(51) **International Patent Classification**[7]: **C12N 15/12,** C07K 14/435, C12N 15/62, G01N 33/52, A61K 31/00, 33/00

(21) **International Application Number:** PCT/US01/04625

(22) **International Filing Date:** 13 February 2001 (13.02.2001)

(25) **Filing Language:** English

(26) **Publication Language:** English

(30) **Priority Data:**
60/184,732   23 February 2000 (23.02.2000)   US

(63) **Related by continuation (CON) or continuation-in-part (CIP) to earlier application:**
US   60/184,732 (CON)
Filed on   23 February 2000 (23.02.2000)

(71) **Applicant** (for all designated States except US): **AURORA BIOSCIENCES CORPORATION** [US/US]; 11010 Torreyana Road, San Diego, CA 92121 (US).

(72) **Inventors; and**

(75) **Inventors/Applicants** (for US only): **NELSON, David** [US/US]; 6540 Windward Ridge Way, San Diego, CA 92121 (US). **ZAMAIRA, Elize** [US/US]; 11890 Arborlake Way, San Diego, CA 92131 (US). **TSIEN, Roger**

[US/US]; 8535 Nottingham Place, La Jolla, CA 92037 (US).

(74) **Agent: HAILE, Lisa, A.**; Gray Cary Ware & Freidenrich, Suite 1600, 4365 Executive Drive, San Diego, CA 92121-2189 (US).

(81) **Designated States** (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, 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, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) **Designated States** (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, 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, GW, ML, MR, NE, SN, TD, TG).

Published:
— with international search report
— before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments

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:** MODIFIED FLUORESCENT PROTEINS

(57) **Abstract:** Functional red fluorescent proteins, nucleic acids encoding them, and methods for their use.

MODIFIED FLUORESCENT PROTEINS

FIELD OF THE INVENTION

The present invention relates generally to functional mutants of red fluorescent proteins, and methods for their use.

5   ## BACKGROUND OF THE INVENTION

Naturally fluorescent proteins are attractive as reporter molecules for cell based assays because of their bright visible fluorescence and ability to be expressed within living cells without the need to add exogenous co-factors or reagents.

10   Fluorescent proteins have been successfully exploited as markers of gene expression, tracers of cell lineage, fusion tags to monitor protein localization within living cells, and as fluorescent donors or acceptors for assays based on the use of fluorescent resonance energy transfer (FRET). Naturally fluorescent proteins have been characterized from a large number of species, however the green fluorescent

15   protein from *Aequorea victoria* is probably the most extensively studied example.

Aequorea green fluorescent protein (GFP) is a stable, proteolysis-resistant single polypeptide chain of 238 residues, and has two absorption maxima at around 395 and 475 nm (Tsien (1998) Annu. Rev. Biochem. $\underline{67}$ 509-544). The relative amplitudes of these two peaks are sensitive to environmental factors (Ward &

20   Bokman (1982) Biochemistry $\underline{21}$:4535-4540, Ward et al. (1982) Photochem. Photobiol. $\underline{35}$ 803-808) and illumination history (A. B. Cubitt et al. (1995) Trends Biochem. Sci. $\underline{20}$ 448-455). Excitation at the primary absorption peak of 395 nm yields an emission maximum at 508 nm with a quantum yield of 0.72-0.85 (Shimomura and Johnson (1962) J. Cell. Comp. Physiol. $\underline{59}$ 223).

25   The fluorophore results from the autocatalytic cyclization of the polypeptide backbone between residues Ser[65] and Gly[67] and oxidation of the α-β bond of Tyr[66] (Cody et al., (1993) Biochemistry $\underline{32}$ 1212-1218, Heim et al.,(1994) Proc. Natl. Acad. Sci. USA $\underline{91}$ 12501-12504). Mutation of Ser[65] to Thr (S65T) simplifies the excitation spectrum to a single peak at 488 nm of enhanced amplitude (Heim et al.,

2

(1995) Nature <u>373</u> 664-665), which no longer gives signs of conformational

isomers. The cDNA for the protein was cloned in 1992 and the protein has been

extensively mutated (D.C. Prasher et al., (1992) Gene <u>111</u> 229-33). Mutagenesis of

GFP has resulted in the creation of a variety of mutants that have distinct spectral

5   properties, improved brightness and enhanced expression and folding in mammalian

cells compared to the native GFP, (SEQ. ID. NO.: 10), Table 1. (*Green*

Fluorescent Proteins, Chapter 2, pages 19 to 47, edited Sullivan and Kay, Academic

Press, U.S. Patent Nos: 5,625,048 to Tsien *et al.*, issued April 29, 1997; 5,777,079

to Tsien *et al.*, issued July 7, 1998; and U.S. Patent No. 5,804,387 to Cormack *et*

10   *al.*, issued September 8, 1998). In many cases, these functional engineered

fluorescent proteins have superior spectral properties to wild-type *Aequorea* GFP,

and are preferred for use herein.

TABLE 1					
Mutants of Aequorea Green Fluorescent Proteins					
Mutations	Common Name	Quantum Yield (Φ) & Molar Extinction (ε)	Excitation & Emission Max	Relative Fluorescence At 37 °C	Sensitivity To Low pH % max F at pH 6
S65T type					
S65T, S72A, N149K, M153T, I167T	Emerald (SEQ. ID. NO.: 28)	Φ = 0.68 ε = 57,500	487 509	100	91
F64L, S65T, V163A		Φ = 0.58 ε = 42,000	488 511	54	43
F64L, S65T	EGFP	Φ = 0.60 ε = 55,900	488 507	20	57
S65T		Φ = 0.64 ε = 52,000	489 511	12	56
Y66H type					
F64L, Y66H, Y145F, V163A	P4-3E	Φ = 0.27 ε = 22,000	384 448	100	N.D.
F64L, Y66H, Y145F		Φ = 0.26 ε = 26,300	383 447	82	57
Y66H, Y145F	P4-3	Φ = 0.3 ε = 22,300	382 446	51	64
Y66H	BFP	Φ = 0.24 ε = 21,000	384 448	15	59
Y66W type					
S65A, Y66W, S72A, N146I, M153T, V163A	W1C	Φ = 0.39 ε = 21,200	435 495	100	82

F64L,S65T, Y66W, N146I, M153T, V163A	W1B	Φ = 0.4 ε = 32,500	434 452 476 (505)	80	71
Y66W, N146I, M153T, V163A	hW7	Φ = 0.42 ε = 23,900	434 452 476 (505)	61	88
Y66W			436 485	N.D.	N.D.
T203Y type					
S65G, S72A, K79R, T203Y	Topaz	Φ = 0.60 ε = 94,500	514 527	100	14
S65G, V68L, S72A, T203Y	10C	Φ = 0.61 ε = 83,400	514 527	58	21
S65G,V68L, Q69K, S72A, T203Y	h10C+	Φ = 0.71 ε = 62,000	516 529	50	54
S65G, S72A, T203H		Φ = 0.78 ε = 48,500	508 518	12	30
S65G, S72A T203F		Φ = 0.70 ε = 65,500	512 522	6	28
T203I type					
T203I, S72A, Y145F	Sapphire	Φ = 0.64 ε = 29,000	395 511	100	90
T203I T202F	H9	Φ = 0.6 ε = 20,000	395 511	13	80

X-ray crystallographic studies have clarified the protein structure and helped to elucidate the effect of mutations, environmental effects, and photochemical events that occur in wild-type and mutant forms of *Aequorea* GFP (Ormo et al., (1996)

5   Science 273 1392-1395, Yang et al., (1996) Nat. Biotechnol. 14 1246-1251, Brejc et al., (1997) Proc. Natl. Acad. Sci. USA 94 2306-2311, Scharnagl et al., (1999) Biophys J. 77 1839-1857, Elsliger et al. (1999) Biochem. 38 5296-5301). These studies have provided a detailed molecular picture of the chromophore structure in *Aequorea* GFP and have enabled a precise understanding of how changes in the

10   electronic environment around the chromophore lead to altered fluorescent properties.

Despite this unique understanding, current efforts to date have failed to create stable, well-defined, red fluorescent mutants of *Aequorea* GFP. Red fluorescent proteins (RFPs) are particularly attractive as fluorescent markers

15   because red light is less phototoxic, is transmitted through tissues more efficiently, and is less scattered than blue or UV light sources. Additionally cells typically exhibit less autofluorescence when illuminated with red light compared to UV light.

4

Recently Anthozoan fluorescent proteins isolated from a number of species of coral (Matz et al., (1999) Nature Biotech. 17 969-973), and these proteins have been the focus of much attention because they exhibit fluorescent emission spectra at red wavelengths.

5   However, the existing wild type Anthozoan fluorescent proteins are not well suited for many applications because of their broad excitation and emission spectra, relatively small stokes shift, and poor quantum yield and molar extinction coefficient when expressed in mammalian cells. The broad excitation spectra result in significant spectral overlap of the red fluorescent protein with the spectra of other available

10   fluorescent proteins, and makes it difficult to efficiently excite the red fluorescent protein without also directly exciting other fluorescent proteins. These factors reduce the effectiveness of the existing red fluorescent proteins for multiplexed analysis and FRET applications.

       The present invention relates to functional red fluorescent proteins that are

15   designed to have improved brightness, reduced spectral cross talk and to be rapidly and efficiently expressed in mammalian cells. Functional red fluorescent proteins are well suited for multiplexed fluorescent analysis, and FRET based applications with existing *Aequorea* fluorescent proteins.

20   **SUMMARY OF THE INVENTION**

       The present invention includes mutants of red fluorescent proteins with improved spectral, and biochemical properties, for use as fluorescent markers and as FRET partners. The functional red fluorescent proteins of the present invention

25   comprise one or more key mutations designed to provide for improved folding, brightness and to create functional red fluorescent proteins that have sharper, more defined excitation and emission peaks when expressed in mammalian cells.

       In one embodiment this invention provides a nucleic acid comprising a nucleotide sequence encoding a functional red fluorescent protein comprising at least

30   one mutation corresponding to positions D59, I60, S62, P63, Q64, F65, Q66, S69, K70, V71, Y72, V73, W93, R95, N98, W143, A145, S146, T147, E148, Y151, G159, I161, K163, G171, S179, Y181, S197, L199, Y214, E215 or R216.

In one aspect, the functional red fluorescent protein exhibits a reduced molar extinction coefficient at 487 nm compared to the wild type Anthozoan red fluorescent protein (SEQ. ID. NO. 7).

5   In one aspect, the functional red fluorescent protein exhibits a reduced molar extinction coefficient at 530 nm compared to the wild type Anthozoan red fluorescent protein (SEQ. ID. NO. 7).

In one aspect, the functional red fluorescent protein exhibits a higher molar extinction coefficient at 583 nm compared to the wild type Anthozoan red fluorescent protein (SEQ. ID. NO. 7).

10   In one aspect, the functional red fluorescent protein is brighter than the wild type Anthozoan red fluorescent protein (SEQ. ID. NO. 7) when excited at 558 nm.

In one aspect, the functional red fluorescent protein is brighter than the wild type Anthozoan red fluorescent protein (SEQ. ID. NO. 7) when expressed in a mammalian cell grown at 37 °C.

15   In another aspect, the functional red fluorescent protein exhibits a higher quantum yield compared to the wild type Anthozoan red fluorescent protein (SEQ. ID. NO. 7).

In one aspect, the functional red fluorescent protein exhibits a faster rate of autocatalytic formation compared to the wild type Anthozoan red fluorescent protein

20   (SEQ. ID. NO. 7).

In one embodiment, the functional red fluorescent protein comprises at least one mutation corresponding to position 59 in SEQ. ID. NO. 7 selected from D59S, D59A, D59H, D59E or D59P.

In one embodiment, the functional red fluorescent protein comprises at least

25   one mutation corresponding to position 60 in SEQ. ID. NO. 7 selected from the group consisting of I60T, I60A, I60C, I60V and I60L.

In one embodiment, the functional red fluorescent protein comprises at least one mutation corresponding to position 62 in SEQ. ID. NO. 7 selected from the group consisting of S62A, S62G, S62C and S62T.

6

In one embodiment, the functional red fluorescent protein comprises at least one mutation corresponding to position 63 in SEQ. ID. NO. 7 selected from the group consisting of P63T, P63H, P63F and P63W.

In one embodiment, the functional red fluorescent protein comprises at least one mutation corresponding to position 64 in SEQ. ID. NO. 7 selected from the group consisting of Q64K, Q64P, Q64T, Q64N and Q64R.

In one embodiment, the functional red fluorescent protein comprises at least one mutation corresponding to position 65 in SEQ. ID. NO. 7 selected from the group consisting of F65L, F65V, F65I, F65M, F65Y and F65W.

In one embodiment, the functional red fluorescent protein comprises at least one mutation corresponding to position 66 in SEQ. ID. NO. 7 selected from the group consisting of Q66R, Q66R, Q66P, Q66K, Q66E, Q66T, Q66A and Q66G.

In one embodiment, the functional red fluorescent protein comprises at least one mutation corresponding to position 69 in SEQ. ID. NO. 7 selected from the group consisting of S69L, S69A, S69V and S69T.

In one embodiment, the functional red fluorescent protein comprises at least one mutation corresponding to position 70 in SEQ. ID. NO. 7 selected from the group consisting of K70M, K70Q, K70L and K70R.

In one embodiment, the functional red fluorescent protein comprises at least one mutation corresponding to position 71 in SEQ. ID. NO. 7 selected from the group consisting of V71C, V71L, V71A and V71L.

In one embodiment, the functional red fluorescent protein comprises at least one mutation corresponding to position 72 in SEQ. ID. NO. 7 selected from the group consisting of Y72F and Y72W.

In one embodiment, the functional red fluorescent protein comprises at least one mutation corresponding to position 73 in SEQ. ID. NO. 7 selected from the group consisting of V73A, V73L, V73S and V73I.

In one embodiment, the functional red fluorescent protein comprises at least one mutation corresponding to position 93 in SEQ. ID. NO. 7 selected from the group consisting of W93L, W93Y, W93C and W93F.

In one embodiment, the functional red fluorescent protein comprises at least one mutation corresponding to position 95 in SEQ. ID. NO. 7 selected from the group consisting of R95K.

In one embodiment, the functional red fluorescent protein comprises at least
5   one mutation corresponding to position 98 in SEQ. ID. NO. 7 selected from the group consisting of N98T, N98D, N98A and N98Q.

In one embodiment, the functional red fluorescent protein comprises at least one mutation corresponding to position 143 in SEQ. ID. NO. 7 selected from the group consisting of W143L, W143F, W143C, W143Y and W143L.

10   In one embodiment, the functional red fluorescent protein comprises at least one mutation corresponding to position 145 in SEQ. ID. NO. 7 selected from the group consisting of A145P, A145S, A145G and A145L.

In one embodiment, the functional red fluorescent protein comprises at least one mutation corresponding to position 146 in SEQ. ID. NO. 7 selected from the
15   group consisting of S146R, S146G, S146N, S146H, S146T, S146A and S146D.

In one embodiment, the functional red fluorescent protein comprises at least one mutation corresponding to position 147 in SEQ. ID. NO. 7 selected from the group consisting of T147N, T147K and T147S.

In one embodiment, the functional red fluorescent protein comprises at least
20   one mutation corresponding to position 148 in SEQ. ID. NO. 7 selected from the group consisting of E148V and E148D.

In one embodiment, the functional red fluorescent protein comprises at least one mutation corresponding to position 151 in SEQ. ID. NO. 7 selected from the group consisting of Y151F, Y151N, Y151D, Y151S, Y151T and Y151A.

25   In one embodiment, the functional red fluorescent protein comprises at least one mutation corresponding to position 159 in SEQ. ID. NO. 7 selected from the group consisting of G159A, G159S and G159V.

In one embodiment, the functional red fluorescent protein comprises at least one mutation corresponding to position 161 in SEQ. ID. NO. 7 selected from the
30   group consisting of I161V, I161V, I161F, I161M and I161L

8

In one embodiment, the functional red fluorescent protein comprises at least one mutation corresponding to position 163 in SEQ. ID. NO. 7 selected from the group consisting of K163I, K163R, K163T, K163E, K163V, K163G and K163A.

In one embodiment, the functional red fluorescent protein comprises at least
5   one mutation corresponding to position 171 in SEQ. ID. NO. 7 selected from the group consisting of G171S and G171A.

In one embodiment, the functional red fluorescent protein comprises at least one mutation corresponding to position 179 in SEQ. ID. NO. 7 selected from the group consisting of S179A, S179P, S179T, S179E, S179Q and S179K.

10   In one embodiment, the functional red fluorescent protein comprises at least one mutation corresponding to position 181 in SEQ. ID. NO. 7 selected from the group consisting of Y181F, Y181W, Y181N and Y181I.

In one embodiment, the functional red fluorescent protein comprises at least one mutation corresponding to position 197 in SEQ. ID. NO. 7 selected from the
15   group consisting of S197Y, S197T, S197N and S197A.

In one embodiment, the functional red fluorescent protein comprises at least one mutation corresponding to position 199 in SEQ. ID. NO. 7 selected from the group consisting of L199I, L199V, L199I and L199A.

In one embodiment, the functional red fluorescent protein comprises at least
20   one mutation corresponding to position 214 in SEQ. ID. NO. 7 selected from the group consisting of Y214F, Y214H and Y214L.

In one embodiment, the functional red fluorescent protein comprises at least one mutation corresponding to position 215 in SEQ. ID. NO. 7 selected from the group consisting of E215G, E215Q and E215R.

25   In one embodiment, the functional red fluorescent protein comprises at least one mutation corresponding to position 216 in SEQ. ID. NO. 7 selected from the group consisting of R216K, R216L, R216C and R216F.

In one embodiment the invention comprises an expression vector, comprising; expression control sequences operatively linked to a nucleic acid molecule encoding a functional red fluorescent protein whose sequence differs from the amino acid sequence of an Anthozoan red fluorescent protein (SEQ. ID. NO. 7) by at least one

5   amino acid substitution corresponding to position D59, I60, S62, P63, Q64, F65, Q66, S69, K70, V71, Y72, V73, W93, R95, N98, W143, A145, S146, T147, E148, Y151, G159, I161, K163, G171, S179, Y181, S197, L199, Y214, E215 or R216.

In another embodiment, the invention includes a recombinant host cell, comprising; a nucleic acid molecule encoding a functional red fluorescent protein

10   whose sequence differs from the amino acid sequence of an Anthozoan red fluorescent protein (SEQ. ID. NO. 7) by at least one amino acid substitution corresponding to position D59, I60, S62, P63, Q64, F65, Q66, S69, K70, V71, Y72, V73, W93, R95, N98, W143, A145, S146, T147, E148, Y151, G159, I161, K163, G171, S179, Y181, S197, L199, Y214, E215 or R216.

15   In yet another embodiment, the invention comprises a functional fluorescent protein, comprising; an amino acid sequence that differs from the amino acid sequence of an Anthozoan red fluorescent protein (SEQ. ID. NO. 7) by at least one amino acid substitution corresponding to position D59, I60, S62, P63, Q64, F65, Q66, S69, K70, V71, Y72, V73, W93, R95, N98, W143, A145, S146, T147, E148, Y151,

20   G159, I161, K163, G171, S179, Y181, S197, L199, Y214, E215 or R216.

In another aspect the invention includes a fusion protein, comprising; a protein of interest operably coupled to a functional red fluorescent protein whose sequence differs from the amino acid sequence of an Anthozoan red fluorescent protein (SEQ. ID. NO. 7) by at least one amino acid substitution corresponding to position D59, I60,

25   S62, P63, Q64, F65, Q66, S69, K70, V71, Y72, V73, W93, R95, N98, W143, A145, S146, T147, E148, Y151, G159, I161, K163, G171, S179, Y181, S197, L199, Y214, E215 or R216.

In one embodiment the invention includes a transgenic organism, comprising; a nucleic acid molecule encoding a functional red fluorescent protein whose sequence

30   differs from the amino acid sequence of an Anthozoan red fluorescent protein (SEQ. ID. NO. 7) by at least one amino acid substitution corresponding to position D59, I60, S62, P63, Q64, F65, Q66, S69, K70, V71, Y72, V73, W93, R95, N98, W143, A145,

10

S146, T. -7, E148, Y151, G159, I161, K163, G171, S179, Y181, S197, L199, Y214, E215 or R216.

In another aspect, the invention includes a method for identifying a protein – protein interaction, comprising;

5   a)   providing a population of cells comprising,

               a functional red fluorescent protein whose sequence differs from the
               amino acid sequence of an Anthozoan red fluorescent protein (SEQ.
               ID. NO. 7) by at least one amino acid substitution corresponding to
               position D59, I60, S62, P63, Q64, F65, Q66, S69, K70, V71, Y72,

10   V73, W93, R95, N98, W143, A145, S146, T147, E148, Y151, G159,
               I161, K163, G171, S179, Y181, S197, L199, Y214, E215 or R216,
               wherein said functional red fluorescent protein is operably coupled to a
               first protein of interest,

       b)   introducing a library of test proteins of interest operably coupled to

15   a functional green fluorescent protein into said population of cells,

                      wherein said functional green fluorescent protein and said
                      functional red fluorescent protein can undergo fluorescence
                      energy transfer (FRET), and
                      wherein each member of said population of cells receives on

20   average one member of said library of test proteins of interest
                      operably coupled to said functional green fluorescent protein,

       c)   screening said population of cells for FRET between said functional
               green fluorescent protein and said functional red fluorescent protein,
               and

25   d)   comparing the FRET in step c) to the FRET in a control cell in the
               absence of said library of test proteins of interest operably coupled to
               said functional green fluorescent protein.

In another embodiment, the invention includes a method for identifying a modulator of protein – protein interactions, comprising;

30   a)   contacting a cell with a test chemical, wherein said cell comprises,

11

i)   a functional red fluorescent protein whose sequence differs
     from the amino acid sequence of an Anthozoan red fluorescent
     protein (SEQ. ID. NO. 7) by at least one amino acid
     substitution corresponding to position D59, I60, S62, P63, Q64,

5
     F65, Q66, S69, K70, V71, Y72, V73, W93, R95, N98, W143,
     A145, S146, T147, E148, Y151, G159, I161, K163, G171,
     S179, Y181, S197, L199, Y214, E215 or R216, wherein said
     functional red fluorescent protein is operably coupled to a first
     protein of interest,

10   ii) a functional green fluorescent protein, wherein said functional
     green fluorescent protein is operably coupled to a second
     protein of interest, and wherein said functional green
     fluorescent protein and said functional red fluorescent protein
     undergo fluorescence energy transfer (FRET) when said first

15
     operably coupled protein of interest and said second operably
     protein of interest associate,

b)   detecting FRET between said functional green fluorescent protein and
     said functional red fluorescent protein in the presence of said test
     chemical, and

20   c)   comparing the FRET in step b) to the FRET in a control cell in the
     absence of said test chemical.

In one aspect of this method, the method further comprises the step of
contacting the cell with an activator prior to the addition of the test chemical. In
another aspect the method further includes the step of detecting the viability of the

25   cell.

In another embodiment the invention includes a test chemical and a
pharmaceutical composition comprising a test chemical identified by the methods
described herein.

BRIEF DESCRIPTION OF THE FIGURES

FIG. 1 Shows the mammalianized RFP created to provide for optimal codon usage and translational initiation in mammalian cells. Restriction sites, for insertion of

5   mutagenic oligonucleotides, are shown above the sequence.

FIG. 2. Shows the retroviral mammalian expression vector ABSC258. In this construct high-level mammalian expression is achieved via the strong viral CMV promoter.

10

FIG. 3. Shows the result of flow cytometry analysis of wild type and RFP expressing NIH3T6 cells.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

15

Definitions

The techniques and procedures are generally performed according to conventional methods in the art and various general references. (Lakowicz, J.R. Topics in Fluorescence Spectroscopy, (3 volumes) New York: Plenum Press (1991),

20   and Lakowicz, J. R. (1996) Scanning Microsc Suppl. $\underline{10}$ 213-24, for fluorescence techniques; Sambrook *et al.* (1989) Molecular Cloning: A Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., for molecular biology methods; Cells: A Laboratory Manual, 1st edition (1998) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., for cell biology methods; Optics Guide

25   5 Melles Griot® Irvine CA, and Optical Waveguide Theory, Snyder & Love published by Chapman & Hall for general optical methods, which are incorporated herein by reference).

"Activity" refers to the enzymatic or non-enzymatic activity capable of modifying an amino acid residue or peptide bond (preferably enzymatic). Such

30   covalent modifications include proteolysis, phosphorylation, dephosphorylation, glycosylation, methylation, sulfation, prenylation and ADP-ribsoylation. The term includes *non-covalent modifications including protein-protein interactions, and the*

13

binding of allosteric, or other modulators or second messengers such as calcium, or cAMP or inositol phosphates to a polypeptide.

Amino acid "substitutions" are defined as one for one amino acid replacements. They are conservative in nature when the substituted amino acid has

5   similar structural and/or chemical properties. Examples of conservative replacements are substitution of a leucine with an isoleucine or valine, an aspartate with a glutamate, or a threonine with a serine.

Amino acid "insertions" or "deletions" are changes to or within an amino acid sequence. They typically fall in the range of about 1 to 5 amino acids. The variation

10   allowed in a particular amino acid sequence may be experimentally determined by producing the peptide synthetically or by systematically making insertions, deletions, or substitutions of nucleotides in the gene sequence using recombinant DNA techniques.

"Animal" as used herein may be defined to include human, domestic (cats,

15   dogs, etc), agricultural (cows, horses, sheep, goats, chicken, fish, etc) or test species (frogs, mice, rats, rabbits, simians, etc).

"Chimeric" molecules are polynucleotides or polypeptides which are created by combining one or more nucleotide sequences of this invention (or their parts) with additional nucleic acid sequence(s). Such combined sequences may be introduced

20   into an appropriate vector and expressed to give rise to a chimeric polypeptide which may be expected to be different from the native molecule in one or more of the following characteristics: cellular location, distribution, ligand- binding affinities, interchain affinities, degradation/turnover rate, signaling, etc.

The terms "cleavage site" or "protease site" refers to the bond cleaved by the

25   protease (e.g. a scissile bond) and typically the surrounding three to four amino acids of either side of the bond.

"Control elements" or "regulatory sequences" are those non-translated regions of the gene or DNA such as enhancers, promoters, introns and 3' untranslated regions which interact with cellular proteins to carry out replication, transcription, and

30   translation. They may occur as boundary sequences or even split the gene. They function at the molecular level and along with regulatory genes are very important in development, growth, differentiation and aging processes.

14

"Corresponds to" refers to a polynucleotide sequence that is homologous (i.e., is identical, not strictly evolutionarily related) to all or a portion of a reference polynucleotide sequence, or that a polypeptide sequence is identical to all or a portion of a reference polypeptide sequence. In contradistinction, the term "complementary

5   to" is used herein to mean that the complementary sequence is homologous to all or a portion of a reference polynucleotide sequence. For illustration, the nucleotide sequence "TATAC" corresponds to a reference sequence "TATAC" and is complementary to a reference sequence "GTATA".

"Derivative" refers to those polypeptides which have been chemically

10   modified by such techniques as ubiquitination, labeling, pegylation (derivatization with polyethylene glycol), and chemical insertion or substitution of amino acids such as ornithine which do not normally occur in human proteins.

The term "engineered protease site" refers to a protease site that has been modified from the naturally existing sequence by at least one amino acid substitution.

15   The term "fluorescent property" refers to any one of the following, the molar extinction coefficient at an appropriate excitation wavelength, the fluorescent quantum efficiency, the shape of the excitation or emission spectrum, the excitation wavelength maximum, or the emission magnitude at any wavelength during, or at one or more times after excitation of the fluorescent moiety, the ratio of excitation

20   amplitudes at two different wavelengths, the ratio of emission amplitudes at two different wavelengths, the excited state lifetime, the fluorescent anisotropy or any other measurable property of a fluorescent moiety and the like. Preferably fluorescent property refers to fluorescence emission, or the fluorescence emission ratio at two or more wavelengths.

25   The term "homolog" refers to two sequences or parts thereof, that are greater than, or equal to 85% identical when optimally aligned using the ALIGN program. Homology or sequence identity refers to the following. Two amino acid sequences are homologous if there is a partial or complete identity between their sequences. For example, 85% homology means that 85% of the amino acids are identical when the

30   two sequences are aligned for maximum matching. Gaps (in either of the two sequences being matched) are allowed in maximizing matching; gap lengths of 5 or less are preferred with 2 or less being more preferred. Alternatively and preferably,

two protein sequences (or polypeptide sequences derived from them of at least 30 amino acids in length) are homologous, as this term is used herein, if they have an alignment score of more than 5 (in standard deviation units) using the program ALIGN with the mutation data matrix and a gap penalty of 6 or greater. See Dayhoff,

5   (1972) in Atlas of Protein Sequence and Structure 5, National Biomedical Research Foundation, 101-110, and Supplement 2 to this volume, pp. 1-10.

An "inhibitor" is a substance that retards or prevents a chemical or physiological reaction or response. Common inhibitors include but are not limited to antisense molecules, antibodies, antagonists and their derivatives.

10   "Isolated" refers to material removed from its original environment (e.g. the natural environment if it is naturally occurring), and thus is altered from its natural state. For example, an isolated polynucleotide could be part of a vector or a composition of matter, or could be contained within a cell, and still be "isolated" because that vector, composition of matter, or particular cell is not the original

15   environment of the polynucleotide.

The term "linker" or "linker moiety" refers to an amino acid, polypeptide or protein sequence that serves to operatively couple a fluorescent protein to a protein of interest or second fluorescent protein. Linkers typically comprise a single polypeptide chain that covalently couples the fluorescent protein to the protein of interest or

20   second fluorescent protein. Linkers may be of any size.

The term "modulates" refers to, either the partial or complete, enhancement or inhibition (e.g. attenuation of the rate or efficiency) of an activity or process.

The term "modulator" refers to a chemical compound (naturally occurring or non-naturally occurring), such as a biological macromolecule (e.g., nucleic acid,

25   protein, non-peptide, or organic molecule), or an extract made from biological materials such as bacteria, plants, fungi, or animal (particularly mammalian, including human) cells or tissues. Modulators are evaluated for potential activity as inhibitors or activators (directly or indirectly) of a biological process or processes (e.g., agonist, partial antagonist, partial agonist, inverse agonist, antagonist, antineoplastic agents,

30   cytotoxic agents, inhibitors of neoplastic transformation or cell proliferation, cell proliferation-promoting agents, and the like) by inclusion in screening assays

16

described herein. The activity of a modulator may be known, unknown or partially known.

"Naturally fluorescent protein" refers to proteins capable of forming a highly fluorescent, intrinsic chromophore either through the cyclization and oxidation of

5   internal amino acids within the protein or via the enzymatic addition of a fluorescent co-factor. Typically such chromophores can be spectrally resolved from weakly fluorescent amino acids such as tryptophan and tyrosine.

An "oligonucleotide" or "oligomer" is a stretch of nucleotide residues which has a sufficient number of bases to be used in a polymerase chain reaction (PCR), a

10   site directed mutagenesis reaction or a cassette to create a desired sequence element. These short sequences are based on (or designed from) genomic or cDNA sequences and are used to amplify, mutate or create particular sequence elements. Oligonucleotides or oligomers comprise portions of a DNA sequence having at least about 10 nucleotides and as many as about 50 nucleotides, preferably about 15 to 30

15   nucleotides. They are chemically synthesized and may also be used as probes.

An "oligopeptide" is a short stretch of amino acid residues and may be expressed from an oligonucleotide. It may be functionally equivalent to and either the same length as or considerably shorter than a "fragment ", "portion ", or "segment" of a polypeptide. Such sequences comprise a stretch of amino acid residues of at least

20   about 5 amino acids and often about 17 or more amino acids, typically at least about 9 to 13 amino acids, and of sufficient length to display biologic and/or immunogenic activity.

The term "operably linked" refers to a juxtaposition wherein the components so described are in a relationship permitting them to function in their intended

25   manner. A control sequence "operably linked" to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under conditions compatible with the control sequences.

The term "operably coupled" refers to a juxtaposition wherein the components so described are either directly or indirectly coupled. Examples of directly coupled

30   components include proteins that are translationally fused together. Examples of indirectly coupled components include proteins that can functionally associate either transiently, or persistently, through a binding interaction.

The term "polynucleotide" refers to a polymeric form of nucleotides of at least
10 bases in length, either ribonucleotides or deoxynucleotides. Modified forms and
analogs of either type of nucleotide are also included, as are ribonucleotides or
deoxynucleotides linked via novel bonds such as those described in U.S. Patent No.

5   5,532,130, European Patent Applications EP 0 839 830, EP 0 742 287, EP 0 285 057
and EP 0 694 559. The term includes single and double stranded forms of nucleotides,
or a mixture of single and double stranded regions. In addition, the polynucleotide
can be composed of triple-stranded regions comprising RNA or DNA or both RNA
and DNA. A polynucleotide may also contain one or more modified bases or DNA or

10   RNA backbones modified for stability or for other reasons. "Modified" bases include,
for example, tritylated bases and unusual bases such as inosine, as well as other
chemical or enzymatic modifications.

The term "polypeptide" refers to amino acids joined to each other by peptide
bonds or modified peptide bonds, i.e. peptide isosteres, and may contain amino acids

15   other than the 20 gene-encoded amino acids. The polypeptides may be modified by
either natural processes, such as posttranslational processing, or by chemical
modification techniques which are well known in the art. Modifications can occur
anywhere in a polypeptide, including the peptide backbone, the amino acid side-
chains and the amino or carboxyl termini. It will be appreciated that the same type of

20   modification may be present in the same or varying degrees at several sites in a given
polypeptide. Also, a given polypeptide may contain many types of modifications.
Modification include acetylation, acylation, ADP-ribosylation, amidation, covalent
attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a
nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative,

25   covalent attachment of a phosphatidylinositol, cross-linking, cyclization, disulfide
bond formation, demethylation, formation of covalent cross-links, formation of
cysteine, formation of pyroglutamate, formylation, gamma-carboxylation,
glycosylation, GPI anchor formation, hydroxylation, iodination, methylation,
myristolyation, oxidation, pergylation, proteolytic processing, phosphorylation,

30   prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of
amino acids to protein such as arginylation. (See Proteins- Structure and Molecular
Properties 2nd Ed., T.E. Creighton, W.H. Freeman and Company, New York (1993);

18

Posttranslational Covalent Modification of Proteins, B.C. Johnson, Ed., Academic
Pres, New York, pp. 1-12 (1983).

A "portion" or "fragment" of a polynucleotide or nucleic acid comprises all or
any part of the nucleotide sequence having fewer nucleotides than about 6 kb,
5   preferably fewer than about 1 kb which can be used as a probe. Such probes may be
labeled with reporter molecules using nick translation, Klenow fill-in reaction, PCR
or other methods well known in the art. After pretesting to optimize reaction
conditions and to eliminate false positives, nucleic acid probes may be used in
Southern, northern or in situ hybridizations to determine whether DNA or RNA
10   encoding the protein is present in a biological sample, cell type, tissue, organ or
organism.

"Probes" are nucleic acid sequences of variable length, preferably between at
least about 10 and as many as about 6,000 nucleotides, depending on use. They are
used in the detection of identical, similar, or complementary nucleic acid sequences.
15   Longer length probes are usually obtained from a natural or recombinant source, are
highly specific and much slower to hybridize than oligomers. They may be single- or
double-stranded and carefully designed to have specificity in PCR, hybridization
membrane-based, or ELISA-like technologies.

The term "recognition motif" refers to all or part of a polypeptide sequence
20   recognized by a post-translational modification activity to enable a polypeptide to
become modified by that post-translational modification activity. Typically, the
affinity of a protein, e.g. enzyme, for the recognition motif is about 1 mM (apparent
K_d), preferably a greater affinity of about 10 μM , more preferably, 1 μM or most
preferably has an apparent K_d of about 0.1 μM. The term is not meant to be limited to
25   optimal or preferred recognition motifs, but encompasses all sequences that can
specifically confer substrate recognition to a peptide. In some embodiments the
recognition motif is a phosphorylated recognition motif (e.g. includes a phosphate
group), or comprises other post-translationally modified residues.

"Recombinant nucleotide variants" are polynucleotides that encode a protein.
30   They may be synthesized by making use of the "redundancy" in the genetic code.
Various codon substitutions, such as the silent changes which produce specific
restriction sites or codon usage-specific mutations, may be introduced to optimize

19

cloning into a plasmid or viral vector or expression in a particular prokaryotic or
eukaryotic host system, respectively.

"Recombinant polypeptide variant" refers to any polypeptide which differs
from a naturally occurring polypeptide by amino acid insertions, deletions and/or

5   substitutions, created using recombinant DNA techniques. Guidance in determining
which amino acid residues may be replaced, added or deleted without abolishing
characteristics of interest may be found by comparing the sequence of a polypeptide
with that of related polypeptides and minimizing the number of amino acid sequence
changes made in highly conserved regions.

10 A "signal or leader sequence" is a short amino acid sequence which is or can
be used, when desired, to direct the polypeptide through a membrane of a cell. Such a
sequence may be naturally present on the polypeptides of the present invention or
provided from heterologous sources by recombinant DNA techniques.

A "standard" is a quantitative or qualitative measurement for comparison.

15 Preferably, it is based on a statistically appropriate number of samples and is created
to use as a basis of comparison when performing diagnostic assays, running clinical
trials, or following patient treatment profiles. The samples of a particular standard
may be normal or similarly abnormal.

The term "stringent hybridization conditions", refers to an overnight

20 incubation at 42 °C in a solution comprising 50 % formamide, 5x SSC (750 mM
NaCl, 75 mM sodium citrate), 50 mM sodium phosphate (pH 7.6), 5x Denhardt's
solution, 10 % dextran sulfate and 20 µg/ml denatured sheared salmon sperm DNA,
followed by washing the filters in 0.1x SSC at about 65 °C. Also contemplated are
nucleic acid molecules that hybridize to the polynucleotides of the present invention

25 at lower stringency hybridization conditions. Changes in the stringency of
hybridization and signal detection are primarily accomplished through the
manipulation of formamide concentration (lower percentages of formamide result in
lower stringency); salt conditions, or temperature. For example, lower stringency
conditions include an overnight incubation at 37 °C in a solution comprising 6x SSPE

30 (20X SSPE=3M NaCl; 0.2M NaH2PO4; 0.02M EDTA, pH 7.4), 0.5% SDS, 30 %
formamide, 100 µg/ml salmon sperm blocking DNA; followed by washes at 50 °C
with 1XSSPE, 0.1% SDS. In addition, to achieve even lower stringency, washes

performed following stringent hybridization can be done at higher salt concentrations
(e.g. 5X SSC). Variation in the above conditions may be accomplished through the
inclusion and / or substitution of alternative blocking reagents used to suppress
background in hybridization experiments. Typical blocking reagents include

5   Denhardt's reagent, BLOTTO, heparin, denatured salmon sperm DNA, and
commercially available proprietary formulations. The inclusion of specific blocking
reagents may require modification of the hybridization conditions described above,
due to problems with compatibility. A polynucleotide which hybridizes only to
polyA+ sequences (such as any 3' terminal polyA+ tract of a cDNA shown in the

10   sequence listing), or to a complementary stretch of T (or U) residues would not be
included in the definition of a "polynucleotide" since such a polynucleotide would
hybridize to any nucleic acid molecule containing a poly (A) stretch, or the
complement thereof.

The term "target" refers to a biochemical entity involved in a biological

15   process. Targets are typically proteins that play a useful role in the physiology or
biology of an organism. A therapeutic chemical binds to a target to alter or modulate
its function. As used herein, targets can include cell surface receptors, G-proteins,
kinases, ion channels, phopholipases, proteases and other proteins mentioned herein.

The term "test chemical" refers to a chemical to be tested by one or more

20   screening method(s) of the invention as a putative modulator. A test chemical can be
any chemical, such as an inorganic chemical, an organic chemical, a protein, a
peptide, a carbohydrate, a lipid, or a combination thereof. Usually, various
predetermined concentrations of test chemicals are used for screening, such as 0.01
micromolar, 1 micromolar and 10 micromolar. Test chemical controls can include the

25   measurement of a signal in the absence of the test compound or comparison to a
compound known to modulate the target.

The term "transgenic" is used to describe an organism that includes exogenous
genetic material within all of its cells. The term includes any organism whose genome has
been altered by *in vitro* manipulation of the early embryo or fertilized egg or by any

30   transgenic technology to induce a specific gene knockout.

The term "transgene" refers any piece of DNA which is inserted by artifice
into a cell, and becomes part of the genome of the organism (*i.e.*, either stably

integrated or as a stable extrachromosomal element) which develops from that cell.
Such a transgene may include a gene which is partly or entirely heterologous (*i.e.*,
foreign) to the transgenic organism, or may represent a gene homologous to an
endogenous gene of the organism. Included within this definition is a transgene

5   created by the providing of an RNA sequence that is transcribed into DNA and then
incorporated into the genome. The transgenes of the invention include DNA
sequences that encode the functional red fluorescent proteins that may be expressed in
a transgenic non-human animal.

The following terms are used to describe the sequence relationships between

10 two or more polynucleotides: "reference sequence", "comparison window", "sequence
identity", "percentage identical to a sequence", and "substantial identity". A
"reference sequence" is a defined sequence used as a basis for a sequence comparison;
a reference sequence may be a subset of a larger sequence, for example, as a segment
of a full-length cDNA or may comprise a complete cDNA or gene sequence.

15 Generally, a reference sequence is at least 20 nucleotides in length, frequently at least
25 nucleotides in length, and often at least 50 nucleotides in length. Since two
polynucleotides may each (1) comprise a sequence (i.e., a portion of the complete
polynucleotide sequence) that is similar between the two polynucleotides, and (2) may
further comprise a sequence that is divergent between the two polynucleotides,

20 sequence comparisons between two (or more) polynucleotides are typically performed
by comparing sequences of the two polynucleotides over a "comparison window" to
identify and compare local regions of sequence similarity. A "comparison window",
as used herein, refers to a conceptual segment of at least 20 contiguous nucleotide
positions wherein a polynucleotide sequence may be compared to a reference

25 sequence of at least 20 contiguous nucleotides and wherein the portion of the
polynucleotide sequence in the comparison window may comprise additions or
deletions (i.e., gaps) of 20 percent or less as compared to the reference sequence
(which does not comprise additions or deletions) for optimal alignment of the two
sequences. Optimal alignment of sequences for aligning a comparison window may

30 be conducted by the local homology algorithm of Smith and Waterman (1981) Adv.
Appl. Math. 2: 482, by the homology alignment algorithm of Needleman and Wunsch
(1970) J. Mol. Biol. 48: 443, by the search for similarity method of Pearson and

22

Lipman (1988) Proc. Natl. Acad. Sci. (U.S.A.) 85: 2444, by computerized

implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the

Wisconsin Genetics Software Package Release 7.0, Genetics Computer Group, 575

Science Dr., Madison, WI), or by inspection, and the best alignment (i.e., resulting in

5   the highest percentage of homology over the comparison window) generated by the

various methods selected. The term "sequence identity" means that two

polynucleotide sequences are identical (i.e., on a nucleotide-by-nucleotide basis) over

the window of comparison. The term "percentage identical to a sequence" is

calculated by comparing two optimally aligned sequences over the window of

10   comparison, determining the number of positions at which the identical nucleic acid

base (e.g., A, T, C, G, U, or I) occurs in both sequences to yield the number of

matched positions, dividing the number of matched positions by the total number of

positions in the window of comparison (i.e., the window size), and multiplying the

result by 100 to yield the percentage of sequence identity. The terms "substantial

15   identity" as used herein denotes a characteristic of a polynucleotide sequence, wherein

the polynucleotide comprises a sequence that has at least 30 percent sequence

identity, preferably at least 50 to 60 percent sequence identity, more usually at least

60 percent sequence identity as compared to a reference sequence over a comparison

window of at least 20 nucleotide positions, frequently over a window of at least 25-50

20   nucleotides, wherein the percentage of sequence identity is calculated by comparing

the reference sequence to the polynucleotide sequence which may include deletions or

additions which total 20 percent or less of the reference sequence over the window of

comparison.

        As applied to polypeptides, the term "substantial identity" means that two

25   peptide sequences, when optimally aligned, such as by the programs GAP or

BESTFIT using default gap weights, share at least 30 percent sequence identity,

preferably at least 40 percent sequence identity, more preferably at least 50 percent

sequence identity, and most preferably at least 60 percent sequence identity.

Preferably, residue positions which are not identical differ by conservative amino acid

30   substitutions. Conservative amino acid substitutions refer to the interchangeability of

residues having similar side chains. For example, a group of amino acids having

aliphatic side chains is glycine, alanine, valine, leucine, and isoleucine; a group of

23

amino acids having aliphatic-hydroxyl side chains is serine and threonine; a group of amino acids having amide-containing side chains is asparagine and glutamine; a group of amino acids having aromatic side chains is phenylalanine, tyrosine, and tryptophan; a group of amino acids having basic side chains is lysine, arginine, and

5   histidine; and a group of amino acids having sulfur-containing side chains is cysteine and methionine. Preferred conservative amino acids substitution groups are: valine-leucine-isoleucine, phenylalanine-tyrosine, lysine-arginine, alanine-valine, glutamic-aspartic, and asparagine-glutamine.

Since the list of technical and scientific terms cannot be all encompassing, any

10   undefined terms shall be construed to have the same meaning as is commonly understood by one of skill in the art to which this invention belongs. Furthermore, the singular forms "a", "an" and "the" include plural referents unless the context clearly dictates otherwise. For example, reference to a "restriction enzyme" or a "high fidelity enzyme" may include mixtures of such enzymes and any other enzymes

15   fitting the stated criteria, or reference to the method includes reference to one or more methods for obtaining cDNA sequences which will be known to those skilled in the art or will become known to them upon reading this specification.

Before the present sequences, variants, formulations and methods for making and using the invention are described, it is to be understood that the invention is not to

20   be limited only to the particular sequences, variants, formulations or methods described. The sequences, variants, formulations and methodologies may vary, and the terminology used herein is for the purpose of describing particular embodiments. The terminology and definitions are not intended to be limiting since the scope of protection will ultimately depend upon the claims.

25

I.   RED FLUORESCENT PROTEINS

Anthozoan fluorescent proteins (SEQ. ID. NOs 1 to 7) isolated from various species of coral display a range of fluorescence properties (**Table 2**) ranging from

30   green fluorescent to red fluorescent emission. Compared to *Aequorea victoria* GFP, the Anthozoan fluorescent proteins exhibit overall sequence identities of between 26 to 30 % identity.

TABLE 2					
Anthozoa Fluorescent Proteins					
Species	Protein Name	Quantum Yield (Φ) & Molar Extinction (ε)	Excitation & Emission Max	Relative Brightness	SEQ. ID. NO.:
Anemonia majano	amFP486	Φ = 0.24 ε = 40,000	458 486	0.43	SEQ. ID. NO.: 1
Zoanthus sp	zFP506	Φ = 0.63 ε = 35,600 Φ = 0.42	496, 506	1.02	SEQ. ID. NO.:2
	zFP538	ε = 20,???	528, 538	0.38	SEQ. ID. NO.:3
Discosoma striata	dsFP483	Φ ? ??? ε ? 23,900	443 483	0.5	SEQ. ID. NO.:4
Discosoma sp "red"	drFP583	Φ =0.23 ε = 22,500	558 583	0.24	SEQ. ID. NO.:5
Clavularia sp	CFP484	Φ =0.48 ε = 35,300	456 484	0.77	SEQ. ID. NO.:6

5   In spite of the relatively low sequence identity, the alignment of the
Anthozoan and *Aequorea* fluorescent proteins is consistent with the possibility that
both types of protein share a common overall structural orientation and protein fold. A
comparison of the sequences reveals a tendency for amino acids to alternate between
hydrophobic and hydrophilic residues along β-strands, and for the conservation of

10   buried hydrophobic core residues, as well as turn motifs.

        Compared to *Aequorea* GFP, the red Anthozoan fluorescent proteins have
relatively low quantum yields and molar extinction coefficients resulting in proteins
that exhibit an overall brightness of approximately one quarter of that of wild type
Aequorea GFP. The broad excitation and emission spectra of the wild type red

15   fluorescent proteins makes it difficult   selectively excite or observe the proteins for
multiplexed analysis or FRET applications.

II.   DESIGN OF FUNCTIONAL RED FLUORESCENT PROTEIN MUTANTS

To design improved mutants of the Anthozoan red fluorescent proteins a
synthetic protein (SEQ. ID. NO. 8 (nucleotide sequence), & SEQ. ID. NO. 9 (amino
5   acid sequence) was constructed which provided for the ability to clone in a series of
oligonucleotides containing randomized nucleic acid sequences at key positions in the
red fluorescent protein (FIG.1).

In order to produce functional red fluorescent proteins capable of high level
expression in mammalian cells, a synthetic gene encoding the coding region was
10   produced. This sequence contained an additional amino acid (valine) after the start
methionine to provide for an optimal Kozak sequence and high level translational
initiation. The synthetic red fluorescent protein (SEQ. ID. NO. 9) was constructed by
systematically replacing the wild-type codons with codons most frequently used in
highly expressed human genes (see U.S. Patent No. 5,795,737, issued August 18,
15   1998). This synthetic gene was assembled from chemically synthesized
oligonucleotides of 70 to 100 bases in length using standard molecular biology
methodology. Single stranded oligonucleotide pools were PCR amplified before
cloning, and the PCR products purified in agarose gels and used as templates in the
next PCR step. Two adjacent fragments were then co-amplified via the use of
20   overlapping sequences at the end of either fragment to build larger fragments. These
fragments which were between 350 and 400 bp in size, were sequentially subcloned to
assemble the entire gene, FIG 1. (Synthetic Genetics, San Diego California) The
synthetic gene (SEQ. ID. NO. 9) was then sequenced, and subcloned into the
retroviral expression vector ABSC258 (FIG. 2).

25   Retroviral expression vectors provide for highly efficient gene transfer to
mammalian cells and stable long-term expression. These characteristics are important
to ensure that libraries of mutant red fluorescent proteins can be efficiently introduced
into mammalian cells and subsequently analyzed and sequenced.

Mutagenesis of the synthetic gene was completed by sub-cloning mutagenic
30   double stranded oligonucleotide sequences into the synthetic gene (SEQ. ID. NO. 9).
These oligonucleotides enabled defined regions of the protein to be targeted for
mutagenesis enabling the conservation of the overall structural framework of the
protein to remain intact. These oligonucleotides (Table 3) were designed to be

cassetted into the engineered restriction sites incorporated during synthesis of the synthetic gene.

TABLE 3													
Relative position in GFP	drFP583	dsFP483, GFP	Degenerate codon bp Upper case indicates 90% probability, lower case indicates 10 % probability			1st row = Amino acids generated from the selected degenerate codon; 2nd row = Codon used; 3rd row = Probability							
58	D59	H, P	G	A	C	D	A	H	P				
			c	c		GAC	GcG	Cac	ccC				
						0.81	0.09	0.09	0.01				
59	I60	T	A	T	C	I	T						
				c		ATC	AcC						
						0.9	0.1						
61	S62	C, V	A	G	C	S	C						
			t			AGC	Tgc						
						0.9	0.1						
62	P63	T	C	C	C	P	T						
			a			CCC	Acc						
						0.9	0.1						
63	Q64	T	C	A	G	Q	K	P	T				
			a	c		CAG	AaG	CgG	acG				
						0.81	0.09	0.09	0.01				
64	F65	L	T	T	C	F	V	I	L				
			a, g		g	TTC	gTg, gTC	Atc	TTg				
						0.72	0.1	0.09	0.08				
65	Q66	S	C	A	G	Q	R	P	K	E	T	A	G
			a, g	c, g		CAG	CgC, agG	CcG	Aag	Gag	acG	gcG	ggG
						0.64	0.09	0.08	0.08	0.08	0.01	0.01	0.01
68	S69	N,V, L	T	C	G	S	L	A	V				
			g	t		TCG	TtG	gCG	gtG				
						0.81	0.09	0.09	0.01				

69	K70	Q, M	A	A	G	K	M	Q	L			
			c	t		AAG	AtG	cAG	ctG			
						0.81	0.09	0.09	0.01			
70	V71	A, C	G	T	C	V	C					
				c		GTC	GcC					
						0.9	0.1					
71	Y72	F	T	A	C	Y	F					
				t		TAC	TtC					
						0.9	0.1					
72	V73	S, A	G	T	G	V	A	L	S			
			t	c		GTG	GcG	tTG	tcG			
						0.81	0.09	0.09	0.01			

Second Mutagenic Section

94	W93	Q	T	G	G	W	L	C	F			
			t	c		TGG	TtG	TGc	Ttc			
						0.81	0.09	0.09	0.01			
96	R95	K	A	G	G	R	K					
			a			AGG	AaG					
						0.9	0.1					
99	N98	H,F,S	A	A	C	N	T	D	A			
			g	c		AAC	AcC	gAC	gcC			
						0.81	0.09	0.09	0.01			

Third Mutagenic Section

145	W143	Y, F	T	G	G	W	L	C	F			
			t	c		TGG	TtG	TGc	Ttc			
						0.81	0.09	0.09	0.01			
147	A145	S, P	G	C	C	A	P	S				
			c, t			GCC	cCC	tCC				
						0.8	0.1	0.1				
148	S146	H, G	A	G	C	S	R	G	N	H	D	
			c, g	a		AGC	cGC	gGC	AaC	caC	gaC	
						0.72	0.09	0.09	0.08	0.01	0.01	

149	T147	N, K	A	C	C	T	N	K					
				a	G	ACC, ACG	AaC	AaG					
						0.9	0.05	0.05					
150	E148	V	G	A	G	E	V						
				t		GAG	GtG						
						0.9	0.1						
153	Y151	M, T	T	A	C	Y	N	D	S	T	A		
			a, g	c		TAC	aAC	gAC	TcC	acC	gcC		
						0.72	0.09	0.09	0.08	0.01	0.01		
Fourth Mutagenic Section													
163	G159	V, A	G	G	C	G	A	V					
				c, t		GGC	GcC	GtC					
						0.8	0.1	0.1					
165	I161	F	A	T	C	I	V	F	M	L			
			t, g		g	ATC	gTC, gTg	tTC	ATg	tTg			
						0.72	0.1	0.09	0.08	0.01			
167	K163	I,T,V	A	A	A	K	I	R	T	E	V	G	A
			g	t,g,c		AAA	AtA	AgA	AcA	gAA	gtA	ggA	gcA
						0.63	0.09	0.09	0.09	0.07	0.01	0.01	0.01
175	G171	S	G	G	C	G	S						
			a			GGC	aGC						
						0.9	0.1						
Fifth Mutagenic Section													
183	S179	Q	T	C	A	S	A	P	T	stop	E	Q	K
			g,c,a	a		TCA	gCA	cCA	aCA	TaA	gaA	caA	aaa
						0.63	0.09	0.09	0.09	0.07	0.01	0.01	0.01

185	Y181	N	T	A	C	Y	F	N	I			
			a	t		TAC	TtC	AaC	atC			
						0.81	0.09	0.09	0.01			
Sixth Mutagenic Section												
203	S197	T, Y	T	C	C	S	Y	T	N			
			a	a		TCC	TaC	aCC	aaC			
						0.81	0.09	0.09	0.01			
205	L199	S	T	T	G	L	S					
				c		TTG	TcG					
						0.9	0.1					
Seventh Mutagenic Section												
221	Y214	L, H	T	A	C	Y	F	H	L			
			c	t		TAC	TtC	cAC	ctC			
						0.81	0.09	0.09	0.01			
222	E215	Q, G	G	A	G	E	G	Q	R			
			c	g		GAG	GgG	cAG	cgG			
						0.81	0.09	0.09	0.01			
223	R216	F	C	G	C	R	L	C	F			
			t	t		CGC	CtC	tGC	ttC			
						0.81	0.09	0.09	0.01			

This approach enables the controlled mutagenesis of key residues in the protein molecule, without the disruption of essential residues, that would otherwise lead to the complete loss of fluorescence. Importantly, the method enables selective control of the first, second, and third position of the codon, thereby enabling the selection of conservative mutations if desired.

To identify key residues in the red fluorescent protein, comparisons were made to known favorable mutations in *Aequorea* GFP, and divergences in the sequences between the various species of Anthozoan GFPs, and particularly the red (drFP583) (SEQ. ID. NO. 5 - nucleic acid, & SEQ. ID. NO. 7 - amino acid) and green (dsFP483) (SEQ. ID. NO. 4) fluorescent proteins from *Discosoma striata*. In **Table 3**, the amino acid positions refer to Aequorea GFP numbering. The corresponding numbering of the equivalent amino acids in the Anthozoan GFPs

30

To maximize the chance of identifying mutations that confer a favorable characteristic, the level of mutagenesis was designed to result in the wild type amino acid being present at each position mutagenized approximately 80% of the time. This approach (often termed soft mutagenesis) helps to avoid the creation of libraries

5   containing mostly non-functional mutants, a situation that can arise if a protein is relatively sensitive to alterations in its amino acid composition.

To ensure that the entire library of mutants was screened, the mutagenesis was completed in a systematic step by step process. This process limited the total diversity in each library to an acceptable value that could be practically screened in mammalian

10   cells via flow cytometry. For example in **Table 3**, the first mutagenic primer has a total diversity of around 1.05×10^6, compared to the diversity of the entire library which is of the order of 3.42×10^{17}. Typical commercially available FACS instrumentation have analysis rates of around $2 - 5 \times 10^4$ cells / second, making a realistic analysis of the entire library impractical in a reasonable time frame. By

15   contrast, a screen of a library of about a million cells is relatively easily accomplished, and can, furthermore, be sorted several times over to ensure that the relatively rare, favorable mutations are identified.

III.   SCREENING OF LIBRARIES

20

Once the mutagenic library of red fluorescent mutants has been subcloned into the retroviral expression vector, a library of retroviral plasmids can be produced using standard packaging cell lines, such as PT67 cells. Supernatant from these cells can then used to infect the mammalian cells in order to express the mutant fluorescent

25   proteins.

Favorable mutants from this step can be identified by FACS analysis based on their improved fluorescence characteristics and increased brightness when expressed in mammalian cells. Typically cells will be selected based on their brightness (fluorescence emission) around 583 nm when excited at around 558 nm.

30   In **FIG. 3**, flow cytometry and cell sorting were conducted using a Becton Dickinson FACSVantage™ SE with a Coherent Innova[R] 70C Spectrum laser producing 60 mW of power at 530.9 nm excitation. The flow cytometer was equipped

with pulse processing and the Macrosort™ flow cell. Fluorescence emission was
detected via a 585/42 nm bandpass emission filter, separated by a 560 nm short path
dichroic mirror. Using the CloneCyt™ Plus integrated deposition system on the
FACSVantage™ SE, single cells were sorted into 96-well microtiter plates based on

5   fluorescence intensity (R3) above cellular autofluorescence from a wild type control
population. In **FIG. 3**, wild type NIH3T6 cells are shown in the upper panel, while
cells transformed with the RFP expression vector ABSC258, are shown in the lower
panel. The R3 region represents cells with higher levels of red fluorescence than
cellular autofluorescence, and these cells were sorted into 96 well plates for further

10   analysis. In this experiment the sort region (R3) = 0.001% of the total population in
the wild type cells, and 1.40% in the RFP transformed cells.

       In addition, multiple rounds of FACS analysis and sorting can be used to
selectively enrich mixed pools of brighter mutants to enable the selection of the best
mutants. An additional aspect of this strategy is to re-sort the fluorescent cells based

15   on their brightness when excited at 488 nm or 530 nm. In this case one would select
for cells with reduced brightness when excited at these wavelengths in order to select
for mutants with narrower, sharper excitation peaks.

       Another useful sort strategy is to analyze the cells relatively rapidly (i.e.
within 24 hours) after transformation in order to identify functional red fluorescent

20   proteins that exhibit more rapid autocatalytic fluorescence development.

       Typically after FACS separation, individual cells, or enriched populations of
cells, can be sorted into culture plates and allowed to recover for a period of about
two weeks. After this period individual cell colonies are typically large enough for
further analysis either by further rounds of FACS or via a 96 well plate reader.

25   Analysis via a plate reader provides for accurate quantification and enables a
determination of the relative magnitude of the excitation peaks at 487 nm, 530 nm and
558 nm in the same sample. Once colonies expressing mutants with improved
characteristics are identified, the sequences of the mutants can be rapidly identified
via PCR based sequencing. This can be achived, for example, by using standard

30   fluorescent dye terminator chemistries on a Perkin Elmer 373 or similar automated
sequencer, using direct sequencing of PCR products as described by Townley *et al.*,
(1997) Genome Res. 7: (3) 293-8. Methods for DNA sequencing are well known in

32

the art and employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE™ (US Biochemical Corp) or Taq polymerase. Methods to extend the DNA from an oligonucleotide primer annealed to the DNA template of interest have been developed for both single- and double-stranded templates. Chain termination

5   reaction products are separated using electrophoresis and detected via their incorporated, labeled precursors.

        Recent improvements in mechanized reaction preparation, sequencing and analysis have permitted expansion in the number of sequences that can be determined per day. Preferably, the process is automated with machines such as the Hamilton

10   Micro Lab 2200 (Hamilton, Reno, Nev.), Peltier Thermal Cycler (PTC200; MJ Research, Watertown Mass.) and the Applied Biosystems Catalyst 800 and 377 and 373 DNA sequencers.

        The best mutants from this first round of mutagenesis may then be used as the starting product for the next round of mutagenesis. As previously described,

15   oligonucleotides containing a reasonable total diversity are selected to ensure that a complete and thorough search of all of the mutants can be rapidly completed.

        After all the mutagenic steps have been completed it is possible to further enhance the fluorescence properties via the use of error prone PCR or Ping pong mutagenesis approaches using methods known in the art to create a highly optimized

20   red fluorescent protein.

        Another mutagenesis step may then be completed by recombining the entire pool of favorable mutants to select the most favorable combinations. In this approach the probability of mutagenesis at each position is approximately 50 %, and all the mutations have an equal probability of incorporation into the template fluorescent

25   protein. The most favorable combinations of mutations are then selected to provide for the greatest improvements in brightness and fluorescent properties.

IV.   USE AS A MARKER OF GENE EXPRESSION AND CELL MOVEMENT

30   Typically the functional red fluorescent proteins of the present invention will be introduced and expressed in target cells via the use of standard molecular biology techniques known in the art.

33

For cell movement studies, expression of the red fluorescent protein will generally be driven via a cell-type specific promoter, in order to be able to selectively monitor the movement of the target cell type. In some cases, for example in cell mixing experiments, it will be preferred for expression to be driven via a constitutive promoter, in other cases it may be preferable to drive expression from an inducible, or developmentally regulated promoter in order to monitor cellular differentiation.

In another embodiment it may be desirable to include additional spectrally resolved fluorescent proteins to simultaneously track both cell movement and differentiation in order to determine both when and where gene expression is modulated. In both cases, nucleic acids in the form of an expression vector including expression control sequences operatively linked to a nucleotide sequence coding for expression of the red fluorescent protein will be used for introducing the proteins into cells. As used, the term "nucleotide sequence coding for expression of" a polypeptide refers to a sequence that, upon transcription and translation of mRNA, produces the polypeptide. This can include sequences containing, *e.g.*, introns. As used herein, the term "expression control sequences" refers to nucleic acid sequences that regulate the expression of a nucleic acid sequence to which it is operatively linked. Expression control sequences are operatively linked to a nucleic acid sequence when the expression control sequences control and regulate the transcription and, as appropriate, translation of the nucleic acid sequence. Thus, expression control sequences can include appropriate promoters, enhancers, transcription terminators, a start codon (*i.e.*, ATG) in front of a protein-encoding gene, splicing signals for introns, IRES sequences (internal ribosome entry site) maintenance of the correct reading frame of that gene to permit proper translation of the mRNA, and stop codons.

Methods that are well known to those skilled in the art can be used to construct expression vectors containing the red fluorescent proteins. These methods include *in vitro* recombinant DNA techniques, synthetic techniques and *in vivo* recombination/genetic recombination. (See, for example, the techniques described in Maniatis, *et al.*,(1989) Cold Spring Harbor Laboratory, N.Y.). Many commercially available expression vectors are available from a variety of sources including Clontech (Palo Alto, CA), Stratagene (San Diego, CA) and Invitrogen (San Diego, CA) as well as many other commercial sources.

34

A contemplated version of the method is to use inducible controlling

nucleotide sequences to produce a sudden increase in the expression of the RFP

construct e.g., by inducing expression of the construct. Examplary inducible systems

include the tetracycline inducible system first described by Bujard and colleagues

5   (Gossen and Bujard (1992) Proc. Natl. Acad. Sci USA 89 5547-5551, Gossen *et al.*

(1995) Science 268 1766-1769) and described in U.S. Patent No 5,464,758.

Transformation of cells

Transformation of a host cell with recombinant DNA may be carried out by

10   conventional techniques as are well known to those skilled in the art. Where the host

is prokaryotic, such as *E. coli*, competent cells that are capable of DNA uptake can be

prepared from cells harvested after exponential growth phase and subsequently treated

by the $CaCl_2$ method by procedures well known in the art. Alternatively, $MgCl_2$ or

RbCl can be used. Transformation can also be performed after forming a protoplast

15   of the host cell or by electroporation.

When the host is an eukaryote, such methods of transfection of DNA as

calcium phosphate co-precipitates, conventional mechanical procedures such as

microinjection, electroporation, insertion of a plasmid encased in liposomes, or virus

vectors may be used. Eukaryotic cells can also be co-transfected with DNA

20   sequences encoding the fusion polypeptide of the invention, and a second foreign

DNA molecule encoding a selectable phenotype, such as the herpes simplex

thymidine kinase gene. Another method is to use an eukaryotic viral vector, such as

simian virus 40 (SV40) or bovine papilloma virus, to transiently infect or transform

eukaryotic cells and express the protein. (*Eukaryotic Viral Vectors*, Cold Spring

25   Harbor Laboratory, Gluzman ed., 1982). Preferably, an eukaryotic host is utilized as

the host cell as described herein.

V.   USE AS A FUSION TAG

30   The functional red fluorescent proteins of this invention are useful to track the

movement of proteins in cells. In this embodiment, a nucleic acid molecule encoding

the fluorescent protein is fused in frame to a nucleic acid molecule encoding the

protein of interest in an expression vector. Upon expression inside the cell, the protein of interest can be localized based on fluorescence. Typically the protein of interest would be coupled to the RFP via a flexible linker to ensure that both the target protein and fluorescent protein functioned correctly and were efficiently folded.

5   Methods for constructing and introducing such fusion proteins are well known in the art and are also discussed above.

In another version, two or more proteins of interest are simultaneously tracked by fusing the first protein with a functional red fluorescent protein, and the second protein fused to a second fluorescent protein, such as one of the proteins listed in

10   **Table 1**. Typically the second fluorescent protein is chosen based on its fluorescent properties so that it can be spectrally resolved from the functional red fluorescent protein.

VI.   USE IN TRANSGENIC ORGANISMS

15

In one embodiment, the invention provides a transgenic non-human organism that expresses a nucleic acid sequence that encodes a functional red fluorescent protein. Because such constructs can be expressed within intact living organisms without the need to add co-factors or reagents, and the red emission passes well

20   through tissues, the red fluorescent proteins enable the monitoring of cell movement and differentiation within the entire, intact, living organism.

In another embodiment, the invention can be used to identify where in specific tissues a particular cell type is located, for example, by expression of a red fluorescent protein from a tissue or cell type specific promoter. In another embodiment it may be

25   desirable to include additional spectrally resolved fluorescent proteins to simultaneously track both cell movement and differentiation in order to determine both when and where gene expression is modulated. Such non-human organisms include vertebrates such as rodents, fish such as Zebrafish, non-human primates and reptiles as well as invertebrates. Preferred non-human organisms are selected from

30   the rodent family including rat and mouse, most preferably mouse. The transgenic non-human organisms of the invention are produced by introducing transgenes into the germline of the non-human organism. Embryonic target cells at various

36

developmental stages can be used to introduce transgenes. Different methods are used depending on the organism and stage of development of the embryonic target cell. In vertebrates, the zygote is the best target for microinjection. In the mouse, the male pronucleus reaches the size of approximately 20 micrometers in diameter, which

5   allows reproducible injection of 1-2 pl of DNA solution. The use of zygotes as a target for gene transfer has a major advantage in that in most cases the injected DNA will be incorporated into the host gene before the first cleavage (Brinster *et al.*, (1985) Proc. Natl. Acad. Sci. USA 82 4438-4442,). As a consequence, all cells of the transgenic non-human animal will carry the incorporated transgene. This will in general also be

10   reflected in the efficient transmission of the transgene to offspring of the founder since 50% of the germ cells will harbor the transgene. Microinjection of zygotes is the preferred method for incorporating transgenes in practicing the invention.

A transgenic organism can be produced by cross-breeding two chimeric organisms which include exogenous genetic material within cells used in

15   reproduction. Twenty-five percent of the resulting offspring will be transgenic *i.e.*, organisms that include the exogenous genetic material within all of their cells in both alleles. 50% of the resulting organisms will include the exogenous genetic material within one allele and 25% will include no exogenous genetic material.

Retroviral infection can also be used to introduce transgene into a non-human

20   organism. In vertebrates, the developing non-human embryo can be cultured *in vitro* to the blastocyst stage. During this time, the blastomeres can be targets for retro viral infection (Jaenich, R., (1976) Proc. Natl. Acad. Sci USA 73 1260-1264,). Efficient infection of the blastomeres is obtained by enzymatic treatment to remove the zona pellucida (Hogan, *et al.* (1986) in Manipulating the Mouse Embryo, Cold Spring

25   Harbor Laboratory Press, Cold Spring Harbor, N.Y.). The viral vector system used to introduce the transgene is typically a replication-defective retrovirus carrying the transgene (Jahner, *et al.*, (1985) Proc. Natl. Acad. Sci. USA 82 6927-6931; Van der Putten, *et al.*, (1985) Proc. Natl. Acad. Sci USA 82 6148-6152). Transfection is easily and efficiently obtained by culturing the blastomeres on a monolayer of

30   virus-producing cells (Van der Putten, *supra*; Stewart, *et al.*, (1987) EMBO J. 6 383-388).

37

Alternatively, infection can be performed at a later stage. Virus or virus-producing cells can be injected into the blastocoele (D. Jahner *et al.*, (1982) Nature 298 623-628). Most of the founders will be mosaic for the transgene since incorporation occurs only in a subset of the cells that formed the transgenic nonhuman

5   animal. Further, the founder may contain various retro viral insertions of the transgene at different positions in the genome that generally will segregate in the offspring. In addition, it is also possible to introduce transgenes into the germ line, albeit with low efficiency, by intrauterine retroviral infection of the midgestation embryo (D. Jahner *et al.*, *supra*). A third type of target cell for transgene introduction for vertebrates is

10 the embryonic stem cell (ES). ES cells are obtained from pre-implantation embryos cultured *in vitro* and fused with embryos (M. J. Evans *et al.* (1981) Nature 292 154-156; M.O. Bradley *et al.*, (1984) Nature 309 255-258; Gossler, *et al.*, (1986) Proc. Natl. Acad. Sci USA 83 9065-9069; and Robertson *et al.*, (1986) Nature 322 445-448). Transgenes can be efficiently introduced into the ES cells by DNA

15 transfection or by retro virus-mediated transduction. Such transformed ES cells can thereafter be combined with blastocysts from a nonhuman animal. The ES cells thereafter colonize the embryo and contribute to the germ line of the resulting chimeric animal. (For review see Jaenisch, R., (1988) Science 240 1468-1474).

In another embodiment, the invention provides a transgenic plant that

20 expresses a nucleic acid sequence that encodes red fluorescent protein. Because such constructs can be specifically expressed, both spatially and temporally, within intact living cells, the invention provides the ability to monitor the spatial distribution of a target cell type, within defined cell populations, tissues, or in the entire transgenic plant.

25   In another embodiment, the approach can be used to specifically identify where in specific tissues a particular gene is expressed, for example by expression of the RFP from tissue specific plant promoters.

In another embodiment it may be desirable to include additional spectrally resolved fluorescent proteins to simultaneously track both cell movement and

30 differentiation in order to determine both when and where gene expression is modulated.

38

Transgenic plants may be produced by any one of a number of methods of plant transformation and regeneration. Numerous methods for plant transformation have been developed, including biological and physical, plant transformation protocols. See, for example, Miki et al., "Procedures for Introducing Foreign DNA

5   into Plants" in Methods in Plant Molecular Biology and Biotechnology, Glick, B.R. and Thompson, J.E. Eds. (CRC Press, Inc. , Boca Raton, 1993) pages 67-88. In addition, expression vectors and in vitro culture methods for plant cell or tissue transformation and regeneration of plants are available. See, for example, Gruber et al., "Vectors for Plant Transformation" in Methods in Plant Molecular Biology and

10   Biotechnology, Glick, B.R. and Thompson, J.E. Eds. (CRC Press, Inc., Boca Raton, 1993) pages 89-119.

The most widely utilized method for introducing an expression vector into plants is based on the natural transformation system of Agrobacterium. See, for example, Horsch et al., (1985) Science 227 1229. A. tumefaciens and A. rhizogenes

15   are plant pathogenic soil bacteria which genetically transform plant cells. The Ti and Ri plasmids of A. tumefaciens and A. rhizogenes, respectively, carry genes responsible for genetic transformation of the plant See, for example, Kado, C.I., Crit. Rev. Plant. Sci. 10: 1 (1991). Descriptions of Agrobacterium vector systems and methods for Agrobacterium-mediated gene transfer are provided by Gruber et al.,

20   supra, Miki et al., supra, and Moloney et al., (1989) Plant Cell Reports 8 238.

Despite the fact the host range for Agrobacterium mediated transformation is broad, some major cereal crop species and gymnosperms have generally been recalcitrant to this mode of gene transfer, even though some success has recently been achieved in rice. Hiei et al., (1994) The Plant Journal 6 271-282. Several methods of

25   plant transformation, collectively referred to as direct gene transfer, have been developed as an alternative to Agrobacterium-mediated transformation.

A generally applicable method of plant transformation is microprojectile-mediated transformation wherein DNA is carried on the surface of microprojectiles measuring 1 to 4 Am. The expression vector is introduced into plant tissues with a

30   biolistic device that accelerates the microprojectiles to speeds of 300 to 600 m/s which is sufficient to penetrate plant cell walls and membranes. Sanford et al., (1987),

Part. Sci. Technol. 5 27, Sanford, J.C., (1988) Trends Biotech. 6 299, Sanford, J.C., (1990) Physiol. Plant 79 206, Klein et al., (1992) Biotechnology 10 268.

Another method for physical delivery of DNA to plants is sonication of target cells. Zhang et al., (1991) BioTechnology 9 996. Alternatively, liposome or

5   spheroplast fusion have been used to introduce expression vectors into plants. Deshayes et al., (1985) EMBO J., 4 2731, Christou et al., (1987) Proc Natl. Acad. Sci. U.S.A. 84 3962. Direct uptake- of DNA into protoplasts using CaCl₂ precipitation, polyvinyl alcohol or poly-Lornithine have also been reported. Hain et al., (1985) Mol.Gen. Genet. 199 161 and Draper et al., (1982) Plant Cell Physiol. 23

10   451. Electroporation of protoplasts and whole cells and tissues have also been described. Donn et al., In Abstracts of VIIth International Congress on Plant Cell and Tissue Culture IAPTC, A2-38, p 53 (1990) ; D'Halluin et al., (1992) Plant Cell 4 1495-1505 and Spencer et al., (1994) Plant Mol. Biol. 24 51-61.

A preferred method is microprojectile-mediated bombardment of immature

15   embryos. The embryos can be bombarded on the embryo axis side to target the meristem at a very early stage of development or bombarded on the scutellar side to target cells that typically form callus and somatic embryos. Targeting of the scutellum using projectile bombardment is well known to those in the art of cereal tissue culture. Klein et al., (1988) BioTechnol. , 6 559-563; Sautter et al., BiolTechnol., 9 1080-1085

20   (1991) ; Chibbar et al., (1991) Genome, 34 435-460. The scutellar origin of regenerable callus from cereals is well known. Green et al., (1975) Crop Sci., 15 417-421; Lu et al., (1982)TAG 62 109-112; and Thomas and Scott, (1985) J. Plant Physiol. 121 159-169 - Targeting the scutellum and then using chemical selection to recover transgenic plants is well established in cereals. D/Halluin et al., Plant Cell 4:

25   1495-1505 (1992) ; Perl et al., MGG 235: 279-284 (1992); Cristou et al., BiolTechnol. 9: 957-962 (1991).

VII.   USE FOR FLUORESCENT RESONANCE ENERGY TRANSFER (FRET)

30   FRET is a general, non-destructive, spectroscopic effect that occurs under certain circumstances (see below) when two fluorophores (a donor fluorophore and acceptor fluorophore) approach closer than about 100 Å. The efficiency of FRET

between the two fluorophores is highly distant dependent, and this fact can be

exploited to monitor the dynamic association of the fluorophores, or two

fluorophore tagged macromolecules. By monitoring FRET between one or more

fluorescent proteins it is possible to develop sensitive, non-invasive, cell based

5   assays for a range of activities including proteolysis (see U.S. Patent No. 5,981,200

issued November 9, 1999), analyte determinations (see U.S. Patent No. 5,998,204

issued December 7, 1999) and protein – protein interactions. FRET is most readily

determined by measuring the relative emissions of the donor and acceptor

fluorophore and then by calculating the emission ratio of these two values. A high

10   degree of FRET is indic·rd by a high value of the ratio of [acceptor emission/donor

emission], and a low degree of FRET is indicated by a low value of this ratio.

FRET may also may determined by measuring the degree of donor fluorescence

quenching, a measurement method that has the important advantage over emission

ratioing in that this value 's indep· · !ent of the concentration the acceptor.

15   The efficiency of F·ET is dependent on the separation distance, the

orientation of the donor and acceptor moieties, the fluorescent quantum yield of the

donor moiety and the energetic overlap with the acceptor moiety. Forster derived the

relationship:

20         $$E = (F^0 - F)/F^0 = R_0^6/(R^6 + R_0^6)$$

where E is the efficiency of FRET, F and F^0 are the fluorescence intensities of the

donor in the presence and absence of the acceptor, respectively, and R is the distance

between the donor and the acceptor. R_0, the distance at which the energy transfer

25   efficiency is 50%, is given (in Å) by

         $$R_0 = 9.79 \times 10^3 (K^2 QJn^{-4})^{1/6}$$

where K^2 is an orientation factor having an average value close to 0.67 for freely

30   mobile donors and acceptors, Q is the quantum yield of the unquenched fluorescent

donor, n is the refractive index of the intervening medium, and J is the overlap

integral, which expresses in quantitative terms the degree of spectral overlap,

$$J = \int_{0\_1}^{\infty} F_1 \lambda^4 d\lambda / \int_0^{\infty} F_1 d\lambda$$

where $\_1$ is the molar absorptivity of the acceptor in M^{-1} cm^{-1} and F_1 is the donor
fluorescence at wavelength λ measured in cm. The dependence of fluorescence
energy transfer on the above parameters has been reported [Forster, T. (1948)
Ann.Physik 2: 55-75; Lakowicz, J.R., Principles of Fluorescence Spectroscopy, New
York:Plenum Press (1983); Herman, B., Resonance energy transfer microscopy, in:
Fluorescence Microscopy of Living Cells in Culture, Part B, Methods in Cell Biology,
Vol 30, ed. Taylor, D.L. & Wang, Y.-L., San Diego: Academic Press (1989), pp. 219-
243; Turro, N.J., Modern Molecular Photochemistry, Menlo Part:
Benjamin/Cummings Publishing Co., Inc. (1978), pp. 296-361], and tables of spectral
overlap integrals are readily available to those working in the field [for example,
Berlman, I.B. Energy transfer parameters of aromatic compounds, Academic Press,
New York and London (1973)].

Accordingly, the functional red fluorescent proteins of the present invention
are intended to have improved brightness, reduced spectral cross talk and to be rapidly
and efficiently expressed in mammalian cells, compared to wild-type Anthozoan
proteins. Specifically such proteins are designed to exhibit reduced excitation in the
region 400 nm to 515 nm, where most *Aequorea* related donor fluorescent proteins are
most efficiently excited, and exhibit an improved molar extinction coefficient when
expressed in mammalian cells. Accordingly such functional red fluorescent proteins
are useful in any methods that involve FRET.

In one embodiment the functional red fluorescent proteins are useful in FRET
based assays for detecting protease activity in which the donor and acceptor
fluorescent proteins are separated by a cleavable linker. In this embodiment a first
fluorescent protein, for example one of the proteins in **Table 1** is selected as the
FRET donor. To optimize the efficiency and detectability of FRET within the tandem
fluorescent protein construct, several factors need to be balanced. The emission
spectrum of the donor moiety should overlap as much as possible with the excitation
spectrum of the acceptor moiety to maximize the overlap integral J. Also, the
quantum yield of the donor moiety and the extinction coefficient of the acceptor

should likewise be as high as possible to maximize R_0. However, the excitation
spectra of the donor and acceptor moieties should overlap as little as possible so that a
wavelength region can be found at which the donor can be excited efficiently without
directly exciting the acceptor. Fluorescence arising from direct excitation of the

5   acceptor is difficult to distinguish from fluorescence arising from FRET. Similarly,
the emission spectra of the donor and acceptor moieties should overlap as little as
possible so that the two emissions can be clearly distinguished. High fluorescence
quantum yield of the acceptor moiety is desirable if the emission from the acceptor is
to be measured either as the sole readout or as part of an emission ratio. In a preferred

10   embodiment, the donor moiety is typically excited by blue light (<500 nm) and
typically emits green light (>500 nm), whereas the acceptor is efficiently excited by
green, but not by blue light, and emits red light (>550 nm), for example, preferred
donors include Sapphire, W1C, W1B, Emerald. Topaz is preferred for functional red
fluorescent proteins that exhibit little or no direct excitation around 500 to 520 nm.

15   For use in measuring protease activity, the donor and acceptor fluorescent
protein moieties are connected through a linker moiety. The linker moiety is
preferably a peptide moiety, but ca.. be another organic molecular moiety as well. In
a preferred embodiment, the linker moiety includes a cleavage recognition site
specific for an enzyme or other cleavage agent of interest. A cleavage site in the

20   linker moiety is useful because when a tandem construct is mixed with the cleavage
agent, the linker is a substrate for cleavage by the cleavage agent. Rupture of the
linker moiety results in separation of the fluorescent protein moieties that is
measurable as a change in FRET.

          When the cleavage agent of interest is a protease, the linker can comprise a

25   peptide containing a cleavage recognition motif for the protease. A recognition motif
for a protease is a specific amino acid sequence recognized by the protease during
proteolytic cleavage. The linker can contain any protease recognition motif known in
the art, or discovered in the future.

          In one embodiment the functional red fluorescent proteins are useful in FRET

30   based assays for detecting the presence of an analyte (See U.S. Patent No. 5,998,204,
issued December 7, 1999). In this case the linker comprising a cleavage site is
replaced by a binding protein moiety. The binding protein moiety has an analyte-

binding region that binds an analyte and causes the tandem construct to change conformation upon exposure to the analyte. The donor fluorescent protein moiety is covalently coupled to the binding protein moiety. The acceptor fluorescent protein moiety, such as a functional red fluorescent protein, is covalently coupled to the

5   binding protein moiety. In the fluorescent indicator, the donor moiety and the acceptor moiety change position relative to each other when the analyte binds to the analyte-binding region, altering fluorescence resonance energy transfer between the donor moiety and the acceptor moiety when the donor moiety is excited. The change in FRET provides an indication of the concentration of the analyte in the sample.

10   In another embodiment the functional red fluorescent proteins are useful for FRET based assays for detecting protein-protein interactions. This approach enables an additional range of post-translational activities to be assayed. In this embodiment, a first protein is typically covalently coupled to donor fluorescent protein (such as a fluorescent protein from **Table 1**), and a second protein is covalently coupled to the

15   acceptor fluorescent protein (such as a functional red fluorescent protein). As previously, the donor and acceptor fluorescent proteins are selected to optimize the degree of FRET. Binding of the first protein to the second protein results in the association of the donor and acceptor fluorescent proteins resulting in an enhancement of the degree of FRET between them. This results in a measurable change in the

20   donor and acceptor emission ratio. This approach thus enables the identification and detection of protein-protein interactions between defined proteins, as well as the ability to detect post-translational modifications that influence these protein-protein interactions.

        Examples of suitable interaction domains include protein-protein interaction

25   domains such as SH2, SH3, PDZ, 14-3-3, WW and PTB domains. Other interaction domains are described in for example, the database of interacting proteins available on the web at http://www.doe-mbi.ucla.edu.

        To identify and characterize the interaction of two test proteins, the method would typically involve; 1) the creation of a first fusion protein comprising the first

30   test protein coupled to the donor fluorescent protein, and a second fusion protein comprising the second test protein coupled to acceptor fluorescent protein; 2) the introduction of the test protein fusion proteins in combination into test cells, and the

44

donor and acceptor fluorescent proteins (without fusion proteins) into control cells; 3) the measurement of the donor and acceptor emission ratios in the control cells and test cells; and 4) comparison of the emission ratio in the control cells, compared to the emission ratio in the test cells.

5   If the cells expressing the fusion proteins exhibits an emission ratio with a significantly altered value compared to the control cells containing the fluorescent proteins alone, then the results indicate that the two proteins do interact under the experimental conditions chosen. Conversely, if the emission ratios in the control cells, and in the test cells are approximately the same (after taking into account

10   differences in relative expression of the fluorescent proteins), then the results indicate that the proteins probably don't interact strongly under the test conditions.

The method also enables the detection and characterization of stimuli (such as receptor stimulation) that cause two proteins to alter their degree of interaction. In this case, a cell line is created that expresses the first and second fusion proteins, as

15   described above, comprising interaction domains that exhibit, or are believed to exhibit post-translational regulated interactions. For example, post-translational modification by phosphorylation of serine or threonine residues can modulate 14-3-3 domain interactions, tyrosine phosphorylation can influence SH2 domain interactions, the redox state can influence disulfide bond formation. The cell line is then exposed

20   to a test stimulus to determine whether the stimulus regulates the interaction of the two proteins. If the stimulus does regulate the interaction of the two proteins, then this will result in a modulation of the coupling of the two fluorescent proteins, subsequently resulting in a modulation of the degree of FRET and hence fluorescence emission ratio in the treated cells, compared to the non-treated cells.

25   The invention is also readily amenable to identifying new protein-protein interactions. For example, where a first protein is known, but the protein(s) with which it interacts are unknown. In this case, a first fusion protein is made between the first protein and the donor fluorescent protein (or acceptor fluorescent protein) and cloned into a suitable expression vector. Second, a library of test proteins, for

30   example isolated from a cDNA expression library, is fused in frame to the acceptor fluorescent protein (or donor fluorescent protein) and subcloned into a second expression vector. Typically the first fusion protein would be then be introduced into

a population of test cells and single clones identified that stably expressed the fusion
protein. The library of test proteins (typically in the form of expression vectors) would
be introduced into the clonal cells, stably expressing the first fusion protein. The
resulting transformed cells would then be screened to identify cells with altered FRET
5   compared to the control cells. Suitable clones expressing the fusion proteins with
modulated FRET, (i.e., altered emission ratios) may then be identified, isolated and
characterized, for example by fluorescence activated cell sorting (FACS™). To
confirm that the altered emission ratio was indeed the result of FRET, and not due to
alterations in the expression level of the acceptor fluorescent protein, secondary
10   measurements of donor emission quenching in the presence and absence of the
acceptor would usually be completed. This could be achieved, for example, by
measuring donor emission before and after photobleaching of the acceptor. Those
library members that display fusion proteins with larger relative changes in emission
ratio may then be identified by the degree to which emission ratio is altered for each
15   library member after exposure to the library of test fusion proteins.

       VIII.   USE FOR DRUG DISCOVERY

              FRET based fluorescence assays are well suited for use with systems and
20   methods that utilize automated and integratable workstations for identifying
modulators, and chemicals having useful activity. Such systems are described
generally in the art (see, U.S. Patent NOs: 4,000,976 to Kramer *et al.* (issued January
4, 1977), 5,104,621 to Pfost *et al.* (issued April 14, 1992), 5,125,748 to Bjornson *et
al.* (issued June 30, 1992), 5,139,744 to Kowalski (issued August 18, 1992),
25   5,206,568 Bjornson *et al.* (issued April 27, 1993), 5,350,564 to Mazza *et al.*
(September 27, 1994), 5,589,351 to Harootunian (issued December 31, 1996), and
PCT Application Nos: WO 93/20612 to Baxter Deutschland GMBH (published
October 14, 1993), WO 96/05488 to McNeil *et al.* (published February 22, 1996),
WO 93/13423 to Agong *et al.* (published July 8, 1993) and U.S. Patent No.
30   5,985,214, issued November 16, 1999.
              Typically, such a system includes: A) a storage and retrieval module
comprising storage locations for storing a plurality of chemicals in solution in

46

addressable chemical wells, a chemical well retriever and having programmable
selection and retrieval of the addressable chemical wells and having a storage capacity
for at least 100,000 addressable wells, B) a sample distribution module comprising a
liquid handler to aspirate or dispense solutions from selected addressable chemical

5   wells, the chemical distribution module having programmable selection of, and
aspiration from, the selected addressable chemical wells and programmable
dispensation into selected addressable sample wells (including dispensation into
arrays of addressable wells with different densities of addressable wells per
centimeter squared) or at locations, preferably pre-selected, on a plate, C) a sample

10   transporter to transport the selected addressable chemical wells to the sample
distribution module and optionally having programmable control of transport of the
selected addressable chemical wells or locations on a plate (including adaptive routing
and parallel processing), and D) a reaction module comprising either a reagent
dispenser to dispense reagents into the selected addressable sample wells or locations

15   on a plate or a fluorescent detector to detect chemical reactions in the selected
addressable sample wells or locations on a plate, and a data processing and integration
module.

        The storage and retrieval module, the sample distribution module, and the
reaction module are integrated and programmably controlled by the data processing

20   and integration module. The storage and retrieval module, the sample distribution
module, the sample transporter, the reaction module and the data processing and
integration module are operably linked to facilitate rapid processing of the
addressable sample wells or locations on a plate. Typically, devices of the invention
can process at least 100,000 addressable wells or locations on a plate in 24 hours.

25   This type of system is described in commonly owned U.S. Patent No. 5,985,214,
issued November 16, 1999. If desired, each separate module is integrated and
programmably controlled to facilitate the rapid processing of liquid samples, as well
as being operably linked to facilitate the rapid processing of liquid samples. In one
embodiment the system provides for a reaction module that is a fluorescence detector

30   to monitor fluorescence. The fluorescence detector is integrated to other workstations
with the data processing and integration module and operably linked with the sample
transporter. Preferably, the fluorescence detector is of the type described herein and

47

can be used for epi-fluorescence. Other fluorescence detectors that are compatible with the data processing and integration module and the sample transporter, if operable linkage to the sample transporter is desired can be used as known in the art or developed in the future. For some embodiments of the invention, particularly for

5   plates with 96, 192, 384 and 864 wells per plate, detectors are available for integration into the system. Such detectors are described in U.S. Patent 5,589,351 (Harootunian), U.S. Patent 5,355,215 (Schroeder), and PCT patent application WO 93/13423 (Akong). Alternatively, an entire plate may be "read" using an imager, such as a Molecular Dynamics Fluor-Imager 595 (Sunnyvale, CA). Multi-well platforms

10   having greater than 864 wells, including 3,456 wells, can also be used in the present invention (see, for example, the PCT Application PCT/US98/11061, filed 6/2/98. These higher density well plates require miniaturized assay volumes that necessitate the use of highly sensitivity assays that do not require washing. The present invention provides such assays as described herein.

15   The screening methods described herein can be made on cells growing in or deposited on solid surfaces. A common technique is to use a microtiter plate well wherein the fluorescence measurements are made by commercially available fluorescent plate readers. One such method is to use cells in Costar 96 well microtiter plates (flat with a clear bottom) and measure fluorescent signal with CytoFluor

20   multiwell plate reader (Perseptive Biosystems, Inc., MA) using two emission wavelengths to record fluorescent emission ratios. In another embodiment, the system comprises a microvolume liquid handling system that uses electrokinetic forces to control the movement of fluids through channels of the system, for example as described in U.S. patent No., 5,800,690 issued September 1, 1998 to Chow *et al.*,

25   European patent application EP 0 810 438 A2 filed May 5 1997, by Pelc *et al.* and PCT application WO 98/00231 filed 24 June 1997 by Parce *et al.* These systems use "chip" based analysis systems to provide massively parallel miniaturized analysis. Such systems are preferred systems of spectroscopic measurements in some instances that require miniaturized analysis.

30

48

A method for identifying a chemical, modulator or a therapeutic

      The present invention can also be used for testing a therapeutic for useful therapeutic activity. A therapeutic is identified by contacting a test chemical suspected of having a modulating activity of a biological process or target with a test

5   cell comprising the constructs of the present invention. Typically the cells are located within at least one well of a multi-well platform. The test chemical can be part of a library of test chemicals that is screened for activity, such as biological activity. The library can have individual members that are tested individually or in combination, or the library can be a combination of individual members. Such libraries can have at

10   least two members, preferably greater than about 100 members or greater than about 1,000 members, more preferably greater than about 10,000 members, and most preferably greater than about 100,000 or 1,000,000 members. After appropriate incubation of the sample with the test cell, an inhibitor of protein synthesis may be added and a substrate for the reporter m . added. At least one optical property

15   (such as fluorescence or absorbance) of the sample is determined and compared to a non-treated control to determine the level of reporter gene expression or activity. If the sample having the test chemical exhibits increased or decreased reporter moiety expression or activity relative to that of the control cell then a candidate modulator has been identified.

20   The candidate modulator can be further characterized and monitored for structure, potency, toxicology, and pharmacology using well-known methods. The structure of a candidate modulator identified by the invention can be determined or confirmed by methods known in the art, such as mass spectroscopy. For putative modulators stored for extended periods of time, the structure, activity, and potency of

25   the putative modulator can be confirmed.

      Depending on the system used to identify a candidate modulator, the candidate modulator will have putative pharmacological activity. For example, if the candidate modulator is found to inhibit a protein tyrosine phosphatase involved, for example in T-cell proliferation *in vitro*, then the candidate modulator would have presumptive

30   pharmacological properties as an immunosuppressant or anti-inflammatory (see, Suthanthiran *et al.*, (1996) Am. J. Kidney Disease, 28 159-172) Such nexuses are known in the art for several disease states, and more are expected to be discovered

over time. Based on such nexuses, appropriate confirmatory *in vitro* and *in vivo* models of pharmacological activity, as well as toxicology, can be selected. The assays, and methods of use described herein, enable rapid pharmacological profiling to assess selectivity and specificity, and toxicity. This data can subsequently be used

5   to develop new candidates with improved characteristics.

Bioavailability and Toxicology of Candidate Modulators

         Once identified, candidate modulators can be evaluated for bioavailability and toxicological effects using known methods (see, Lu, Basic Toxicology, Fundamentals,

10   Target Organs, and Risk Assessment, Hemisphere Publishing Corp., Washington (1985); U.S. Patent Nos: 5,196,313 to Culbreth (issued March 23, 1993) and U.S. Patent No. 5,567,952 to Benet (issued October 22, 1996). For example, toxicology of a candidate modulator can be established by determining *in vitro* toxicity towards a cell line, such as a mammalian i.e. human, cell line. Candidate modulators can be

15   treated with, for example, tissue extracts, such as preparations of liver, such as microsomal preparations, to determine increased or decreased toxicological properties of the chemical after being metabolized by a whole organism. The results of these types of studies are often predictive of toxicological properties of chemicals in animals, such as mammals, including humans.

20   The toxicological activity can be measured using reporter genes that are activated during toxicological activity or by cell lysis (see WO 98/13353, published 4/2/98). Preferred reporter genes produce a fluorescent or luminescent translational product (such as, for example, a Green Fluorescent Protein (see, for example, U.S. Patent No. 5,625,048 to Tsien *et al.*, issued 4/29/98; U.S. Patent No. 5,777,079 to

25   Tsien *et al.*, issued 7/7/98; WO 96/23810 to Tsien, published 8/8/96; WO 97/28261, published 8/7/97; PCT/US97/12410, filed 7/16/97; PCT/US97/14595, filed 8/15/97)) or a translational product that can produce a fluorescent or luminescent product (such as, for example, beta-lactamase (see, for example, U.S. Patent No. 5,741,657 to Tsien, issued 4/21/98, and WO 96/30540, published 10/3/96)), such as an enzymatic

30   degradation product. Cell lysis can be detected in the present invention as a reduction in a fluorescence signal from at least one photon-producing agent within a cell in the presence of at least one photon reducing agent. Such toxicological determinations can

50

be made using prokaryotic or eukaryotic cells, optionally using toxicological

profiling, such as described in PCT/US94/00583, filed 1/21/94 (WO 94/17208),

German Patent No 69406772.5-08, issued 11/25/97; EPC 0680517, issued 11/12/94;

U.S. Patent No. 5,589,337, issued 12/31/96; EPO 651825, issued 1/14/98; and U.S.

5   Patent No. 5,585,232, issued 12/17/96).

        Alternatively, or in addition to these *in vitro* studies, the bioavailability and

toxicological properties of a candidate modulator in an animal model, such as mice,

rats, rabbits, or monkeys, can be determined using established methods (see, Lu, supra

(1985); and Creasey, Drug Disposition in Humans, The Basis of Clinical

10   Pharmacology, Oxford University Press, Oxford (1979), Osweiler, Toxicology,

Williams and Wilkins, Baltimore, MD (1995), Yang, Toxicology of Chemical

Mixtures; Case Studies, Mechanisms, and Novel Approaches, Academic Press, Inc.,

San Diego, CA (1994), Burrell *et al.*, Toxicology of the Immune System; A Human

Approach, Van Nostrand Reinhld, Co. (1997), Niesink *et al.*, Toxicology; Principles

15   and Applications, CRC Press, Boca Raton, FL (1996)). Depending on the toxicity,

target organ, tissue, locus, and presumptive mechanism of the candidate modulator,

the skilled artisan would not be burdened to determine appropriate doses, LD_{50} values,

routes of administration, and regimes that would be appropriate to determine the

toxicological properties of the candidate modulator. In addition to animal models,

20   human clinical trials can be performed following established procedures, such as

those set forth by the United States Food and Drug Administration (USFDA) or

equivalents of other governments. These toxicity studies provide the basis for

determining the therapeutic utility of a candidate modulator *in vivo*.

25   *Efficacy of Candidate Modulators*

        Efficacy of a candidate modulator can be established using several art-

recognized methods, such as *in vitro* methods, animal models, or human clinical trials

(see, Creasey, supra (1979)). Recognized *in vitro* models exist for several diseases or

conditions. For example, the ability of a chemical to extend the life-span of HIV-

30   infected cells *in vitro* is recognized as an acceptable model to identify chemicals

expected to be efficacious to treat HIV infection or AIDS (see, Daluge *et al.*, (1995)

Antimicro. Agents Chemother. 41 1082-1093). Furthermore, the ability of

51

cyclosporin A (CsA) to prevent proliferation of T-cells *in vitro* has been established

as an acceptable model to identify chemicals expected to be efficacious as

immunosuppressants (see, Suthanthiran *et al.*, supra, (1996)). For nearly every class

of therapeutic, disease, or condition, an acceptable *in vitro* or animal model is

5   available. Such models exist, for example, for gastro-intestinal disorders, cancers,

cardiology, neurobiology, and immunology. In addition, these *in vitro* methods can

use tissue extracts, such as preparations of liver, such as microsomal preparations, to

provide a reliable indication of the effects of metabolism on the candidate modulator.

Similarly, acceptable animal models may be used to establish efficacy of chemicals to

10   treat various diseases or conditions. For example, the rabbit knee is an accepted

model for testing chemicals for efficacy in treating arthritis (see, Shaw and Lacy, J.

(1973) Bone Joint Surg. (Br) $\underline{55}$ 197-205. Hydrocortisone, which is approved for use

in humans to treat arthritis, is efficacious in this model which confirms the validity of

this model (see, McDonough, (1982) Phys. Ther. $\underline{62}$ 835-839). When choosing an

15   appropriate model to determine efficacy of a candidate modulator, the skilled artisan

can be guided by the state of the art to choose an appropriate model, dose, and route

of administration, regime, and endpoint and as such would not be unduly burdened.

        In addition to animal models, human clinical trials can be used to determine

the efficacy of a candidate modulator in humans. The USFDA, or equivalent

20   governmental agencies, have established procedures for such studies (see,

www.fda.gov).

Selectivity of Candidate Modulators

        The *in vitro* and *in vivo* methods described above also establish the selectivity

25   of a candidate modulator. It is recognized that chemicals can modulate a wide variety

of biological processes or be selective. Panels of cells, each containing constructs

with varying specificity, based on the red fluorescent proteins of the present

invention, can be used to determine the specificity of the candidate modulator.

Selective modulators are preferable because they have fewer side effects in the

30   clinical setting. The selectivity of a candidate modulator can be established *in vitro*

by testing the toxicity and effect of a candidate modulator on a plurality of cell lines

that exhibit a variety of cellular pathways and sensitivities. The data obtained from

52

these *in vitro* toxicity studies can be extended into *in vivo* animal model studies, including human clinical trials, to determine toxicity, efficacy, and selectivity of the candidate modulator suing art-recognized methods.

5   *An identified chemical, modulator, or therapeutic and compositions*

The invention includes compositions, such as novel chemicals, and therapeutics identified by at least one method of the present invention as having activity by the operation of methods, systems or components described herein. Novel chemicals, as used herein, do not include chemicals already publicly known in the art

10   as of the filing date of this application. Typically, a chemical would be identified as having activity from using the invention and then its structure can be revealed from a proprietary database of chemical structures or determined using analytical techniques such as mass spectroscopy.

One embodiment of the invention is a chemical with useful activity,

15   comprising a chemical identified by the method described above. Such compositions include small organic molecules, nucleic acids, peptides and other molecules readily synthesized by techniques available in the art and developed in the future. For example, the following combinatorial compounds are suitable for screening: peptoids (PCT Publication No. WO 91/19735, 26 Dec. 1991), encoded peptides (PCT

20   Publication No. WO 93/20242, 14 Oct. 1993), random bio-oligomers (PCT Publication WO 92/00091, 9 Jan. 1992), benzodiazepines (U.S. Patent No. 5,288,514), diversomeres such as hydantoins, benzodiazepines and dipeptides (Hobbs DeWitt, S. *et al.*, (1993) Proc. Nat. Acad. Sci. USA 90 6909-6913), vinylogous polypeptides (Hagihara *et al.*, (1992) J. Amer. Chem. Soc. 114 6568), nonpeptidal

25   peptidomimetics with a Beta-D-Glucose scaffolding (Hirschmann, R. *et al.*, (1992) J. Amer. Chem. Soc. 114 9217-9218), analogous organic syntheses of small compound libraries (Chen, C. *et al.*, (1994) J. Amer. Chem. Soc. 116 2661), oligocarbamates (Cho, C.Y. *et al.*, (1993) Science 261: 1303), and/or peptidyl phosphonates (Campbell, D.A. *et al.*, (1994) J. Org. Chem. 59 658). See, generally, Gordon, E. M.

30   *et al.*, (1994). J. Med Chem. 37 1385. The contents of all of the aforementioned publications are incorporated herein by reference.

53

The present invention also encompasses the identified compositions in a
pharmaceutical composition comprising a pharmaceutically acceptable carrier
prepared for storage and subsequent administration, which have a pharmaceutically
effective amount of the products disclosed above in a pharmaceutically acceptable

5   carrier or diluent. Acceptable carriers or diluents for therapeutic use are well known
in the pharmaceutical art, and are described, for example, in Remington's
Pharmaceutical Sciences, Mack Publishing Co. (A.R. Gennaro edit. 1985).
Preservatives, stabilizers, dyes and even flavoring agents may be provided in the
pharmaceutical composition. For example, sodium benzoate, acsorbic acid and esters

10   of p-hydroxybenzoic acid may be added as preservatives. In addition, antioxidants
and suspending agents may be used.

The compositions of the present invention may be formulated and used as
tablets, capsules or elixirs for oral administration; suppositories for rectal
administration; sterile solutions, suspensions for injectable administration; and the

15   like. Injectables can be prepared in conventional forms, either as liquid solutions or
suspensions, solid forms suitable for solution or suspension in liquid prior to
injection, or as emulsions. Suitable excipients are, for example, water, saline,
dextrose, mannitol, lactose, lecithin, albumin, sodium glutamate, cysteine
hydrochloride, and the like. In addition, if desired, the injectable pharmaceutical

20   compositions may contain minor amounts of nontoxic auxiliary substances, such as
wetting agents, pH buffering agents, and the like. If desired, absorption enhancing
preparations (e.g., liposomes) may be utilized.

The pharmaceutically effective amount of the composition required as a dose
will depend on the route of administration, the type of animal being treated, and the

25   physical characteristics of the specific animal under consideration. The dose can be
tailored to achieve a desired effect, but will depend on such factors as weight, diet,
concurrent medication and other factors which those skilled in the medical arts will
recognize. In practicing the methods of the invention, the products or compositions
can be used alone or in combination with one another or in combination with other

30   therapeutic or diagnostic agents. These products can be utilized *in vivo*, ordinarily in
a mammal, preferably in a human, or *in vitro*. In employing them *in vivo*, the
products or compositions can be administered to the mammal in a variety of ways,

including parenterally, intravenously, subcutaneously, intramuscularly, colonically, rectally, nasally or intraperitoneally, employing a variety of dosage forms. Such methods may also be applied to testing chemical activity *in vivo*.

As will be readily apparent to one skilled in the art, the useful *in vivo* dosage 5 to be administered and the particular mode of administration will vary depending upon the age, weight and mammalian species treated, the particular compounds employed, and the specific use for which these compounds are employed. The determination of effective dosage levels, that is the dosage levels necessary to achieve the desired result, can be accomplished by one skilled in the art using routine 10 pharmacological methods. Typically, human clinical applications of products are commenced at lower dosage levels, with dosage level being increased until the desired effect is achieved. Alternatively, acceptable *in vitro* studies can be used to establish useful doses and routes of administration of the compositions identified by the present methods using established pharmacological methods.

15 In non-human animal studies, applications of potential products are commenced at higher dosage levels, with dosage being decreased until the desired effect is no longer achieved or adverse side effects disappear. The dosage for the products of the present invention can range broadly depending upon the desired affects and the therapeutic indication. Typically, dosages may be between about 10 20 mg/kg and 100 mg/kg body weight, and preferably between about 100 µg/kg and 10 mg/kg body weight. Administration is preferably oral on a daily basis.

The exact formulation, route of administration and dosage can be chosen by the individual physician in view of the patient's condition. (See e.g., Fingl *et al.*, in The Pharmacological Basis of Therapeutics, 1975). It should be noted that the 25 attending physician would know how to and when to terminate, interrupt, or adjust administration due to toxicity, or to organ dysfunctions. Conversely, the attending physician would also know to adjust treatment to higher levels if the clinical response were not adequate (precluding toxicity). The magnitude of an administrated dose in the management of the disorder of interest will vary with the severity of the condition 30 to be treated and to the route of administration. The severity of the condition may, for example, be evaluated, in part, by standard prognostic evaluation methods. Further, the dose and perhaps dose frequency, will also vary according to the age, body

55

weight, and response of the individual patient. A program comparable to that discussed above may be used in veterinary medicine.

Depending on the specific conditions being treated, such agents may be formulated and administered systemically or locally. Techniques for formulation and · administration may be found in Remington's Pharmaceutical Sciences, 18th Ed., Mack Publishing Co., Easton, PA (1990). Suitable routes may include oral, rectal, transdermal, vaginal, transmucosal, or intestinal administration; parenteral delivery, including intramuscular, subcutaneous, intramedullary injections, as well as intrathecal, direct intraventricular, intravenous, intraperitoneal, intranasal, or intraocular injections.

For injection, the agents of the invention may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks' solution, Ringer's solution, or physiological saline buffer. For such transmucosal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art. Use of pharmaceutically acceptable carriers to formulate the compounds herein disclosed for the practice of the invention into dosages suitable for systemic administration is within the scope of the invention. With proper choice of carrier and suitable manufacturing practice, the compositions of the present invention, in particular, those formulated as solutions, may be administered parenterally, such as by intravenous injection. The compounds can be formulated readily using pharmaceutically acceptable carriers well known in the art into dosages suitable for oral administration. Such carriers enable the compounds of the invention to be formulated as tablets, pills, capsules, liquids, gels, syrups, slurries, suspensions and the like, for oral ingestion by a patient to be treated.

Agents intended to be administered intracellularly may be administered using techniques well known to those of ordinary skill in the art. For example, such agents may be encapsulated into liposomes, then administered as described above. All molecules present in an aqueous solution at the time of liposome formation are incorporated into the aqueous interior. The liposomal contents are both protected from the external micro-environment and, because liposomes fuse with cell membranes, are efficiently delivered into the cell cytoplasm. Additionally, due to

56

their hydrophobicity, small organic molecules may be directly administered intracellularly.

Pharmaceutical compositions suitable for use in the present invention include compositions wherein the active ingredients are contained in an effective amount to
5   achieve its intended purpose. Determination of the effective amounts is well within the capability of those skilled in the art, especially in light of the detailed disclosure provided herein. In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into
10   preparations which can be used pharmaceutically. The preparations formulated for oral administration may be in the form of tablets, dragees, capsules, or solutions. The pharmaceutical compositions of the present invention may be manufactured in a manner that is itself known, for example, by means of conventional mixing, dissolving, granulating, dragee-making, levitating, emulsifying, encapsulating,
15   entrapping, or lyophilizing processes.

Pharmaceutical formulations for parenteral administration include aqueous solutions of the active compounds in water-soluble form. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or
20   synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes. Aqueous injection suspensions may contain substances that increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Optionally, the suspension may also contain suitable stabilizers or agents that increase the solubility of the compounds to allow for the preparation of highly concentrated
25   solutions.

Pharmaceutical preparations for oral use can be obtained by combining the active compounds with solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable auxiliaries, if desired, to obtain tablets or dragee cores. Suitable excipients are, in particular, fillers such as
30   sugars, including lactose, sucrose, mannitol, or sorbitol; cellulose preparations such as, for example, maize starch, wheat starch, rice starch, potato starch, gelatin, gum tragacanth, methyl cellulose, hydroxypropylmethyl-cellulose, sodium

57

carboxymethylcellulose, and/or polyvinylpyrrolidone (PVP). If desired, disintegrating agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, or alginic acid or a salt thereof such as sodium alginate. Dragee cores are provided with suitable coatings. For this purpose, concentrated sugar solutions may

5   be used, which may optionally contain gum arabic, talc, polyvinyl pyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for identification or to characterize different combinations of active compound doses. For this purpose, concentrated sugar

10   solutions may be used, which may optionally contain gum arabic, talc, polyvinyl pyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for identification or to characterize different combinations of active compound doses. Such formulations can be made

15   using methods known in the art (see, for example, U.S. Patent Nos. 5,733,888 (injectable compositions); 5,726,181 (poorly water soluble compounds); 5,707,641 (therapeutically active proteins or peptides); 5,667,809 (lipophilic agents); 5,576,012 (solubilizing polymeric agents); 5,707,615 (anti-viral formulations); 5,683,676 (particulate medicaments); 5,654,286 (topical formulations); 5,688,529 (oral

20   suspensions); 5,445,829 (extended release formulations); 5,653,987 (liquid formulations); 5,641,515 (controlled release formulations) and 5,601,845 (spheroid formulations).

All publications and patents mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described

25   method and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the above-described modes for

30   carrying out the invention which are obvious to those skilled in the field of molecular biology or related fields are intended to be within the scope of the following claims.

58

WHAT IS CLAIMED IS:

1.   A nucleic acid molecule, comprising; a nucleotide sequence encoding a
      functional red fluorescent protein whose sequence differs from the amino acid
      sequence of an Anthozoan red fluorescent protein (SEQ. ID. NO. 7) by at least

5   one amino acid substitution at position D59, I60, S62, P63, Q64, F65, Q66,
      S69, K70, V71, Y72, V73, W93, R95, N98, W143, A145, S146, T147, E148,
      Y151, G159, I161, K163, G171, S179, Y181, S197, L199, Y214, E215 or
      R216, wherein said functional red fluorescent protein has a different
      fluorescent property compared to said Anthozoan red fluorescent protein

10   (SEQ. ID. NO. 7).

2.   The nucleic acid molecule of claim 1, wherein said functional red fluorescent
      protein exhibits a reduced molar extinction coefficient at 487 nm compared to
      said Anthozoan red fluorescent protein (SEQ. ID. NO. 7).

15

3.   The nucleic acid molecule of claim 1, wherein said functional red fluorescent
      protein exhibits a reduced molar extinction coefficient at 530 nm compared to
      said Anthozoan red fluorescent protein (SEQ. ID. NO. 7).

20
4.   The nucleic acid molecule of claim 1, wherein said functional red fluorescent
      protein exhibits a higher molar extinction coefficient at 583 nm compared to
      said Anthozoan red fluorescent protein (SEQ. ID. NO. 7).

25   5.   The nucleic acid molecule of claim 1, wherein said functional red fluorescent
      protein is brighter than said Anthozoan red fluorescent protein (SEQ. ID. NO.
      7) when excited at 558 nm.

6.   The nucleic acid molecule of claim 1, wherein said functional red fluorescent
30   protein is brighter than said Anthozoan red fluorescent protein (SEQ. ID. NO.
      7) when expressed in a mammalian cell grown at 37 °C.

7.   The nucleic acid molecule of claim 1, wherein said functional red fluorescent protein exhibits a higher quantum yield compared to said Anthozoan red fluorescent protein (SEQ. ID. NO. 7).

5   8.   The nucleic acid molecule of claim 1, wherein said functional red fluorescent protein exhibits a faster rate of autocatalytic formation compared to said Anthozoan red fluorescent protein (SEQ. ID. NO. 7).

9.   A nucleic acid molecule, comprising; a nucleotide sequence encoding a

10   functional red fluorescent protein whose sequence differs from the amino acid sequence of an Anthozoan red fluorescent protein (SEQ. ID. NO. 7) by at least one amino acid substitution at position 59, wherein said functional red fluorescent protein has a different fluorescent property compared to said Anthozoan red fluorescent protein (SEQ. ID. NO. 7).

15

10.   The nucleic acid molecule of claim 9, wherein said at least one amino acid substitution at position 59 is selected from the group consisting of D59S, D59A, D59H, D59E and D59P.

20

11.   A nucleic acid molecule, comprising; a nucleotide sequence encoding a functional red fluorescent protein whose sequence differs from the amino acid sequence of Anthozoan red fluorescent protein (SEQ. ID. NO. 7) by at least one amino acid substitution at position 60, wherein said functional red

25   fluorescent protein has a different fluorescent property compared to said Anthozoan red fluorescent protein (SEQ. ID. NO. 7).

12.   The nucleic acid molecule of claim 11, wherein said at least one amino acid substitution at position 60 is selected from the group consisting of I60T, I60A,

30   I60C, I60V and I60L.

13.   A nucleic acid molecule, comprising; a nucleotide sequence encoding a
       functional red fluorescent protein whose sequence differs from the amino acid
       sequence of Anthozoan red fluorescent protein (SEQ. ID. NO. 7) by at least
       one amino acid substitution at position 62, wherein said functional red
5   fluorescent protein has a different fluorescent property compared to said
       Anthozoan red fluorescent protein (SEQ. ID. NO. 7).

14.   The nucleic acid molecule of claim 13, wherein said at least one amino acid
       substitution at position 62 is selected from the group consisting of S62A,
10   S62G, S62C and S62T.

15.   A nucleic acid molecule, comprising; a nucleotide sequence encoding a
       functional red fluorescent protein whose sequence differs from the amino acid
       sequence of Anthozoan red fluorescent protein (SEQ. ID. NO. 7) by at least
15   one amino acid substitution at position 63, wherein said functional red
       fluorescent protein has a different fluorescent property compared to said
       Anthozoan red fluorescent protein (SEQ. ID. NO. 7).

16.   The nucleic acid molecule of claim 15, wherein said at least one amino acid
20   substitution at position 63 is selected from the group consisting of P63T,
       P63H, P63F and P63W.

17.   A nucleic acid molecule, comprising; a nucleotide sequence encoding a
       functional red fluorescent protein whose sequence differs from the amino acid
25   sequence of Anthozoan red fluorescent protein (SEQ. ID. NO. 7) by at least
       one amino acid substitution at position 64, wherein said functional red
       fluorescent protein has a different fluorescent property compared to said
       Anthozoan red fluorescent protein (SEQ. ID. NO. 7).

30   18.   The nucleic acid molecule of claim 17, wherein said at least one amino acid
       substitution at position 64 is selected from the group consisting of Q64K,
       Q64P, Q64T, Q64N and Q64R.

19.   A nucleic acid molecule, comprising; a nucleotide sequence encoding a functional red fluorescent protein whose sequence differs from the amino acid sequence of Anthozoan red fluorescent protein (SEQ. ID. NO. 7) by at least one amino acid substitution at position 65, wherein said functional red fluorescent protein has a different fluorescent property compared to said Anthozoan red fluorescent protein (SEQ. ID. NO. 7).

20.   The nucleic acid molecule of claim 19, wherein said at least one amino acid substitution at position 65 is selected from the group consisting of F65L, F65V, F65I, F65M, F65Y and F65W.

21.   A nucleic acid molecule, comprising; a nucleotide sequence encoding a functional red fluorescent protein whose sequence differs from the amino acid sequence of Anthozoan red fluorescent protein (SEQ. ID. NO. 7) by at least one amino acid substitution at position 66, wherein said functional red fluorescent protein has a different fluorescent property compared to said Anthozoan red fluorescent protein (SEQ. ID. NO. 7).

22.   The nucleic acid molecule of claim 21, wherein said at least one amino acid substitution at position 66 is selected from the group consisting of Q66R, Q66R, Q66P, Q66K, Q66E, Q66T, Q66A and Q66G.

23.   A nucleic acid molecule, comprising; a nucleotide sequence encoding a functional red fluorescent protein whose sequence differs from the amino acid sequence of Anthozoan red fluorescent protein (SEQ. ID. NO. 7) by at least one amino acid substitution at position 69, wherein said functional red fluorescent protein has a different fluorescent property compared to said Anthozoan red fluorescent protein (SEQ. ID. NO. 7).

24.   The nucleic acid molecule of claim 23, wherein said at least one amino acid substitution at position 69 is selected from the group consisting of S69L, S69A, S69V and S69T.

25.   A nucleic acid molecule, comprising; a nucleotide sequence encoding a
      functional red fluorescent protein whose sequence differs from the amino acid
      sequence of Anthozoan red fluorescent protein (SEQ. ID. NO. 7) by at least
      one amino acid substitution at position 70, wherein said functional red
5   fluorescent protein has a different fluorescent property compared to said
      Anthozoan red fluorescent protein (SEQ. ID. NO. 7).

26.   The nucleic acid molecule of claim 25, wherein said at least one amino acid
      substitution at position 70 is selected from the group consisting of K70M,
10   K70Q, K70L and K70R.

27.   A nucleic acid molecule, comprising; a nucleotide sequence encoding a
      functional red fluorescent protein whose sequence differs from the amino acid
      sequence of Anthozoan red fluorescent protein (SEQ. ID. NO. 7) by at least
15   one amino acid substitution at position 71, wherein said functional red
      fluorescent protein has a different fluorescent property compared to said
      Anthozoan red fluorescent protein (SEQ. ID. NO. 7).

28.   The nucleic acid molecule of claim 27, wherein said at least one amino acid
20   substitution at position 71 is selected from the group consisting of V71C,
      V71L, V71A and V71I.

29.   A nucleic acid molecule, comprising; a nucleotide sequence encoding a
      functional red fluorescent protein whose sequence differs from the amino acid
25   sequence of Anthozoan red fluorescent protein (SEQ. ID. NO. 7) by at least
      one amino acid substitution at position 72, wherein said functional red
      fluorescent protein has a different fluorescent property compared to said
      Anthozoan red fluorescent protein (SEQ. ID. NO. 7).

30   30.   The nucleic acid molecule of claim 29, wherein said at least one amino acid
      substitution at position 72 is selected from the group consisting of Y72F and
      Y72W.

31.   A nucleic acid molecule, comprising; a nucleotide sequence encoding a
functional red fluorescent protein whose sequence differs from the amino acid
sequence of Anthozoan red fluorescent protein (SEQ. ID. NO. 7) by at least
one amino acid substitution at position 73, wherein said functional red
5   fluorescent protein has a different fluorescent property compared to said
Anthozoan red fluorescent protein (SEQ. ID. NO. 7).

32.   The nucleic acid molecule of claim 31, wherein said at least one amino acid
substitution at position 73 is selected from the group consisting of V73A,
10   V73L, V73S and V73I.

33.   A nucleic acid molecule, comprising; a nucleotide sequence encoding a
functional red fluorescent protein whose sequence differs from the amino acid
sequence of Anthozoan red fluorescent protein (SEQ. ID. NO. 7) by at least
15   one amino acid substitution at position 93, wherein said functional red
fluorescent protein has a different fluorescent property compared to said
Anthozoan red fluorescent protein (SEQ. ID. NO. 7).

34.   The nucleic acid molecule of claim 33, wherein said at least one amino acid
20   substitution at position 93 is selected from the group consisting of W93L,
W93Y, W93C and W93F.

35.   A nucleic acid molecule, comprising; a nucleotide sequence encoding a
functional red fluorescent protein whose sequence differs from the amino acid
25   sequence of Anthozoan red fluorescent protein (SEQ. ID. NO. 7) by at least
one amino acid substitution at position 95, wherein said functional red
fluorescent protein has a different fluorescent property compared to said
Anthozoan red fluorescent protein (SEQ. ID. NO. 7).

30 36.   The nucleic acid molecule of claim 35, wherein said at least one amino acid
substitution at position 95 is selected from the group consisting of R95K.

37.   A nucleic acid molecule, comprising; a nucleotide sequence encoding a functional red fluorescent protein whose sequence differs from the amino acid sequence of Anthozoan red fluorescent protein (SEQ. ID. NO. 7) by at least one amino aci̇ substitution at position 98, wherein said functional red fluorescent pr.ẹin has a different fluorescent property compared to said Anthozoan red fluorescent protein (SEQ. ID. NO. 7).

38.   The nucleic acid molecule of claim 37, wherein said at least one amino acid substitution at position 98 is selected from the group consisting of N98T, N98D, N98A and N98Q.

39.   A nucleic acid molecule, comprising; a nucleotide sequence encoding a functional red fluorescent protein whose sequence differs from the amino acid sequence of Anthozoan red fluorescent protein (SEQ. ID. NO. 7) by at least one amino acid substitution at position 143, wherein said functional red fluorescent protein has a different fluorescent property compared to said Anthozoan red fluorescent protein (SEQ. ID. NO. 7).

40.   The nucleic acid molecule of claim 39, wherein said at least one amino acid substitution at position 143 is selected from the group consisting of W143L, W143F, W143C, W143Y and W143L.

41.   A nucleic acid molecule, comprising; a nucleotide sequence encoding a functional red fluorescent protein whose sequence differs from the amino acid sequence of Anthozoan red fluorescent protein (SEQ. ID. NO. 7) by at least one amino acid substitution at position 145, wherein said functional red fluorescent protein has a different fluorescent property compared to said Anthozoan red fluorescent protein (SEQ. ID. NO. 7).

42.   The nucleic acid molecule of claim 41, wherein said at least one amino acid substitution at position 145 is selected from the group consisting of A145P, A145S, A145G and A145L.

43.   A nucleic acid molecule, comprising; a nucleotide sequence encoding a
        functional red fluorescent protein whose sequence differs from the amino acid
        sequence of Anthozoan red fluorescent protein (SEQ. ID. NO. 7) by at least
        one amino acid substitution at position 146, wherein said functional red
5   fluorescent protein has a different fluorescent property compared to said
        Anthozoan red fluorescent protein (SEQ. ID. NO. 7).

44.   The nucleic acid molecule of claim 43, wherein said at least one amino acid
        substitution at position 146 is selected from the group consisting of S146R,
10   S146G, S146N, S146H, S146T, S146A and S146D

45.   A nucleic acid molecule, comprising; a nucleotide sequence encoding a
        functional red fluorescent protein whose sequence differs from the amino acid
        sequence of Anthozoan red fluorescent protein (SEQ. ID. NO. 7) by at least
15   one amino acid substitution at position 147, wherein said functional red
        fluorescent protein has a different fluorescent property compared to said
        Anthozoan red fluorescent protein (SEQ. ID. NO. 7).

46.   The nucleic acid molecule of claim 45, wherein said at least one amino acid
20   substitution at position 147 is selected from the group consisting of T147N,
        T147K and T147S.

47.   A nucleic acid molecule, comprising; a nucleotide sequence encoding a
        functional red fluorescent protein whose sequence differs from the amino acid
25   sequence of Anthozoan red fluorescent protein (SEQ. ID. NO. 7) by at least
        one amino acid substitution at position 148, wherein said functional red
        fluorescent protein has a different fluorescent property compared to said
        Anthozoan red fluorescent protein (SEQ. ID. NO. 7).

30   48.   The nucleic acid molecule of claim 46, wherein said at least one amino acid
        substitution at position 148 is selected from the group consisting of E148V
        and E148D.

49.   A nucleic acid molecule, comprising; a nucleotide sequence encoding a
       functional red fluorescent protein whose sequence differs from the amino acid
       sequence of Anthozoan red fluorescent protein (SEQ. ID. NO. 7) by at least
       one amino acid substitution at position 151, wherein said functional red
5   fluorescent protein has a different fluorescent property compared to said
       Anthozoan red fluorescent protein (SEQ. ID. NO. 7).

50.   The nucleic acid molecule of claim 49, wherein said at least one amino acid
       substitution at position 151 is selected from the group consisting of Y151F,
10   Y151N, Y151D, Y151S, Y151T and Y151A.

51.   A nucleic acid molecule, comprising; a nucleotide sequence encoding a
       functional red fluorescent protein whose sequence differs from the amino acid
       sequence of Anthozoan red fluorescent protein (SEQ. ID. NO. 7) by at least
15   one amino acid substitution at position 159, wherein said functional red
       fluorescent protein has a different fluorescent property compared to said
       Anthozoan red fluorescent protein (SEQ. ID. NO. 7).

52.   The nucleic acid molecule of claim 51, wherein said at least one amino acid
20   substitution at position 159 is selected from the group consisting of G159A.
       G159S and G159V.

53.   A nucleic acid molecule, comprising; a nucleotide sequence encoding a
       functional red fluorescent protein whose sequence differs from the amino acid
25   sequence of Anthozoan red fluorescent protein (SEQ. ID. NO. 7) by at least
       one amino acid substitution at position 161, wherein said functional red
       fluorescent protein has a different fluorescent property compared to said
       Anthozoan red fluorescent protein (SEQ. ID. NO. 7).

30 54.   The nucleic acid molecule of claim 53, wherein said at least one amino acid
       substitution at position 161 is selected from the group consisting of I161V,
       I161V, I161F, I161M and I161L

55. A nucleic acid molecule, comprising; a nucleotide sequence encoding a functional red fluorescent protein whose sequence differs from the amino acid sequence of Anthozoan red fluorescent protein (SEQ. ID. NO. 7) by at least one amino acid substitution at position 163, wherein said functional red fluorescent protein has a different fluorescent property compared to said Anthozoan red fluorescent protein (SEQ. ID. NO. 7).

56. The nucleic acid molecule of claim 55, wherein said at least one amino acid substitution at position 163 is selected from the group consisting of K163I, K163R, K163T, K163E, K163V, K163G and K163A.

57. A nucleic acid molecule, comprising; a nucleotide sequence encoding a functional red fluorescent protein whose sequence differs from the amino acid sequence of Anthozoan red fluorescent protein (SEQ. ID. NO. 7) by at least one amino acid substitution at position 171, wherein said functional red fluorescent protein has a different fluorescent property compared to said Anthozoan red fluorescent protein (SEQ. ID. NO. 7).

58. The nucleic acid molecule of claim 57, wherein said at least one amino acid substitution at position 171 is selected from the group consisting of G171S and G171A.

59. A nucleic acid molecule, comprising; a nucleotide sequence encoding a functional red fluorescent protein whose sequence differs from the amino acid sequence of Anthozoan red fluorescent protein (SEQ. ID. NO. 7) by at least one amino acid substitution at position 179, wherein said functional red fluorescent protein has a different fluorescent property compared to said Anthozoan red fluorescent protein (SEQ. ID. NO. 7).

60. The nucleic acid molecule of claim 59, wherein said at least one amino acid substitution at position 179 is selected from the group consisting of S179A, S179P, S179T, S179E, S179Q and S179K.

61.   A nucleic acid molecule, comprising; a nucleotide sequence encoding a
        functional red fluorescent protein whose sequence differs from the amino acid
        sequence of Anthozoan red fluorescent protein (SEQ. ID. NO. 7) by at least
        one amino acid substitution at position 181, wherein said functional red
5   fluorescent protein has a different fluorescent property compared to said
        Anthozoan red fluorescent protein (SEQ. ID. NO. 7).

62.   The nucleic acid molecule of claim 61, wherein said at least one amino acid
        substitution at position 181 is selected from the group consisting of Y181F,
10   Y181W, Y181N and Y181I.

63.   A nucleic acid molecule, comprising; a nucleotide sequence encoding a
        functional red fluorescent protein whose sequence differs from the amino acid
        sequence of Anthozoan red fluorescent protein (SEQ. ID. NO. 7) by at least
15   one amino acid substitution at position 197, wherein said functional red
        fluorescent protein has a different fluorescent property compared to said
        Anthozoan red fluorescent protein (SEQ. ID. NO. 7).

64.   The nucleic acid molecule of claim 63, wherein said at least one amino acid
20   substitution at position 197 is selected from the group consisting of S197Y,
        S197T, S197N and S197A.

65.   A nucleic acid molecule, comprising; a nucleotide sequence encoding a
        functional red fluorescent protein whose sequence differs from the amino acid
25   sequence of Anthozoan red fluorescent protein (SEQ. ID. NO. 7) by at least
        one amino acid substitution at position 199, wherein said functional red
        fluorescent protein has a different fluorescent property compared to said
        Anthozoan red fluorescent protein (SEQ. ID. NO. 7).

30 66. The nucleic acid molecule of claim 65, wherein said at least one amino acid
        substitution at position 199 is selected from the group consisting of L199I,
        L199V, L199I and L199A.

67.   A nucleic acid molecule, comprising; a nucleotide sequence encoding a
        functional red fluorescent protein whose sequence differs from the amino acid
        sequence of Anthozoan red fluorescent protein (SEQ. ID. NO. 7) by at least
        one amino acid substitution at position 214, wherein said functional red
5   fluorescent protein has a different fluorescent property compared to said
        Anthozoan red fluorescent protein (SEQ. ID. NO. 7).

68.   The nucleic acid molecule of claim 67, wherein said at least one amino acid
        substitution at position 214 is selected from the group consisting of Y214F,
10   Y214H and Y214L.

69.   A nucleic acid molecule, comprising; a nucleotide sequence encoding a
        functional red fluorescent protein whose sequence differs from the amino acid
        sequence of Anthozoan red fluorescent protein (SEQ. ID. NO. 7) by at least
15   one amino acid substitution at position 215, wherein said functional red
        fluorescent protein has a different fluorescent property compared to said
        Anthozoan red fluorescent protein (SEQ. ID. NO. 7).

70.   The nucleic acid molecule of claim 69, wherein said at least one amino acid
20   substitution at position 215 is selected from the group consisting of E215G,
        E215Q and E215R.

71.   A nucleic acid molecule, comprising; a nucleotide sequence encoding a
        functional red fluorescent protein whose sequence differs from the amino acid
25   sequence of Anthozoan red fluorescent protein (SEQ. ID. NO. 7) by at least
        one amino acid substitution at position 216, wherein said functional red
        fluorescent protein has a different fluorescent property compared to said
        Anthozoan red fluorescent protein (SEQ. ID. NO. 7).

30 72.   The nucleic acid molecule of claim 71, wherein said at least one amino acid
        substitution at position 216 is selected from the group consisting of R216K,
        R216L, R216C and R216F.

73.   An expression vector, comprising; expression control sequences operatively
linked to a nucleic acid molecule encoding a functional red fluorescent protein
whose sequence differs from the amino acid sequence of an Anthozoan red
fluorescent protein (SEQ. ID. NO. 7) by at least one amino acid substitution at

5   position D59, I60, S62, P63, Q64, F65, Q66, S69, K70, V71, Y72, V73, W93,
R95, N98, W143, A145, S146, T147, E148, Y151, G159, I161, K163, G171,
S179, Y181, S197, L199, Y214, E215 or R216, wherein said functional red
fluorescent protein has a different fluorescent property compared to said
Anthozoan red fluorescent protein (SEQ. ID. NO. 7).

10
74.   A recombinant host cell, comprising; a nucleic acid molecule encoding a
functional red fluorescent protein whose sequence differs from the amino acid
sequence of an Anthozoan red fluorescent protein (SEQ. ID. NO. 7) by at least
one amino acid substitution at position D59, I60, S62, P63, Q64, F65, Q66,

15   S69, K70, V71, Y72, V73, W93, R95, N98, W143, A145, S146, T147, E148,
Y151, G159, I161, K163, G171, S179, Y181, S197, L199, Y214, E215 or
R216, wherein said functional red fluorescent protein has a different
fluorescent property compared to said Anthozoan red fluorescent protein
(SEQ. ID. NO. 7).

20
75.   A functional fluorescent protein, comprising; an amino acid sequence that
differs from the amino acid sequence of an Anthozoan red fluorescent protein
(SEQ. ID. NO. 7) by at least one amino acid substitution at position D59, I60,
S62, P63, Q64, F65, Q66, S69, K70, V71, Y72, V73, W93, R95, N98, W143,

25   A145, S146, T147, E148, Y151, G159, I161, K163, G171, S179, Y181, S197,
L199, Y214, E215 or R216, wherein said functional red fluorescent protein
has a different fluorescent property compared to said Anthozoan red
fluorescent protein (SEQ. ID. NO. 7).

76. A fusion protein, comprising; a protein of interest operably coupled to a functional red fluorescent protein whose sequence differs from the amino acid sequence of an Anthozoan red fluorescent protein (SEQ. ID. NO. 7) by at least one amino acid substitution at position D59, I60, S62, P63, Q64, F65, Q66,

5   S69, K70, V71, Y72, V73, W93, R95, N98, W143, A145, S146, T147, E148, Y151, G159, I161, K163, G171, S179, Y181, S197, L199, Y214, E215 or R216, wherein said functional red fluorescent protein has a different fluorescent property compared to said Anthozoan red fluorescent protein (SEQ. ID. NO. 7).

10

77. A transgenic organism, comprising; a nucleic acid molecule encoding a functional red fluorescent protein whose sequence differs from the amino acid sequence of an Anthozoan red fluorescent protein (SEQ. ID. NO. 7) by at least one amino acid substitution at position D59, I60, S62, P63, Q64, F65, Q66,

15   S69, K70, V71, Y72, V73, W93, R95, N98, W143, A145, S146, T147, E148, Y151, G159, I161, K163, G171, S179, Y181, S197, L199, Y214, E215 or R216, wherein said functional red fluorescent protein has a different fluorescent property compared to said Anthozoan red fluorescent protein (SEQ. ID. NO. 7).

20

78. A method for identifying a protein – protein interaction, comprising;

a)   providing a population of cells comprising,
        a functional red fluorescent protein whose sequence differs from the
        amino acid sequence of an Anthozoan red fluorescent protein (SEQ.

25   ID. NO. 7) by at least one amino acid substitution at position D59, I60,
        S62, P63, Q64, F65, Q66, S69, K70, V71, Y72, V73, W93, R95, N98,
        W143, A145, S146, T147, E148, Y151, G159, I161, K163, G171,
        S179, Y181, S197, L199, Y214, E215 or R216, wherein said
        functional red fluorescent protein has a different fluorescent property

30   compared to said Anthozoan red fluorescent protein (SEQ. ID. NO. 7),
        and wherein said functional red fluorescent protein is operably coupled
        to a first protein of interest,

b)   introducing a library of test proteins of interest operably coupled to

a functional green fluorescent protein into said population of cells,

wherein said functional green fluorescent protein and said functional

red fluorescent protein can undergo fluorescence energy transfer

5   (FRET), and

wherein each member of said population of cells recieves on avearge

one member of said library of test proteins of interest operably coupled

to said functional green fluorescent protein,

c)   screening said population of cells for FRET between said functional

10   green fluorescent protein and said functional red fluorescent protein,

and

d)   comparing the FRET in step c) to the FRET in a control cell in the

absence of said library of test proteins of interest operably coupled to

said functional green fluorescent protein.

15

79.   A method for identifying a modulator of protein – protein interactions,

comprising;

a)   contacting a cell with a test chemical, wherein said cell comprises,

i)   a functional red fluorescent protein whose sequence differs

20   from the amino acid sequence of an Anthozoan red fluorescent

protein (SEQ. ID. NO. 7) by at least one amino acid

substitution at position D59, I60, S62, P63, Q64, F65, Q66,

S 9, K70, V71, Y72, V73, W93, R95, N98, W143, A145,

S146, T147, E148, Y151, G159, I161, K163, G171, S179,

25   Y181, S197, L199, Y214, E215 or R216, wherein said

functional red fluorescent protein has a different fluorescent

property compared to said Anthozoan red fluorescent protein

(SEQ. ID. NO. 7), and wherein said functional red fluorescent

protein is operably coupled to a first protein of interest,

30   ii)   a functional green fluorescent protein, wherein said functional

green fluorescent protein is operably coupled to a second

protein of interest, and wherein said functional green

fluorescent protein and said functional red fluorescent protein

undergo fluorescence energy transfer (FRET) when said first

operably coupled protein of interest and said second operably

protein of interest associate,

5   b)   detecting FRET between said functional green fluorescent protein and

said functional red fluorescent protein in the presence of said test

chemical, and

c)   comparing the FRET in step b) to the FRET in a control cell in the

absence of said test chemical.

10

80.   The method of claim 79, further comprising the step of contacting said cell

with an activator prior to the addition of said test chemical.

81.   The method of claim 79, further comprising the step of detecting the viability

15   of said cell

82.   A test chemical identified by the method of claims 79.

83.   A pharmaceutical composition comprising a test chemical identified by the

20   method of claim 79.

84.   The pharmaceutical composition of claim 79, further comprising a

pharaceutically acceptable carrier.

```
                    Apol
                   wwwwww
                    EcoRI
                   wwwwww
                            PaeR7I
                           wwwwwww
              Tsp509
              wwwww
                    XhoI
                   wwwwwww
       AcsI                              Met Val Arg Ser Ser Lys Asn Val Ile Lys Glu-
+1    wwwww
  1   CCGAATTCTC GAGCCACCAT GGTGAGGAGC AGCAAGAACG TGATCAAGGA
      GGCTTAAGAG CTCGGTGGTA CCACTCCTCG TCGTTCTTGC ACTAGTTCCT

                            Avill
                           wwwwwwww
              Rcal            FspI
             wwwwwww         wwwwwwww
              BspHl           AosI
             wwwwwww         wwwwwww
+1   -Glu Phe Met   Arg  Phe Lys Val Arg Met Glu Gly Thr Val Asn Gly His Glu Phe-
 51   GTTCATGAGG TTCAAGGTGC GCATGGAGGG CACCGTGAAC GGCCACGAGT
      CAAGTACTCC AAGTTCCACG CGTACCTCCC GTGGCACTTG CCGGTGCTCA

+1   -Phe Glu Ile   Glu  Gly Glu  Gly  Glu Gly Arg Pro Tyr Glu Gly His Asn Thr
101   TCGAGATCGA GGGCGAGGGC GAGGGCAGGC CCTACGAGGG CCACAACACC
      AGCTCTAGCT CCCGCTCCCG CTCCCGTCCG GGATGCTCCC GGTGTTGTGG

              BfrI
             wwwwwwwww
           HindIll
          wwwwwww
           AflII                                          BsmFl
          wwwwwww                                        wwwwww
+1   Val Lys Leu Lys Val Thr Lys Gly Gly Pro Leu Pro Phe Ala Trp Asp Ile-
151   GTGAAGCTTA AGGTGACCAA GGGCGGCCCC CTGCCCTTCG CCTGGGACAT
      CACTTCGAAT TCCACTGGTT CCCGCCGGGG GACGGGAAGC GGACCCTGTA

           BanII
          wwwwwww
+1   -Ile Leu Ser Pro Gln Phe Gln Tyr Gly Ser Lys Val Tyr Val Lys His Pro Ala-
201   CCTGAGCCCC CAGTTCCAGT ACGGCAGCAA GGTGTACGTG AAGCACCCCG
      GGACTCGGGG GTCAAGGTCA TGCCGTCGTT CCACATGCAC TTCGTGGGGC

                            CelII
                           wwwwwwwww
                            EspI
                           wwwwwwwww
                            Bpu1102I
                           wwwwwwww
+1   -Ala Asp Ile  Pro Asp Tyr Lys Lys Leu Ser Phe Pro Glu Gly Phe Lys Trp
251   CCGACATCCC CGACTACAAG AAGCTGAGCT TCCCCGAGGG CTTCAAGTGG
      GGCTGTAGGG GCTGATGTTC TTCGACTCGA AGGGGCTCCC GAAGTTCACC

+1    Glu Arg Val Met Asn Phe Glu Asp Gly Gly Val Val Thr Val Thr Gln Asp-
301   GAGAGGGTGA TGAACTTCGA GGACGGCGGC GTGGTGACCG TGACCCAGGA
      CTCTCCCACT ACTTGAAGCT CCTGCCGCCG CACCACTGGC ACTGGGTCCT

              SfcI
             wwwwwwwww
           PstI
          wwwwwwww
+1   -Asp Ser Ser Leu Gln Asp Gly Cys Phe Ile Tyr Lys Val Lys Phe Ile Gly Val-
351   CAGCAGCCTG CAGGACGGCT GCTTCATCTA CAAGGTGAAG TTCATCGGCG
      GTCGTCGGAC GTCCTGCCGA CGAAGTAGAT GTTCCACTTC AAGTAGCCGC

                                                    BpuAl
                                                   wwwwwwww
                                        StaNI        Bbsl
                                       wwwwwww      wwwwwww
+1   -Val Asn Phe Pro Ser Asp Gly Pro Val Met Gln Lys Lys Thr Met Gly Trp
401   TGAACTTCCC CAGCGACGGC CCCGTGATGC AGAAGAAGAC CATGGGCTGG
      ACTTGAAGGG GTCGCTGCCG GGGCACTACG TCTTCTTCTG GTACCCGACC
```

FIG. 1

```
                                                    FnuDII
                                                    ᴧᴧᴧᴧᴧᴧ
             StuI                                    MvnI
             ᴧᴧᴧᴧᴧᴧᴧᴧ                                ᴧᴧᴧᴧᴧᴧ
                                                     ThaI                              XhoII
             AatI                    HaeII          ᴧᴧᴧᴧᴧᴧ                              ᴧ
             ᴧᴧᴧᴧᴧᴧ                  ᴧᴧᴧᴧᴧᴧᴧ         BstUI                             BstYI
             ᴧᴧᴧᴧᴧᴧ                  ᴧᴧᴧᴧᴧᴧᴧ         ᴧᴧᴧᴧᴧᴧ                             ᴧ
   +1    Glu Ala Ser  Thr  Glu Arg Leu  Tyr Pro  Arg Asp Gly Val  Leu  Lys Gly Glu
   451   GAGGCCTCCA CCGAGCGCCT GTACCCCCGC GACGGCGTGC TGAAGGGCGA
         CTCCGGAGGT GGCTCGCGGA CATGGGGGCG CTGCCGCACG ACTTCCCGCT

             BstYI
             ᴧᴧᴧᴧᴧᴧ
             XhoII
             ᴧᴧᴧᴧᴧᴧ
             AlwI                                              SexAI
             ᴧᴧᴧᴧᴧᴧ                                            ᴧᴧᴧᴧᴧᴧᴧᴧᴧ
   +1    -Glu Ile His  Lys  Ala Leu Lys  Leu  Lys Asp Gly  Gly His  Tyr  Leu Val Glu Phe-
   501   GATCCACAAG GCCCTGAAGC TGAAGGACGG CGGCCACTAC CTGGTGGAGT
         CTAGGTGTTC CGGGACTTCG ACTTCCTGCC GCCGGTGATG GACCACCTCA

                          BstXI                          PvuII
                          ᴧᴧᴧᴧᴧᴧᴧᴧᴧᴧᴧᴧᴧᴧ                  ᴧᴧᴧᴧᴧᴧᴧ
                              MluNI                                      NciI
                              ᴧᴧᴧᴧᴧᴧᴧ                                    ᴧᴧᴧᴧᴧᴧ
                              MscI                       MspA1I  MspI
                              ᴧᴧᴧᴧᴧᴧᴧ                     ᴧᴧᴧᴧᴧᴧᴧ  ᴧᴧᴧᴧᴧᴧ
                              BalI                       BsgI    HpaII
   +1    -Phe Lys Ser  Ile  Tyr Met Ala  Lys  Lys Pro  Val Gln Leu  Pro  Gly Tyr  Tyr
   551   TCAAGTCCAT CTACATGGCC AAGAAGCCCG TGCAGCTGCC CGGCTACTAC
         AGTTCAGGTA GATGTACCGG TTCTTCGGGC ACGTCGACGG GCCGATGATG

   +1    Tyr Val Asp  Ser  Lys Leu Asp  Ile  Thr Ser  His Asn Glu  Asp  Tyr Thr Ile-
   601   TACGTGGACT CCAAGCTGGA CATCACCAGC CACAACGAGG ACTACACCAT
         ATGCACCTGA GGTTCGACCT GTAGTGGTCG GTGTTGCTCC TGATGTGGTA

                                                                            Sal
                                                                            ᴧ
                                   AvaII                                    AccI
                                   ᴧᴧᴧᴧᴧᴧᴧ                                   ᴧ
   +1    -Ile Val Glu  Gln  Tyr Glu Arg  Thr  Glu Gly  Arg His His  Leu  Phe Leu  ***
   651   CGTGGAGCAG TACGAGAGGA CCGAGGGCAG GCACCACCTG TTCCTGTGAG
         GCACCTCGTC ATGCTCTCCT GGCTCCCGTC CGTGGTGGAC AAGGACACTC

                      HpaI
                      ᴧᴧᴧᴧᴧᴧᴧᴧᴧ
             SalI
             ᴧᴧᴧᴧᴧᴧ
             AccI
             ᴧᴧᴧᴧᴧᴧ
   701   TCGACGTTAA CCC
         AGCTGCAATT GGG
```
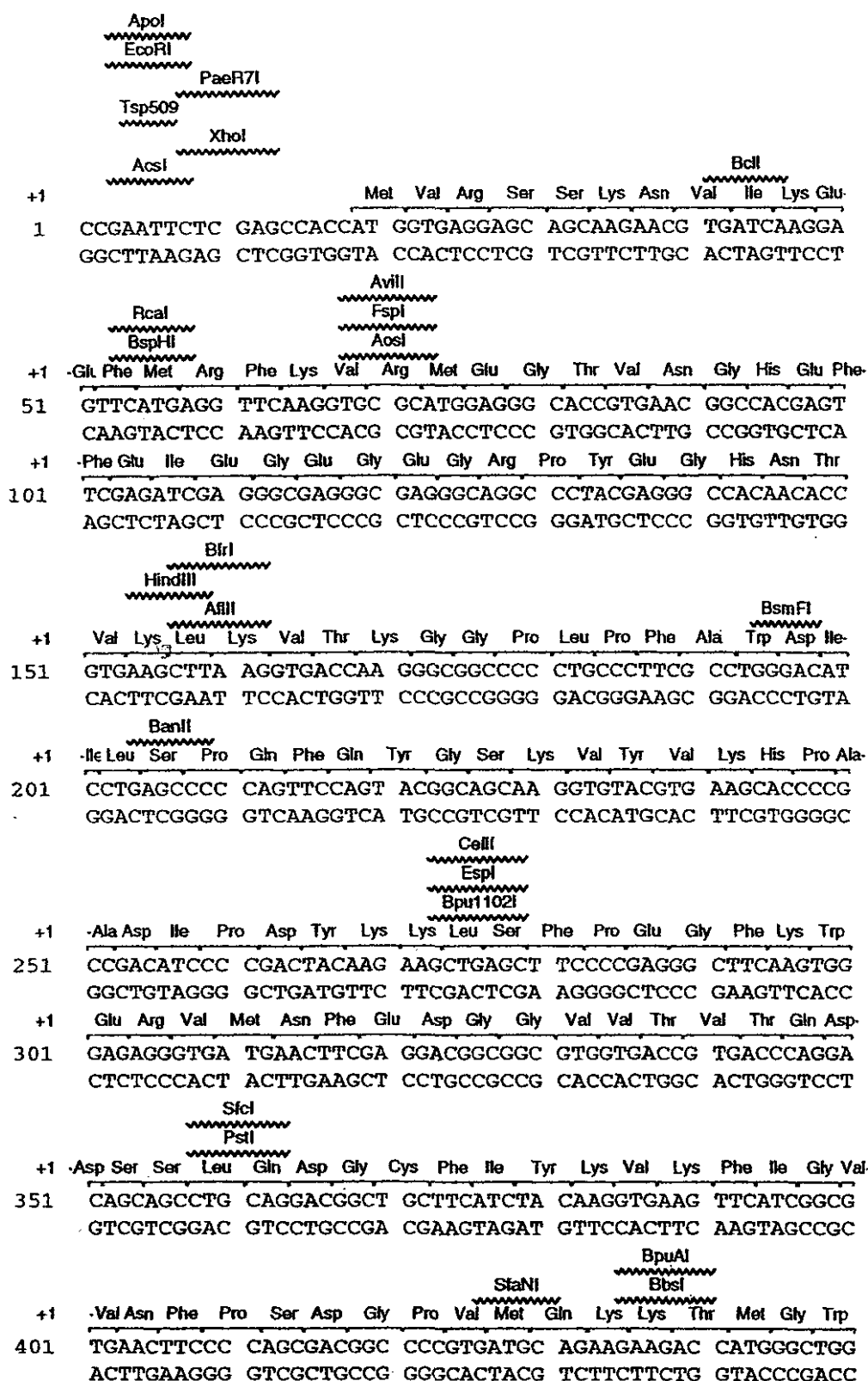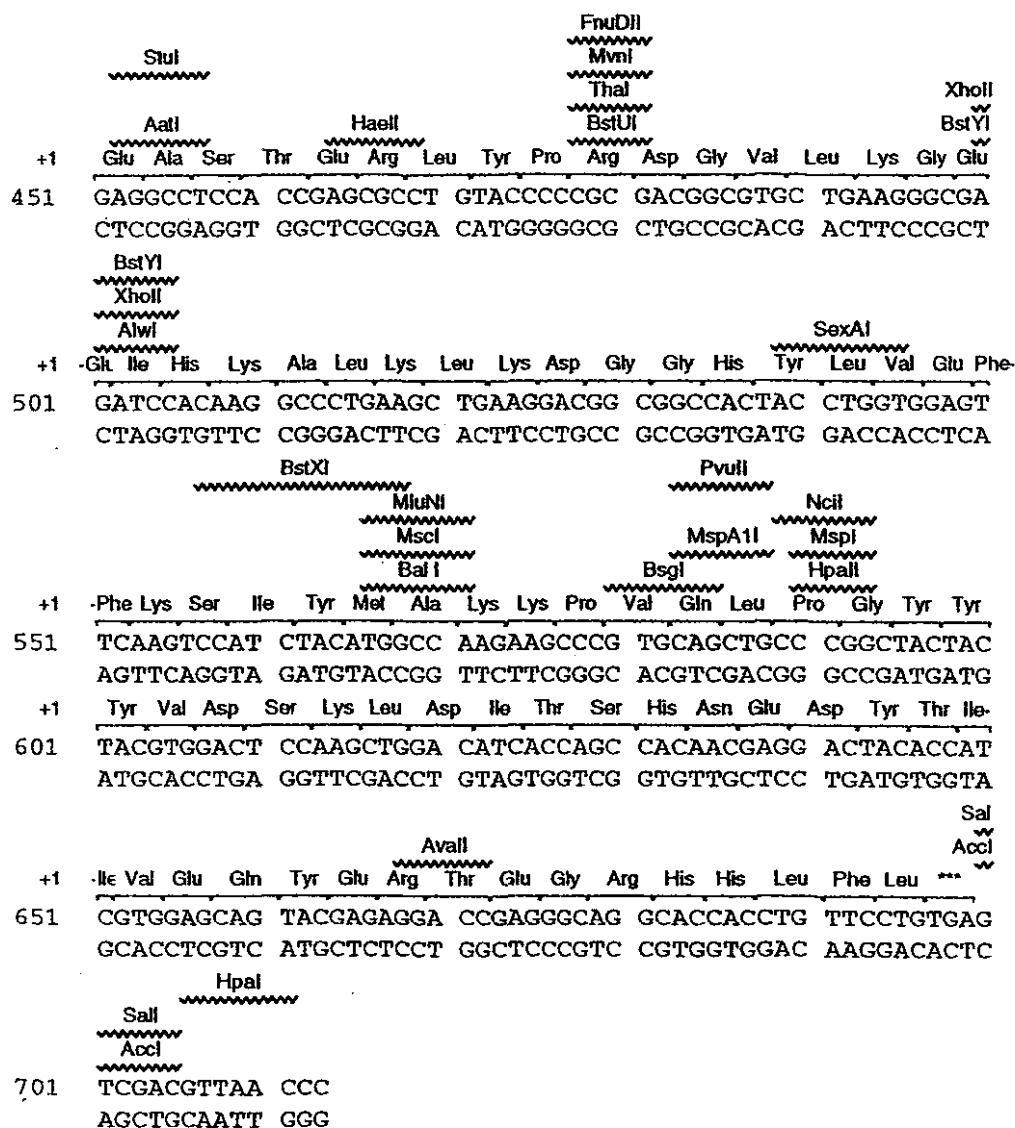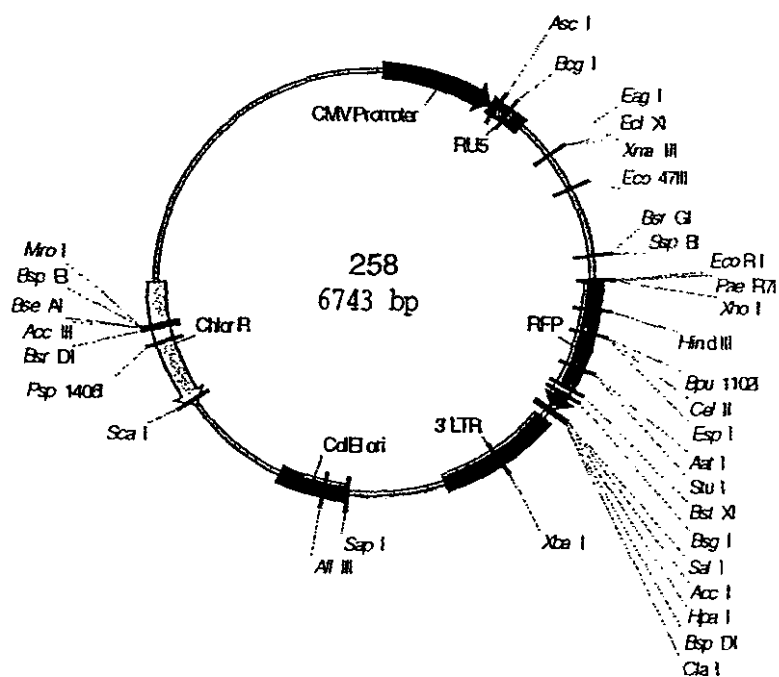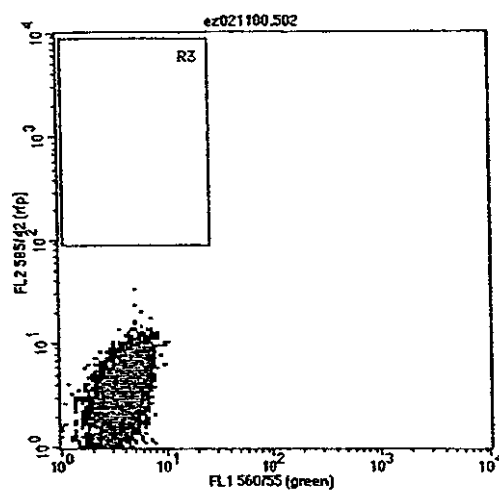
FIG. 1 continued

FIG. 2

Wild Type

Cells Transformed with the retroviral
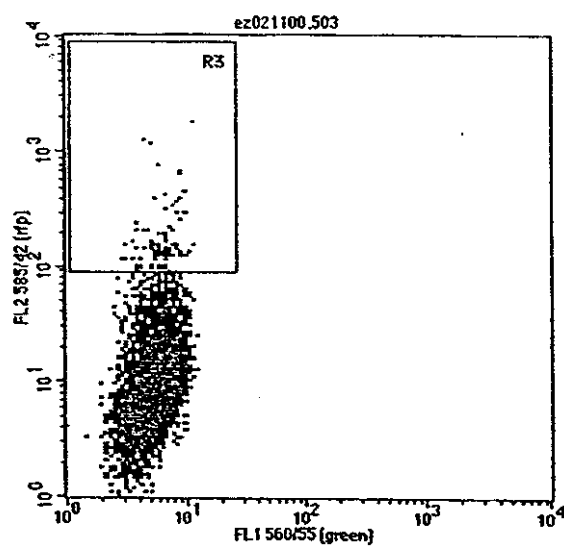expression plasmid containing synthetic RFP

FIG. 3

SEQUENCE ID. LISTING

(1) GENERAL INFORMATION:

(iii) NUMBER OF SEQUENCES: 10

5

(2) INFORMATION FOR SEQ. ID. NO.: 1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 690 base pairs

(B) TYPE: nucleic acid

10   (C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

15   (B) LOCATION: 1....690

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO.1:

ATG GCT CTT TCA AAC AAG TTT ATC GGA GAT GAC ATG AAA ATG ACC TAC CAT

20

ATG GAT GGC TGT GTC AAT GGG CAT TAC TTT ACC GTC AAA GGT GAA GGC AAC

GGG AAG CCA TAC GAA GGG ACG CAG ACT TCG ACT TTT AAA GTC ACC ATG GCC

25   AAC GGT GGG CCC CTT GCA TTC TCC TTT GAC ATA CTA TCT ACA GTG TTC AAA

TAT GGA AAT CGA TGC TTT ACT GCG TAT CCT ACC AGT ATG CCC GAC TAT TTC

AAA CAA GCA TTT CCT GAC GGA ATG TCA TAT GAA AGG ACT TTT ACC TAT GAA

30   GAT GGA GGA GTT GCT ACA GCC AGT TGG GAA ATA AGC CTT AAA GGC AAC TGC

TTT GAG CAC AAA TCC ACG TTT CAT GGA GTG AAC TTT CCT GCT GAT GGA CCT

35   GTG ATG GCG AAG AAG ACA ACT GGT TGG GAC CCA TCT TTT GAG AAA ATG ACT

GTC TGC GAT GGA ATA TTG AAG GGT GAT GTC ACC GCG TTC CTC ATG CTG CAA

GGA GGT GGC AAT TAC AGA TGC CAA TTC CAC ACT TCT TAC AAG ACA AAA AAA

40   CCG GTG ACG ATG CCA CCA AAC CAT GTG GTG GAA CAT CGC ATT GCG AGG ACC

GAC CTT GAC AAA GGT GGC AAC AGT GTT CAG CTG ACG GAG CAC GCT GTT GCA

45   CAT ATA ACC TCT GTT GTC CCT TTC TGA

(2) INFORMATION FOR SEQ. ID. NO.2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 696 base pairs

(B) TYPE: nucleic acid

5   (C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

10   (B) LOCATION: 1....696

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO.2:

ATG GCT CAG TCA AAG CAC GGT CTA ACA AAA GAA ATG ACA ATG AAA TAC CGT

15

ATG GAA GGG TGC GTC GAT GGA CAT AAA TTT GTG ATC ACG GGA GAG GGC ATT

GGA TAT CCG TTC AAA GGG AAA CAG GCT ATT AAT CTG TGT GTG GTC GAA GGT

20   GGA CCA TTG CCA TTT GCC GAA GAC ATA TTG TCA GCT GCC TTT AAC TAC GGA

AAC AGG GTT TTC ACT GAA TAT CCT CAA GAC ATA GTT GAC TAT TTC AAG AAC

TCG TGT CCT GCT GGA TAT ACA TGG GAC AGG TCT TTT CTC TTT GAG GAT GGA

25   GCA GTT TGC ATA TGT AAT GCA GAT ATA ACA GTG AGT GTT GAA GAA AAC TGC

ATG TAT CAT GAG TCC AAA TTT TAT GGA GTG AAT TTT CCT GCT GAT GGA CCT

30   GTG ATG AAA AAG ATG ACA GAT AAC TGG GAG CCA TCC TGC GAG AAG ATC ATA

CCA GTA CCT AAG CAG GGG ATA TTG AAA GGG GAT GTC TCC ATG TAC CTC CTT

CTG AAG GAT GGT GGG CGT TTA CGG TGC CAA TTC GAC ACA GTT TAC AAA GCA

35

AAG TCT GTG CCA AGA AAG ATG CCG GAC TGG CAC TTC ATC CAG CAT AAG CTC

ACC CGT GAA GAC CGC AGC GAT GCT AAG AAT CAG AAA TGG CAT CTG ACA GAA

40   CAT GCT ATT GCA TCC GGA TCT GCA TTG CCC TGA

(2) INFORMATION FOR SEQ. ID. NO.3:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 696 base pairs

5   (B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

10   (A) NAME / KEY: Coding Sequence

(B) LOCATION: 1...696

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

15   ATG GCT CAT TCA AAG CAC GGT CTA AAA GAA GAA ATG ACA ATG AAA TAC CAC

ATG GAA GGG TGC GTC AAC GGA CAT AAA TTT GTG ATC ACG GGC GAA GGC ATT

GGA TAT CCG TTC AAA GGG AAA CAG ACT ATT AAT CTG TGT GTG ATC GAA GGG

20   GGA CCA TTG CCA TTT TCC GAA GAC ATA TTG TCA GCT GGC TTT AAG TAC GGA

GAC AGG ATT TTC ACT GAA TAT CCT CAA GAC ATA GTA GAC TAT TTC AAG AAC

25   TCG TGT CCT GCT GGA TAT ACA TGG GGC AGG TCT TTT CTC TTT GAG GAT GGA

GCA GTC TGC ATA TGC AAT GTA GAT ATA ACA GTG AGT GTC AAA GAA AAC TGC

ATT TAT CAT AAG AGC ATA TTT AAT GGA ATG AAT TTT CCT GCT GAT GGA CCT

30   GTG ATG AAA AAG ATG ACA ACT AAC TGG GAA GCA TCC TGC GAG AAG ATC ATG

CCA GTA CCT AAG CAG GGG ATA CTG AAA GGG GAT GTC TCC ATG TAC CTC CTT

35   CTG AAG GAT GGT GGG CGT TAC CGG TGC CAG TTC GAC ACA GTT TAC AAA GCA

AAG TCT GTG CCA AGT AAG ATG CCG GAG TGG CAC TTC ATC AGC ATC AAG CTC

CTC CGT GAA GAC CGC AGC GAT GCT AAG AAT CAG AAG TGG CAG CTG ACA GAG

40   CAT GCT ATT GCA TTC CCT TCT GCC TTG GCC TGA

(2) INFORMATION FOR SEQ. ID. NO.4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 699 base pairs

(B) TYPE: nucleic acid

5   (C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

10   (B) LOCATION: 1...699

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

ATG AGT TGT TCC AAG AGT GTG ATC AAG GAA GAA ATG TTG ATC GAT CTT CAT

15

CTG GAA GGA ACG TTC AAT GGG CAC TAC TTT GAA ATA AAA GGC AAA GGA AAA

GGA CAG CCT AAT GAA GGC ACC AAT ACC GTC ACG CTC GAG GTT ACC AAG GGT

20   GGA CCT CTG CCA TTT GGT TGG CAT ATT TTG TGC CCA CAA TTT CAG TAT GGA

AAC AAG GCA TTT GTC CAC CAC CCT GAC AAC ATA CAT GAT TAT CTA AAG CTG

TCA TTT CCG GAG GGA TAT ACA TGG GAA CGG TCC ATG CAC TTT GAA GAC GGT

25

GGC TTG TGT TGT ATC ACC AAT GAT ATC AGT TTG ACA GGC AAC TGT TTC TAC

TAC GAC ATC AAG TTC ACT GGC TTG AAC TTT CCT CCA AAT GGA CCC GTT GTG

30   CAG AAG AAG ACA ACT GGC TGG GAA CCG AGC ACT GAG CGT TTG TAT CCT CGT

GAT GGT GTG TTG ATA GGA GAC ATC CAT CAT GCT CTG ACA GTT GAA GGA GGT

GGT CAT TAC GCA TGT GAC ATT AAA ACT GTT TAC AGG GCC AAG AAG GCC GCC

35

TTG AAG ATG CCA GGG TAT CAC TAT GTT GAC ACC AAA CTG GTT ATA TGG AAC

AAC GAC AAA GAA TTC ATG AAA GTT GAG GAG CAT GAA ATC GCC GTT GCA CGC

40   CAC CAT CCG TTC TAT GAG CCA AAG AAG GAT AAG TAA

5

(2) INFORMATION FOR SEQ. ID. NO.5:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 678 base pairs

(B) TYPE: nucleic acid

5   (C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

10   (B) LOCATION: 1....678

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO.5:

```
ATG AGG TCT TCC AAG AAT GTT ATC AAG GAG TTC ATG AGG TTT AAG GTT CGC
ATG

GAA GGA ACG GTC AAT GGG CAC GAG TTT GAA ATA GAA GGC GAA GGA GAG GGG
AGG

CCA TAC GAA GGC CAC AAT ACC GTA AAG CTT AAG GTA ACC AAG GGG GGA CCT
TTG

CCA TTT GCT TGG GAT ATT TTG TCA CCA CAA TTT CAG TAT GGA AGC AAG GTA
TAT

GTC AAG CAC CCT GCC GAC ATA CCA GAC TAT AAA AAG CTG TCA TTT CCT GAA
GGA

TTT AAA TGG GAA AGG GTC ATG AAC TTT GAA GAC GGT GGC GTC GTT ACT GTA
ACC

CAG GAT TCC AGT TTG CAG GAT GGC TGT TTC ATC TAC AAG GTC AAG TTC ATT
GGC

GTG AAC TTT CCT TCC GAT GGA CCT GTT ATG CAA AAG AAG ACA ATG GGC TGG
GAA

GCC AGC ACT GAG CGT TTG TAT CCT CGT GAT GGC GTG TTG AAA GGA GAG ATT
CAT

AAG GCT CTG AAG CTG AAA GAC GGT GGT CAT TAC CTA GTT GAA TTC AAA AGT
ATT

TAC ATG GCA AAG AAG CCT GTG CAG CTA CCA GGG TAC TAC TAT GTT GAC TCC
AAA

CTG GAT ATA ACA AGC CAC AAC GAA GAC TAT ACA ATC GTT GAG CAG TAT GAA
AGA

ACC GAG GGA CGC CAC CAT CTG TTC CTT TAA
```

15

20

25

30

35

40

45

50

(2) INFORMATION FOR SEQ. ID. NO.6:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 801 base pairs

(B) TYPE: nucleic acid

5   (C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

10   (B) LOCATION: 1......801

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO.6:

ATG AAG TGT AAA TTT GTG TTC TGC CTG TCC TTC TTG GTC CTC GCC ATC ACA

AAC GCG AAC ATT TTT TTG AGA AAC GAG GCT GAC TTA GAA GAG AAG ACA TTG

AGA ATA CCA AAA GCT CTA ACC ACC ATG GGT GTG ATT AAA CCA GAC ATG AAG

ATT AAG CTG AAG ATG GAA GGA AAT GTA AAC GGG CAT GCT TTT GTG ATC GAA

GGA GAA GGA GAA GGA AAG CCT TAC GAT GGG ACA CAC ACT TTA AAC CTG GAA

GTG AAG GAA GGT GCG CCT CTG CCT TTT TCT TAC GAT ATC TTG TCA AAC GCG

TTC CAG TAC GGA AAC AGA GCA TTG ACA AAA TAC CCA GAC GAT ATA GCA GAC

TAT TTC AAG CAG TCG TTT CCC GAG GGA TAT TCC TGG GAA AGA ACC ATG ACT

TTT GAA GAC AAA GGC ATT GTC AAA GTG AAA AGT GAC ATA AGC ATG GAG GAA

GAC TCC TTT ATC TAT GAA ATT CGT TTT GAT GGG ATG AAC TTT CCT CCC AAT

GGT CCG GTT ATG CAG AAA AAA ACT TTG AAG TGG GAA CCA TCC ACT GAG ATT

ATG TAC GTG CGT GAT GGA GTG CTG GTC GGA GAT ATT AGC ATC TCT CTG TTG

CTG GAG GGA GGT GGC CAT TAC CGA TGT GAC TTC AAA AGT ATT TAC AAA GCA

AAA AAA GTT GTC AAA TTG CCA GAC TAT CAC TTT GTG GAC CAT CGC ATT GAG

ATC TTG AAC CAT GAC AAG GAT TAC AAC AAA GTA ACG CTG TAT GAG AAT GCA

GTT GCT CGC TAT TCT TTG CTG CCA AGT CAG GCC TAG

45

(2) INFORMATION FOR SEQ. ID. NO.7:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 225 amino acids

(B) TYPE: amino acid

5   (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

10   (B) LOCATION: 1......225

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO.7:

    Met Arg Ser Ser Lys Asn Val Ile Lys Glu Phe Met Arg Phe Lys Val Arg
    Met Glu

15

    Gly Thr Val Asn Gly His Glu Phe Glu Ile Glu Gly Glu Gly Glu Gly Arg
    Pro Tyr

    Glu Gly His Asn Thr Val Lys Leu Lys Val Thr Lys Gly Gly Pro Leu Pro
20   Phe Ala

    Trp Asp Ile Leu Ser Pro Gln Phe Gln Tyr Gly Ser Lys Val Tyr Val Lys
    His Pro

25   Ala Asp Ile Pro Asp Tyr Lys Lys Leu Ser Phe Pro Glu Gly Phe Lys Trp
    Glu Arg

    Val Met Asn Phe Glu Asp Gly Gly Val Val Thr Val Thr Gln Asp Ser Ser
    Leu Gln
30

    Asp Gly Cys Phe Ile Tyr Lys Val Lys Phe Ile Gly Val Asn Phe Pro Ser
    Asp Gly

    Pro Val Met Gln Lys Lys Thr Met Gly Trp Glu Ala Ser Thr Glu Arg Leu
35   Tyr Pro

    Arg Asp Gly Val Leu Lys Gly Glu Ile His Lys Ala Leu Lys Leu Lys Asp
    Gly Gly

40   His Tyr Leu Val Glu Phe Lys Ser Ile Tyr Met Ala Lys Lys Pro Val Gln
    Leu Pro

    Gly Tyr Tyr Tyr Val Asp Ser Lys Leu Asp Ile Thr Ser His Asn Glu Asp
    Tyr Thr
45

    Ile Val Glu Gln Tyr Glu Arg Thr Glu Gly Arg His His Leu Phe Leu

8

(2) INFORMATION FOR SEQ. ID. NO.8:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 681 base pairs

(B) TYPE: nucleic acid

5   (C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: synthetic

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

10   (B) LOCATION: 1......681

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO.8:

    ATG GTG AGG AGC AGC AAG AAC GTG ATC AAG GAG TTC ATG AGG TTC AAG GTG

15 CGC ATG GAG GGC ACC GTG AAC GGC CAC GAG TTC GAG ATC GAG GGC GAG GGC

    GAG GGC AGG CCC TAC GAG GGC CAC AAC ACC GTG AAG CTT AAG GTG ACC AAG

    GGC GGC CCC CTG CCC TTC GCC TGG GAC ATC CTG AGC CCC CAG TTC CAG TAC

    GGC AGC AAG GTG TAC GTG AAG CAC CCC GCC GAC ATC CCC GAC TAC AAG AAG

    CTG AGC TTC CCC GAG GGC TTC AAG TGG GAG AGG GTG ATG AAC TTC GAG GAC

20 GGC GGC GTG GTG ACC GTG ACC CAG GAC AGC AGC CTG CAG GAC GGC TGC TTC

    ATC TAC AAG GTG AAG TTC ATC GGC GTG AAC TTC CCC AGC GAC GGC CCC GTG

    ATG CAG AAG AAG ACC ATG GGC TGG GAG GCC TCC ACC GAG CGC CTG TAC CCC

    CGC GAC GGC GTG CTG AAG GGC GAG ATC CAC AAG GCC CTG AAG CTG AAG GAC

    GGC GGC CAC TAC CTG GTG GAG TTC AAG TCC ATC TAC ATG GCC AAG AAG CCC

25 GTG CAG CTG CCC GGC TAC TAC TAC GTG GAC TCC AAG CTG GAC ATC ACC AGC

    CAC AAC GAG GAC TAC ACC ATC GTG GAG CAG TAC GAG AGG ACC GAG GGC AGG

    CAC CAC CTG TTC CTG TGA

(2) INFORMATION FOR SEQ. ID. NO.9:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 226 amino acids

(B) TYPE: amino acid

5   (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

10   (B) LOCATION: 1......226

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO.9:

15   Met Val Arg Ser Ser Lys Asn Val Ile Lys Glu Phe Met Arg Phe Lys Val Arg Met

Glu Gly Thr Val Asn Gly His Glu Phe Glu Ile Glu Gly Glu Gly Glu Gly Arg Pro

Tyr Glu Gly His Asn Thr Val Lys Leu Lys Val Thr Lys Gly Gly Pro Leu Pro Phe

20   Ala Trp Asp Ile Leu Ser Pro Gln Phe Gln Tyr Gly Ser Lys Val Tyr Val Lys His

Pro Ala Asp Ile Pro Asp Tyr Lys Lys Leu Ser Phe Pro Glu Gly Phe Lys Trp Glu

Arg Val Met Asn Phe Glu Asp Gly Gly Val Val Thr Val Thr Gln Asp Ser Ser Leu

25   Gln Asp Gly Cys Phe Ile Tyr Lys Val Lys Phe Ile Gly Val Asn Phe Pro Ser Asp

Gly Pro Val Met Gln Lys Lys Thr Met Gly Trp Glu Ala Ser Thr Glu Arg Leu Tyr

30   Pro Arg Asp Gly Val Leu Lys Gly Glu Ile His Lys Ala Leu Lys Leu Lys Asp Gly

Gly His Tyr Leu Val Glu Phe Lys Ser Ile Tyr Met Ala Lys Lys Pro Val Gln Leu

Pro Gly Tyr Tyr Tyr Val Asp Ser Lys Leu Asp Ile Thr Ser His Asn Glu Asp Tyr

35   Thr Ile Val Glu Gln Tyr Glu Arg Thr Glu Gly Arg His His Leu Phe Leu

(2) INFORMATION FOR SEQ. ID. NO. 10

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 720 base pairs

(B) TYPE: nucleic acid

5   (C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

10   (B) LOCATION: 1....720

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO. 10:


```
     ATG GTG AGC AAG GGC GAG GAG CTG TTC ACC GGG GTG GTG CCC ATC CTG GTC
15
     GAG CTG GAC GGC GAC GTA AAC GGC CAC AAG TTC AGC GTG TCC GGC GAG GGC

     GAG GGC GAT GCC ACC TAC GGC AAG CTG ACC CTG AAG TTC ATC TGC ACC ACC

20   GGC AAG CTG CCC GTG CCC TGG CCC ACC CTC GTG ACC ACC TTC TCC TAC GGC

     GTG CAG TGC TTC AGC CGC TAC CCC GAC CAC ATG AAG CAG CAC GAC TTC TTC

     AAG TCC GCC ATG CCC GAA GGC TAC GTC CAG GAG CGC ACC ATC TTC TTC AAG
25
     GAC GAC GGC AAC TAC AAG ACC CGC GCC GAG GTG AAG TTC GAG GGC GAC ACC

     CTG GTG AAC CGC ATC GAG CTG AAG GGC ATC GAC TTC AAG GAG GAC GGC AAC

30   ATC CTG GGG CAC AAC CTG GAG TAC AAC TAC AAC AGC CAC AAC GTC TAT ATC

     ATG GCC GAC AAG CAG AAG AAC GGC ATC AAG GTG AAC TTC AAG ATC CGC CAC

     AAC ATC GAG GAC GGC AGC GTG CAG CTC GCC GAC CAC TAC CAG CAG AAC ACC
35
     CCC ATC GGC GAC GGC CCC GTG CTG CTG CCC GAC AAC CAC TAC CTG AGC ACC

     CAG TCC GCC CTG AGC AAA GAC CCC AAC GAG AAG CGC GAT CAC ATG GTC CTG

40   CTG GAG TTC GTG ACC GCC GCC GGG ATC ACT CTC GGC ATG GAC GAG CTG TAC

     AAG TAA
```

事項編號	:	9880304
參考編號	:	WO 01/62919 A1
名稱	:	改良螢光蛋白
摘要	:	可改變用途的紅色螢光蛋白、它們的核酸編碼，以及它們的使用方法。