Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 12501–12504, December 1994
Biochemistry

Wavelength mutations and posttranslational autoxidation of green fluorescent protein

(*Aequorea victoria*/blue fluorescent protein/*Escherichia coli*/imidazolidinone)

ROGER HEIM*, DOUGLAS C. PRASHER†, AND ROGER Y. TSIEN*‡

*Howard Hughes Medical Institute, University of California, San Diego, La Jolla, CA 92093-0647; and †U.S. Dept. of Agriculture, Animal and Plant Health Inspection Service, Otis Methods Development Center, Otis Air National Guard Base, MA 02542

Communicated by Eric R. Kandel, August 19, 1994

ABSTRACT **The green fluorescent protein (GFP) of the jellyfish *Aequorea victoria* is an unusual protein with strong visible absorbance and fluorescence from a *p*-hydroxybenzylidene-imidazolidinone chromophore, which is generated by cyclization and oxidation of the protein's own Ser-Tyr-Gly sequence at positions 65–67. Cloning of the cDNA and heterologous expression of fluorescent protein in a wide variety of organisms indicate that this unique posttranslational modification must be either spontaneous or dependent only on ubiquitous enzymes and reactants. We report that formation of the final fluorophore requires molecular oxygen and proceeds with a time constant (≈4 hr at 22°C and atmospheric pO_2) independent of dilution, implying that the oxidation does not require enzymes or cofactors. GFP was mutagenized and screened for variants with altered spectra. The most striking mutant fluoresced blue and contained histidine in place of Tyr-66. The availability of two visibly distinct colors should significantly extend the usefulness of GFP in molecular and cell biology by enabling *in vivo* visualization of differential gene expression and protein localization and measurement of protein association by fluorescence resonance energy transfer.**

Proteins are often labeled with fluorescent tags to detect their localization and sometimes their conformational changes both *in vitro* and in intact cells. Such labeling is essential both for immunofluorescence and for fluorescence analog cytochemistry, in which the biochemistry and trafficking of proteins are monitored after microinjection into living cells (1). Traditionally, fluorescence labeling is done by purifying proteins and then covalently conjugating them to reactive derivatives of organic fluorophores. The stoichiometry and locations of dye attachment are often difficult to control, and careful repurification of the proteins is usually necessary. If the proteins are to be used inside living cells, a final challenging step is to get them across the plasma membrane via micropipet techniques or various methods of reversible permeabilization.

An alternative would be to devise molecular biological means to generate fluorescent proteins. The natural UV fluorescence of tryptophan residues is of little use except in the simplest *in vitro* samples, because the wavelengths are too short and tryptophan is too ubiquitous for any one protein to stand out in a complex biological mixture. A more promising strategy would be to concatenate the gene for the nonfluorescent protein of interest with the gene for a naturally fluorescent protein and express the fusion product. The most highly fluorescent proteins known are the phycobiliproteins (2), but their fluorescence depends entirely on correct enzymatic insertion of a difficult-to-obtain tetrabilin chromophore into a large apoprotein (3). The green fluorescent protein (GFP) from the jellyfish *Aequorea victoria* is a much smaller molecule of 238 amino acids, whose natural function seems to be to convert the blue chemiluminescence of the Ca^{2+}-sensitive photoprotein aequorin into green emission (4). GFP's absorption bands in the blue and emission peak in the green do not arise from a distinct cofactor but rather from an internal *p*-hydroxybenzylideneimidazolidinone chromophore generated by cyclization and oxidation of a Ser-Tyr-Gly sequence at residues 65–67 (5). When the gene for GFP was first cloned (6), whether jellyfish-specific enzymes were required for this posttranslational modification was unknown. Heterologous expression of the gene in *Escherichia coli* (7, 8), *Caenorhabditis elegans* (7), *Saccharomyces cerevisiae* (R.H., S. D. Emr and R.Y.T., unpublished data), and *Drosophila melanogaster* (9) showed that additional *Aequorea*-specific enzymes were not required because the protein became brightly fluorescent in all these organisms.

The ability to generate fluorescence *in situ* by expressing the gene for GFP has opened up tremendous possibilities for continuously monitoring gene expression, cell developmental fates, and protein trafficking in living, minimally perturbed cells, tissues, and organisms. Nevertheless, major questions about GFP itself remain. What is the mechanism of fluorophore formation? How does fluorescence relate to protein structure? Can its fluorescence properties be tailored and improved—in particular, to provide a second distinguishable color for comparison of independent proteins and gene expression events? This study provides initial answers.

MATERIALS AND METHODS

The coding region of a clone of *A. victoria gfp10* cDNA (6) was amplified by the PCR (7) to create *Nde* I and *Bam*HI sites at the 5′ and 3′ ends, respectively, and cloned behind the T7 promoter of pGEMEX-2 (Promega), replacing most of the T7 gene *10*. The resulting plasmid was transformed into *E. coli* strain JM109(DE3), and high-level expression was achieved by growing the cultures at 24°C to saturation without induction by isopropyl β-D-thiogalactoside. Random mutagenesis of the *gfp* cDNA was done by hydroxylamine treatment (10) or by increasing the error rate of the PCR with 0.1 mM $MnCl_2$, 50 μM dATP, and 200 μM of dGTP, dCTP, and dTTP (11). The product was ligated into pGEMEX-2 and subsequently transformed into strain JM109(DE3). Colonies on agar were visually screened for different emission colors and ratios of brightness when excited at 475 vs. 395 nm, supplied by a xenon lamp and grating monochromator for which the output beam was expanded to illuminate an entire culture dish.

To prepare soluble extracts, cells from a 1.5-ml suspension were collected, washed, and resuspended in 150 μl of 50 mM Tris·HCl, pH 8.0/2 mM EDTA. Lysozyme and DNase I were added to 0.2 mg/ml and 20 μg/ml, respectively, and the samples were incubated on ice until lysis occurred (1–2 hr).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: GFP, green fluorescent protein.

‡To whom reprint requests should be addressed.

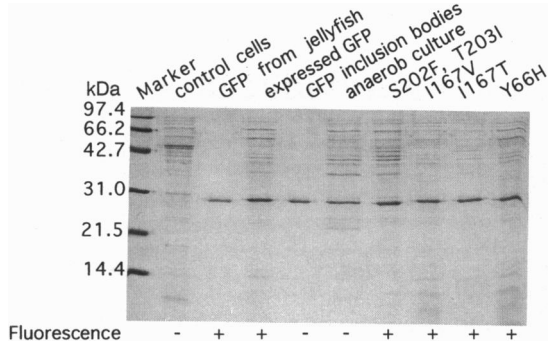

FIG. 1. GFP expression in *E. coli*: Variations in fluorescence properties despite equalized apoprotein contents. Different versions of GFP are compared by gel electrophoresis and Coomassie blue staining. Soluble extracts of *E. coli* expressing GFP show a predominant band that is absent in extracts from control cells and has the same electrophoretic mobility as native GFP isolated from the jellyfish *A. victoria* (5). Inclusion bodies of expressing cells consist mainly of nonfluorescent GFP, which has the same mobility as soluble GFP. Nonfluorescent soluble GFP of anaerobically grown (anerob) cultures is also a major band with correct mobility. Soluble extracts of the mutated clones H9 (S202F, T203I), P9 (I167V), P11 (I167T), and P4 (Y66H) again contain a dominant protein with essentially the same molecular mass.

The lysates were then clarified by centrifuging at 12,000 × *g* for 15 min. Anaerobic cultures were grown in GasPak pouches (Becton Dickinson). Inclusion bodies were obtained as described (12).

Excitation and emission spectra were measured with 1.8-nm bandwidths, and the nonscanning wavelength was set to the appropriate peak. Excitation spectra were corrected with a rhodamine B quantum counter, whereas emission spectra (except for mutation Y66W) were corrected for monochromator and detector efficiencies by using manufacturer-supplied correction spectra.

RESULTS AND DISCUSSION

GFP was expressed in *E. coli* under the control of a T7 promoter for quantitative analysis of the properties of the recombinant protein. Gel electrophoresis under denaturing conditions showed protein of the expected molecular mass (27 kDa) as a dominant band (Fig. 1), which could be quantified simply by densitometry of staining with Coomassie blue. Soluble recombinant GFP proved to have identical spectra and the same or even slightly more fluorescence per mol of protein as GFP purified from *A. victoria*, showing that the soluble protein in *E. coli* undergoes correct folding and oxidative cyclization with as high an efficiency as in the jellyfish. The bacteria also contained inclusion bodies consisting of protein indistinguishable from jellyfish or soluble recombinant protein on denaturing gels (Fig. 1). However, this material was completely nonfluorescent, lacked the visible absorbance bands of the chromophore, and did not become fluorescent when solubilized and subjected to protocols that renature GFP (13). Therefore protein from inclusion bodies seemed unable to generate the internal chromophore.

An interesting intermediate stage in protein maturation could be generated by growing the bacteria anaerobically. The soluble protein again looked the same as GFP on denaturing gels (Fig. 1) but was nonfluorescent. In this case, fluorescence gradually developed after admission of air, even when fresh protein synthesis was blocked by using puromycin and tetracycline. Evidently the soluble nonfluorescent protein synthesized under anaerobic conditions was ready to become fluorescent once atmospheric oxygen was readmitted. The fluorescence per protein molecule approached its final asymptotic value with a single-exponential time course and a rate constant of $0.24 \pm 0.06\ hr^{-1}$ (at 22°C), measured either in intact cells with protein-synthesis inhibitors or in lysates in which the soluble proteins and cofactors were $\approx 10^5$ more dilute than in cells. Such pseudo-first-order kinetics strongly suggest that no enzymes or cofactors are necessary for the final step of fluorophore formation in GFP. A tentative molecular interpretation is presented in Fig. 2. If the newly translated apoprotein evades precipitation into inclusion bodies, the amino group of Gly-67 might cyclize onto the carbonyl group of Ser-65 to form an imidazolidin-5-one, where the process would stop were O_2 absent. The new N=C double bond would be expected to promote dehydrogenation to form a conjugated chromophore; imidazolidin-5-ones undergo autoxidative formation of double bonds at the 4-position (14, 15), which would complete the fluorophore (Fig. 2, upper right). Because fluorophore formation requires at least one step with a time constant of ≈4 hr, use of GFP as a reporter protein to monitor faster changes in promoter activity may be problematic.

A major question in protein photophysics is how a single chromophore can give widely different spectra depending on its local protein environment. This question has received the most attention with respect to the multiple colors of visual pigments based on retinal (16) but is also important in GFP. The GFP from *Aequorea* and that of the sea pansy *Renilla reniformis* share the same chromophore, yet *Aequorea* GFP

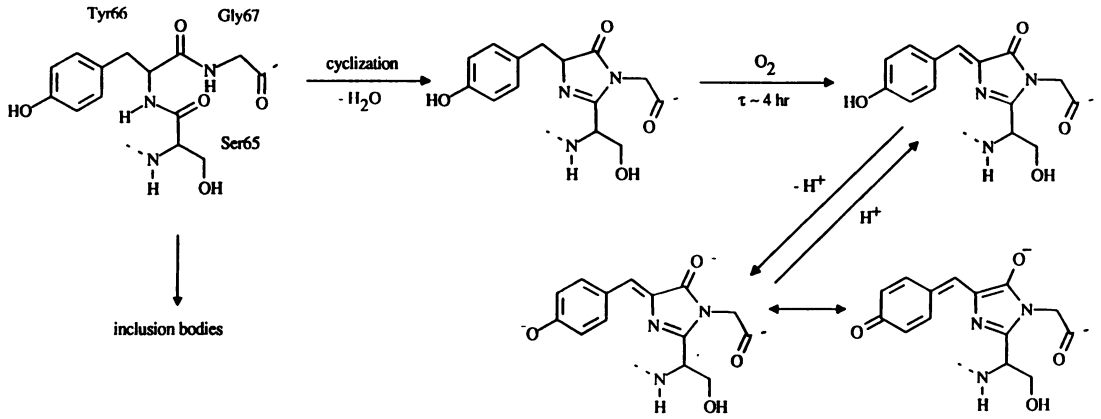

FIG. 2. Proposed biosynthetic scheme for the chromophore of GFP. The freshly translated protein (upper left) could be trapped by inclusion bodies or remain soluble and nonfluorescent (upper center) until oxidation by O_2, which would dehydrogenate Tyr-66 to form the fluorophore (upper right). The protonated and deprotonated species (upper and lower right) may be responsible for the 395- and 470- to 475-nm excitation peaks, respectively. The excited states of phenols are much more acidic than their ground states, so that emission would come only from the deprotonated species.

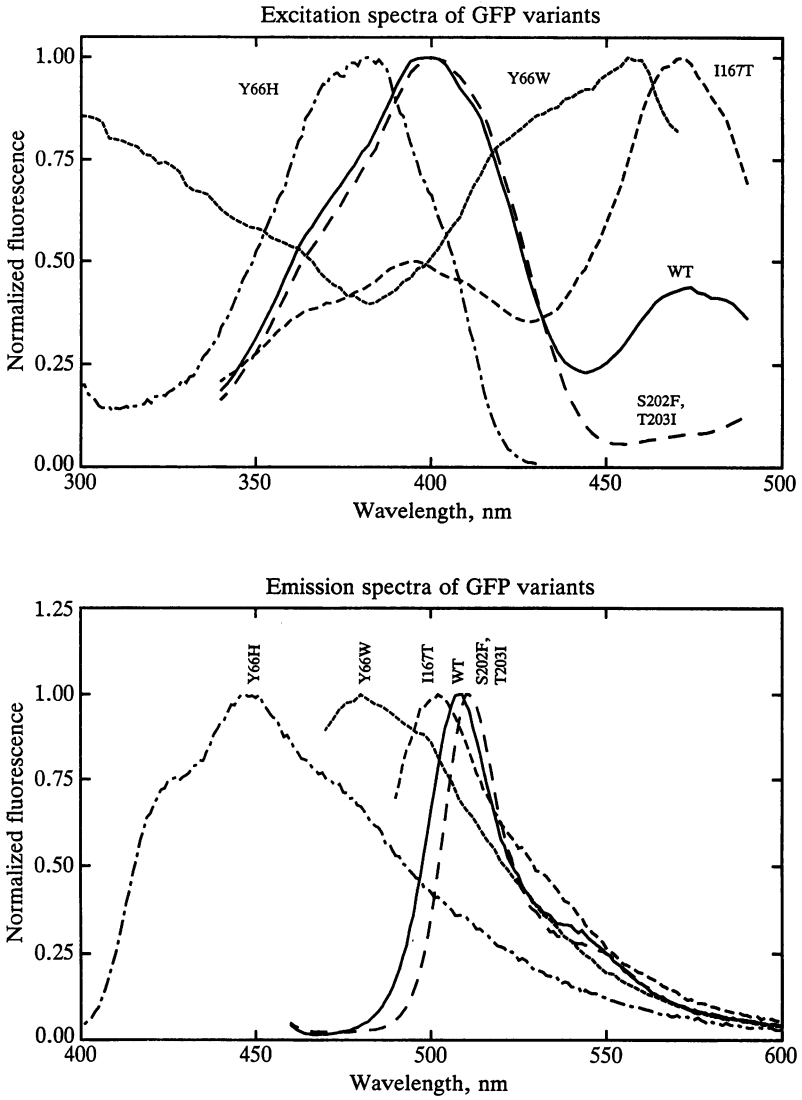

FIG. 3. Excitation and emission spectra of Wild-type and mutant GFPs. ——, Wild-type; — —, mutations S202F, T203I; – – –, mutation I167T; -------, mutation Y66W; —·—, mutation Y66H. Samples were soluble fractions from *E. coli* expressing the proteins at high level, except for mutant Y66W, which was obtained in very low yield and measured on intact cells. Autofluorescence was negligible for all spectra except those of mutant Y66W, whose excitation spectrum below 380 nm may be contaminated by autofluorescence. All amplitudes have been arbitrarily normalized to a maximum value of 1.0. To compare brightnesses at equal protein concentrations, see Table 1.

has two absorbance peaks at 395 and 475 nm, whereas *Renilla* GFP has only a single absorbance peak at 498 nm, with ≈5.5-fold greater monomer extinction coefficient than the major 395-nm peak of the *Aequorea* protein (17). The isolated chromophore and denatured protein at neutral pH do not match either native protein (5). For many practical applications, the spectrum of *Renilla* GFP would be preferable to that of *Aequorea* because wavelength discrimination between different fluorophores and detection of resonance energy transfer are easier when the component spectra are tall and narrow rather than low and broad. Furthermore, the longer wavelength excitation peak (475 nm) of *Aequorea* GFP is almost ideal for fluorescein filter sets and is resistant to photobleaching but has lower amplitude than the shorter wavelength peak at 395 nm, which is more susceptible to photobleaching (7). For all these reasons, it would be interesting to convert the *Aequorea* GFP excitation spectrum to a single peak, preferably at longer wavelengths. The cDNA was therefore subjected to random mutagenesis by hydroxylamine treatment or PCR. Approximately six thousand bacterial colonies on agar plates were illuminated with alternating 395- and 475-nm excitation and visually screened for altered excitation properties or emission colors. Although this number of colonies falls far short of saturating the possible mutations of a protein of 238 residues, interesting variants have already appeared. Three mutants were found with significant alterations in the ratio of the two main excitation peaks (Fig. 3 and Table 1). Compared with wild-type GFP, mutant H9 had increased fluorescence at 395-nm excitation, whereas mutants P9 and P11 were more fluorescent at 475-nm excitation, mutant P11 being the best species for fluorescein filters. The mutations were sequenced and recombined with the wild-type gene in different ways to eliminate neutral mutations and assign the fluorescence effects to single-amino acid substitutions (except for mutant H9, where two neighboring mutations have not yet been separated). The mutations all lay in the C-terminal part of the protein (Table 1), remote in primary sequence from the chromophore formed from residues 65–67. Determination of the three-dimensional structure of GFP should enable detailed interpretation of these spectral perturbations. One possibility is that the mutations at Ile-167 shift a positive charge slightly closer to the phenolic group of the fluorophore; this shift should both increase the percentage of phenolic anion, which is probably the species responsible for the 470- to 475-nm

Table 1. Characteristics of mutated vs. wild-type GFP

GFP	Mutation	Excitation maxima,* nm	Emission maxima,† nm	Relative fluorescence,‡ %
Wild type	None	396 (476)	508 (503)	(=100)
Mutant H9	Ser-202 → Phe, Thr-203 → Ile	398	511	117§
Mutant P9	Ile-167 → Val	471 (396)	502 (507)	166¶
Mutant P11	Ile-167 → Thr	471 (396)	502 (507)	188¶
Mutant P4	Tyr-66 → His	382	448	57‖
Mutant W	Tyr-66 → Trp	458	480	ND

*Values in parentheses are lower-amplitude peaks.
†Primary values were seen when exciting at the main excitation peak; values in parentheses were seen when illuminating at the lower-amplitude excitation peak.
‡Equal amounts of protein were used based on densitometry of gels stained with Coomassie blue (Fig. 1). ND, not done.
§Emission maxima of spectra recorded at excitation 395 nm were compared.
¶Emission maxima of spectra recorded at excitation 475 nm were compared.
‖Emission spectrum of mutant P4 recorded at 378-nm excitation was integrated and compared with the integrated emission spectrum of wild type recorded at 475-nm excitation; both excitation and emission characteristics were corrected.

excitation peak, and shift the emission peak to shorter wavelengths. However, the hypothesized ionizable phenolic group would have to be buried inside the protein at normal pH because the ratio of 471- to 396-nm peaks in the mutants could not be further affected by external pH until it was raised to 10, just below the threshold for denaturation. The pH-sensitivity of wild-type GFP is similar (18).

A fourth mutant, P4, was excitable by UV light and fluoresced bright blue in contrast to the green of wild-type protein (Fig. 4). The excitation and emission maxima were hypsochromically shifted by 14 and 60 nm, respectively, from those of wild-type GFP. The mutated DNA was sequenced and found to contain five amino acid substitutions, only one of which proved to be critical, Tyr-66 → His in the center of the chromophore. The surprising tolerance for substitution at this key residue prompted further site-directed mutagenesis to tryptophan and phenylalanine at this position. Tryptophan gave excitation and emission wavelengths intermediate between tyrosine and histidine (Fig. 3 and Table 1) but was only weakly fluorescent, perhaps due to inefficiency of folding or chromophore formation, whereas phenylalanine gave no detectable fluorescence.

The availability of several forms of GFP with such different excitation and emission maxima [the most distinguishable pair being mutant P4 (Y66H) vs. mutant P11 (I167T)] should facilitate two-color assessment of differential gene expression, developmental fate, or protein trafficking. It may also be possible to use these GFP variants analogously to fluorescein and rhodamine to tag interacting proteins or subunits whose association could then be monitored dynamically in intact cells by fluorescence resonance energy transfer (19, 20). Such fluorescence labeling via gene fusion would be site-specific and would eliminate the present need to purify and label proteins *in vitro* and microinject them into cells.

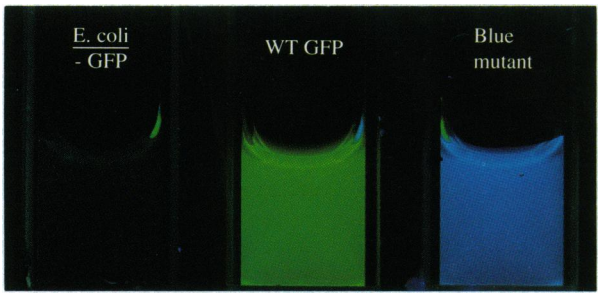

FIG. 4. *E. coli* producing either no GFP (*Left*), wild-type GFP (*Center*), or the P4 mutant (*Right*) in which Tyr-66 is replaced by histidine. Approximately equal quantities of bacteria were suspended in 1-cm cuvettes, excited by a 365-nm transilluminator, and photographed through a low-fluorescence filter passing wavelengths >450 nm. The bacteria producing no GFP were transformed with vector alone. Flecks of green in the two flanking cuvettes are stray reflections from the center cuvette.

1. Wang, Y. L. & Taylor, D. L., eds. (1989) *Methods Cell Biol.* **29.**
2. Peck, K., Stryer, L., Glazer, A. N. & Mathies, R. A. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 4087–4091.
3. Fairchild, C. D. & Glazer, A. N. (1994) *J. Biol. Chem.* **269,** 8686–8694.
4. Ward, W. W. (1979) in *Photochemical and Photobiological Reviews*, ed. Smith, K. C. (Plenum, New York), Vol. 4, pp. 1–57.
5. Cody, C. W., Prasher, D. C., Westler, W. M., Prendergast, F. G. & Ward, W. W. (1993) *Biochemistry* **32,** 1212–1218.
6. Prasher, D. C., Eckenrode, V. K., Ward, W. W., Prendergast, F. G. & Cormier, M. J. (1992) *Gene* **111,** 229–233.
7. Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W. & Prasher, D. C. (1994) *Science* **263,** 802–805.
8. Inouye, S. & Tsuji, F. I. (1994) *FEBS Lett.* **341,** 277–280.
9. Wang, S. & Hazelrigg, T. (1994) *Nature (London)* **369,** 400–403.
10. Sikorski, R. S. & Boeke, J. D. (1991) *Methods Enzymol.* **194,** 302–318.
11. Muhlrad, D., Hunter, R. & Parker, R. (1992) *Yeast* **8,** 79–82.
12. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY), 2nd Ed., pp. 17.37–17.41.
13. Ward, W. W. & Bokman, S. H. (1982) *Biochemistry* **21,** 4535–4540.
14. Kidwai, A. R. & Devasia, G. M. (1962) *J. Org. Chem.* **27,** 4527–4531.
15. Kjaer, A. (1953) *Acta Chem. Scand.* **7,** 1030–1035.
16. Merbs, S. L. & Nathans, J. (1992) *Science* **258,** 464–466.
17. Ward, W. W. (1981) in *Bioluminescence and Chemiluminescence*, eds. DeLuca, M. A. & McElroy, W. D. (Academic, New York), pp. 235–242.
18. Ward, W. W., Prentice, H. J., Roth, A. F., Cody, C. W. & Reeves, S. C. (1982) *Photochem. Photobiol.* **35,** 803–808.
19. Adams, S. R., Harootunian, A. T., Buechler, Y. J., Taylor, S. S. & Tsien, R. Y. (1991) *Nature (London)* **349,** 694–697.
20. Adams, S. R., Bacskai, B. J., Taylor, S. S. & Tsien, R. Y. (1993) in *Fluorescent Probes for Biological Activity of Living Cells: A Practical Guide*, ed. Mason, W. T. (Academic, New York), pp. 133–149.