGGA GGGCTCTTCG CCGACATCGC CTCCCACCCC TGGCAGGCTG
 GGGAGGACGA GGTGGACGCC GGACTCTGTC ATGTCGGTGC GAGTCAAAGC GTAGTTTCCT CCCGAGAAGC GGCTGTAGCG GAGGTGGGG ACCGTCCGAC
 apyI[dcm+] fnu4HI
 bslI bsaJI
 mnlI bslI
 bbvI

FIG. 31

FIG. 3I

2901

3001

3101

FIG. 3J

[illegible]

13 / 81

FIG. 3K

```

styI
aciI      fnu4HI   sau96I
fnu4HI    bgII    nlaIII
sfiI ncoI haeIII/palI
haeIII/palI
eaeI dsalI asuI      aluI
cfrI bsaJI          fnu4HI
                        bbvI
3601 GATGGCCGCC ATGGCCCAAC TTGTTTATTG CAGCTTATAA TGGTTACAAA TAAAGCAATA GCATCACAAA TTTCACAAAT AAAGCATTTT TTTCACCTGCA bsmI
CTACCGGCGG TACCGGGTGG AACAAATAAC GTCGAATATT ACCAATGTTT ATTTCGTTAT CGTAGTGTTT AAAGTGTTTA TTTCGTAAAA AAAGTGACGT

sau3AI
mboI/ndeII[dam-]
dpmI[dam+]
dpmII[dam-]
pvuI/bspCI
mcrI
taqI[dam-] tru9I
claiI/bsp106[dam-]
sau3AI mseI fnu4HI haeI styI
mboI/ndeII[dam-] bbvI ncoI
dpmI[dam+] xmnI hinPI dsalI haeIII/palI
dpmII[dam-] aseI/asnI/vspI bsaJI
nlaIII alwI[dam-] asp700 hhaI/cfoI nlaIII mnlI
3701 TTCTAGTTGT GGTGTGTCCA AACTCATCAA TGTATCTTAT CATGTCTGGA TCGATCGGGA ATTAATTGG CGCAGCACCA TGGCCTGAAA TAACCTCTGA
AAGATCAACA CCAACACAGT TTGAGTAGTT ACATAGAATA GTACAGACCT AGCTAGCCCT TAATTAAGCC GCGTCGTGGT ACCGGACTTT ATTGGAGACT

rsal
csp6I
nlaIV
kpnI
hgiCI
banI
asp718 mnlI
acc65I ddeI aciI
mnlI
3801 AAGAGGAACT TGGTTAGGTA CCTTCTGAGG CGGAAAGAAC CAGCTGTGGA ATGTGTGTCA GTTAGGGTGT GGAAAGTCCC CAGGCTCCCC AGCAGGCAGA
TTCTCCTTGA ACCAATCCAT GGAAGACTCC GCCTTTCTTG GTCGACACCT TACACACAGT CAATCCCACA CCTTTCAGGG GTCCGAGGGG TCGTCCGTCT

```

14 / 81

FIG. 3L

```

          sfanI      nlaIV      scrFI      scrFI      mvaI      mvaI      ppul0I      ppul0I
          nsII/avaIII      nsII/avaIII      nsII/avaIII      nsII/avaIII      nsII/avaIII      nsII/avaIII
          nlaIII      nlaIII      nlaIII      nlaIII      nlaIII      nlaIII      nlaIII      nlaIII
          sphI      sphI      sphI      sphI      sphI      sphI      sphI      sphI
          nspI      nspI      nspI      nspI      nspI      nspI      nspI      nspI
          nspHI      nspHI      nspHI      nspHI      nspHI      nspHI      nspHI      nspHI

3901 AGTATGCAAA GCATGCATCT CAATTAGTCA GCAACCAGGT GTGGAAAGTC CCCAGGCTCC CCAGCAGGCA GAAGTATGCA AAGCATGCAT CTCAAATTAGT
TCATACCGTTT CGTACGTAGA GTTAATCAGT CGTTGGTCCA CACCTTTCAG GGGTCCGAGG GGTGCTCCGT CTTCATACGT TTCGTACGTA GAGTTAATCA

          nlaIII      nlaIII      nlaIII      nlaIII      nlaIII      nlaIII      nlaIII      nlaIII
          styI      styI      styI      styI      styI      styI      styI      styI
          ncoI      ncoI      ncoI      ncoI      ncoI      ncoI      ncoI      ncoI
          bslI      bslI      bslI      bslI      bslI      bslI      bslI      bslI
          dsal      dsal      dsal      dsal      dsal      dsal      dsal      dsal
          aciI      aciI      aciI      aciI      aciI      aciI      aciI      aciI
          bsaJI      bsaJI      bsaJI      bsaJI      bsaJI      bsaJI      bsaJI      bsaJI

4001 CAGCAACCAT AGTCCGCGC CTAACCTCGC CCATCCCGCC CCTAACTCCG CCCATTCTCC GCCCATGCGC TGACTAATTT TTTTATTTTA
GTCGTTGGTA TCAGGGCGGG GATTGAGCG GGTAGGCGG GGTGCAAGGC GGTCAAGGC GGTCAAGGC GGTCAAGGC ACTGATTAAA AAAAATAAAT

          rmaI      rmaI      rmaI      rmaI      rmaI      rmaI      rmaI      rmaI
          styI      styI      styI      styI      styI      styI      styI      styI
          bsaJI      bsaJI      bsaJI      bsaJI      bsaJI      bsaJI      bsaJI      bsaJI
          blnI      blnI      blnI      blnI      blnI      blnI      blnI      blnI
          avrII      avrII      avrII      avrII      avrII      avrII      avrII      avrII
          haeIII/palI      haeIII/palI      haeIII/palI      haeIII/palI      haeIII/palI      haeIII/palI      haeIII/palI      haeIII/palI
          stuI      stuI      stuI      stuI      stuI      stuI      stuI      stuI
          haeI      haeI      haeI      haeI      haeI      haeI      haeI      haeI
          mnlI      mnlI      mnlI      mnlI      mnlI      mnlI      mnlI      mnlI
          maeI      maeI      maeI      maeI      maeI      maeI      maeI      maeI
          aluI      aluI      aluI      aluI      aluI      aluI      aluI      aluI
          msel      msel      msel      msel      msel      msel      msel      msel
          aluI      aluI      aluI      aluI      aluI      aluI      aluI      aluI
          hpaI      hpaI      hpaI      hpaI      hpaI      hpaI      hpaI      hpaI
          hincII/hindII      hincII/hindII      hincII/hindII      hincII/hindII      hincII/hindII      hincII/hindII      hincII/hindII      hincII/hindII

4101 TGCAGAGGCC GAGGCCGCT CGCCTCTGA GCTATTCCAG AAGTAGTGAG GAGGCTTTT TGGAGGCCCTA GGCTTTTGTCA AAAAGCTGTT AACAGCTTGG
ACGTCTCCG CTCCGGCGGA GCCGGAGACT CGATAGGTC TTTCATCACTC CTCCGAAAAA ACCTCCGAT CCGAAAAACGT TTTTCGACAA TTGTCGAACC

          scrFI      scrFI      scrFI      scrFI      scrFI      scrFI      scrFI      scrFI
          mvaI      mvaI      mvaI      mvaI      mvaI      mvaI      mvaI      mvaI
          ecorII      ecorII      ecorII      ecorII      ecorII      ecorII      ecorII      ecorII
          dsav      dsav      dsav      dsav      dsav      dsav      dsav      dsav
          bstNI      bstNI      bstNI      bstNI      bstNI      bstNI      bstNI      bstNI
          apyI[dcM+]      apyI[dcM+]      apyI[dcM+]      apyI[dcM+]      apyI[dcM+]      apyI[dcM+]      apyI[dcM+]      apyI[dcM+]
          bsaJI      bsaJI      bsaJI      bsaJI      bsaJI      bsaJI      bsaJI      bsaJI
          maeIII      maeIII      maeIII      maeIII      maeIII      maeIII      maeIII      maeIII
          mnlI      mnlI      mnlI      mnlI      mnlI      mnlI      mnlI      mnlI
          aluI      aluI      aluI      aluI      aluI      aluI      aluI      aluI
          bsrI      bsrI      bsrI      bsrI      bsrI      bsrI      bsrI      bsrI
          maeIII      maeIII      maeIII      maeIII      maeIII      maeIII      maeIII      maeIII
          bsaJI      bsaJI      bsaJI      bsaJI      bsaJI      bsaJI      bsaJI      bsaJI
          fnu4HI      fnu4HI      fnu4HI      fnu4HI      fnu4HI      fnu4HI      fnu4HI      fnu4HI
          mnlI      mnlI      mnlI      mnlI      mnlI      mnlI      mnlI      mnlI
          maeI      maeI      maeI      maeI      maeI      maeI      maeI      maeI
          foki      foki      foki      foki      foki      foki      foki      foki
          nspBII      nspBII      nspBII      nspBII      nspBII      nspBII      nspBII      nspBII

4201 CACTGGCCGT CGTTTACAA CGTCGTGACT GGGAAAACCC TGGCGTTACC CAACTTAATC GCCTTGACG ACATCCCCC TTCGCCAGCT GCGGTAATAG
GTGACCGCA GCAAAATGTT GCAGCACTGA CCTTTTGG ACCGCAATGG GTTGAATTAG CGGAACGTCG TGTAGGGGGG AAGCGGTGCA CCGCATTATC

```

FIG. 3M

[illegible]

FIG. 3N

4801	CGCGATTTCG	GCCTATTGGT	TAAAAAATGA	GCTGATTATA	CAAAAAATTTA	ACGGCAATTT	TAACAAAATA	TTAACGTTTA	CAATTTTATG	GTGCACTCTC
	CGGCTAAAGC	CGGATAACCA	ATTTTACT	CGACTAAAT	GTTTTAAAT	TGCGCTTAA	ATTGTTTAT	AATGCAAT	GTTAAATAC	CACGTGAGAG
	haeIII/palI	tru9I	mseI	tru9I	mseI	apoI	apoI	sspI	mseI	
	alul	mseI	apoI	bsh1236I						
	tru9I	apoI	tru9I	apoI	tru9I	maeII	psp1406I			
	maeII/mvnI	thaI								
	bspl286	bsiHKA1	bmyI	ddeI	apaLI/snoI	alw44I/snoI	GTGCACTCTC			
	hgiAI/aspHI									
4901	AGTACAATCT	GCTCTGATGC	CGCATAGTTA	AGCCAACTCC	GCTATCGCTA	CGTGACTGGG	TCATGGCTGC	GCCCCGACAC	CCGCTGACGC	
	TCATGTTAGA	CGAGACTACG	CGGTATCAAT	TCGGTTGAGG	CGATAGCGAT	GCACTGACCC	AGTACCGACG	CGGGGCTGTG	GGCGACTGCG	
	rsal	fnu4HI	tru9I	mseI	aciI	bsaAI	tthlII/aspI	bbvI	aciI	hgal
	mspI	hpaII	scrFI	ncii	dsav	esp3I	fnu4HI	bbvI		
	cauII	aciI	alul	bslI	cauII	alul	nlaIII	mnII	hphI	
	drdI	gccCTGACGG	GCTTGCTGTC	TCCCGGCATC	CGCTTACAGA	CAAGCTGTGA	CCGTCTCCGG	GAGCTGCATG	TGTCAGAGGT	TTTCACCGTC
	CGGACTGCC	CGAACAGACG	AGGGCCGCTAG	GCGAATGTCT	GTTCGACACT	GGCAGAGGCC	CTCGACGCTAC	ACAGTCTCCA	AAAGTGGCAG	TAGTGGCTTT
	thaI	fnuDII/mvnI	bstUI	bsh1236I	hinPI	hhaI/cfoI	thaI	fnuDII/mvnI	bstUI	
	hhaI/cfoI	thaI	mnII	haeIII/palI	mnII	saue6I	asuI	ecoO109I/draII	bbsI	
	mbolI	bpuAI	bbsI	agATATCTTG	AAGACGAAAG	GGCCTCGTGA	TACGCCCTATT	TTTATAGGTT	AATGTCATGA	TAATAATGTT
	CGCGCGAGGC	AGTATCTTG	AAGACGAAAG	GGCCTCGTGA	TACGCCCTATT	TTTATAGGTT	AATGTCATGA	TAATAATGTT	TTCTTAGACG	TCAGGTGGCA
	CGCGCTCCG	TCATAAGAAC	TTCTGCTTTC	CCGAGCACT	ATGCGGATAA	AAATATCCAA	TTACAGTACT	ATTATTACCA	AAGATCTGC	AGTCCACCGT
5101	CGCGCTCCG	TCATAAGAAC	TTCTGCTTTC	CCGAGCACT	ATGCGGATAA	AAATATCCAA	TTACAGTACT	ATTATTACCA	AAGATCTGC	AGTCCACCGT
	bspl236I	hhaI/cfoI	thaI	mnII	haeIII/palI	mnII	saue6I	asuI	ecoO109I/draII	bbsI
	mbolI	bpuAI	bbsI	agATATCTTG	AAGACGAAAG	GGCCTCGTGA	TACGCCCTATT	TTTATAGGTT	AATGTCATGA	TAATAATGTT
	CGCGCTCCG	TCATAAGAAC	TTCTGCTTTC	CCGAGCACT	ATGCGGATAA	AAATATCCAA	TTACAGTACT	ATTATTACCA	AAGATCTGC	AGTCCACCGT
	bspl236I	hhaI/cfoI	thaI	mnII	haeIII/palI	mnII	saue6I	asuI	ecoO109I/draII	bbsI
	mbolI	bpuAI	bbsI	agATATCTTG	AAGACGAAAG	GGCCTCGTGA	TACGCCCTATT	TTTATAGGTT	AATGTCATGA	TAATAATGTT
	CGCGCTCCG	TCATAAGAAC	TTCTGCTTTC	CCGAGCACT	ATGCGGATAA	AAATATCCAA	TTACAGTACT	ATTATTACCA	AAGATCTGC	AGTCCACCGT
	bspl236I	hhaI/cfoI	thaI	mnII	haeIII/palI	mnII	saue6I	asuI	ecoO109I/draII	bbsI
	mbolI	bpuAI	bbsI	agATATCTTG	AAGACGAAAG	GGCCTCGTGA	TACGCCCTATT	TTTATAGGTT	AATGTCATGA	TAATAATGTT
	CGCGCTCCG	TCATAAGAAC	TTCTGCTTTC	CCGAGCACT	ATGCGGATAA	AAATATCCAA	TTACAGTACT	ATTATTACCA	AAGATCTGC	AGTCCACCGT
	bspl236I	hhaI/cfoI	thaI	mnII	haeIII/palI	mnII	saue6I	asuI	ecoO109I/draII	bbsI
	mbolI	bpuAI	bbsI	agATATCTTG	AAGACGAAAG	GGCCTCGTGA	TACGCCCTATT	TTTATAGGTT	AATGTCATGA	TAATAATGTT
	CGCGCTCCG	TCATAAGAAC	TTCTGCTTTC	CCGAGCACT	ATGCGGATAA	AAATATCCAA	TTACAGTACT	ATTATTACCA	AAGATCTGC	AGTCCACCGT
	bspl236I	hhaI/cfoI	thaI	mnII	haeIII/palI	mnII	saue6I	asuI	ecoO109I/draII	bbsI
	mbolI	bpuAI	bbsI	agATATCTTG	AAGACGAAAG	GGCCTCGTGA	TACGCCCTATT	TTTATAGGTT	AATGTCATGA	TAATAATGTT
	CGCGCTCCG	TCATAAGAAC	TTCTGCTTTC	CCGAGCACT	ATGCGGATAA	AAATATCCAA	TTACAGTACT	ATTATTACCA	AAGATCTGC	AGTCCACCGT
	bspl236I	hhaI/cfoI	thaI	mnII	haeIII/palI	mnII	saue6I			

17 / 81

FIG. 30

```

nlaIV
aciI
thai
fnuDII/mvnI
bstUI
bsh1236I
hinPI
hhaI/cfoI

5201 CTTTTCGGG AAATGTGCG GGAACCCCTA TTTGTTTATT TTTCTAAATA CATTCAAATA TAACCCCTGAT AAATGCTTCA
GAAAAGCCCC TTACACGCG CTTGGGGAT AAACAATAA AAAGATTAT GTAAAGTTTAT ACATAGGCGA GTACTCTGTT ATTGGGACTA TTTACGAAGT

rcaI
bspHI
bsrBI bsmAI
aciI nlaIII
fnu4HI
aciI

5301 ATAATATTGA AAAAGGAAGA GTATGAGTAT TCAACATTTC CGTGTGCGCC TTATTCCCTT TTTTGGCGCA TTTTGCCTTC CTGTTTTTGC TCACCCAGAA
TATTATAACT TTTTCTTCT CATACTCATA AGTTGTAAG GCACAGCGGG AATAAGGGA AAAACGCCGT AAACGGGAG GACAAAAACG AGTGGGTCTT

hgiAI/aspHI
bsp1286
bsiHKA1
sau3AI mboI/ndeII[dam-]
dpnI[dam+] bmyI
dpnII[dam-]
eco57I apaLI/snoI
hphI sfanI mboII[dam-] alw4I/snoI maeIII taqI alwI[dam-] aciI bstYI/xhoII
hphI bstYI/xhoII
bsrI dpnII[dam-] alwI[dam-]
bstYI/xhoII
dpnI[dam+]
dpnII[dam-]
sau3AI nspBII
sau3AI mboI/ndeII[dam-]
dpnI[dam+]
dpnII[dam-]
alwI[dam-]
bstYI/xhoII
aciI bstYI/xhoII
AGTTTTCGCC
TGGGACCACT TTCAATTTCT ACGACTTCTA GTCAACCCAC GTGCTCACCC AATGTAGCTT GACCTAGAGT TGTGCGCATT CTAGGAATC TCAAAAGCGG

maeII
psp1406I
xmnI
asp700
mboII
ccgAAGAAGC TTTTCCAATG ATGAGCACTT TTAAGTTCT GCTATGTGGC GCGGTATTAT CCCGTGATGA CGCCGGGCAA GAGCAACTCG GTCGCGGCAT
GGCTTCTTGC AAAAGGTTAC TACTCGTGAA AATTTCAGA CGTACACCG CGCCATAATA GGGCACTACT GCGGCCCGTT CTCGTTGAGC CAGCGCGGTA

scrFI
nciI
mspI
hpaII
dsav
hinII/acyI
hgaI caulI
ahaII/bsaHI
bcgI mcrI fnu4HI
aciI
fnu4HI
bbvI
nlaIII

5501 ACACATTCT CAGAAATGACT TGGTTGAGTA CTCACCACTC ACAGAAAAGC ATCTTACGGA TGGCATGACA GTAAGAGAAT TATGCAGTGC TGCCATAACC
TGTGATAAGA GTCTTACTGA ACCAACTCAT GAGTGGTCAG TGCTTTTTCG TAGAATGCCT ACCGTAAGTGT CATCTCTCTTA ATACGTCACG ACGGTATTGG

```

18 / 81

FIG. 3P

sau96I
 avaiI
 sau3AI asuI
 mboI/ndeII[dam-]
 dpnI[dam+]
 dpnII[dam-]
 pvuI/bspCI
 mcrI mnlI
 aluI aciI
 nlaIII
 sau3AI maeIII
 mboI/ndeII[dam-]
 dpnI[dam+]
 dpnII[dam-]
 nlaIII alwI[dam-]
 5701 ATGAGTGATA AACTCGGCG CAACTTACTT CTGACAAACGA TCGGAGGACC GAAGGAGCTA ACCGCTTTTT TGCACAACAT GGGGATCAT GTAACTCGCC
 TACTCACTAT TGTGACGCCG GTTGAATGAA GACTGTGCT AGCCTCCTGG CTTCCTCGAT TGGCGAAAAA ACGTGTGTA CCCCTAGTA CATTGAGCGG

hinPI
 hhaI/cfoI
 mstI
 avilI/fspI bsrI
 maeII
 pspl406I
 msel
 5801 TTGATCGTTG GGAACCGGAG CTGAATGAAG CCATACCAAA CGACGAGCGT GACACCACGA TGCCAGCAGC AATGGCAACA ACGTTGCCGA AACTATTAAAC
 AACTAGCAAC CCTTGGCCTC GACTTACTTC GGTATGGTTT GCTGCTCGCA CTGTGTGCT ACGTGTGCT TACCGTTGT TGCAACGCGT TTGATAATTG

mspI
 hpaII
 scrFI
 aluI nciI
 rmaI dsav
 maeI cauII
 fnu4HI
 sbfNI
 bbvI
 sau96I
 haeIII/palI
 hinPI asuI mspI
 hhaI/cfoI hpaII
 5901 TGGCGAATA CTTACTCTAG CTTCCCGGCA ACAAATAAATA GACTGGATGG AGCGGATAA AGTTCAGGA CCACCTCTGC GCTCGGCCCT TCCGGCTGGC
 ACCGCTTGAT GAATGAGATC GAAGGGCCGT TGTTAATTAT CTGACCTACC TCCGCCTATT TCAACGTCCT GGTGAAGACG CGAGCCGGGA AGCCGACCG

mspI
 hpaII
 cfr10I
 nlaIV hphI
 gsuI/bpmI
 bsaI bsh1236I
 bbvI bsrI asuI
 mnlI
 6001 TGGTTTATTG CTGATAAATC TGGAGCCGGT GAGCGTGGGT CTCGCGGTAT CATTGCAGCA CTGGGCCAG ATGGTAAGCC CTCCCCTATC GTAGTTATCT
 ACCAAATAAC GACTATTAG ACCTCGGCA CTGCAACCA GAGCGCCATA GTAACGTCGT GACCCCGTC TACCATTGG GAGGCGATAG CATCAATAGA

pleI
 hinfi
 eam1105I
 foki
 sau3AI
 nlaIV
 mboI/ndeII[dam-]
 dpnI[dam+]
 hgiCI
 banI mnlI msel
 tru9I
 maeIII
 6101 ACACGACGGG GAGTCAGGCA ACTATGGATG AACGAATAG ACAGATCGCT GAGATAGGTG CCTCACTGAT TAAGCATGG TAACGTGTCAG ACCAAGTTTA
 TGTGCTGCCC CTCAGTCCGT TGATACCTAC TTGCTTTATC TGCTAGCGA CTCTATCCAC GGAGTGACTA ATTCGTAACC ATTGACAGTC TGGTTCAAAT

19 / 81

FIG. 3Q

```

rmaI      sau3AI
sau3AI hphI mboI/ndeII[dam-]
mboI/ndeII[dam-]
dpmI[dam+] dpmI[dam+]
dpmII[dam-] dpmII[dam-]
tru9I bstYI/xhoII alwI[dam-] maeII
mseI      tru9I mseI alwI[dam-] bstYI/xhoII rcaI tru9I
ahaIII/draI mseI ahaIII/draI maeI mboII[dam-] bspHI mseI
6201 CTCATATATA CTTTAGATTG ATTTAAAGGA TCTAGGTGAA GATCCTTTT GATAATCTCA TGACCAAAAT CCCTTAACGT
GAGTATATAT GAAATCTAAC TAAATTTTGA AGTAAATTTT AAATTTTCT AGATCCACTT CTAGGAAAAA CTATTAGAGT ACTGGTTTTA GGAATTTGCA

sau3AI
mboI/ndeII[dam-]
dpmI[dam+] sau3AI
dpmII[dam-] mboI/ndeII[dam-] thaI fnuDII/mvnI
bstYI/xhoII dpmI[dam+] bstUI
sau3AI alwI[dam-] dpmII[dam-] bsh1236I
mboI/ndeII[dam-] alwI[dam-] hinPI fnu4HI
dpmI[dam+] mboII[dam-] bstYI/xhoII hhaI/cfoI bbvI
dpmII[dam-]
6301 GAGTTTTCGT TCCACTGAGC GTCAGACCCC GTAGAAAAGA TCAAGGATC TTCTTGAGAT CCTTTTTC TCGCGGTAAT CTGCTGCTTG CAAACAAAAA
CTCAAAAGCA AGTGACTCG CAGTCTGGG CATCTTTTCT AGTTCTCTAG AAGAACTCTA GGAATAAAAG ACGGCAATTA GACGACGAAC GTTTGTTTTT

sau3AI
mboI/ndeII[dam-]
dpmI[dam+]
dpmII[dam-]
alwI[dam-]
aciI      mspI      hinPI
aciI nspBII hpaII aluI      bsrI      eco57I hhaI/cfoI
6401 AACCACCGCT ACCAGCGGTG GTTTGTTTGC CGGATCAAGA GCTACCAACT CTTTTCCGA AGTAACTGG CTTACAGCAGA GCGCAGATAC CAAATACTGT
TTGGTGGCGA TGSTGCGCAC CAAACAAACG GCCTAGTTCT CGATGGTTGA GAAAAGGCT TCCATTGACC GAAATCGTCT CGCGTCTATG GTTTATGACA

aciI      haeIII/palI
aciI      haeI
bslI      haeI
6501 CCTTCTAGTG TAGCCGTAGT TAGGCCACCA CTTCAAGAAC TCTGTAGCAC CGCCTACATA CCTCGCTCTG CTAATCCTGT TACCAGTGGC TGCTGCCAGT
GGAAGATCAC ATCCGGTGGT GAAGTTCTTG AGACATCGTG GCGGATGTAT GGAGCGAGAC GATTAGGACA ATGGTCACCG ACGACGGTCA

```

FIG. 3R

[illegible]

```
>length: 7360
```

22 / 81

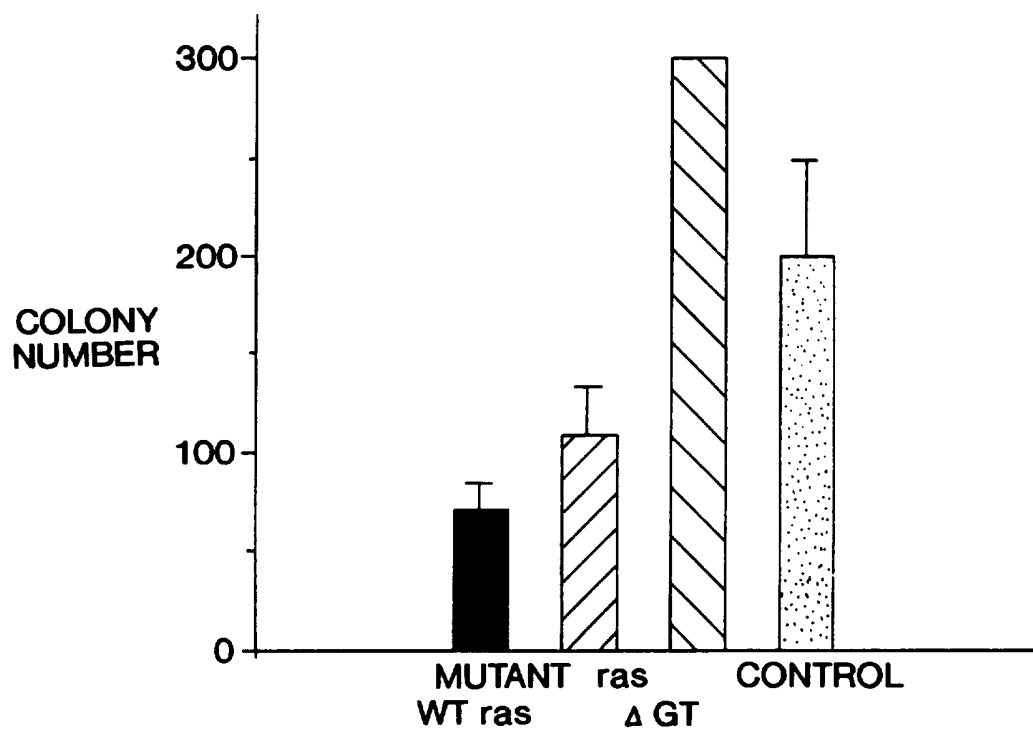


FIG. 4

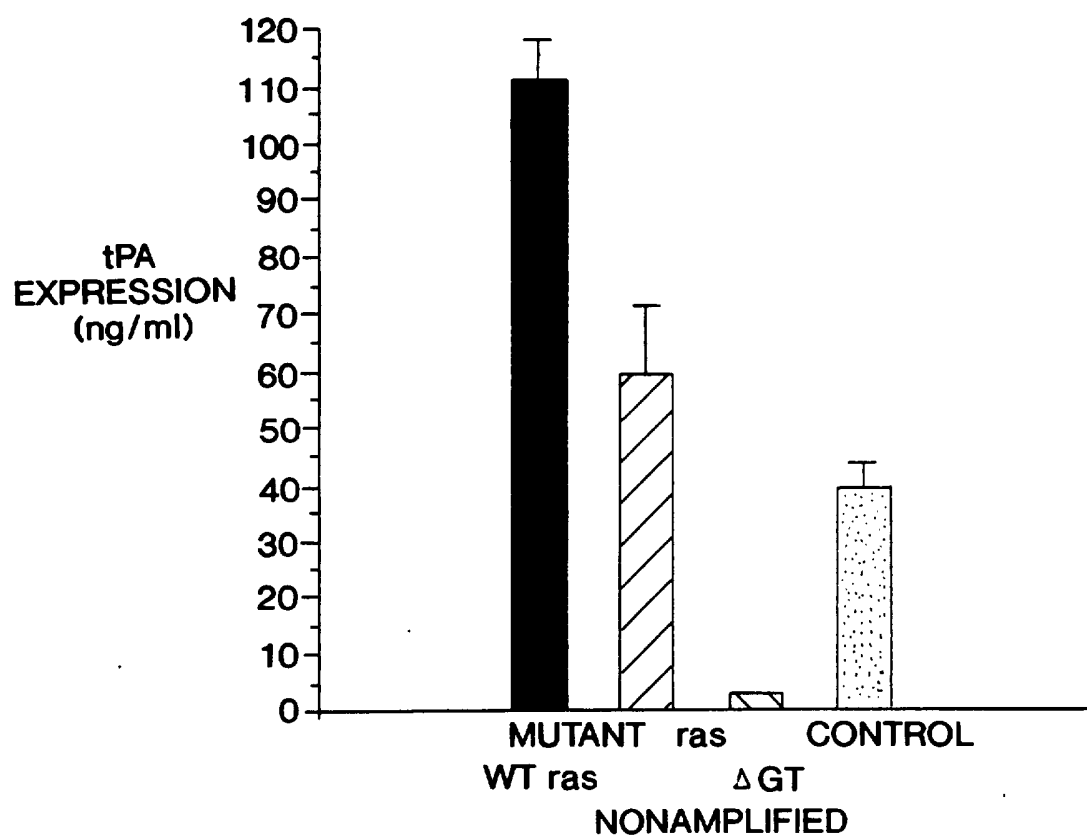


FIG. 5A

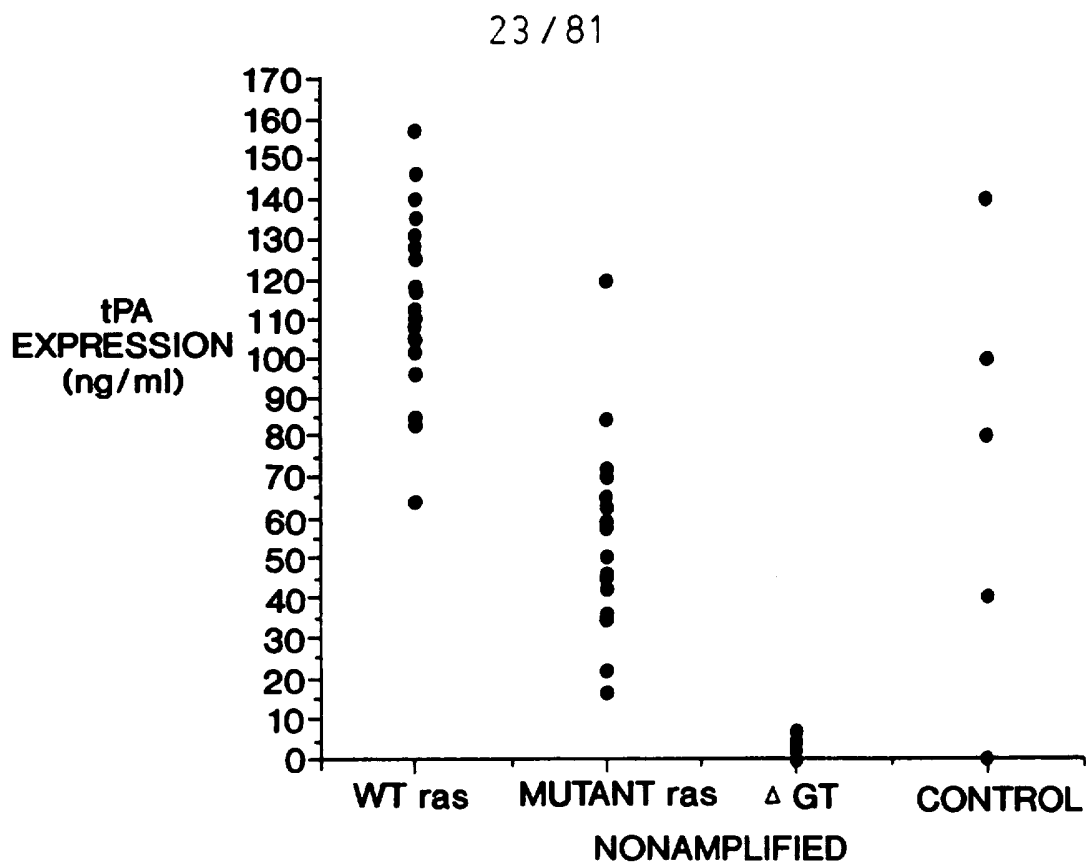


FIG. 5B

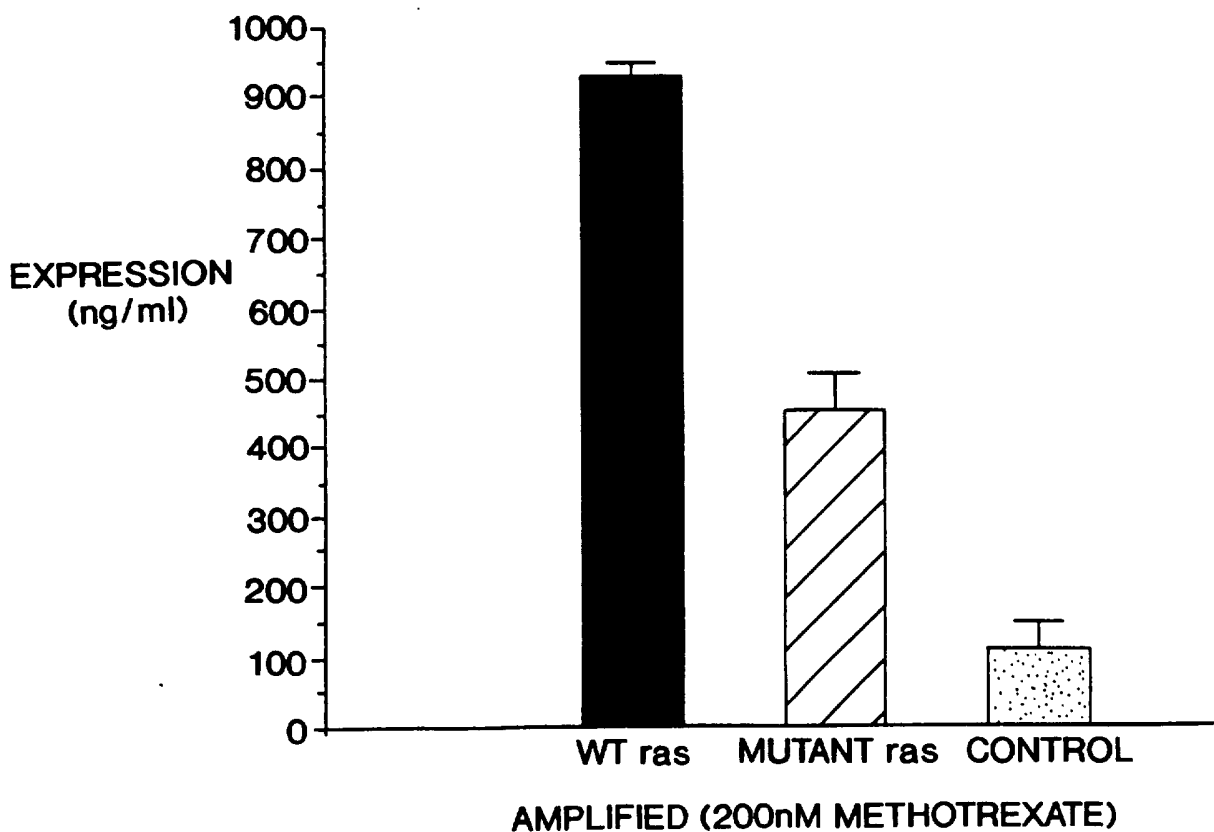


FIG. 5C

[illegible]

[illegible]

FIG. 6C

[illegible]

27 / 81

FIG. 6D

```

1101 CAAGTAAAGT AGACATGGTT TGGATAGTCG GAGGCAGTTC TGTTTACCAG GAAGCCATGA ATCAACCAGG CCACCTTAGA CTCTTGTGA CAAGGATCAT
    GTTCATTTC TCTGTACCAA ACCTATCAGC CTCCGTCAAG ACAATGCTC TAGTTGGTCC GGTGGATCT GAGAAACACT GTTCCTAGTA
    accI nlaIII mnlI apyI[dcm+] hinfI apyI[dcm+] hinfI maeIII alwI[dam-] dpnII[dam-]
    scrFI mvaI ecorII dsav pleI bstNI nlaIII bstNI ddeI
    haeIII/palI haeI mvaI mboI/ndeII[dam-] dpnI[dam+] sau3AI
    scrFI mvaI ecorII dsav pleI bstNI nlaIII bstNI ddeI
    haeIII/palI haeI mvaI mboI/ndeII[dam-] dpnI[dam+] sau3AI

1201 GCAGGAATTT GAAAGTGACA CGTTTTTCCC AGAATTTGAT TTGGGGAAT ATAAACCTCT CCCAGAATAC CCAGCGTCC TCTCTGAGGT CCAGGAGGAA
    CGTCCTTAAA CTTTCACTGT GCAAAAAGGG TCTTTAACTA TATTTGGAGA GGTCTTATG GGTCCGACAG AGAGACTCCA GTTCCTCCTT
    maeII aflIII maeIII apoI maeIII mnlI bsajI hgaI ddeI apyI[dcm+]
    scrFI mvaI ecorII dsav bstNI bsliI asuI mnlI
    hinfI/acyI ahaII/bsaHI scrFI mvaI
    mvaI ecorII dsav
    ecorII sau96I
    dsav avall
    bstNI bsliI asuI mnlI
    apyI[dcm+] mnlI bstNI
    bsajI hgaI ddeI apyI[dcm+]

1301 AAAGGCATCA AGTATAAGTT TGAAGTCTAC GAGAAGAAAG ACTAACAGGA AGATGCTTTC AAGTTCTCTG CTCCCTCCT AAAGCTATGC ATTTTATATA
    TTTCCTAGT TCATATTCAA ACTTCAGATG CTCTTCTTTC TGATTGCTCT TCTACGAAAG TTCAAGAGAC GAGGGGAGGA TTTTCGATACG TAAAAATATT
    sfaNI accI mboII mboII mnlI aluI
    ppulOI nsii/avaIII

1401 GACCATGGA CTTTGTCTGG CTTTAGACCC CTTTGGCTTC GTTAGAACGC GGCTACAATT AATACATAAC CTTATGTATC ATACACATAG ATTTAGGTGA
    CTGGTACCCT GAAAACGACC GAAATCTGGG CAATCTTGG CCGATGTTAA TTATGTATTG GAATACATAG TATGTGTATC TAAATCCACT
    nlaIII styI ncoI dsalI bsajI
    fnu4HI aciI thalI fnuDII/mvni tru9I bstUI mseI
    bsajI styI bsajI bsh1236I aseI/asnI/vspI
    maeIII hphI

```

28 / 81

FIG. 6E

```

      haeIII/palI
      eaeI
      cfrI
      mspI
      hpaII
      scrFI
      nciI
      ecoRI dsav
      taqI apoI caulI
      claiI/bsp106 bsaJI aluI
      apyI[dcM+] mnlI bsaJI
      gsuI/bpmI hincII/hindII
      maeIII
      foki
      scfI
      CACTATAGAA TAACATCCAC TTTGCCCTTTC TCTCCACAGG TGTCACCTCA GGTCAACTGC ACCTCGGTTT CATCGATTGA ATTCCCGGC CATAGCTGTC
      GTGATATCTT ATTGTAGGTG AACGGAAAG AGAGGTGTCC ACAGTGAGGT CCAGTTGACG TGGAGCCAAG ATAGCTAACT TAAGGGGCCG GTATCGACAG
      1501

      gsuI/bpmI
      scrFI
      mvaI
      ecoRII
      dsav
      bstNI
      apyI[dcM+] mnlI bsaJI
      gsuI/bpmI hincII/hindII
      maeIII
      foki
      scfI
      CACTATAGAA TAACATCCAC TTTGCCCTTTC TCTCCACAGG TGTCACCTCA GGTCAACTGC ACCTCGGTTT CATCGATTGA ATTCCCGGC CATAGCTGTC
      GTGATATCTT ATTGTAGGTG AACGGAAAG AGAGGTGTCC ACAGTGAGGT CCAGTTGACG TGGAGCCAAG ATAGCTAACT TAAGGGGCCG GTATCGACAG
      1501

      mnlI
      haeIII/palI
      sau96I
      asuI
      nlaIII
      TGGCATGGGC CTCTCCACCG TGCCTGACCT GCTGCTGCCG CTGGTGTCTCC TGGAGCTGTT GGTGGGAATA TACCCTCAG GGGTTATTGG ACTGGTCCCT
      ACCGTACCCG GAGAGGTGGC ACGGACTGGA CGACGACGGC GACCACGAGG ACCTCGACAA CCACCCTTAT ATGGGGAGTC CCCAATAACC TGACCAGGGA
      1601

      xmaI
      maeI
      styI
      bsaJI
      bliI
      avrII
      mboII
      earI/ksp632I
      bsaJI
      styI
      bsaJI
      mnlI
      taqI
      rsaI
      csp6I
      nlaIV
      1701 CACCTAGGGG ACAGGGAGAA GAGAGATAGT GTGTGTCCCC AAGGAATAA TATCCACCCT CAAAATAATT CGATTTCCTG TACCAAGTGC CACAAAGGAA
      GTGGATCCCC TGTCCCTCTT CTCTCTATCA CACACAGGGG TTCCCTTTTAT ATAGGTGGGA GTTTTATTAA GCTAAACGAC ATGGTTCAGG GTGTTTCCTT

```

29 / 81

FIG. 6F

```

scrFI      nciI      mspI      hpaII      dsav      cauII
xmaI/pspAI
smaI      scrFI      nciI      dsav      cauII      bsII
sau96I      haeIII/palI
asuI
scrFI      mvaI bsaJI      ecorII      dsav
rsal      bstNI bsaJI      scfI      pstI      bsgI      nlaIV      fnu4HI      aciI      hphI      eco57I      alwNI      ddeI      mnlI
csp6I      bspI407I      CCAGGCCCGG      GGCAGGATAC      GGACTGCAGG      GAGTGTGAGA      GCGGCTCCTT      CACCGCTTCA      GAAAACCCACC      TCAGACACTG
1801 CCTACTTGTA CAATGACTGT      GGTCCGGGCC      CCGTCCTATG      CCTGACGTCC      CTCACACTCT      CGCCGAGGAA      GTGGCGAAGT      CTTTGTGTGG      AGTCTGTGAC
GGATGAACAT GTTACTGACA

fnu4HI      aluI      pvuII      nspBII      ddeI      mnlI      bbvI
1901 CCTCAGCTGC TCCAAATGCC      GAAAGGAAT      GGGTCAGGTG      GAGATCTCTT      CTTGCACAGT      GGACCCGGGAC      ACCGTGTGTG      GCTGCAGGAA      GAACCACTAC
GGAGTCGACG      AGGTTTACGG      CTTTCCCTTTA      CCCAGTCCAC      CTCATAGAGAA      GAACGTGTCA      CTTGGCCCTG      TGGCACACAC      CGACGTCTT      CTTGTCTATG

```

hgiAI/asphI
bsp1286
bsiHKA I
bmy I
pm I

alut

bsp1286

[illegible]

31 / 81

FIG. 6H

```

          eam1105I
          sau96I
          scrFI
          mvaI    avaII
          ecoRII
          dsaV
          bstNI   asuI      mboII mboII
          bsaJI   mnlI      bpuAI earI/ksp632I    styI
          apyI[dcM+] bbsI mnlI    bsaJI
          2401 CTTGTGACAC ACCTCCCCCA TGCCACCGGT GCCCAGCACC TGAACCTCTG GGAGGACCGT CAGTCTTCCT CTCCCCCCCC AAACCCCAAGG ATACCCCTTAT
              GAACACTGTG TGGAGGGGGT ACGGGTGCCA CGGGTCGTGG ACTTGAGGAC CCTCCTGGCA GTCAGAAGGA GAAGGGGGGT TTTGGGTTCC TATGGGAATA

          sau96I
          nlaIV
          avaII
          asuI
          mspI      maeII
          hpaII     pmlI
          scrFI     mnlI bsaAI
          nciI      ddeI maeIII
          dsaV      eco8II bbrPI
          cauII     bsu36I/mstII/sauI      maeII
          2501 GATTTCCTCGG ACCCCTGAGG TCACGTGCGT GGTGGTGGAC GTAGGCACG AAGACCCCGA GTCCAGTTC AAGTGGTACG TGGACGGCGT GGAGGTGCAT
              CTAAAGGGCC TGGGACTCC AGTGACGCA CCACCACCTG CACTCGGTGC TTCTGGGGCT CCAGGTCAAG TTCACCATGC ACCTGCCGCA CCTCCACGTA

          acil
          thal
          fnuDII/mvnI
          bstUI
          bsh1236I
          sacII/sstII
          nspBII
          kspI
          dsaI
          bsaJI
          acil
          fnu4HI mnlI      maeII
          2601 AATGCCAAGA CAAAGCCCGG GGAGGAGCAG TTCAACAGCA CGTTCCTGT GGTACCGTC CTCACCGTCC TGCACAGGA CTGGCTGAAC GGCAAGGAGT
              TTACGGTTCT GTTCGGCGC CCTCCTCGTC AAGTTGTCGT GCAAGGCACA CCAGTCGCAG GAGTGGCAGG ACGTGGTCT GACCGACTTG CCGTTCCTCA
          scrFI
          mvaI bsrI
          ecoRII
          dsaV
          mnlI   ecoNI   bstNI
          hgaI   hphI   bslI   apyI[dcM+]
          rsaI   csp6I
  
```

FIG. 61

FIG. 6I

2701	ACAAAGTGCAA	GGTCTCCAAC	AAAGCCCTCC	CAGCCCCCAT	CGAGAAACC	ATCTCCAAA	CCAAAGGACA	CCCCCGAGAA	CCACAGGTGT	ACACCTGCC	bslI
	TGTTACAGTT	CCAGAGGTTG	TTTCGGGAGG	GTCGGGGGTA	GCTCTTTTGG	TAGAGGTTTT	GGTTTCTGT	CGGGGCTCTT	GGTGTCACA	TGTGGACGG	
	bsmAI	bsaI	mnII	taqI	avaI						rsal
	scriFI	nciI	mspI	hpaII	dsav	cauII	xnaI/pspAI	smaI	scriFI	nciI	mspI
	bsaI	bsaI	mnII	taqI	avaI						bspl407I
	scriFI	nciI	mspI	hpaII	dsav	cauII	xnaI/pspAI	smaI	scriFI	nciI	mspI
	bsaI	bsaI	mnII	taqI	avaI						bspl407I
	scriFI	nciI	mspI	hpaII	dsav	cauII	xnaI/pspAI	smaI	scriFI	nciI	mspI
	bsaI	bsaI	mnII	taqI	avaI						bspl407I
	scriFI	nciI	mspI	hpaII	dsav	cauII	xnaI/pspAI	smaI	scriFI	nciI	mspI
	bsaI	bsaI	mnII	taqI	avaI						bspl407I
	scriFI	nciI	mspI	hpaII	dsav	cauII	xnaI/pspAI	smaI	scriFI	nciI	mspI
	bsaI	bsaI	mnII	taqI	avaI						bspl407I
	scriFI	nciI	mspI	hpaII	dsav	cauII	xnaI/pspAI	smaI	scriFI	nciI	mspI
	bsaI	bsaI	mnII	taqI	avaI						bspl407I
	scriFI	nciI	mspI	hpaII	dsav	cauII	xnaI/pspAI	smaI	scriFI	nciI	mspI
	bsaI	bsaI	mnII	taqI	avaI						bspl407I
	scriFI	nciI	mspI	hpaII	dsav	cauII	xnaI/pspAI	smaI	scriFI	nciI	mspI
	bsaI	bsaI	mnII	taqI	avaI						bspl407I
	scriFI	nciI	mspI	hpaII	dsav	cauII	xnaI/pspAI	smaI	scriFI	nciI	mspI
	bsaI	bsaI	mnII	taqI	avaI						bspl407I
	scriFI	nciI	mspI	hpaII	dsav	cauII	xnaI/pspAI	smaI	scriFI	nciI	mspI
	bsaI	bsaI	mnII	taqI	avaI						bspl407I
	scriFI	nciI	mspI	hpaII	dsav	cauII	xnaI/pspAI	smaI	scriFI	nciI	mspI
	bsaI	bsaI	mnII	taqI	avaI						bspl407I
	scriFI	nciI	mspI	hpaII	dsav	cauII	xnaI/pspAI	smaI	scriFI	nciI	mspI
	bsaI	bsaI	mnII	taqI	avaI						bspl407I
	scriFI	nciI	mspI	hpaII	dsav	cauII	xnaI/pspAI	smaI	scriFI	nciI	mspI
	bsaI	bsaI	mnII	taqI	avaI						bspl407I
	scriFI	nciI	mspI	hpaII	dsav	cauII	xnaI/pspAI	smaI	scriFI	nciI	mspI
	bsaI	bsaI	mnII	taqI	avaI						bspl407I
	scriFI	nciI	mspI	hpaII	dsav	cauII	xnaI/pspAI	smaI	scriFI	nciI	mspI
	bsaI	bsaI	mnII	taqI	avaI						bspl407I
	scriFI	nciI	mspI	hpaII	dsav	cauII	xnaI/pspAI	smaI	scriFI	nciI	mspI
	bsaI	bsaI	mnII	taqI	avaI						bspl407I
	scriFI	nciI	mspI	hpaII	dsav	cauII	xnaI/pspAI	smaI	scriFI	nciI	mspI
	bsaI	bsaI	mnII	taqI	avaI						bspl407I
	scriFI	nciI	mspI	hpaII	dsav	cauII	xnaI/pspAI	smaI	scriFI	nciI	mspI
	bsaI	bsaI	mnII	taqI	avaI						bspl407I
	scriFI	nciI	mspI	hpaII	dsav	cauII	xnaI/pspAI	smaI	scriFI</		

33 / 81

FIG. 6J

FIG. 6J

```

rmaI
mnlI
sau3AI
mboI/ndeII[dam-]
dpmI[dam+]
dpmII[dam-] taqI
alwI[dam-] salI scfI
nlaIV maeI hincII/hindII
bstYI/xhoII accI pstI
bamHI xbaI pleI bsgI
alwI[dam-] hinfI bspMI
GGGGATCCTC TAGAGTCGAC CTGCAGAAGC TGGCCGCCA TGGCCCAACT TGTATTATGC AGCTTATAAT GGTACAAAT AAAGCAATAG CATCACAAAT
CCCCTAGGAG ATCTCAGCTG GACGTCTTCG AACCGGCGGT ACCGGGTTGA ACAAATAACG TCGAATATTA CCAATGTTTA TTTCGTATC GTAGTGTTTA

rmaI
bsmI maeI
rmaI
TTCACAAATA AAGCATTTT TTCACTGCAT TCTAGTTGTG GTTTGTCCAA ACTCATCAAT GTATCTTATC ATGTCTGGAT CGATCGGGAA TTAATTGGC
AAGTGTTTAT TTCGTAAAAA AAGTGACGTA AGATCAACAC CAAACAGGTT TGAGTAGTTA CATAGAAATAG TACAGACCTA GCTAGCCCTT AATTAAAGCCG

haeI
styI
ncol
dsal haeIII/pali
fnu4HI nlaIII
bbvI bsaJI
GCAGCACCTT GGCCTGAAAT AACCTCTGAA AGAGGAACCTT GGTAGGTAC CTTCGTAGGC GGAAGAACC AGCTGTGGAA TGTGTGTGAG TTAGGTGTG
CGTCGTGGTA CCGGACTTTA TTGGAGACTT TCTCTTGAA CCAATCCATG GAAGACTCCG CCTTCTTGG TCGACACCTT ACACACAGTC AATCCCACAC
3301

```

34 / 81

[illegible]

[illegible]

FIG. 6M

[illegible]

FIG. 6N

FIG. 6N

```

nlaIV
aciI
thai
fnuDII/mvnI
bstUI
bsh1236I
hinPI
hhaI/cfoI
hinII/acyI
ahaII/bsaHI
aatII
ddei maeII
4701 AATAATGGTT TCTTAGACGT CAGGTGGCAC TTTTCGGGGA AATGTGCGC GAACCCCTAT TTGTTTATTT TTCTAAATAC ATTCAAATAT GTATCCGCTC
TTATTACCAA AGAATCTGCA GTCCACCGTG AAAAGCCCTT TTACACGCGC CTTGGGGATA AACAAATAAA AAGATTATG TAAGTTTATA CATAGGCGAG
rcaI
bspHI
bsrBI
aciI nlaIII
4801 ATGAGACAAT AACCTTGATA AATGCTTCAA TAATATTGAA AAAGGAAGAG TATGAGTAT CAACATTTCC GTGTGCGCCT TATTCCCTTT TTGCGGCGAT
TACTCTGTTA TTGGGACTAT TTACGAAGTT ATTATAACTT TTTCTTCTC ATACTCATAA GTTGTAAGG CACAGCGGA ATAAGGAAA AAACGCCGTA
fnu4HI
aciI
bsmAI
sspI
earI/ksp632I
4801 ATGAGACAAT AACCTTGATA AATGCTTCAA TAATATTGAA AAAGGAAGAG TATGAGTAT CAACATTTCC GTGTGCGCCT TATTCCCTTT TTGCGGCGAT
TACTCTGTTA TTGGGACTAT TTACGAAGTT ATTATAACTT TTTCTTCTC ATACTCATAA GTTGTAAGG CACAGCGGA ATAAGGAAA AAACGCCGTA
hgiAI/aspHI
bsp1286
sau3AI
bsiHKAi
mboI/ndeII[dam-]
dpmI[dam+] bmyI
dpmII[dam-]
eco57I
apaLI/snoI
bsrI dpmII[dam-]
4901 TTGCGCTTCC TGTTTTTGCT CACCCAGAAA CGCTGGTGAA AGTAAAGAT GCTGAAGATC AGTTGGGTGC ACAGTGGGT TACATCGAAC TGGATCTCAA
AAACGGAAGG ACAAAAACGA GTGGTCTTT GCGACCACTT TCATTTTCTA CGACTTCTAG TCAACCCACG TGCTCACCCA ATGTAGCTTG ACCTAGAGTT
hphI
hphI
sfaNI mboII[dam-] alw44I/snoI maeIII taqI alwI[dam-]
5001 TTGCGCTTCC TGTTTTTGCT CACCCAGAAA CGCTGGTGAA AGTAAAGAT GCTGAAGATC AGTTGGGTGC ACAGTGGGT TACATCGAAC TGGATCTCAA
AAACGGAAGG ACAAAAACGA GTGGTCTTT GCGACCACTT TCATTTTCTA CGACTTCTAG TCAACCCACG TGCTCACCCA ATGTAGCTTG ACCTAGAGTT
sau3AI
maeII
psp1406I
xmnI
asp700
mboII
hphI
bstYI/xhoII
nsbII
bstYI/xhoII
5001 CAGCGGTAAG ATCCTTGAGA GTTTTCGCCC CGAAGAACGT TTTCCTGAGA GTTTTCGCCC CGAAGAACGT TTTCCTGAGA GTTTTCGCCC CGAAGAACGT
GTGCGCATTC TAGGAACTCT CAAAAGCGGG GCTTCTTGCA AAAGGTACT ACTCGTGAAA ATTTCAAGAC GATACACCGC GCCATAATAG GGCACCTACTG
aciI
thai
fnuDII/mvnI
bstUI
bsh1236I
hinPI
hhaI/cfoI
hinII/acyI
hgiI
ahaII/bsaHI
5101 GCCGGCAAG AGCAACTCGG TCGCCGCATA CACTATTCTC AGAATGACTT GGTTGAGTAC TCACCAAGTCA CAGAAGACA TCTTACGGAT GGCATGACAG
CGGCCGTTT TCGTTGAGCC AGCGCGGTAT GTGATAAGAG TCTTACTGAA CCAACTCATG AGTGGTCACT GTCTTTTCGT AGAATGCCTA CCGTACTGTC
rsal
csp6I
bsrI
scaI hphI maeIII
sfanI
foki
nlaIII
5101 GCCGGCAAG AGCAACTCGG TCGCCGCATA CACTATTCTC AGAATGACTT GGTTGAGTAC TCACCAAGTCA CAGAAGACA TCTTACGGAT GGCATGACAG
CGGCCGTTT TCGTTGAGCC AGCGCGGTAT GTGATAAGAG TCTTACTGAA CCAACTCATG AGTGGTCACT GTCTTTTCGT AGAATGCCTA CCGTACTGTC

```

SUBSTITUTE SHEET (RULE 26)

40 / 81

FIG. 6Q

```

        scrFI      aciI      nspBII      fnu4HI      bbvI      mspI      hpaII      bsaWI      maeIII      hinPI      mcrl      hhaI/cfoI
        nciI      mspI      hpaII      dsav      cauII      pleI      hinFI      hhaI/cfoI      hhaI/cfoI      haeII
        fnu4HI      bbvI      bsrI      fnu4HI      bbvI      alwNI      bsrI      fnu4HI      bbvI      mspI      hpaII      bsaWI      maeIII      hinPI      mcrl      hhaI/cfoI
        alwNI      bsrI      fnu4HI      bbvI      maeIII      ACCAGTGGCT      GCTGCCAGTG      GCGATAAGTC      GTGTCTTACC      GGGTTGGAAT      CAAGACGATA      GTTACCGGAT      AAGCGGCAGC      GGTGGGGCTG
        ATTAGGACAA      TGGTCACCGA      CGACGGTCAC      CGCTATTTCAG      CACAGAATGG      CCAACCTGA      GTTCTGCTAT      CAATGGCCTA      TTCCGGCTCG      CCAGCCCGAC

        hgiAI/aspHI
        bsp1286
        bsiHKAI
        bmyI
        apaLI/snoI
        alw44I/snoI      aluI
        6201 AACGGGGGT      TCGTGCACAC      AGCCAGCTT      GGAGCGAACG      ACCTACACCG      AACTGAGATA      CCTACAGCGT      GAGCATTGAG      AAAGCGCCAC      GCTTCCCGAA
        TTGCCCCCA      AGCAGGTGTG      TCGGTCGAA      CCTCGCTTGC      TGGATGTGGC      TTGACTCTAT      GGATGTGCA      CTCGTAACTC      TTTCGGGTG      CGAAGGGCTT

        scrFI      mval      ecorII      dsav      bstNI      bsaJI      aluI      apyI[dcn+]      apyI[dcn+]
        mval      scrFI      ecorII      mval      ecorII      dsav      bstNI      bsaJI      aluI      apyI[dcn+]      apyI[dcn+]
        hinPI      mnII      hhaI/cfoI      GCGACGAGG      GCGACGAGG      GCGACGAGG      GCGACGAGG      GCGACGAGG      GCGACGAGG      GCGACGAGG      GCGACGAGG      GCGACGAGG      GCGACGAGG      GCGACGAGG      GCGACGAGG
        6301 GGGAGAAAGG      CGGACAGGTA      TCCGGTAAGC      GGCAGGGTCG      GAACAGGAGA      GCGACGAGG      GCGACGAGG      GCGACGAGG      GCGACGAGG      GCGACGAGG      GCGACGAGG      GCGACGAGG      GCGACGAGG      GCGACGAGG      GCGACGAGG
        CCTCTTTCC      GCCTGTCCAT      AGGCCATTG      CCGTCCCAGC      CTTGTCTCT      GCGGTGCTCC      CTGGAAGTC      CCCCTTTGCG      GACCATAGAA      ATATCAGGAC

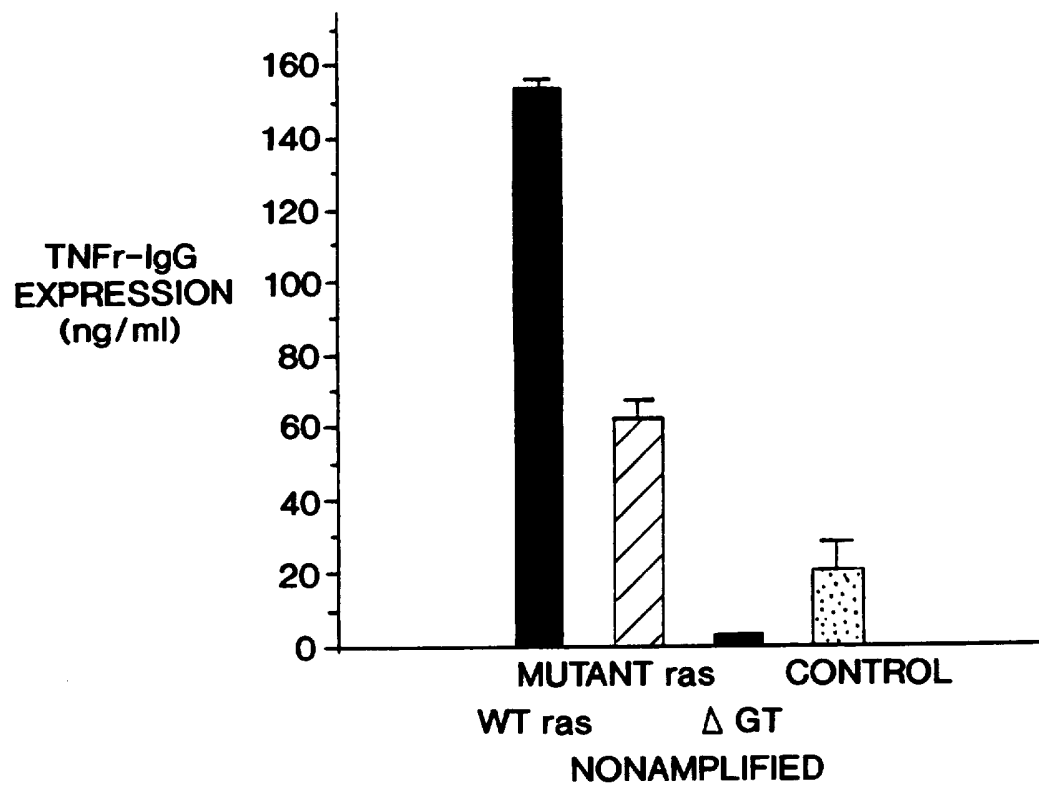
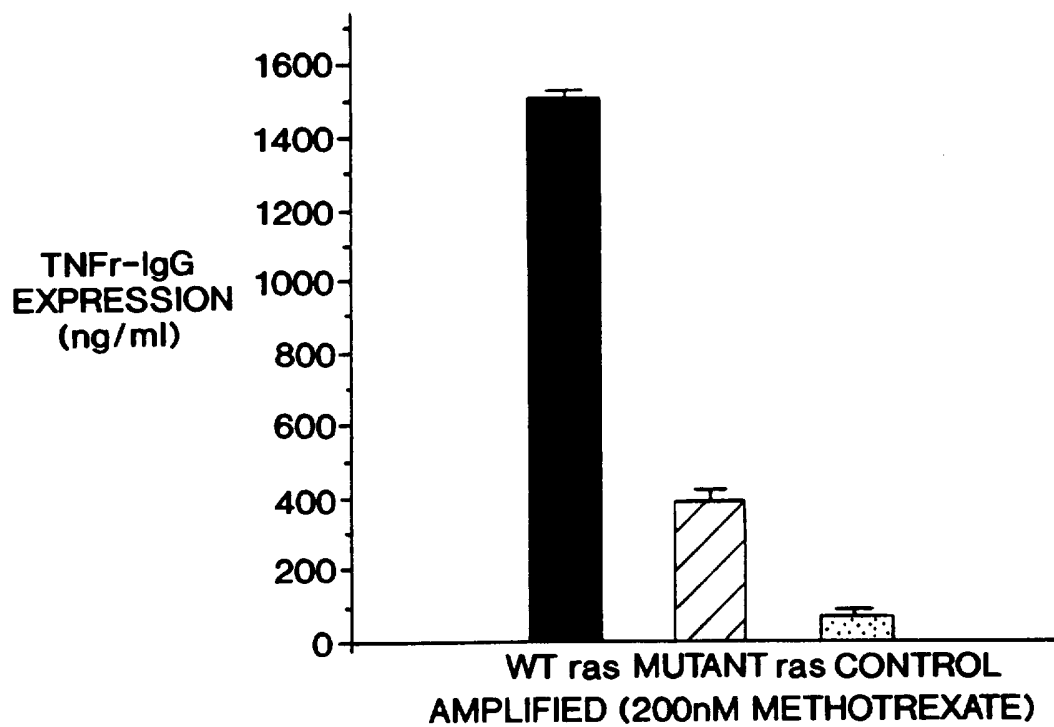
        haeIII/palI
        fnu4HI      aciI      thaI      bsII      fnuDII/mvnl      bstUI      bsh1236I      nlaIV
        nmlI      drdI      hgaI      taqI      sfaNI      aciI      nlaIV
        6401 TCGGGTTTCG      CCACCTCTGA      CTTGAGCGTC      GATTTTGTG      ATGCTCGTCA      GGGGGCGGA      GCCTATGGA      AAAGCGCCAGC      AACGCGGCT      TTTTACGGTT
        AGCCCCAAGC      GGTGGAGACT      GAACTCGCAG      CTAAAAACAC      TACGAGCAGT      CCGCCGCT      CCGGCGGTG      TTGCGCCGA      AAAATGCCAA

```

FIG. 6R

```
>length: 6889
```

42 / 81

**FIG. 7A****FIG. 7B**

43 / 81

FIG. 8

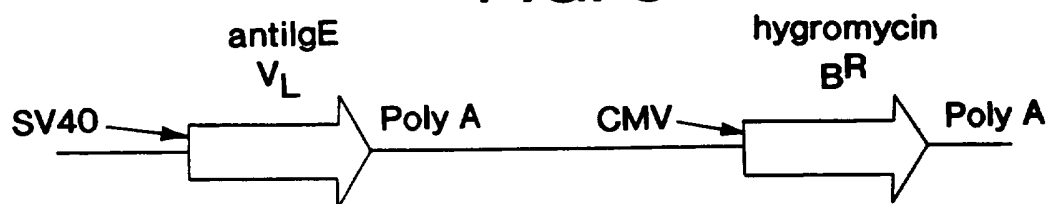


FIG. 11

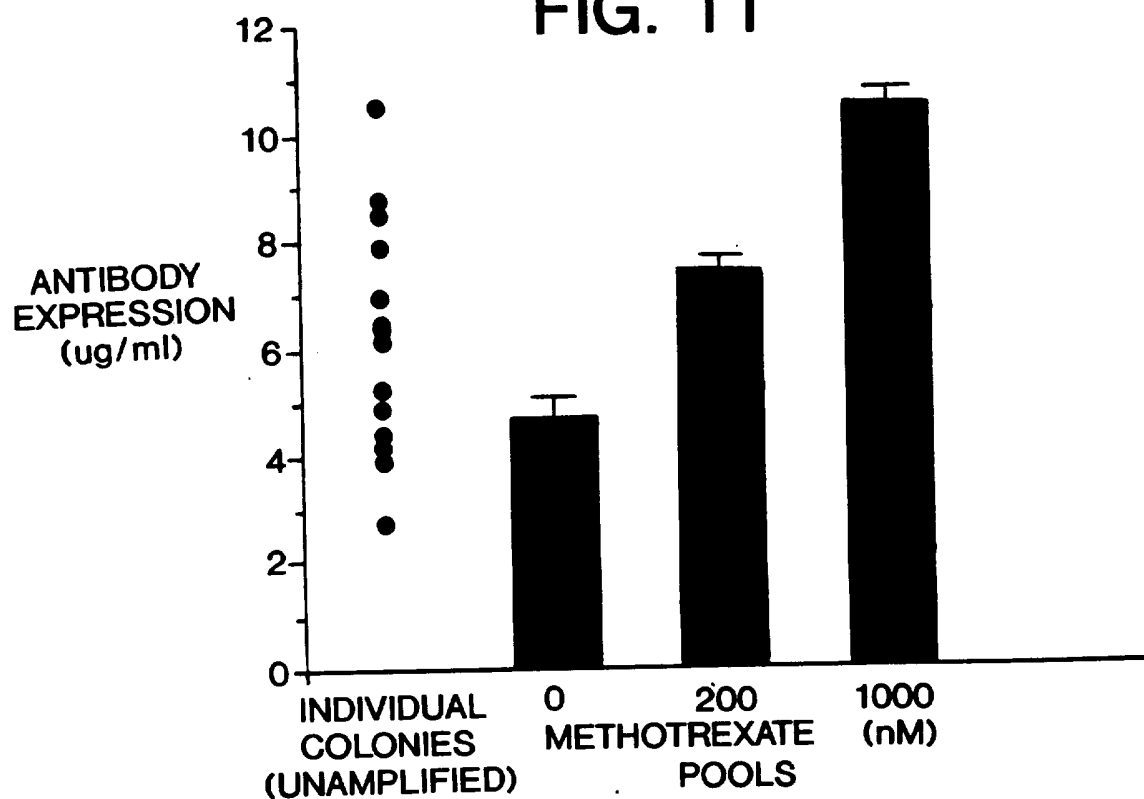


FIG. 9A

[illegible]

SUBSTITUTE SHEET (RULE 26)

FIG. 9C

[illegible]

47 / 81

FIG. 9D

```

sau96I      sau96I
avaII       avaII
asuI        asuI
scrFI       scrFI
mvaI        mvaI
ecorII      ecorII
dsav        dsav
bstNI       bstNI
            apyI[dcn+]
            mnlI      taqI      ecorI
            bsaJI     clai/bspl06
            bslI      bsaJI
1201 ATACGATTGA GGTGACACTA TAGATAACAT CCACCTTGCC TTTCTCTCCA CAGGTGTCCA CTCCCAGGTC CAACCTGCACC TCGGTTCTAT CGATTGAATT
TATGCTAAAT CCACCTGTGAT ATCTATTGTA GGTGAAACGG AAGAGAGAGT GTCCACAGGT GAGGTCTCCAG GTTGACGTGG AGCCAAGATA GCTAACTTAA
            maeIII
            hphI      scfI      foki
            bslI      bsaJI
            rsal      gsul/bpmI      pvuII      pleI      haeI      hinfI      acil      haeIII/palI
            bsaJI     nlaIII     foki      nspBII      hinfI      haeI      haeIII/palI
1301 CCACCATGGG ATGGTCATGT ATCATCCTTT TTCTAGTAGC AACTGCAACT GGAGTACATT CAGAAGTTCA GCTGGTGGAG TCTGGCGGTG GCCTGGTGCA
GGTGGTACCC TACCAGTACA TAGTAGGAAA AAGATCATCG TTGACGTTGA CCTCATGTAA GTCTTCAAGT CGACCACCTC AGACCGCCAC CGGACCACGT
            nlaIII
            styI
            pflMI
            ncoI
            dsal      bslI      foki
            bsaJI     nlaIII     foki      nmaI      maeI
            rsal      gsul/bpmI      pvuII      pleI      haeI      hinfI      acil      haeIII/palI
            bstNI     fnu4HI
            apyI[dcn+]
            scrFI
            mvaI
            ecorII
            dsav
            bstNI     fnu4HI
            apyI[dcn+]

```

48 / 81

FIG. 9E

FIG. 9E

SUBSTITUTE SHEET (RULE 26)

50 / 81

FIG. 9G

```

scrFI      hinPI      nlaIV
mvaI       nari       kasI
ecorII     hinII/acyI hgiCI
econI      haelI      bsp1286 hgiAI/aspHI
dsav       banI       bsiHKA I mspI
bstNI      cfr10I     hpaII scrFI
bsII       bsaWI tth111I/aspI hmyI nciI
          bslI ageI maeIII fnu4HI
          bsaII tth111I/aspI acII apaLI/snoI dsav
          bslI ageI maeIII ddeI hhaI/cfoI nspBII alw44I/snoI cauII scfI
1801 GGCTGCCTGG TCAAGGACTA CTTCCCCGAA CCGGTGACGG TGTCTGTGAA CTCAGGCGCC CTGACCAGCG GCGTGCACAC CTTCCTGGCT GTCTACAGT
CCGACGGACC AGTTCCTGAT GAAGGGGCTT GGCACACTGCC ACAGCACCTT GAGTCCCGGG GACTGGTGGC GCACGTGTG GAAGGGCCGA CAGGATGTCA

          fnu4HI      bsvI
          ddeI pleI
          mnlI hinfI
          eco8II
          bsu36I/mstII/sauI ddeI bsvI mnlI fnu4HI maeIII bsp1286 rmaI bsp1286
1901 CCTCAGGACT CTACTCCCTC AGCAGCGTGG TGACTGTGCC CTCTAGCAGC TTGGGGACCC AGACCTACAT CTGCAACGTG AATCACAAGC CCAGCAACAC
GGAGTCCTGA GATGAGGGAG TCGTCGCACC ACTGACACGG GAGATCGTGG AACCCGTGGG TCTGGATGTA GAGGTGTCAC TTAGTGTTCG GGTCTGTGTC

          nlaIV      hgiCI
          fnu4HI      bgiCI
          mnlI      bsvI      banI
          mnlI fnu4HI maeIII bsp1286 rmaI bsp1286
          mnlI maeI aluI bmyI
          maeI
          maeII      hinfI
          styI      bsaJI
          eam1105I
          sau96I
          scrFI
          mvaI      avaII
          ecoRII
          dsav
          bstNI      asuI      mboII mboII
          bsaJI      nlaIV      bpuAI earI/ksp632I
          apyI[dcn+]      bbsI mnlI
          apyI[dcn+]      GAACTCCTGG GGGGACCGTC AGTCTTCCTC
          CTGAGGACC CCCCTGGCAG CCGCTGGCAG TCAGAAGGAG
          bsp1286
          hgiJII
          bsp1286
          bmyI
          banII
          maeIII
          nspI
          nspHI
          ACTGACAAA ACTGACAAA ACTGACAAA ACTGACAAA ACTGACAAA ACTGACAAA ACTGACAAA ACTGACAAA ACTGACAAA ACTGACAAA
          AAGAAAGTTG AGCCCAAATC TTGTGACAAA TTGTGACAAA TTGTGACAAA TTGTGACAAA TTGTGACAAA TTGTGACAAA TTGTGACAAA TTGTGACAAA
          TTCTTTCAAC TCGGGTTTAG AACACTGTTT TGAGTGTGTA CCGGIGGCAC GGGTGTGGA GGGTGTGGA GGGTGTGGA GGGTGTGGA GGGTGTGGA
          GTTECACCTG TTCTTTCAAC TCGGGTTTAG AACACTGTTT TGAGTGTGTA CCGGIGGCAC GGGTGTGGA GGGTGTGGA GGGTGTGGA GGGTGTGGA GGGTGTGGA

```

51 / 81

FIG. 9H

```

sau96I
nlaIV
avaII
mspI
scrFI
nciI
sau3AI hpaII
mboI/ndeII(dam-)
dpmI(dam+)
nlaIII dsav
nlaIII mnlI nspI
rcal caulI ddeI maeIII
bspHI(dam-) asuI eco8II nspHI
mnlI dpmII(dam-) bsu36I/mstII/sauI maeII
styI
bsaJI
2101 TTCCCCCAA AACCCAAGGA CACCTCATG ATCTCCCGGA CCCCTGAGT CACATGCGTG GTGGTGGACG TGAGCCACGA AGACCTGAG GTCAAGTTCA
AAGGGGGTT TTGGGTTCTT GTGGGAGTAC TAGAGGGCCT GGGGACTCCA GTGTACGCAC CACCACCTGC ACTCGGTGCT TCTGGGACTC CAGTTCAAGT

aciI
thai
fnuDI/mvnI
bstUI
bsh1236I
sacII/sttII
nsp8II
kspI
dsaI
bsaJI
maeII
rsaI
csp6I
mnlI
fnu4HI mnlI csp6I bsaAI
rsaI maeII mnlI hgaI hphI bslI
2201 ACTGGTACGT GGACGGCGTG GAGGTGCATA ATGCCAAGAC AAAGCCGCGG GAGGAGCAGT ACAACAGCAC GTACCGTGTG GTACCGTCC TCACCGTCT'
TGACCATGCA CCTGCCGCAC CTCCACGTAT TACGGTTCTG TTTCGGCGCC CTCCTCGTCA TGTGTGCTG CATGGCACAC CAGTGGCAGG AGTGGCAGGA

scrFI
mval bsrI
ecoRII
dsav
bstNI
apyI(dcm+)
2301 GCACCAAGGAC TGGCTGAATG GCACGAGTA CAAGTGCAAG GTCTCCAACA AGCCCTCCC AGCCCCATC GAGAAACCA TCTCCAAGC CAAAGGGCAG
CGTGGTCTTG ACCGACTTAC CGTTCCTCAT GTTCACGTC CAGAGTTGT TTGGGAGGG TCGGGGTAG CTCTTTTGGT AGAGTTTGG GTTTCCTGTC
fnu4HI
bbvI

```

52 / 81

FIG. 9I

```

scrFI      nciI      mspI      hpaII      dsav      caulI      xmaI/pspAI
smalI      scrFI      mvaI      ecorII      dsav      bstNI      apyI[dcn+]
nciI      dsav      cauII      foki      mboII      bsli      bsajI
rsal      csp6I      bsp1407I      avaI      earI/ksp632I      sexAI      bspMI      apyI[dcn+]
avaI      CCCCAGAAC      CACAGGTGTA      CACCTGCCC      CCATCCGGG      AAGAGATGAC      CAAGAACCAG      GTCAGCCTGA      CCTGCTGGT      CAAAGGCTT      CTTCCAGCG
GGGCTCTTG      GTGTCCACAT      GTGGGACGG      GGATAGGCC      TTCTCTACTG      GTTCTTGGTC      CAGTCGGACT      GGACGGACCA      GTTCCGAAG      ATAGGTCGC

2401      dsal      bsli      bsajI      mspI      fnu4HI      bbvI      hpaII      mnlI      nlaIV      mboII      scfI      aluI
ACATGCCGT      GGAGTGGGAG      AGCAATGGC      AGCCGGAGAA      CAACTACAAG      ACCAGCCTC      CCGTCTGGA      CTCGACGGC      TCCTCTTCC      TCTACAGCAA
TGATAGCGCA      CCTCACCTC      TCCTTACCG      TCGGCTCTT      GTTGATGTT      TGTGCGGAG      GGCACGACCT      GAGGCTGGC      AGGAAGAAG      AGATGCTGT

2501      dsal      bsajI      hphI      bspMI      bspMI      fnu4HI      bbvI      asp700      nlaIII      sfaNI      mnlI      pleI      hinFI      nlaIV      mboII      scfI      aluI
GCTCACCGTG      GACAAGAGCA      GGTGGCAGCA      GGGGAACGT      TTCTCATGT      CCGTGATGCA      TGAGGCTCTG      CACAACCACT      ACACGACAA      GAGCCTCTCC
CGAGTGGCAC      CTGTTCTCGT      CCACCGTCGT      CCCCTTGCAG      AAGAGTACGA      GGCACCTACGT      ACTCCGAGAC      GTGTTGGTGA      TGTGCTCTT      CTCGGAGAGG

2601      scrFI      nciI      mspI      hpaII      dsav      cauII      bsmAI      styI      acil      fnu4HI      sau96I      sfiI      ncoI      haeIII/palI
taqI      salI      rmaI      hincII/hindII      maeI      accI      scfI      haeIII/palI      eaeI      dsal      asuI      aluI      fnu4HI      bbvI      maeIII
GTAATGAGT      GCGACGGCC      TAGAGTCGAC      CTGCAGAGC      TTGGCCGCC      TGCCCCAACT      TGTATTATGC      AGCTTATAAT      GGTACAAAT
GACAGAGCC      CATTTACTCA      CGCTGCCGG      ATCTCAGCTG      GAGTCTTTC      AACCGCGGT      ACCGGTTCG      ACAAATAACG      TCGAATATTA      CCAATGTTTA

```

FIG. 9J

[illegible]

FIG. 9K

[illegible]

FIG. 9L

[illegible]

FIG. 9M

FIG. 9M

4101	ACAACTGCT	CTGATGCGC	ATAGTAAAG	CAACTCCGT	ACII	bsaAI	tthlII	aspi	bbvI	aciI	hhaI/cfoI	hnu4HI	maeIII	hinPI	hnuDII/mvnI	bstUI	nspBII	bsh1236I	thai
4201	CTGACGGGCT	TGCTGCTCC	CGGCATCCG	TTACAGACAA	GCTGTGACCG	alul	maeIII	cauII	bbvI	nlaIII	dsav	alul	nspHI	bsaAI	nspI	bsmAI	scrFI	nciI	thai
4301	GGGAGGCAGT	ATTCTTTGAAG	ACGAAAGGC	CTCGTGATAC	GCTTTTCCCG	GAGCACTATG	CGGATAAAA	TATCCAATTA	CAGTACTATT	ATTACCAAG	AATCTGCAGT	CCACCGTGAA	maeII	hinII/acyI	ahaII/bsaHI	ddeI	aatII	mnII	hphI
4401	TTGCGGGAAA	TGTGCGCGGA	ACCCCTATT	GTTTATTTT	CTAAATACAT	TCAATATGT	ATCCGTCAT	GAGACAATA	CCCTGATAA	TGCTTCAATA	sspi	rcal	bspHI	bsrBI	bsmAI	aciI	nlaIII	hnu4HI	hphI
4501	ATATTGAAA	AGGAAGACTA	TGAGTATCA	ACATTTCCGT	GTCGCCCTTA	TTCCCTTTT	TGCGGCAAT	TGCCCTCCCTG	TTTTTGCTCA	CCGAGAAACG	hnuDII/mvnI	bstUI	nspBII	bsh1236I	thai	hnuDII/mvnI	bstUI	nspBII	bsh1236I
4601	TATAACTTTT	TCCTTCTCAT	ACTCATAAGT	TGTAAGGCA	CAGCGGGAAT	AAGGAAAAA	ACGCCGTA	CTCTGTTAT	GGGACTATT	ACGAAGTAT	hnuDII/mvnI	bstUI	nspBII	bsh1236I	thai	hnuDII/mvnI	bstUI	nspBII	bsh1236I
4701	GGGAGGCAGT	ATTCTTTGAAG	ACGAAAGGC	CTCGTGATAC	GCTTTTCCCG	GAGCACTATG	CGGATAAAA	TATCCAATTA	CAGTACTATT	ATTACCAAG	AATCTGCAGT	CCACCGTGAA	maeII	hinII/acyI	ahaII/bsaHI	ddeI	aatII	mnII	hphI
4801	GGGAGGCAGT	ATTCTTTGAAG	ACGAAAGGC	CTCGTGATAC	GCTTTTCCCG	GAGCACTATG	CGGATAAAA	TATCCAATTA	CAGTACTATT	ATTACCAAG	AATCTGCAGT	CCACCGTGAA	maeII	hinII/acyI	ahaII/bsaHI	ddeI	aatII	mnII	hphI
4901	GGGAGGCAGT	ATTCTTTGAAG	ACGAAAGGC	CTCGTGATAC	GCTTTTCCCG	GAGCACTATG	CGGATAAAA	TATCCAATTA	CAGTACTATT	ATTACCAAG	AATCTGCAGT	CCACCGTGAA	maeII	hinII/acyI	ahaII/bsaHI	ddeI	aatII	mnII	hphI
5001	GGGAGGCAGT	ATTCTTTGAAG	ACGAAAGGC	CTCGTGATAC	GCTTTTCCCG	GAGCACTATG	CGGATAAAA	TATCCAATTA	CAGTACTATT	ATTACCAAG	AATCTGCAGT	CCACCGTGAA	maeII	hinII/acyI	ahaII/bsaHI	ddeI	aatII	mnII	hphI
5101	GGGAGGCAGT	ATTCTTTGAAG	ACGAAAGGC	CTCGTGATAC	GCTTTTCCCG	GAGCACTATG	CGGATAAAA	TATCCAATTA	CAGTACTATT	ATTACCAAG	AATCTGCAGT	CCACCGTGAA	maeII	hinII/acyI	ahaII/bsaHI	ddeI	aatII	mnII	hphI
5201	GGGAGGCAGT	ATTCTTTGAAG	ACGAAAGGC	CTCGTGATAC	GCTTTTCCCG	GAGCACTATG	CGGATAAAA	TATCCAATTA	CAGTACTATT	ATTACCAAG	AATCTGCAGT	CCACCGTGAA	maeII	hinII/acyI	ahaII/bsaHI	ddeI	aatII	mnII	hphI
5301	GGGAGGCAGT	ATTCTTTGAAG	ACGAAAGGC	CTCGTGATAC	GCTTTTCCCG	GAGCACTATG	CGGATAAAA	TATCCAATTA	CAGTACTATT	ATTACCAAG	AATCTGCAGT	CCACCGTGAA	maeII	hinII/acyI	ahaII/bsaHI	ddeI	aatII	mnII	hphI
5401	GGGAGGCAGT	ATTCTTTGAAG	ACGAAAGGC	CTCGTGATAC	GCTTTTCCCG	GAGCACTATG	CGGATAAAA	TATCCAATTA	CAGTACTATT	ATTACCAAG	AATCTGCAGT	CCACCGTGAA	maeII	hinII/acyI	ahaII/bsaHI	ddeI	aatII	mnII	hphI
5501	GGGAGGCAGT	ATTCTTTGAAG	ACGAAAGGC	CTCGTGATAC	GCTTTTCCCG	GAGCACTATG	CGGATAAAA	TATCCAATTA	CAGTACTATT	ATTACCAAG	AATCTGCAGT	CCACCGTGAA	maeII	hinII/acyI	ahaII/bsaHI	ddeI	aatII	mnII	hphI
5601	GGGAGGCAGT	ATTCTTTGAAG	ACGAAAGGC	CTCGTGATAC	GCTTTTCCCG	GAGCACTATG	CGGATAAAA	TATCCAATTA	CAGTACTATT	ATTACCAAG	AATCTGCAGT	CCACCGTGAA	maeII	hinII/acyI	ahaII/bsaHI	ddeI	aatII	mnII	hphI
5701	GGGAGGCAGT	ATTCTTTGAAG	ACGAAAGGC	CTCGTGATAC	GCTTTTCCCG	GAGCACTATG	CGGATAAAA	TATCCAATTA	CAGTACTATT	ATTACCAAG	AATCTGCAGT	CCACCGTGAA	maeII	hinII/acyI	ahaII/bsaHI	ddeI	aatII	mnII	hphI
5801	GGGAGGCAGT	ATTCTTTGAAG	ACGAAAGGC	CTCGTGATAC	GCTTTTCCCG	GAGCACTATG	CGGATAAAA	TATCCAATTA	CAGTACTATT	ATTACCAAG	AATCTGCAGT	CCACCGTGAA	maeII	hinII/acyI	ahaII/bsaHI	ddeI	aatII	mnII	hphI
5901	GGGAGGCAGT	ATTCTTTGAAG	ACGAAAGGC	CTCGTGATAC	GCTTTTCCCG	GAGCACTATG	CGGATAAAA	TATCCAATTA	CAGTACTATT	ATTACCAAG	AATCTGCAGT	CCACCGTGAA	maeII	hinII/acyI	ahaII/bsaHI	ddeI	aatII	mnII	hphI
6001	GGGAGGCAGT	ATTCTTTGAAG	ACGAAAGGC	CTCGTGATAC	GCTTTTCCCG														

FIG. 9N

[illegible]

58 / 81

FIG. 90

```

mspi      hinpi      bsrI      msI
hpaII     mstI       avIII/fspI  tru9I
bsaWI     maeII hhaI/cfoI  mseI
nlaIV     aluI      fnu4HI  psp1406I
5001 ATCGTTGGG ACCGGAGCTG AATGAAGCCA TACCAACAGA CGAGCGTGAC ACCACGATGC CAGCAGCAAT GGCAACAACG TTGGCGAAAC TATTAACCTGG
TAGCAACCTT TGGCCTCGAC TTACTTCGGT ATGGTTTGGT GCTGCGACTG TGCTGCTACG GTCGTCGTTA CCGTTGTTGC AACGCGTTG ATAATTGACC

mspi      bglI
hpaII     sau96I
scrFI     haeIII/palI
aluI      foki      sau96I
rmaI      msel bsrI  avall  hinPI asuI mspi
maeI      aseI/asnI/vspI mnlI  hhaI/cfoI hpaII
5101 CGAACTACTT ACTCTAGCTT CCCGGCAACA ATTAATAGAC TGGATGGAGG CGGATAAAGT TGCAGGACCA CTTCTGCGCT CGGCCCTTCC GGCTGGCTGG
GCTTGATGAA TGAGATCGAA GGGCCGTTGT TAATTATCTG ACCTACCTCC GCCTATTCA ACCTCTGCT GAAGACGCGA GCCGGGAAGG CCGACCGACC

mspi      thal      haeIII/palI
hpaII     fnuDII/mvnI  sau96I
cfr10I    bstUI      nlaIV
nlaIV     hphI      fnu4HI  bsrI  asuI  mnlI
gsuI/bpmI bsmAI acII  bbvI bsrI  asuI
5201 TTTATTGCTG ATAAATCTGG AGCCGGTGAG CGTGGGTCTC GCGTATCAT TGCAGCACTG GGGCCAGATG GTAAGCCCTC CCGTATCGTA GTTATCTACA
AAATAACGAC TATTAGACC TCGGCCACTC GCACCCAGAG CGCCATAGTA ACCTCTGAC CCCGCTCTAC CATTCGGGAG GGCATAGCAT CAATAGATGT

dclI
sau3AI    nlaIV
mboI/ndeII[dam-] mnlI
dpmI[dam+] hgiCI      tru9I
dpmII[dam-] bsnI      msel      maeIII
5301 CGACGGGGAG TCAGGCAACT ATGGATGAAC GAAATAGACA GATCGCTGAG ATAGGTGCTT CACTGATTAA GCATTGGTAA CTGTCAGACC AAGTTTACTC
GCTGCCCTC AGTCCGTTGA TACCTACTTG CTTTATCTGT CTAGCGACTC TATCCACGGA GTGACTAATT CGTAACCAAT GACAGTCTGG TTCAAAATGAG

rmaI      sau3AI
sau3AI    hphI  mboI/ndeII[dam-]
mboI/ndeII[dam-]
dpmI[dam+] dpmI[dam+]
tru9I      dpmII[dam-] dpmII[dam-]
ahaIII/draI maeI  alwI[dam-]
tru9I      bstVI/xhoII bstVI/xhoII  nlaIII
mseI       msel  alwI[dam-] rcaI      maeII
ahaIII/draI msel msel mboII[dam-] bspHI      tru9I
5401 ATATATACTT TAGATTGATT TAAAACTTCA TTTTAAATTT AAAAGGATCT AGGTGAAGAT CCTTTTGTAT AATCTCATGA CCAAAATCCC TTAACGTGAG
TATATATGAA ATCTAACTAA ATTTTGAAGT AAAAATTAA TTTTCCTCTA TCCACTTCTA GGAAAACTA TTAGAGTACT GGTTTTAGGG AATTGCACCTC

```

59 / 81

FIG. 9P

```

sau3AI
mboI/ndeII[dam-]
dpmI[dam+] sau3AI          thaI
dpmII[dam-] mboI/ndeII[dam-]
bstYI/xhoII dpmI[dam+]      fnuDII/mvnI
sau3AI alwI[dam-] dpmII[dam-] bstUI
mboI/ndeII[dam-] alwI[dam-] bsh1236I
dpmI[dam+] mboII[dam-]      hinPI          fnu4HI
dpmII[dam-] bstYI/xhoII     hhaI/cfoI     bbvI
5501 TTTTGGTTCC ACTGAGCGTC AGACCCCGTA GAAAGATCTT TTTTCTCTGC GCGTAATCTG CTGCTTGCAA ACAAAAAAAC
AAAAGCAAGG TGACTCGCAG TCTGGGGCAT CTTTCTAGT TTCCTAGAAG AACTCTAGGA AAAAAAGACG CGCATTAGAC GACGAACGTT TGTTTTTTTG

sau3AI
mboI/ndeII[dam-]
dpmI[dam+]
dpmII[dam-]
alwI[dam-]
5601 CACCGTACC AGCGGTGGTT TGTGTCGGG ATCAAGAGCT ACCAATCTTT TTTCCGAAGG TAACTGGCTT CAGCAGACG CAGATACCAA ATACTGTCTT
GTGGCGATGG TCGCCACCAA ACAACGGCC TAGTTCTCGA TGGTTGAGAA AAAGGCTTCC ATTGACCGAA GTGCTCTCGC GTCTATGGTT TATGACAGGA

aciI      nspBII      haeIII/palI      haeI      haeIII      eco57I      hhaI/cfoI      hinPI
aciI      nspBII      haeI      haeI      haeIII      eco57I      hhaI/cfoI      hinPI
5701 TCTAGTGTAG CCGTAGTTAG GCCACCACTT CAAGAACTCT GTAGCACCGC CTACATACCT CGCTCTGCTA ATCCTGTTAC CAGTGGCTGC TGCCAGTGGC
AGATCACATC GGCATCAATC CGGTGGTGAA GTTCTTGAGA CATCGTGGCG GATGTATGGA GCGAGACGAT TAGGACAATG GTCACCGACG ACGGTCAACG

rmaI      maeI      haeIII/palI      haeI      haeIII      eco57I      hhaI/cfoI      hinPI
rmaI      maeI      haeI      haeI      haeIII      eco57I      hhaI/cfoI      hinPI
5801 GATAAGTCGT GTCTTACCGG GTTGACTCA AGACGATAGT TACCGATAA GCGCGACCGG TCGGGGTGAA CGGGGGGTTC GTCCACACAG CCCAGTGTGG
CTATTTCAGCA CAGAATGGCC CAACCTGAGT TCTGTCTATCA ATGCGCTATT CCGCGTCGCC AGCCCGACTT GCCCCCCAAG CACGTGTGTC GGTTCGAACC

```

FIG. 9Q

[illegible]

61 / 81

FIG. 9R

```

thai
fndDII/mvni
bstUI
bsh1236I
hinPI
hhaI/cfoI
thai
fndDII/mvni
bstUI
bsh1236I haeIII/palI
bsII eaeI tfil asel/asnI/vspi
mnII aciI cfrI hinfi msel nspBII
sapi hinPI
mboII hhaI/cfoI
earI/ksp632I
mnII aciI haeII
sapi hinPI
mboII hhaI/cfoI
earI/ksp632I
mnII aciI haeII
6301 AGTGAGCGAG GAAGCGGAG AGCGCCCAAT ACGCAACCG CCTCTCCCG CGCGTTGGCC GAATCATTAA TCCAGCTGGC ACGACAGGTT TCCCGACTGG
TCACTCGCTC CTTCCGCTTC TCGCGGGTTA TCGGTTTGGC GGAGAGGGGC GCGCAACCG CTAAGTAATT AGGTCGACCG TGCTGTCCAA AGGGCTGACC
scrFI
mval
ecorII
dsav
nlaIV bstNI
hgiCI apyI[dcn+]
banI bsaJI
GGCACCACCCAG GCTTTACACT TTATGCTTCC GGCTCGTATG TTGTGTGGAA
CCGTGGGGTC CGAAATGTGA AATACGAAGG CCGAGCATAC AACACACCTT
tru9I
mseI
asel/asnI/vspi
xmnl
asp700
nlaIII
aluI
6401 AAAGCGGGCA GTGAGCGCAA CGCAATTAAT GTGAGTTACC TCACTCATTG CACTCAATGG AGTGAGTAAT CCGTGGGGTC CGAAATGTGA AATACGAAGG CCGAGCATAC AACACACCTT
TTTCGCCCGT CACTCGCGTT GCGTTAATTA CACTCAATGG AGTGAGTAAT CCGTGGGGTC CGAAATGTGA AATACGAAGG CCGAGCATAC AACACACCTT
tru9I
mseI
asel/asnI/vspi
xmnl
asp700
nlaIII
aluI
6501 TTGTGAGCGG ATAACAATTT CACACAGGAA ACAGCTATGA CCATGATTAC GAATTAA
AACACTCGCC TATTGTTAAA GTGTGTCCTT TGTCGATACT GGTACTAATG CTTAATT

```

>length: 6557

62 / 81

FIG. 10A

```

aluI
sstI
sacI
hgiJII
hgiAI/aspHI
ecII136II
bsp1286
bsiHKA1
bmyI
banII
taqI
1 TTCGAGCTCG CCGGACATTG ATTATTGACT AGTTATTAAAT AGTAATCAAT TACGGGGTCA TTAGTTCATA GCCCATATAT GGAGTTCCGC GTTACATAAC
AAGCTCGAGC GGGCTGTAAC TAATAACTGA TCAATAAATTA TCATTAGTTA ATGCCCCAGT AATCAAGTAT CCGGTATATA CCTCAAGGCG CAATGTATTG
201 TTGACGTCAA TGGGTGGAGT ATTTACGGTA AACTGCCAC TGGCAGTAC ATCAAGTGTG TCATATGCCA AGTACGCCC CTATTGACGT CAATGACGGT
AACTGCAGTT ACCACCTCA TAAATGCCAT TTGACGGGTG AACCGTCATG TAGTTCACAT AGTATACGGT TCATGCGGGG GATAACTGCA GTTACTGCCA
301 AAATGGCCG CCTGGCATTG TGCCAGTAC ATGACCTTAT GGGACTTTCC TACTGGCAG TACATCTACG TATTAGTCAT CGCTATTACC ATGGTGATGC
TTTACCGGC GGACCGTAAT ACGGGTCATG TACTGGAATA CCCTGAAAGG ATGAACCGTC ATGTAGATGC ATAATCAGTA GCGATAATGG TACCACTACG

```

Restriction sites (indicated by vertical lines) and their corresponding sequences are listed below:

Enzyme	Sequence
aluI	AGCT
sstI	GAATTC
sacI	GAATTC
hgiJII	GAATTC
hgiAI/aspHI	GAATTC
ecII136II	GAATTC
bsp1286	GAATTC
bsiHKA1	GAATTC
bmyI	GAATTC
banII	GAATTC
taqI	GAATTC
1	TTCGAGCTCG CCGGACATTG ATTATTGACT AGTTATTAAAT AGTAATCAAT TACGGGGTCA TTAGTTCATA GCCCATATAT GGAGTTCCGC GTTACATAAC
201	TTGACGTCAA TGGGTGGAGT ATTTACGGTA AACTGCCAC TGGCAGTAC ATCAAGTGTG TCATATGCCA AGTACGCCC CTATTGACGT CAATGACGGT
301	AAATGGCCG CCTGGCATTG TGCCAGTAC ATGACCTTAT GGGACTTTCC TACTGGCAG TACATCTACG TATTAGTCAT CGCTATTACC ATGGTGATGC

FIG. 10B

[illegible]

FIG. 10C

SUBSTITUTE SHEET (RULE 26)

FIG. 10D

[illegible]

FIG. 10E

[illegible]

67 / 81

FIG. 10F

```

1801  ACGCAATCGT  CGATCCGGA  GCCGGGACTG  TCGGGCGTAC  ACAAATCGCC  CGCAGAGCG  CGGCCGTCTG  GACCGATGGC  TGTGTAGAAG  TACTCGCCGA
      TGCCTTAGCA  GGCTAGGCT  CGGCCCTGAC  AGCCCGCTG  TGTCTTTCG  GCGTCTTGC  GCGGCGAGAC  CTGGCTACCG  ACACATCTTC  ATGAGCGGCT

      nlaIV
      mspI      scrFI
      hpaII     nciI
      bslI      mspI
      mroI      hpaII
      bspMI     hpaII
      bspEI[dam-]
      bsaWI     dsav
      accIII[dam-]
      sau3AI    cauII
      mboI/ndeII[dam-]
      dpnI[dam+]
      dpnII[dam-]
      alwI[dam-]

      rsaI      csp6I
      csp6I     rsal
      csp6I     csp6I
      scaI      scaI

      haeIII/palI
      mcrI       eagi/xmaIII/ecI XI
      eaeI
      cfrI
      fnu4HI
      aciI
      thal
      fnuDII/mvnI
      bstUI
      bsh1236I  sau96I
      hinPI     avaiI
      hhaI/cfoI  asuI
      aciI
      aciaATCGCC  CGCAGAGCG  CGGCCGTCTG  GACCGATGGC  TGTGTAGAAG  TACTCGCCGA
      TGTCTTTCG  GCGTCTTGC  GCGGCGAGAC  CTGGCTACCG  ACACATCTTC  ATGAGCGGCT

      scrFI
      nciI
      mspI
      hpaII
      dsav
      xmaI/pspAI
      smaI
      scrFI
      nciI
      dsav
      cauII
      bsaJI
      aval
      bsaJI

      sau3AI
      mboI/ndeII[dam-]
      dpnI[dam+]
      dpnII[dam-]
      alwI[dam-]
      nlaIV cauII
      bstVI/xhoII
      bamHI bsaJI ecoRI
      alwI[dam-] apoI
      clai/bsp106 bsaJI

      mcrI
      bslI
      sfaNI
      mnlI
      bsaJI
      hgaI
      ahaII/bsaHI
      TAGTGGAAC  CGACGCCCCA  GCACTCGTCC  GAGGGCAAG  GAATAGAGTA  GATGCCGACC  GAAGGATCCC  CGGGGAATTC  AATCGATGGC  CGCCATGGCC
      ATCACCTTTG  GCTGCGGGGT  CGTGAGCAGG  CTCCCGTTTC  CTTATCTCAT  CTACGGCTGG  CTTCTCTAGG  GCGGCTACCG  GCGGTACCGG

      haeIII/palI
      aciI
      fnu4HI
      bglI nlaIII
      sfiI styI
      eaeI ncoI
      cfrI dsal
      tagI haeIII/palI
      clai/bsp106 bsaJI
      AATCGATGGC  CGCCATGGCC
      TTAGCTACCG  GCGGTACCGG

```

68 / 81

FIG. 10G

```

2001 CAACTTGTTT ATTCAGCTT ATAATGGTTA CAAATAAAGC AATAGCATCA CAAATTTCAC AAATAAAGCA TTTTTCAC TGCATTCTAG TTGTGTTT
    aluI fnu4HI bbsI maeI rmaI bsmI maeI
    GTTGAACAAA TAACGTCGAA TATTACCAAT GTTATTTCG TATCGTAGT GTTAAAGTG TTTATTTCGT AAAAAAAGTG ACGTAAGATC AACACCAAC
    sau3AI mboI/ndeII[dam-] dpnI[dam+] dpnII[dam-] pvuI/bspCI mcrI
    tagI[dam-] tru9I clai/bsp106[dam-] sau3AI mseI fnu4HI styI haeIII/palI haeI
    mboI/ndeII[dam-] dpnI[dam+] xmnI hspI ncoI bbsI ncoI
    dpnII[dam-] asel/asnI/vspI bsaJI hhaI/cfoI nlaIII mnlI
    nlaIII alwI[dam-] asp700 TCGGCGCAGC ACCATGGCCT GAAATAACCT CTGAAAAGAGG AACTTGTTA
2101 TCCAAACTCA TCAATGTATC TTATCATGTC GGAATTAAT TCGGCGCAGC ACCATGGCCT GAAATAACCT CTGAAAAGAGG AACTTGTTA
    rsal csp6I nlaIV kpnI hgiCI bsaI aluI pvuII nspBII
    asp718 acc65I ddeI aciI
    AGGTTTGAGT AGTTACATAG AATAGTACAG ACCTAGCTAG CCCTTAATTA AGCCGGCTCG TGGTACCGA CTTATTGGA GACTTCTCC TTGAACCAAT
    nlaIV scrFI mvaI ecorII dsav bstNI apyI[dcM+] bsaJI
    sfaNI ppulOI nsII/avaIII nlaIII sphI nspI nspHI
    2201 CCAATGGAAGA CTCCGCCCTT CTGGTCGAC ACCTACACA CAGTCAATCC CACACCTTC CACACCTTC AGGGTCCGA GGGTCCGTCC GTCCTCATAC GTTTCGTACG
    scrFI mvaI ecorII dsav bstNI apyI[dcM+] sexAI
    nlaIV
    2301 ATCTCAATTA GTCAGCAACC AGGTGTGAA AGTCCCGCAGG CTCCCGCAGG GGCAGAGTA TGCAAGCAT GCATCTCAAT TAGTCAGCAA CCATAGTCCC
    TAGAGTTAAT CAGTCGTTGG TCCACACCTT TCCAGGGTCC GAGGGTCCGT CCGTCTTCAT ACGTTTCGTA CGTAGAGTA ATCAGTCGTT GGTATCAGG
    aciI
    sphI nspI sfaNI nspHI
    ppulOI nsII/avaIII nlaIII
  
```

69 / 81

FIG. 10H

fnu4HI
 bglI
 sfiI
 haeIII/palI
 mnlI
 haeIII/palI
 mnlI bsaJI aciI
 TTTATGCCAGA GGCCGAGGCC
 AAATACGTCT CCGGCTCCGG
 nlaIII
 styI
 ncoI
 bslI dsal
 aciI bsaJI
 CTCCGCCCA TGGCTGACTA ATTTTMTTTA
 TTTATGCCAGA GGCCGAGGCC
 AAATACGTCT CCGGCTCCGG
 scrFI
 nciI
 mspI
 hpaII
 dsav
 haeIII/palI
 mcrI
 eagI/xmaIII/ecI XI
 eaeI
 cfrI
 mspI caulI
 hpaII
 aluI aluI
 hpaII
 TGCAGGAGCT TTTTGGAGG CCTAGGCTTT TGCAGGAGCT TAGCTTATCC GGCCGGGAAC GGTCGATTGG
 CCGAGCCGGA GACTCGATAA GGTCTTCATC ACTCCTCCGA AAAAACCTCC GGATCCGAAA ACCTTTTTCG ATCGAATAGG CCGGCCCTTG CCACGTAACC
 haeIII/palI
 mnlI ddeI
 bsaJI mnlI aluI
 GCCTCGGCT CTGAGCTATT CCAGAAGTAG TGAGGAGGCT TTTTGGAGG CCTAGGCTTT TGCAGGAGCT TAGCTTATCC GGCCGGGAAC GGTCGATTGG
 CCGAGCCGGA GACTCGATAA GGTCTTCATC ACTCCTCCGA AAAAACCTCC GGATCCGAAA ACCTTTTTCG ATCGAATAGG CCGGCCCTTG CCACGTAACC
 tfiI
 hinfI
 aciI
 thalI
 fnuDII/mvnI
 bstUI
 bsh1236I
 AACGCGGATT CCGCGTGCCA AGAGTCAGGT AAGTACCGCC TATAGAGTCT ATAGGCCAC CCCCTGGCT TCGTTAGAAC GCGGTACAA TTAATACATA
 TTGCGCCTAA GGGGCACGGT TCTCAGTCCA TTCTAGTGGG ATATCTCAGA TATCCGGGTG GGGGAACCGA AGCAATCTTG CCGCGATGTT AATTATGTAT
 bstXI
 sau96I
 haeIII/palI
 asuI
 bsaJI
 ATAGGCCAC CCCCTGGCT TCGTTAGAAC GCGGTACAA TTAATACATA
 TTGCGCCTAA GGGGCACGGT TCTCAGTCCA TTCTAGTGGG ATATCTCAGA TATCCGGGTG GGGGAACCGA AGCAATCTTG CCGCGATGTT AATTATGTAT

FIG. 10I

SUBSTITUTE SHEET (RULE 26)

71 / 81

FIG. 10J

```

mspI      mspI      gsuI/bpmI      mspI      mboI/ndeII[dam-]
hpaII     hpaII     scrFI      mvaI      dpnII[dam+]
bslI      bslI      haeIII/palI      haeIII      alwI[dam-]
bsaWI     bsaWI     aciI      ecorII      nlaIV
sau3AI    sau3AI    fnu4HI      mnlI      bstYI/xhoII
                        bstNI      bstNI      bamHI
                        dsav      dsav      alwI[dam-]
                        bstUI      bstUI      alwI[dam-]
                        bsh1236I      bsh1236I      alwI[dam-]
                        csp6I      csp6I      alwI[dam-]
                        hinfI      hinfI      alwI[dam-]
3001 TATCAACAGA AACCAGGAAA AGCTCCGAAA CTACTGATTT AGCGGGCTC GTACCTGGAG TCTGGAGTCC CTTCTCGCTT CTCTGGATCC GGTTCCTGGGA
ATAGTTGTCT TTGCTCCTTT TCGAGGCTTT GATGACTAAA TCGGCGCGGAG CATGGACCTC AGACCTCAGG GAAGAGCGAA GAGACCTAGG CCAAGACCCCT
                        fnu4HI      mboII      fnu4HI      mboII      fnu4HI      mboII      fnu4HI      mboII
                        bbvI      bpuAI      bbvI      bpuAI      bbvI      bpuAI      bbvI      bpuAI
                        scfI      bbsI      scfI      bbsI      scfI      bbsI      scfI      bbsI
                        pstI      mspI      pstI      mspI      pstI      mspI      pstI      mspI
                        bsgI      hpaII      bsgI      hpaII      bsgI      hpaII      bsgI      hpaII
3101 CGGATTTTAC TCTGACCATC AGCAGTCTGC AGCCGGAAGA CTTGGCAACT TATTACTGTC AGCAAGTCA CGAGGATCCG TACACATTTC GACAGGGTAC
GCCTAAAGTG AGACTGGTAG TCGTCAGACG TCGGCTTCTT GAAGCGTTGA ATAATGACAG TCGTTTCAGT GCTCCTAGGC ATGTGTAAAC CTGTCCCATG
sau3AI    sau3AI    mboI/ndeII[dam-]      mboI/ndeII[dam-]      mboI/ndeII[dam-]      mboI/ndeII[dam-]
                        dpnII[dam+]      dpnII[dam+]      dpnII[dam+]      dpnII[dam+]
                        dpnII[dam-]      dpnII[dam-]      dpnII[dam-]      dpnII[dam-]
                        alwI[dam-]      alwI[dam-]      alwI[dam-]      alwI[dam-]
                        nlaIV      nlaIV      nlaIV      nlaIV
                        bstYI/xhoII      bstYI/xhoII      bstYI/xhoII      bstYI/xhoII
                        bamHI      bamHI      bamHI      bamHI
                        mnlI      mnlI      mnlI      mnlI
                        csp6I      csp6I      csp6I      csp6I
                        asp718      asp718      asp718      asp718
                        acc65I      acc65I      acc65I      acc65I
3201 CAAGGTGGAG ATCAAACGAA CTGTGGCTGC ACCATCTGTC TTCACTCTCC CGCATCTGA TGAGCAGTTG AAATCTGGA CTGCCCTCTGT TGTGTGCTG
GTTCCACCTC TAGTTTGCTT GACACCGACG TGGTAGACAG AAGTAGAAGG GCGGTAGACT ACTCGTCAAC TTAGACCTT GACGGAGACA ACACACGGAC
                        mnlI      mnlI      mnlI      mnlI      mnlI      mnlI      mnlI      mnlI
                        scrFI      scrFI      scrFI      scrFI      scrFI      scrFI      scrFI      scrFI
                        mvaI      mvaI      mvaI      mvaI      mvaI      mvaI      mvaI      mvaI
                        ecorII      ecorII      ecorII      ecorII      ecorII      ecorII      ecorII      ecorII
                        dsav      dsav      dsav      dsav      dsav      dsav      dsav      dsav
                        bstNI      bstNI      bstNI      bstNI      bstNI      bstNI      bstNI      bstNI
                        apyI[dcm+]      apyI[dcm+]      apyI[dcm+]      apyI[dcm+]
                        mnlI      mnlI      mnlI      mnlI
                        bslI      bslI      bslI      bslI
                        maeIII      maeIII      maeIII      maeIII
                        bsaJI      bsaJI      bsaJI      bsaJI
3301 CTGAATAACT TCTATCCCAG AGAGGCCAAA GTACAGTGGA AGGTGGATAA CGCCCTCCAA TCGGGTAACT CCCAGGAGAG TGTACACAGC CAGGACAGCA
GACTTATTGA AGATAGGGTC TCTCCGGTTT CATGTACACCT TCACCTATT GCGGGAGGTT AGCCCATTTA GGGTCTCTC ACAGTGTCTC GTCCTGTCTG
xmnI      xmnI      xmnI      xmnI      xmnI      xmnI      xmnI      xmnI
asp700     asp700     asp700     asp700     asp700     asp700     asp700     asp700

```

[illegible]

SUBSTITUTE SHEET (RULE 26)

sau96I

mlaIII

aciI haeIII/palI

fnu4HI **asuI**

11bq Iltys

fi If you

el dsai

bsa,JI

taql haeIII/pall

TTAA GCTTCGATCG CCGCATCG

11AA GCCTCGATGG CCGCCTATGG
AATT CGAAGCTACC GGCCTATACC

ATAATCGAAGCTAACC GCGGTAAC

mai

bsmI maeI

TTCA CTGCATTCTA GTTGTGGT

AAGT GACGTAAGAT CAACACCA

FIG. 10L

3701 CGGGAATTAA TTCGGCGCAG CACCATGGCC TGAATAAACC TCTGAAAGAG GAACTTGGTT AGGTACCTTC TGAGGCGGAA AGAACCAGCT GTGGAATGTG
 GCCCTTAATT AAGCCGCGTC GTGGTACCGG ACTTTATTGG AGACTTTCTC CTTGAACCAA TCCATGGAAG ACTCCGCTT TCTTGGTGA CACCTTACAC
 3801 TGTCAGTTAG GGTGTGAAA GTCCCCAGGC TCCCCAGCAG GCAGAAGTAT GCAAAGCATG CATCTCAATT AGTCAGCAAC CAGGTGTGGA AAGTCCCCAG
 ACAGTCAATC CCACACCTTT CAGGGTCCG AGGGTCCGTC CGTCTTCATA CGTTTCGTAC GTAGAGTTAA TCAGTCTGTG GTCCACACCT TTCAGGGGTC
 3901 GGTCCCCAGC AGGCAGAGT ATGCAAGCA TGCATCTCAA TTAGTCAGCA ACCATAGTCC CGCCCCTAAC TCCGCCCATC CCGCCCCCTAA CTCGCCCCAG
 CGAGGGGTCG TCCGTCTTCA TACGTTTCGT ACGTAGAGTT AATCAGTCGT AATCAGTCGT TGGTATCAGG GCGGGGATTG AGCGGGGTAG GCGGGGGATT GAGCGGGGTC
 4001 TTCCGCCCAT TCTCCGCCCC ATGGCTGACT AATTTTTTTT ATTTATGACG AGGCCGAGGC CGCCTCGGCC TCTGAGCTAT TCCAGAAGTA GTGAGGAGGC
 AAGCGGGTA AGAGGCGGGG TACCGACTGA TTAATAAATAA TAAATACGTC TCCGGTCCG GCGGAGCCGG AGACTCGATA AGGTCTTCAT CACTCTCCG

FIG. 10M

[illegible]

FIG. 10N

mspI nlaIV hgiJII mnlI
 hpaII bsp1286 nlaIV hgiCI
 nael bmyI banII taqI hphI
 cfr10I aluI nlaIV
 4501 GCCGGCTTC CCCGTCAAGC TCTAAATCGG GGGCTCCCTT TAGGTTCCG ATTATAGCT TTAGCGCAC ACCACCCCAA AAAACTTGAT TTGGGTGATG
 CGGCCGAAAG GGGCAGTTCC AGATTTAGCC CCGGAGGAA ATCCCAAGGC TAAATCAGCA AATGCCGTGG AGCTGGGTT TTTTGAAC TAACCCACTAC

maeII haeIII/palI pleI
 draIII sau96I drdI hinfI maeII msel hinfI bsrI
 bsaI asuI
 4601 GTTACAGTAG TGGGCCATCG CCTGTATAGA CGGTTTTTCG CCTTTGACG TTGGAGTCCA CGTTCTTTAA TAGTGGACTC TTGTTCCAAA CTGGAACAAC
 CAAGTGCATC ACCCGGTAGC GGGACTATCT GCCAAAAGC GGGAACTGC AACCTCAGGT GCAAGAAATT ATCACCTGAG AACAAAGTTT GACCTTGTG

bslI
 bslI avaI
 4701 ACTCAACCT ATCTCGGGT ATTCTTTTGA TTTATAAGG ATTGTGCGA TTTCCGCCCTA TTGGTTAAA AATGAGCTGA TTTAACAAA ATTTAAGCG
 TGAGTTGGA TAGAGCCCGA TAAGAAAAC TAATATCCC TAAACCGCT AAAGCCGAT AACCAATTTT TTAATCGACT AAATGTTTT TAAATTCGC

bslI
 bslI ddeI
 hgiAI/aspHI
 bsp1286 bsiHKAI
 bmyI apaLI/snoI rsaI fnu4HI
 alw44I/snoI csp6I sfaNI msel
 4801 AATTTTACA AATATTAAC GTTTACAATT TTATGGTGCA CTCTCAGTAC AATCTGCTCT GATCCGCTAT AGTTAAGCCA ACTCCGCTAT CGTACGTGA
 TTAAATGTT TTTATAATTG CAAATGTTAA AATACCAGT GAGAGTCTATG TTAGACGAGA CTACGGCGTA TCAATTCGCT TGAGGCGATA GCGATGCACT

maeII
 tru9I msel
 tth111I/aspI
 maeII
 bsaAI bsrI
 4901 CTGGGTGATG GCTGGCCCC GACACCGCC AACACCGCT GACGGGCCCT GACGGGCTTG TCTGCTCCG GCATCCGCTT ACAGACAAGC TGTGACCGTC
 GACCCAGTAC CGACCGGGG CTGTGGGGG TTGTGGCGGA CTGCCGGAAC AGACAGGGC CGTAGGCGAA TGTCTGTTG ACTGTCAG

hinPI
 fnu4HI
 bbvI
 nlaIII hhaI/cfoI aciI
 mspI hpaII
 scrFI
 nciI aciI
 dsav sfaNI
 cauII foki
 drdI
 bsmAI
 esp3I
 bslI
 aluI maeII

76 / 81

FIG. 100

```

scrFI      thai      fnuDII/mvnI
nciI       bstUI
mspI       bsh1236I
hpaII      hinPI
dsav aluI  hhaI/cfoI
          nspI       thai
          fnu4HI     fnuDII/mvnI
          cauII bbvI nlaIII mnlI hphI hphI hphI hphI
          TCCGGGAGCT GCATGTGTCA GAGGTTTCA CCGTCATCAC CGAACGCGC GAGGAGTAT TCTTGAAGC GAAAGGGCCT CGTGATACGC CTATTTTAT
          AGGCCCTCGA CGTACACAGT CTCCAAAAGT GGCAGTAGTG GCTTTGCGG CTCCGTCATA AGAACTTCTG CTTTCCCGGA GCACTATGCG GATAAAAATA
5001

          nlaIII      maeII      hinII/acyI      ahaII/bsaHI
          tru9I rcaI    hinPI/aciI      hhaI/cfoI
          msei bspHI    ddel aatII
          AGGTTAATGT CATGATAATA ATGGTTTCTT AGACGTCAGG TGGCACTTTT CGGGGAAATG TGCGCGGAAC CCTATTGTGT TTATTTTCT AAATACATTC
          TCCAATTACA GTACTATTAT TACCAAAGAA TCTGCAGTCC ACCGTGAAAA GCCCTTTTAC ACGCGCCTTG GGGATAAACA AATAAAAAGA TTTATGTAAG
5101

          rcaI      bspHI
          bsrBI bsmAI
          aciI nlaIII
          5201 AAATATGTAT CCGTCATGA GACAATAACC CTGATAAATG CTTCATAAT ATTGAAAAAG GAAGAGTATG AGTATTCAAC ATTTCCGTGT CGCCCTTATT
          TTTATACATA GCGGAGTACT CTGTTATTGG GACTATTTC GAAAGTTATTA TAACTTTTTC CTTCTCATAC TCATAAGTTG TAAAGGCACA GCGGGAATAA

          fnu4HI      aciI
          hphI      hphI      hphI      hphI      hphI      hphI      hphI      hphI      hphI      hphI
          5301 CCCTTTTTTG CGGCATTTTG CCTTCTCTGT TTTGCTCACC CAGAAACGCT GGTGAAAGTA AAAGATGCTG AAGATCAGTT GGTGACAGA GTGGGTATCA
          GGGAAAAAAC GCCGTAAAC GGAAGGACAA AAACGAGTGG GTCTTTGCGA CCACITTCAT TTTCTACGAC TTCTAGTCAA CCCACGTGCT CACCCAATGT

```

FIG. 10P

SUBSTITUTE SHEET (RULE 26)

FIG. 10Q

[illegible]

FIG. 10T

chaf
 fndDII/mvnI
 bstUI
 bsh1236I
 hinPI
 hhaI/cfoI
 thaI
 fndDII/mvnI
 bstUI
 bsh1236I
 bslI
 aciI
 TCTCCCGCG
 TCTCCCGCG
 AGAGGGCGC

81 / 81

7001 GTGAGCTGAT ACCGCTCGCC GCAGCCGAAC GACCGAGCGC AGCGAGTCAG TGAGCGAGGA AGCCCAATAC GCAAACCGCC TCTCCCGCG
 CACTCGACTA TGGCGAGCGG CGTCGGCTTG CTGCTCGC TCGCTCAGTC ACTCGCTCTC GCGGCTTATG CGTTGGCGG AGAGGGCGC

7101 CGTTGGCCGA TTCAATTAATC CAGCTGGCAC GACAGGTTTC CCGACTGGAA AGCGGGCAGT GAGCGCAACG CAATTAATGT GAGTTACCTC ACTCATTTAGG
 GCAACCGGCT AAGTAATTAG GTCGACCGTG CTGTCCAAAG GGCTGACCTT TCGCCCGTCA CTGCGGTTGC GTTAATTACA CTCATGGAG TGAGTAATCC

7201 CACCCAGGC TTACACTTT ATGCTTCCG CTGCTATGTT GTGTGGAATT GTGAGCGGAT AACAAATTCA CACAGGAAAC AGCTATGACC ATGATTACGA
 GTGGGGTCCG AAATGTGAAA TACGAAGGCC GAGCATACAA CACACCTTAA CACTCGCCTA TTGTTAAAGT GTGCTCTTG TCGATACTGG TACTAATGCT

>length: 7305

INTERNATIONAL SEARCH REPORT

International Application No
PC1/US 95/09576

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/64 C12N15/67 C12N15/85 C12N9/72 C12N5/10

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DNA CLONING, VOLUME III, EDITED BY D.M. GLOVER, 1987 IRL PRESS, OXFORD, GB;, pages 189-212, A.M.C. BROWN AND M.R.D. SCOTT 'Retroviral vectors'	1-3,7,8
Y	see page 192, line 7 - page 196, line 5; figures 2,3 --- -/--	5,6, 9-12, 16-21

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents:

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

- *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- *&* document member of the same patent family

Date of the actual completion of the international search

23 November 1995

Date of mailing of the international search report

08.12.95

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Hornig, H

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 95/09576

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CELL, vol. 37, no. 3, July 1984 CELL PRESS,CAMBRIDGE,MA,US;; pages 1053-1062, C.L. CEPKO ET AL. 'Construction and applications of a highly transmissible murine retrovirus shuttle vector' cited in the application	1-3,7,8
Y	pZIP-Neo SV(B)1 see figure 1 ---	5,6, 9-12, 16-21
Y	MOL. CELL. BIOL., vol. 5, no. 3, March 1985 ASM WASHINGTON, DC,US, pages 431-437, A.D. MILLER ET AL. 'Generation of helper-free amphotrophic retroviruses that transduce a dominant-acting, methotrexate-resistant dihydrofolate reductase gene' see page 432, right column, line 5 - page 436, right column, line 7; figure 1 ---	5,6, 9-12, 16-21
Y	WO,A,94 05784 (US) 17 March 1994 see the whole document ---	5,6, 9-12, 16-21
Y	EP,A,0 215 548 (ZYMOGENETICS INC ;UNIV WASHINGTON (US)) 25 March 1987 see the whole document ---	5,6, 9-12, 16-21
A	WO,A,92 17566 (GENENTECH INC) 15 October 1992 cited in the application see the whole document ---	1-21
A	WO,A,90 12025 (UNIV LELAND STANFORD JUNIOR) 18 October 1990 cited in the application see the whole document ---	1-21
A	EP,A,0 260 148 (GENENTECH INC) 16 March 1988 see the whole document ---	1-21
A	EP,A,0 160 457 (GENENTECH INC) 6 November 1985 cited in the application see the whole document ---	1-21

-/--

INTERNATIONAL SEARCH REPORT

Int ional Application No
PCT/US 95/09576

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	PROC. NATL.ACAD SCI., vol. 86, February 1989 NATL. ACAD SCI.,WASHINGTON,DC,US;; pages 1041-1045, M. VIVAUD ET AL. 'A 5' splice-region G-C mutation in exon 1 of the human beta-globin gene inhibits pre-mRNA splicing: A mechanism for beta+-thalassemia' see the whole document -----	1-4

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US 95/09576

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO-A-9405784	17-03-94	AU-B- 4839493	29-03-94
EP-A-0215548	25-03-87	US-A- 4968626	06-11-90
		DE-A- 3688900	23-09-93
		DE-T- 3688900	09-12-93
		JP-A- 62111690	22-05-87
		US-A- 4959318	25-09-90
		US-A- 5073609	17-12-91
		US-A- 5302529	12-04-94
WO-A-9217566	15-10-92	AU-B- 660671	06-07-95
		AU-A- 1580692	02-11-92
		JP-T- 6506356	21-07-94
WO-A-9012025	18-10-90	US-A- 5043270	27-08-91
EP-A-0260148	16-03-88	AU-B- 613316	01-08-91
		AU-B- 7831787	19-05-88
		DE-A- 3730599	07-07-88
		FR-A- 2603899	18-03-88
		GB-A, B 2197321	18-05-88
		US-A- 5024939	18-06-91
		JP-A- 63152986	25-06-88
EP-A-0160457	06-11-85	AU-B- 601358	13-09-90
		AU-B- 4134585	24-10-85
		AU-B- 5295890	30-08-90
		EP-A- 0385558	05-09-90
		HK-A- 8395	27-01-95
		JP-A- 60243023	03-12-85
		JP-A- 6040942	15-02-94
		NO-B- 174934	25-04-94
		SG-A- 3994	10-06-94
		US-A- 4965199	23-10-90