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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
8 August 2002 (08.08.2002)

PCT

(10) International Publication Number
WO 02/060941 A2

(51) International Patent Classification[7]: C07K 14/435

(21) International Application Number: PCT/US02/02500

(22) International Filing Date: 29 January 2002 (29.01.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/264,932 29 January 2001 (29.01.2001) US

(71) Applicant: **ANTICANCER, INC.** [US/US]; 7917 Ostrow Street, San Diego, CA 92111-3604 (US).

(72) Inventors: **ZHAO, Ming**; 8020 Avenida Navidad, Apt. 47, San Diego, CA 92122 (US). **XU, Mingxu**; 3203 Via Alicante, Apt.1, La Jolla, CA 92037 (US). **JIANG, Ping**; 8056 Camino Huerta, San Diego, CA 92122 (US). **YANG, Meng**; 12631 El Camino Real, Apt. 2105, San Diego, CA 92130 (US).

(74) Agents: **MURASHIGE, Kate, H.** et al.; Morrison & Foerster LLP, 3811 Valley Centre Drive, Suite 500, San Diego, CA 92130-2332 (US).

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

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

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

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

(54) Title: IMPROVED FLUORESCENT PROTEINS

MVSKQILKNTGLQEIMSFKVNLEGVVNNHVFTMEGCGKGNILFGNQ
LVQIRVTKGAPLPFAFDILSPAFQYGNRTFTKYPEDISDFFIQSFPAGF
VYERTLRYEDGGLVEIRSDINLIEEMFVYRVEYKGRNFPNDGPVMK
KTITGLQPSFEVVYMNDGVLVGQVILVYRLNSGKFYSCHMRTLMKS
KGVVKDFPEYHFIQHRLEKTYVEDGGFVEQHETAIAQLTSLGKPLGS
LHEWV

(57) Abstract: Improved forms of fluorescent protein with high fluorescence and low toxicity are disclosed.

WO 02/060941 A2

IMPROVED FLUORESCENT PROTEINS

Cross-Reference to Related Applications

This application claims priority under 35 U.S.C. § 119(e) from provisional application 60/264,932 filed 29 January 2001. The contents of this application are incorporated herein by reference.

Technical Field

The invention relates to new forms of "green fluorescent protein" and their uses. Specifically, the invention is directed to a particular GFP and variants thereof which are bright and nontoxic.

Background Art

Green fluorescent protein (GFP) was initially isolated from *Aequorea victoria* by Chalfie, U.S. patent 5,491,084. Modifications were made to the amino acid sequence to enhance brightness as reported by Ward and Chalfie in PCT publication WO 95/21191. Tsien, as disclosed in U.S. patents 5,625,048 and 5,777,079, provided modified forms of this protein which exhibited differing spectral characteristics and provided fluorescence of various colors. In addition, modifications were made to the nucleotide sequence encoding these proteins to make the sequence compatible with human cells. In addition, PCT publication WO 99/49019 published 30 September 1999 provides some sequence information regarding green fluorescent protein expressed from genes isolated from *Renilla* and *Ptilocarpus*. Gurskaya, N.G., *et al., BMC Biochem* (2002) 2:5 describes mutations which change the spectrum of emission of fluorescent protein from coral.

The above documents, each of which is incorporated herein by reference in its entirety, demonstrate that variations in the amino acid sequence of a protein which exhibits fluorescence upon excitation with radiation of shorter wavelength than the fluorescent wavelength provide a range of color choice and intensity. The fluorescent proteins have found wide use both in scientific research and in the production of novelty items, such as toys. Because the only requirements for fluorescence are irradiation with a suitable wavelength and because the fluorescence is visible to the naked eye, these proteins have proved convenient markers and have inspired whimsical applications.

It has now been found that additional variants of fluorescent protein have improved brightness and exhibit low toxicity. These proteins can also be modified to fluoresce in a variety of colors and to vary in intensity.

Disclosure of the Invention

The invention is directed to compositions and methods which employ fluorescent proteins that are related by homology to the protein encoded by the nucleotide sequence set forth in Figure 1B. The exemplified GFP contains, in comparison to known fluorescent proteins, conservative substitutions and several substitutions which render it less acidic at the N-terminal portion but more acidic at the C-terminal portion. The invention is directed to compositions and methods related to a group of variants which are slightly less acidic in the N-terminal approximately half of the sequence and slightly more acidic at the C-terminal approximately half of the sequence. The proteins of the invention are non-toxic, and cells containing them can survive for at least 4 weeks and for 3, 4 or 6 months.

Thus, in one aspect, the invention is directed to the variant proteins *per se* and to methods to use the variants. In other aspects, the invention is directed to recombinant materials which encode the variants and methods to produce the variants using these materials, as well as other applications of the recombinant materials themselves.

Brief Description of the Drawings

Figure 1A shows the deduced amino acid sequence of a variant of the invention designated herein A/C. Figure 1B shows the nucleotide sequence of a nucleic acid, putatively isolated from *R. reniformis*, which encodes the amino acid sequence of Figure 1A.

Figures 2A, 2B and 2C show nucleotide sequences based on the nucleotide sequence of Figure 1B which are modified to conform to codon usage preferences for *Saccharomyces cerevisiae*, *Escherichia coli* and *Bifidobacterium longum*, respectively.

Figure 3 shows a comparison of the amino acid sequence of the invention variant labeled A/C with a green fluorescent protein whose gene was cloned from *R. mulleri*.

Figures 4A and 4B show diagrams of a retroviral vector for production of the variants in various tumor cell lines. The vector is provided to packaging cells and the resulting virions used to infect tumor cell lines.

Figures 5A-5C show images of the fluorescent protein expressed in the packaging cell line PT67, the melanoma cell line B16F0, and the prostate cancer cell line PC3.

Figure 6A-6D show images of tumors grown from H460, a lung cancer cell line, PC3, a prostate cancer cell line, CAPAN-1, a pancreatic cancer cell line and RKO, a colon cancer cell line.

Figures 7A and 7B show the nucleotide sequence encoding a fluorescent protein from coral, at positions 289-964.

Modes of Carrying Out the Invention

Improved "green fluorescent proteins" (GFP) and recombinant materials which encode them are provided. This permits the use of a bright, nontoxic label to monitor gene expression, to label various cells, and to monitor the progress of metastases as described, for example, in PCT publication WO 98/49336, incorporated herein by reference. The improved fluorescent proteins of the invention offer more sensitive methods to assess these phenomena while remaining nontoxic to cells and entire organisms. Thus, the proteins of the present invention is useful in a variety of art known methods which employ the known forms of GFP described above. Production of GFP in general and use of recombinant materials as well as the GFP itself are well known in the art in view of the extensive literature describing the previously known forms of this fluorescent protein.

A protein of the amino acid sequence shown in Figure 1A, designated A/C herein, emits green fluorescence and has the brightness and nontoxic properties stated above. However, as is known in the art, these "green" fluorescent proteins may be modified so that they fluoresce in various colors in the visible spectrum. Thus, by suitably modifying the amino acid sequence of the protein set forth in Figure 1A, red, yellow, blue, or other color fluorescence may also be obtained. In addition, the brightness of the fluorescence can be varied by making small changes to the amino acid sequence. The nature of such modifications is helpfully described in, for example, U.S. patent 5,777,079 incorporated herein by reference above. As described, modifications to the serine residue which is found at position 66 of the A/C sequence can be replaced by alanine, leucine, cysteine, valine, isoleucine or threonine to obtain proteins with red shifted spectra which are generally brighter as compared to the unmodified form of A/C. Other modifications that appear to affect brightness or fluorescence wavelength include those at position 67. The chromophore appears to be focused on positions 66-68 of the A/C protein, which

correspond to positions 65-67 of the *Aequorea* wildtype GFP protein discussed in the '079 patent. Thus, mutations at positions 66-68 are particularly important in modifying the properties of the protein.

Thus, the invention includes mutants having substitutions at any of positions 66-68 and particularly at position 66.

Additional modifications of 1-4 amino acids elsewhere in the molecule are also permitted; a minimal number of mutations of this type is insufficient to convert the A/C amino acid sequence into that of any known "green fluorescent protein" and has minimal impact on the properties of the molecule. Thus, the proteins of the invention include fluorescent proteins which are at least 90% homologous, preferably 95% homologous, and more preferably 99% homologous to the amino acid sequence shown in Figure 1A. Particularly preferred are fluorescent proteins having the amino acid sequence shown in Figure 1A and variants thereof which have at least one amino acid substitution in positions 66-68, preferably in position 66. Also preferred are modifications analogous to those described for the proteins from coral, described by Gurskaya, cited above.

The nucleotide sequence encoding the fluorescent protein variants of the invention may be expressed in a wide variety of cells. Expression systems suitable for production of proteins from recombinant systems are by now conventional for prokaryotes, eukaryotes such as yeast and fungi, higher plants, animal cells, including vertebrate cells, mammalian cells, and especially human cells, and a variety of cell lines. The appropriate expression system and vector depends on the nature of the host and the application intended. In addition to providing an appropriate expression system, the nucleotide sequence may be modified to convert it to a preferred codon usage for the intended host. The nucleotide sequence shown in Figure 1B, thus, may contain one or more of the modifications shown in Figures 2A, 2B and 2C for expression in the indicated hosts, *Saccharomyces cerevisiae*, *Escherichia coli* and *Bifidobacterium longum*, respectively. Thus, the sequence for expression in *Saccharomyces cerevisiae* may contain 1-230 silent base changes; that for *E. coli* from 1-94 silent base changes and that for *Bifidobacterium longum* 1-21 base changes. All intermediate numbers of base changes are also included within the scope of the invention.

The fluorescent proteins of the invention and the recombinant materials encoding them may be applied in a wide variety of uses as is set forth in detail in PCT publication

WO99/49019, cited above, and incorporated herein by reference. This publication describes a many uses, including analytical, research, diagnostic, and commercial uses.

Thus, the object of the production of the fluorescent proteins of the invention may be to obtain the protein itself for use in various compositions and articles of manufacture. These fluorescent proteins may be used in various items such as toys, dolls, card games, paints, textiles, balloons, cosmetics, and foodstuffs or any other article or composition designed to glow. The protein itself may be produced for incorporation into these articles and compositions. The fluorescent proteins of the invention may also be combined with other materials which fluoresce or emit light, such as luciferase. The '019 publication describes compositions in which other luminescent biological materials are combined in the same composition with fluorescent proteins so that rather than effecting excitation by irradiation from an external source, the irradiating wavelengths are generated *in situ* by the luminescent combined material.

More serious uses of the green fluorescent protein focus on its value as a research tool. In one embodiment, the fluorescent proteins of the invention may be fused or otherwise coupled to antibodies directed to target tissues in plants or animals. For example, it may be desirable to label tumors in animals and to follow metastases by coupling the fluorescent label to the tumor. The fluorescent proteins of the invention may be prepared as conjugates with moieties which are able to target tissues or cells. Typical targeting moieties are specific binding partners for a material displayed on tissues or cells. Typical targeting moieties are antibodies and ligands for receptors.

Alternatively, the production of a fusion protein containing a green fluorescent protein can be used to monitor expression of the coupled protein. In addition, as described, for example, by Yang, M, *et al., Cancer Res.* (1998) 58:4217-4221 and by Yang, M, *et al., Cancer Res.* (1999) 59:731-736 and as reviewed by Hoffman, R.M., *Cancer & Metastasis Reviews* (1999) 17:271-277, fluorescent proteins may be generated in tumors and used to monitor metastasis.

The fluorescent proteins of the invention may also be used to label reagents in assays such as immunoassays. In one example, a sandwich assay may be employed wherein one specific binding partner to an analyte is a capture moiety which immobilizes the analyte and a second specific binding partner is used to label the immobilized analyte. The fluorescent proteins of the invention may be used directly as a label on the labeling

binding partner or on a secondary binding partner such as, for example, the use of a second antibody-bearing label to couple to a first antibody directly bound to analyte.

Expression of a protein can also be monitored by fusing the nucleotide sequence encoding the protein to a nucleotide sequence encoding the fluorescent protein of the invention. Expression of the protein of interest may then be determined by monitoring the fluorescence generated as the fusion protein is produced. Alternatively, the capacity of a promoter or other control sequence to effect expression can be monitored by placing a nucleotide sequence encoding the invention fluorescent protein in operable linkage therewith. Again, fluorescence is generated by virtue of production of the fluorescent protein, thereby indicating that the expression controls are operable.

The foregoing applications are merely exemplary. In general, the fluorescent proteins of the invention can be used in any application where fluorescent labels are employed, including assay modifications such as fluorescence polarization and fluorescence quenching assays. The fluorescent proteins of the invention have the additional advantage of the capability of being generated *in situ* so that their presence or absence or amount can be used as an index of expression as well as to provide an internal source of fluorescence in cells. Thus, the course of tumor metastasis, bacterial or viral infection, or the movement of cells of any type within a plant or animal organism can be traced using the fluorescent proteins of the invention.

The following example is intended to illustrate but not to limit the invention.

Example 1

Expression of Improved Fluorescent Protein

A nucleic acid molecule encoding the fluorescent protein having the amino acid sequence shown as A/C in Figure 1A was obtained, labeled *R. reniformis* GFP, from Stratagene (San Diego, CA). The nucleotide sequence encoding the protein was determined and is shown in Figure 1B.

The nucleotide sequence is optionally modified as shown in Figures 2A-2C by silent base substitutions so as to optimize expression in various host organisms.

The amino acid sequence of the fluorescent protein, in comparison to the fluorescent protein in the art to which applicants believe it is most closely related, is shown

in Figure 3. The A/C protein is approximately 86% identical to the amino acid sequence of a protein encoded by a nucleotide sequence isolated from *R. mulleri*. The chromophore portions of these proteins, positions 66-68 of A/C and positions 65-67 of *R. mulleri*, are identical.

The nucleotide sequence of Figure 1B is inserted into a retroviral vector under control of a 3′ LTR promoter (Figure 4A). In this vector, derived from pFB (Stratagene) shown in Figure 4B, the multiple cloning site region is shown. The resultant vector, designated pFB-Rmv GFP, was used to modify PT-67 packaging cells. The packaging cells were cultured, producing a bright green signal due to the production of the GFP protein. See Figure 5A. High intensity was maintained for over 10 days, indicating the GFP is nontoxic.

Example 2

Preparation of Cell Lines Expressing A/C Fluorescent Protein

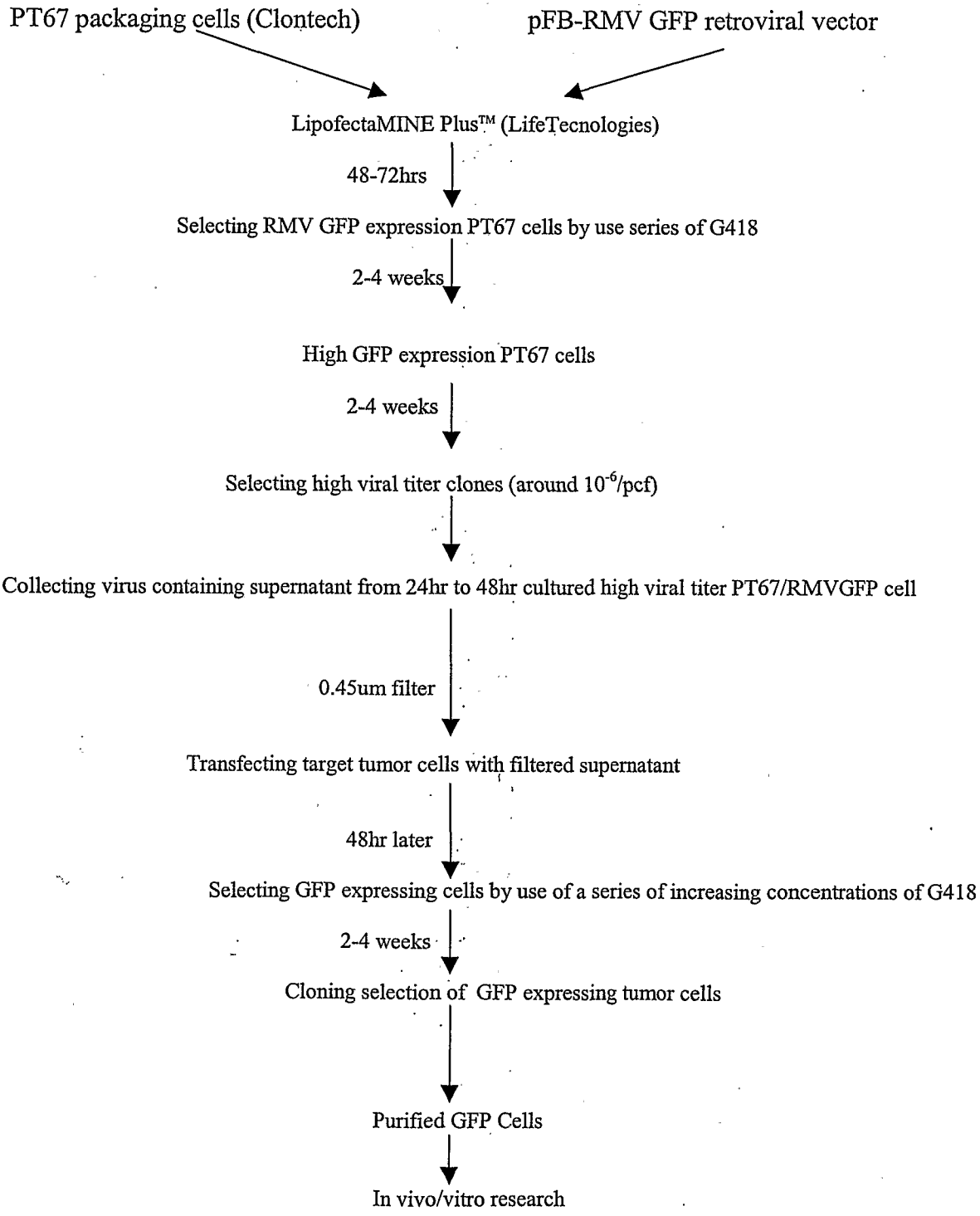
A number of tumor cell lines were infected with the packaged virions prepared in the PT-67 packing cell line and are as follows: The breast cancer cell line MX-1; the prostate cancer cell lines MDA-PCA2B and PC3; the pancreatic cancer cell line CAPAN-1; the colon cancer cell line RKO; the sarcoma cell line MES-SA/DX5; the lung cancer cell lines H460, Lewis lung, and A549 and the melanoma line B16F0. Several of these are illustrated. The fluorescence of these cell lines is illustrated in Figures 5B (B16F0) and 5C (PC3). The fluorescence is maintained for weeks.

Example 3

Tumor Labeling

Tumors were grown from the transformed cell lines by subcutaneous injection in immunocompromised mice and the progress of the tumor followed by monitoring fluorescence. Illustrative results are shown in Figures 6A-6D.

Preparation of GFP cell lines by use of pFB-RMV GFP expression vector



Claims

1. A fluorescent protein comprising the amino acid sequence of the A/C protein of Figure 1A or a fluorescent protein having at least 90% homology with said A/C protein.

2. The fluorescent protein of claim 1, which has at least 95% homology with the A/C protein of Figure 1A.

3. The fluorescent protein of claim 2, which has at least 99% homology with the A/C protein of Figure 1A.

4. The fluorescent protein of claim 1, which comprises the amino acid sequence of the A/C protein of Figure 1A.

5. The fluorescent protein of claim 1, wherein at least 1 amino acid in positions 66-68 has been replaced by a different amino acid.

6. The fluorescent protein of claim 5, wherein the amino acid at position 66 has been replaced.

7. A nucleic acid molecule comprising a nucleotide sequence which encodes the fluorescent protein of claim 1.

8. A nucleic acid molecule comprising a nucleotide sequence which encodes the fluorescent protein of claim 4.

9. The nucleic acid molecule of claim 8, which comprises the nucleotide sequence set forth in Figure 1B, optionally containing 1 or more nucleotide substitutions as set forth in Figure 2A, 2B or 2C.

10. A recombinant expression system which comprises a nucleotide sequence encoding the fluorescent protein of claim 1 operably linked to control sequences for its expression.

11. A recombinant host cell which comprises the expression system of claim 10.

12. A method to produce a fluorescent protein, which method comprises culturing the cells of claim 11, wherein said nucleotide sequence is expressed to produce said fluorescent protein, and optionally recovering said fluorescent protein.

13. Antibodies specifically immunoreactive with the fluorescent protein of claim 1.

14. A conjugate comprising the protein of claim 1 coupled to a targeting moiety.

15. The conjugate of claim 14, wherein said targeting moiety is an antibody or a ligand for a receptor.

16. The expression system of claim 10, wherein said nucleotide sequence encoding the fluorescent protein is fused to a nucleotide sequence encoding additional protein.

17. A method to monitor the production of a protein, which method comprises

providing an expression system which comprises a first nucleotide sequence encoding said protein fused to a second nucleotide sequence encoding the fluorescent protein of claim 1;

placing said expression system in an environment in which expression is to be monitored; and

assessing generation of fluorescence, whereby generation of fluorescence indicates expression of said protein.

18. A method to evaluate the activity of a promoter, which method comprises providing a nucleic acid which comprises said promoter operably linked to a nucleotide sequence encoding fluorescent protein of claim 1;

placing said nucleic acid in an environment in which the activity of the promoter is to be evaluated;

monitoring appearance and amount of fluorescence;

wherein the appearance and amount of fluorescence indicates the activity of the promoter.

19. A method to label tissue, which method comprises contacting said tissue with the conjugate of claim 14, wherein said targeting moiety is specific for said tissue.

20. A method to assess the position and progression of tumors and their metastases, which method comprises modifying said tumors to express the fluorescent protein of claim 1 and observing the position of fluorescence.

21. An improved method to conduct an immunoassay, wherein said immunoassay comprises entrapping analyte or an analog thereof in a sandwich comprising a first specific binding partner for said analyte coupled to a solid support and a second specific binding partner for said analyte comprising a label wherein the improvement comprises employing as a label the fluorescent protein of claim 1.

MVSKQILKNTGLQEIMSFKVNLEGVVNNHVFTMEGCGKGNILFGNQ
LVQIRVTKGAPLPFAFDILSPAFQYGNRTFTKYPEDISDFFIQSFPAGF
VYERTLRYEDGGLVEIRSDINLIEEMFVYRVEYKGRNFPNDGPVMK
KTITGLQPSFEVVYMNDGVLVGQVILVYRLNSGKFYSCHMRTLMKS
KGVVKDFPEYHFIQHRLEKTYVEDGGFVEQHETAIAQLTSLGKPLGS
LHEWV

1A

ATGGTGAGCAAGCAGATCCTGAAGAACACCGGCCTGCAGGAGATCATGAGCTTCAAGGTGAACCTGGAGGGCGTGG
TGAACAACCACGTGTTCACCATGGAGGGCTGCGGCAAGGGCAACATCCTGTTCGGCAACCAGCTGGTGCAGATCCG
CGTGACCAAGGGCGCCCCCCTGCCCTTCGCCTTCGACATCCTGAGCCCCGCCTTCCAGTACGGCAACCGCACCTTC
ACCAAGTACCCCGAGGACATCAGCGACTTCTTCATCCAGAGCTTCCCCGCCGGCTTCGTGTACGAGCGCACCCTGC
GCTACGAGGACGGCGGCCTGGTGGAGATCCGCAGCGACATCAACCTGATCGAGGAGATGTTCGTGTACCGCGTGGA
GTACAAGGGCCGCAACTTCCCCAACGACGGCCCCGTGATGAAGAAGACCATCACCGGCCTGCAGCCCAGCTTCGAG
GTGGTGTACATGAACGACGGCGTGCTGGTGGGCCAGGTGATCCTGGTGTACCGCCTGAACAGCGGCAAGTTCTACA
GCTGCCACATGCGCACCCTGATGAAGAGCAAGGGCGTGGTGAAGGACTTCCCCGAGTACCACTTCATCCAGCACCG
CCTGGAGAAGACCTACGTGGAGGACGGCGGCTTCGTGGAGCAGCACGAGACCGCCATCGCCCAGCTGACCAGCCTG
GGCAAGCCCCTGGGCAGCCTGCACGAGTGGGTGTAA

1B

Optimized for *Sacchromyces cerevisiae*

ATG GTT TCT AAA CAA ATT TTG AAA AAT ACT GGT TTG
CAA GAA ATT ATG TCT TTT AAA GTT AAT TTG GAA GGT
GTT GTT AAT AAT CAT GTT TTT ACT ATG GAA GGT TGT
GGT AAA GGT AAT ATT TTG TTT GGT AAT CAA TTG GTT
CAA ATT AGA GTT ACT AAA GGT GCT CCA TTG CCA TTT
GCT TTT GAT ATT TTG TCT CCA GCT TTT CAA TAT GGT AAT
AGA ACT TTT ACT AAA TAT CCA GAA GAT ATT TCT GAT
TTT TTT ATT CAA TCT TTT CCA GCT GGT TTT GTT TAT GAA
AGA ACT TTG AGA TAT GAA GAT GGT GGT TTG GTT GAA
ATT AGA TCT GAT ATT AAT TTG ATT GAA GAA ATG TTT
GTT TAT AGA GTT GAA TAT AAA GGT AGA AAT TTT CCA
AAT GAT GGT CCA GTT ATG AAA AAA ACT ATT ACT GGT
TTG CAA CCA TCT TTT GAA GTT GTT TAT ATG AAT GAT
GGT GTT TTG GTT GGT CAA GTT ATT TTG GTT TAT AGA
TTG AAT TCT GGT AAA TTT TAT TCT TGT CAT ATG AGA
ACT TTG ATG AAA TCT AAA GGT GTT GTT AAA GAT TTT
CCA GAA TAT CAT TTT ATT CAA CAT AGA TTG GAA AAA
ACT TAT GTT GAA GAT GGT GGT TTT GTT GAA CAA CAT
GAA ACT GCT ATT GCT CAA TTG ACT TCT TTG GGT AAA
CCA TTG GGT TCT TTG CAT GAA TGG GTT TAA

There are 230 silent base changes which correspond to *Sacchromyces cerevisiae* codon-usage preferences.

2A

Optimized for *Escherichia coli*

ATG GTG AGC AAA CAG ATT CTG AAA AAC ACC GGC CTG CAG GAA ATT ATG AGC TTT AAA GTG AAC CTG GAA GGC GTG GTG AAC AAC CAT GTG TTT ACC ATG GAA GGC TGC GGC AAA GGC AAC ATT CTG TTT GGC AAC CAG CTG GTG CAG ATT CGC GTG ACC AAA GGC GCG CCC CTG CCC TTT GCG TTT GAT ATT CTG AGC CCC GCG TTT CAG TAT GGC AAC CGC ACC TTT ACC AAA TAT CCC GAA GAT ATT AGC GAT TTT TTT ATT CAG AGC TTT CCC GCG GGC TTT GTG TAT GAA CGC ACC CTG CGC TAT GAA GAT GGC GGC CTG GTG GAA ATT CGC AGC GAT ATT AAC CTG ATT GAA GAA ATG TTT GTG TAT CGC GTG GAA TAT AAA GGC CGC AAC TTT CCC AAC GAT GGC CCC GTG ATG AAA AAA ACC ATT ACC GGC CTG CAG CCC AGC TTT GAA GTG GTG TAT ATG AAC GAT GGC GTG CTG GTG GGC CAG GTG ATT CTG GTG TAT CGC CTG AAC AGC GGC AAA TTT TAT AGC TGC CAT ATG CGC ACC CTG ATG AAA AGC AAA GGC GTG GTG AAA GAT TTT CCC GAA TAT CAT TTT ATT CAG CAT CGC CTG GAA AAA ACC TAT GTG GAA GAT GGC GGC TTT GTG GAA CAG CAT GAA ACC GCG ATT GCG CAG CTG ACC AGC CTG GGC AAA CCC CTG GGC AGC CTG CAT GAA TGG GTG TAA

There are 94 silent base changes which correspond to *Escherichia coli* codon-usage preferences.

2B

Optimized for *Bifidobacterium longum*

ATG GTG TCC AAG CAG ATC CTG AAG AAC ACC GGC CTG
CAG GAG ATC ATG TCC TTC AAG GTG AAC CTG GAG GGC
GTG GTG AAC AAC CAC GTG TTC ACC ATG GAG GGC TGC
GGC AAG GGC AAC ATC CTG TTC GGC AAC CAG CTG GTG
CAG ATC CGC GTG ACC AAG GGC GCC CCG CTG CCG TTC
GCC TTC GAC ATC CTG TCC CCG GCC TTC CAG TAC GGC
AAC CGC ACC TTC ACC AAG TAC CCG GAG GAC ATC TCC
GAC TTC TTC ATC CAG TCC TTC CCG GCC GGC TTC GTG
TAC GAG CGC ACC CTG CGC TAC GAG GAC GGC GGC CTG
GTG GAG ATC CGC TCC GAC ATC AAC CTG ATC GAG GAG
ATG TTC GTG TAC CGC GTG GAG TAC AAG GGC CGC AAC
TTC CCG AAC GAC GGC CCG GTG ATG AAG AAG ACC ATC
ACC GGC CTG CAG CCG TCC TTC GAG GTG GTG TAC ATG
AAC GAC GGC GTG CTG GTG GGC CAG GTG ATC CTG GTG
TAC CGC CTG AAC TCC GGC AAG TTC TAC TCC TGC CAC
ATG CGC ACC CTG ATG AAG TCC AAG GGC GTG GTG AAG
GAC TTC CCG GAG TAC CAC TTC ATC CAG CAC CGC CTG
GAG AAG ACC TAC GTG GAG GAC GGC GGC TTC GTG GAG
CAG CAC GAG ACC GCC ATC GCC CAG CTG ACC TCC CTG
GGC AAG CCG CTG GGC TCC CTG CAC GAG TGG GTG TAA

There are 21 silent base changes which correspond to *Bifidobacterium longum* codon-usage preferences.

2C

```
                    M
A/C,          1 2( VSKQILKNTGLQEIMSFKVNLEGVVNNHVFTMEGCGKGNILFGNQLVQIRVTKGAPLPFA
Mulleri,      1 MSKQILKNTCLQEVMSYKVNLEGIVNNHVFTMEGCGKGNILFGNQLVQIRVTKGAPLPFA
                ******** *** ** ****** ************************************

A/C,         62 FDILSPAFQYGNRTFTKYPEDISDFFIQSFPAGFVYERTLRYEDGGLVEIRSDINLIEEM
Mulleri,     61 FBIVSPAFQYGNRTFTKYPNDISDYFIQSFPAGFMYERTLRYGDGGLVEIRSDINLIEDK
                * * *************** **** ********* ******* ***************

A/C,        122 FVYRVEYKGRNFPNDGPVMKKTITGLQPSFEVVYMNDGVLVGQVILVYRLNSGKFYSCHM
Mulleri,    121 FVYRVEYKESNFPDDGPVMQKTILGIEPSFGAMYMNNGVLVGEVILVYKLNSGKYYSCHM
                ********  *** ***** *** *  ***   *** ***** ***** ***** *****

A/C,        182 RTLMKSKGVVKDFPEYHFIQHRLEKTYVEDGGFVEQHETAIAQLTSLGKPLGSLHEWV
Mulleri,    181 KTLMKSKGVVKEFPSYHFIEHRLEKTYVEDGGFVEQHETAIAQMTSIGKPLGSLHEWV
                ********** ** **** ********************** ** ***********
```

3

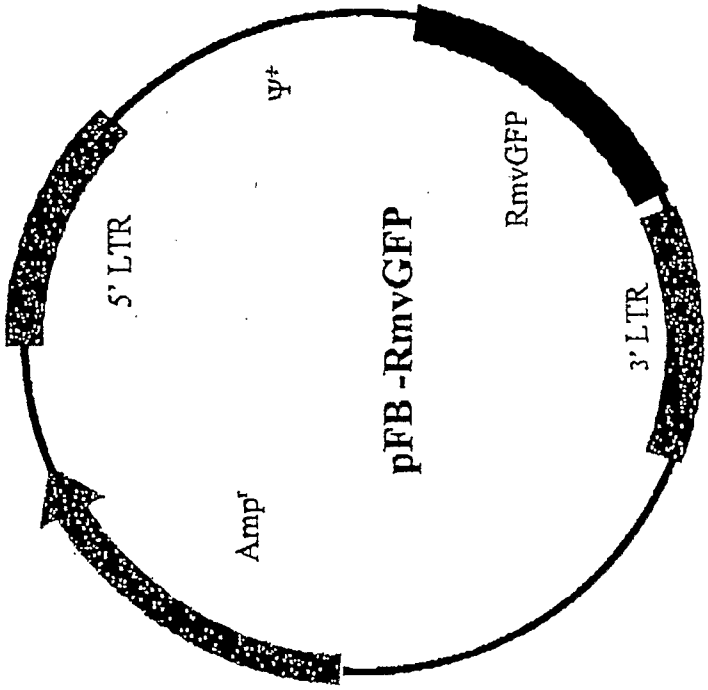
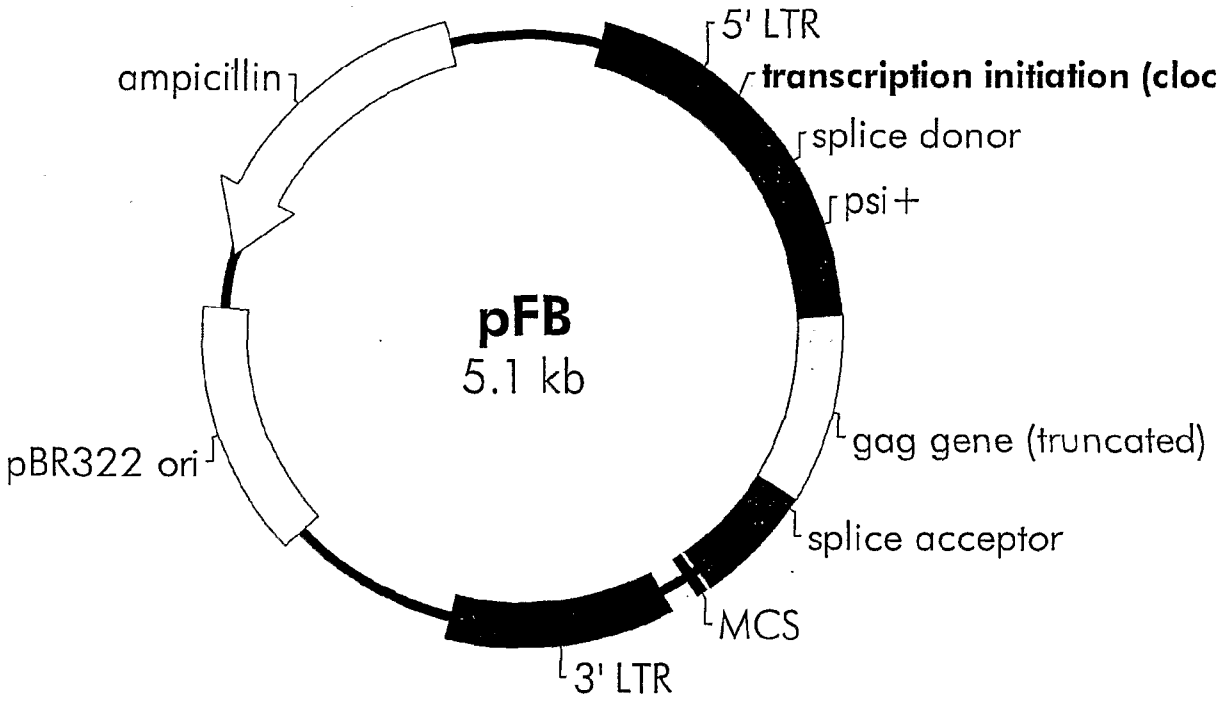


Figure 4A

5′-long terminal repeat 209–760
transcription initiation **(clockwise)** 616
splice donor 818–822
ψ+ packaging signal 760–2046
gag gene (truncated) 1236–1723
splice acceptor 1751–1753
multiple cloning site 2057–2086
3′-long terminal repeat 2163–2756
pBR322 origin 3237–3904
ampicillin resistance (bla) ORF 4055–4912

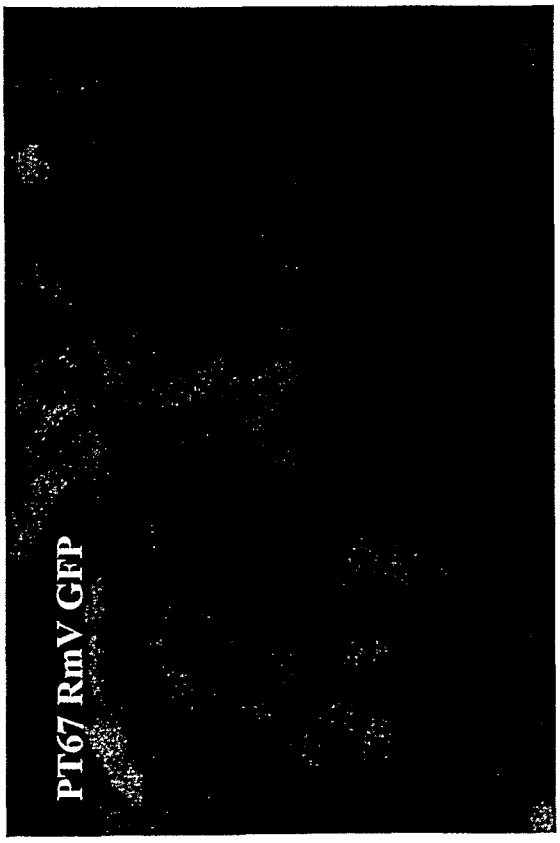


pFB Multiple Cloning Site Region
(sequence shown 2057–2086)

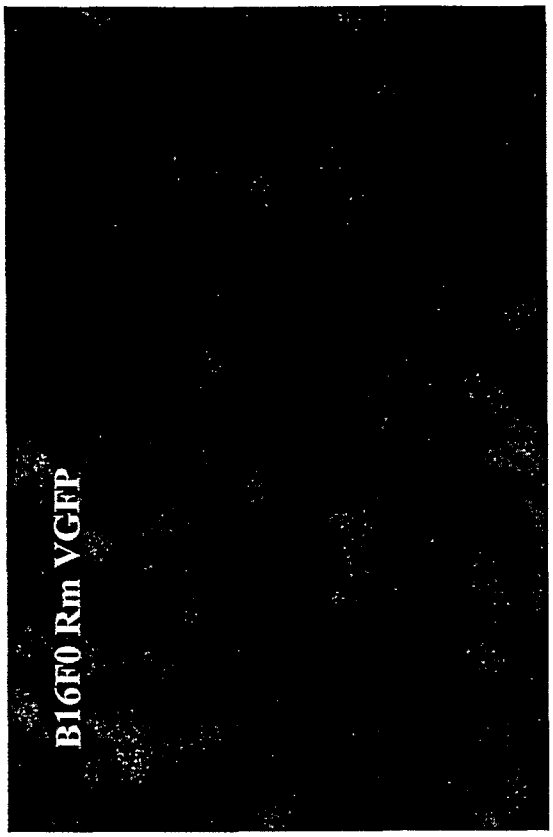
Sal I EcoR I BamH I Xho I Not I

GTCGACGAATTCGGATCCTCGAGCGGCCGC

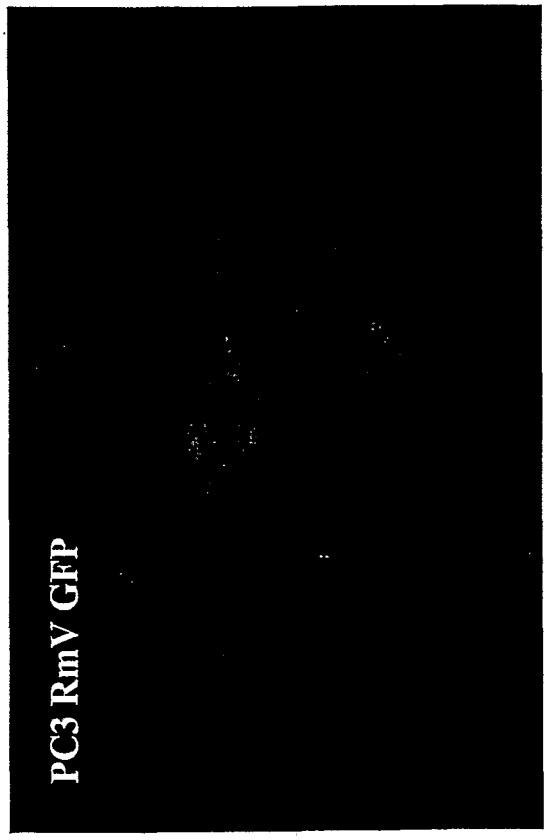
4B



5A.



5B



5C

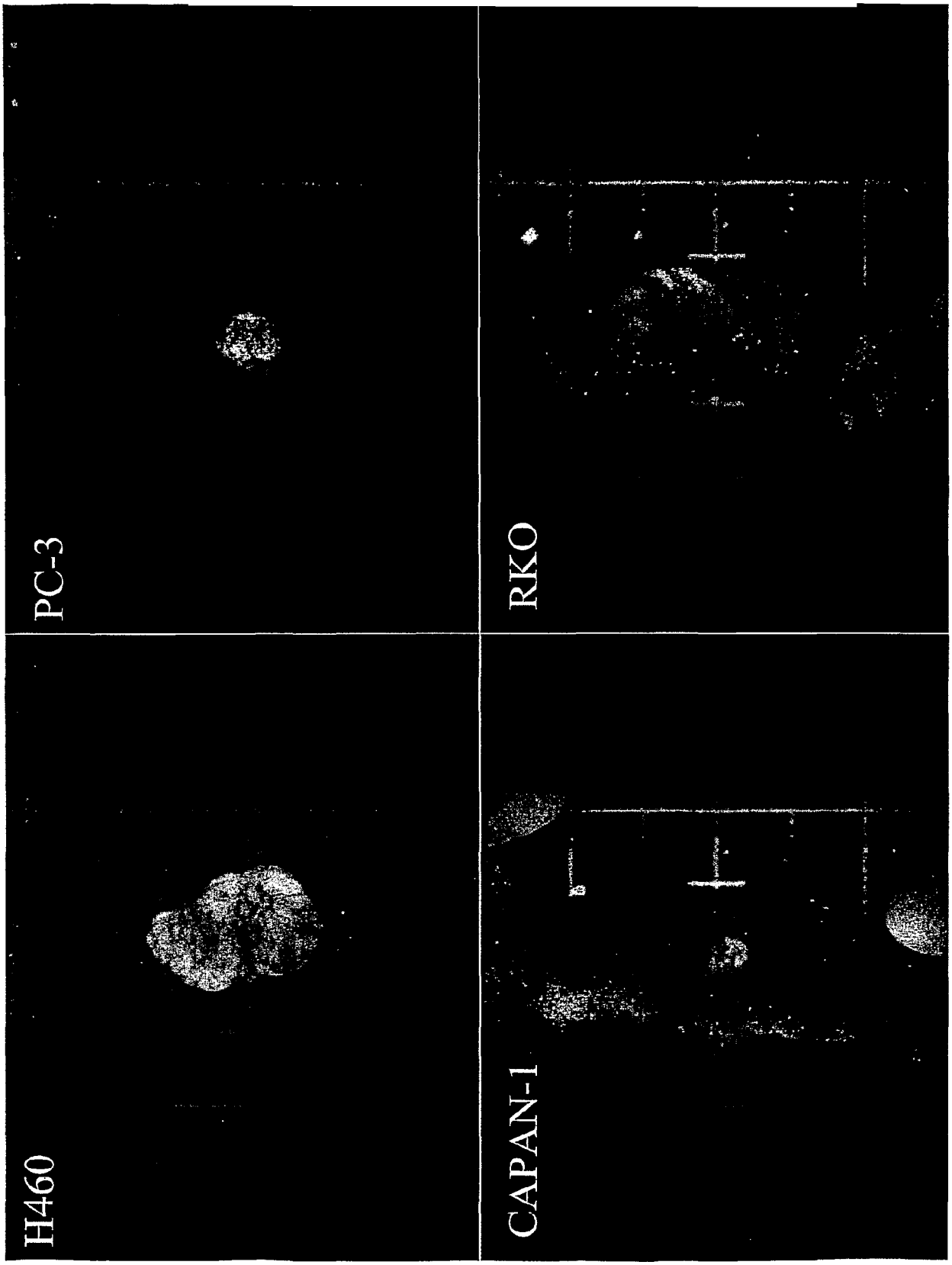


Figure 6

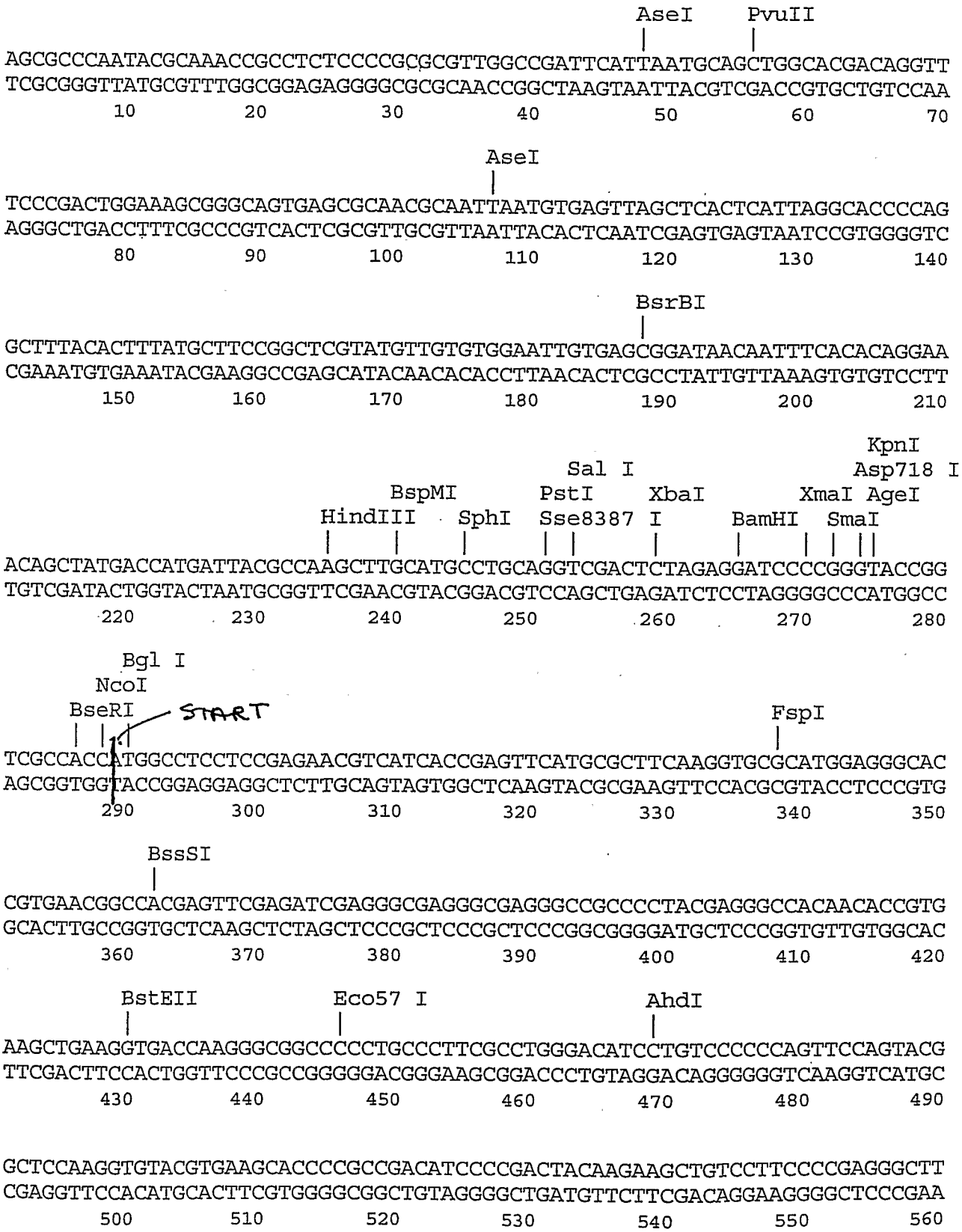

AseI
PvuII
AGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTT
TCGCGGGTTATGCGTTTGGCGGAGAGGGGCGCGCAACCGGCTAAGTAATTACGTCGACCGTGCTGTCCAA
10 20 30 40 50 60 70
AseI
TCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAG
AGGGCTGACCTTTCGCCCGTCACTCGCGTTGCGTTAATTACACTCAATCGAGTGAGTAATCCGTGGGGTC
80 90 100 110 120 130 140
BsrBI
GCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAA
CGAAATGTGAAATACGAAGGCCGAGCATACAACACACCTTAACACTCGCCTATTGTTAAAGTGTGTCCTT
150 160 170 180 190 200 210
KpnI
Sal I
Asp718 I
BspMI
PstI
XbaI
XmaI AgeI
HindIII
SphI
Sse8387 I
BamHI
SmaI
ACAGCTATGACCATGATTACGCCAAGCTTGCATGCCTGCAGGTCGACTCTAGAGGATCCCCGGGTACCGG
TGTCGATACTGGTACTAATGCGGTTCGAACGTACGGACGTCCAGCTGAGATCTCCTAGGGGCCCATGGCC
220 230 240 250 260 270 280
Bgl I
NcoI
BseRI
START
FspI
TCGCCACCATGGCCTCCTCCGAGAACGTCATCACCGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGCAC
AGCGGTGGTACCGGAGGAGGCTCTTGCAGTAGTGGCTCAAGTACGCGAAGTTCCACGCGTACCTCCCGTG
290 300 310 320 330 340 350
BssSI
CGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCCACAACACCGTG
GCACTTGCCGGTGCTCAAGCTCTAGCTCCCGCTCCCGCTCCCGGCGGGGATGCTCCCGGTGTTGTGGCAC
360 370 380 390 400 410 420
BstEII
Eco57 I
AhdI
AAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCCCAGTTCCAGTACG
TTCGACTTCCACTGGTTCCCGCCGGGGGACGGGAAGCGGACCCTGTAGGACAGGGGGGTCAAGGTCATGC
430 440 450 460 470 480 490
GCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCCGACTACAAGAAGCTGTCCTTCCCCGAGGGCTT
CGAGGTTCCACATGCACTTCGTGGGGCGGCTGTAGGGGCTGATGTTCTTCGACAGGAAGGGGCTCCCGAA
500 510 520 530 540 550 560

Fig 7A

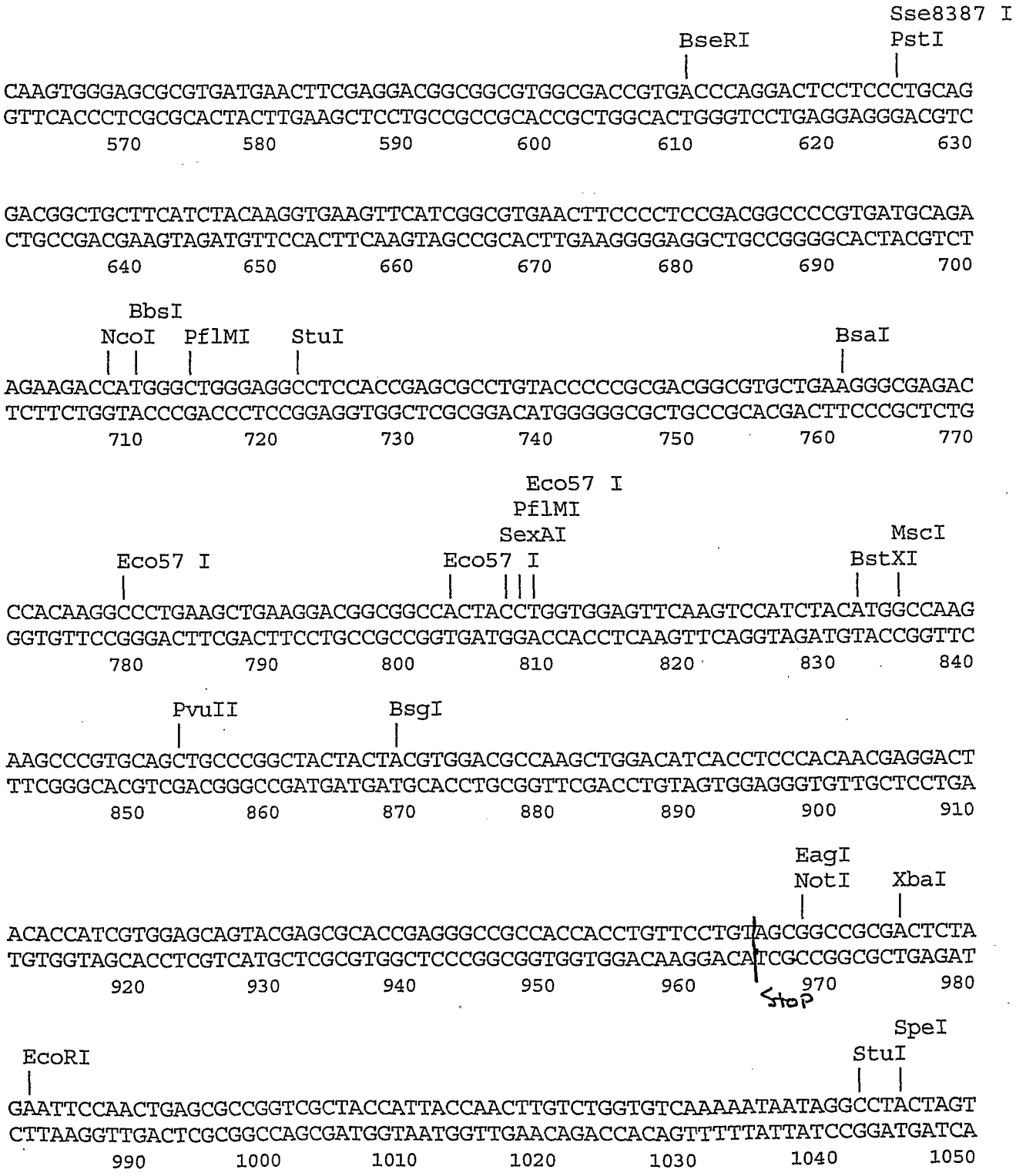

Sse8387 I
BseRI
PstI
CAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGCGACCGTGACCCAGGACTCCTCCCTGCAG
GTTCACCCTCGCGCACTACTTGAAGCTCCTGCCGCCGCACCGCTGGCACTGGGTCCTGAGGAGGGACGTC
570 580 590 600 610 620 630
GACGGCTGCTTCATCTACAAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTGATGCAGA
CTGCCGACGAAGTAGATGTTCCACTTCAAGTAGCCGCACTTGAAGGGGAGGCTGCCGGGGCACTACGTCT
640 650 660 670 680 690 700
BbsI
NcoI
PflMI
StuI
BsaI
AGAAGACCATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAGAC
TCTTCTGGTACCCGACCCTCCGGAGGTGGCTCGCGGACATGGGGGCGCTGCCGCACGACTTCCCGCTCTG
710 720 730 740 750 760 770
Eco57 I
PflMI
SexAI
Eco57 I
Eco57 I
MscI
BstXI
CCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATCTACATGGCCAAG
GGTGTTCCGGGACTTCGACTTCCTGCCGCCGGTGATGGACCACCTCAAGTTCAGGTAGATGTACCGGTTC
780 790 800 810 820 830 840
PvuII
BsgI
AAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACGCCAAGCTGGACATCACCTCCCACAACGAGGACT
TTCGGGCACGTCGACGGGCCGATGATGATGCACCTGCGGTTCGACCTGTAGTGGAGGGTGTTGCTCCTGA
850 860 870 880 890 900 910
EagI
NotI
XbaI
ACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCCGCCACCACCTGTTCCTGTAGCGGCCGCGACTCTA
TGTGGTAGCACCTCGTCATGCTCGCGTGGCTCCCGGCGGTGGTGGACAAGGACATCGCCGGCGCTGAGAT
920 930 940 950 960 970 980
Stop
SpeI
EcoRI
StuI
GAATTCCAACTGAGCGCCGGTCGCTACCATTACCAACTTGTCTGGTGTCAAAAATAATAGGCCTACTAGT
CTTAAGGTTGACTCGCGGCCAGCGATGGTAATGGTTGAACAGACCACAGTTTTTATTATCCGGATGATCA
990 1000 1010 1020 1030 1040 1050

Fig 7B

专利名称(译)	荧光蛋白		
公开(公告)号	EP1356050A2	公开(公告)日	2003-10-29
申请号	EP2002709205	申请日	2002-01-29
[标]申请(专利权)人(译)	抗癌公司		
申请(专利权)人(译)	抗癌，INC.		
当前申请(专利权)人(译)	抗癌，INC.		
[标]发明人	ZHAO MING XU MINGXU JIANG PING YANG MENG		
发明人	ZHAO, MING XU, MINGXU JIANG, PING YANG, MENG		
IPC分类号	A61K49/00 C07K14/435 C12N15/12 C07K16/18 C12N5/10 C12N15/62 G01N33/53		
CPC分类号	A61K49/0047 C07K14/43595		
代理机构(译)	FISHER，ADRIAN JOHN		
优先权	60/264932 2001-01-29 US		
外部链接	Espacenet		

摘要(译)

公开了具有高荧光和低毒性的荧光蛋白的形式。