JOURNAL OF VIROLOGY, July 1996, p. 4646–4654 Vol. 70, No. 7
0022-538X/96/$04.00+0

A "Humanized" Green Fluorescent Protein cDNA Adapted for High-Level Expression in Mammalian Cells

SERGEI ZOLOTUKHIN,[1,2] MARK POTTER,[1,2] WILLIAM W. HAUSWIRTH,[1,2,3] JOHN GUY,[3,4] AND NICHOLAS MUZYCZKA[1,2]*

Department of Molecular Genetics and Microbiology,[1] Gene Therapy Center,[2] Department of Ophthalmology,[3] and Department of Neurology,[4] University of Florida, Gainesville, Florida 32610

Received 11 January 1996/Accepted 26 March 1996

We constructed gfp_h, a synthetic version of the jellyfish *Aequorea victoria* green fluorescent protein (*gfp*) cDNA that is adapted for high-level expression in mammalian cells, especially those of human origin. A total of 92 base substitutions were made in 88 codons in order to change the codon usage within the *gfp10* coding sequence so that it was more appropriate for expression in mammalian cells. We also describe a series of versatile recombinant adeno-associated virus and adenovirus vectors for delivery and expression of genes into mammalian cells and, using these vectors, demonstrate the efficient transduction and expression of the gfp_h gene in the human cell line 293 and also in vivo, within neurosensory cells of guinea pig eye. Cells infected with recombinant adeno-associated virus–GFP_H can be readily sorted by fluorescence-activated cell sorting, suggesting that the newly designed gfp_h gene could be widely used as a reporter in many gene delivery technologies, including human gene therapy.

Adeno-associated virus (AAV) is gaining acceptance as a vector to deliver genes into a wide variety of cell types (17, 28). There are many advantages of using AAV, including the apparent absence of pathogenicity, high viability of the virion, the potential for site-specific integration, long-term expression of the delivered gene, and relative independence of infectivity from host chromosome replication and cell cycling. One disadvantage of AAV is the limited packaging size of the virus, which cannot exceed 5,000 nucleotides. Most AAV vectors currently available for preclinical studies carry a reporter gene, usually the *Escherichia coli* β-galactosidase or neomycin phosphotransferase gene. Both of these reporter genes are quite bulky and occupy much of the limited space of the AAV genome. Additionally, detection protocols for these gene products are cumbersome.

The usefulness of the jellyfish *gfp10* gene as a reporter in prokaryotes and animals (4) opened new perspectives in gene delivery technologies. Green fluorescent protein (GFP) from the jellyfish *Aequorea victoria* is a protein of 238 amino acids which absorbs blue light (major peak at 395 nm) and emits green light (major peak at 509 nm) (27, 29, 36). The GFP hexapeptide chromophore starts at amino acid 64 and is derived from the primary amino acid sequence through the cyclization of serine-dehydrotyrosine-glycine within this hexapeptide (6, 32). The light-stimulated GFP fluorescence is species independent and does not require any cofactors, substrates, or additional gene products from *A. victoria* (4), thus allowing detection in living cells. The small size of *gfp10* and the easy real-time detection of the product make it an ideal reporter for most viral vectors, AAV included. However, our initial attempt to show the expression of the jellyfish GFP reporter gene delivered into a human cell by a recombinant AAV (rAAV) was unsuccessful. We hypothesized that one of the reasons for the low expression of GFP was the poor translation efficiency of the mRNA in the human cell environment, which is characterized by a set of isoacceptor tRNAs that are different than those used in the jellyfish. We therefore synthesized a synthetic version of the jellyfish GFP, substituting preferentially used codons in the human genome for the rarely used codons in the original *gfp10* cDNA. Using the newly designed "humanized" *gfp* (gfp_h) gene, we constructed a series of recombinant AAV and adenovirus (Ad) vectors that are capable of efficiently expressing the reporter gene in cells of mammalian origin.

* Corresponding author. Phone: (352) 392-8541. Fax: (352) 392-3133. Electronic mail address: muzyczka@medmicro.med.ufl.edu.

MATERIALS AND METHODS

Synthesis of the gfp_h cDNA. The gfp_h cDNA was synthesized by assembling mutually priming synthetic oligonucleotides (Fig. 1). The gene was divided into eight segments of approximately equal length, and four pairs of oligonucleotides were synthesized, each pair consisting of two overlapping oligonucleotides with a short stretch of overlap (underlined in Fig. 1), one coding for the sense strand and the other coding for the antisense strand. After annealing and extension with Sequenase, pairs 1 and 2 were digested with *Eae*I, whereas pairs 3 and 4 were digested with *Bam*HI. The digested products were then ligated in two separate reactions: pairs 1 and 2 and pairs 3 and 4. Ligation products of the desired length were purified on a 5% polyacrylamide gel under nondenaturing conditions. Both DNA fragments were then digested with *Eco*RII and ligated to each other. The final product was amplified by PCR, using a pair of oligonucleotides partially complementary to the gfp_h cDNA (boldface in the sequences presented below) and containing the restriction sites *Not*I, *Xba*I, and *Hin*dIII (underlined in the sequences presented below) for cloning. The sequence of the upstream primer, which included a Kozak consensus sequence (18), was 5′-TGCTCTAGAGCG GCCGCCGCCACC**ATGAGCAAGGGCGAGGAACTG**-3′; the downstream primer sequence was 5′-CGGAAGCTTGCGGCCGC**TCACTTGTACAGCTCGTCCAT**-3′. After digestion of the PCR product with *Xba*I and *Hin*dIII, the DNA fragment was cloned into pBS(+) (Stratagene) and sequenced. Several independent clones were isolated and sequenced. These clones had mutations in the coding sequence which presumably either occurred during PCR amplification or were present in the oligonucleotides. Portions of these clones were then spliced together to produce the final gfp_h gene that encoded a wild-type amino acid sequence. The resulting construct, designated pBS-GFP_H1, contained the coding sequence for wild-type GFP. Two mutants were constructed in the pBS-GFP_H background by site-directed PCR mutagenesis. One of these converted Ser-65 to Thr and was called pBS-GFP_H2; the other converted Tyr-66 to His and was called pBS-GFP_HB.

Construction of rAAV vector plasmids. Briefly, the *gfp10* sequence was subcloned into the *Not*I site of pCMVβ (Clontech) after digestion of the parent plasmid TU#65 (4) with *Age*I and *Eco*RI, filling in the ends with Klenow DNA polymerase, and adding *Not*I linkers. The resulting plasmid, designated pCMV green, was then used as a template to amplify in a PCR the transcription cassette containing the cytomegalovirus (CMV) promoter, the simian virus 40 (SV40) intron, the *gfp10* cDNA, and the SV40 polyadenylation signal. The upstream PCR primer complementary to the CMV promoter also included an overhang that contained the *Bgl*II, *Eco*RI, and *Kpn*I sites. The downstream PCR primer,

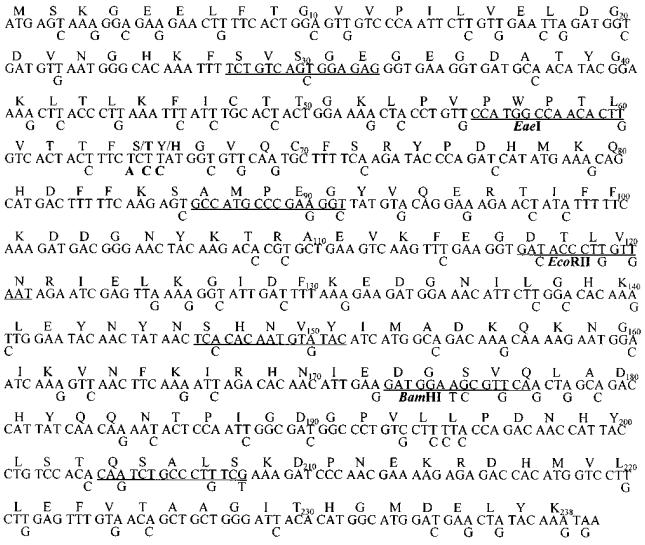


FIG. 1. Nucleotide sequence of the *gfp10* cDNA and the deduced amino acid sequence. Above each codon is the single-letter designation for the amino acid. The mutations introduced in the gfp_h sequence are shown below the substituted nucleotide of *gfp10*. The horizontal lines underline overlap regions of mutually priming oligonucleotides used to synthesize the gfp_h cDNA. The sites of the restriction enzymes used to assemble extended pairs of oligonucleotides are shown in boldface. The codons mutated to produce the Ser-65→Thr mutation, which produces higher fluorescence yield, and the Tyr-66→His mutation, which produces blue fluorescence, are shown in boldface.

complementary to the polyadenylation signal, included a *Sal*I site overhang. The polyadenylation signal of the bovine growth hormone gene was amplified in another PCR using plasmid pRc/CMV (Invitrogen) as the template. The upstream primer in this reaction contained a *Sal*I site overhang, and the downstream primer contained a *Bgl*II site. After purification of the PCR products on a 1% agarose gel, the respective fragments were digested with *Sal*I and ligated to each other via the exposed *Sal*I ends. The ligation product was gel purified and digested with *Bgl*II. The 160-bp *Bgl*II-*Pst*I fragment, containing the AAV terminal repeat, was isolated by gel purification from plasmid $pTR_{BR}(+)$ (30). This fragment had been subcloned into $pTR_{BR}(+)$ from the previously described plasmid *dl*3-94 (26). It was then ligated to both ends of the *Bgl*II-digested cassette, containing the CMV promoter, SV40 intron, *gfp10* cDNA, SV40 poly (A) site, and bovine growth hormone poly(A) site. The ligation product was then cut with *Pst*I and subcloned into plasmid pBS(+) (Stratagene), which had been modified by converting the *Pvu*II sites at positions 766 and 1148 into *Pst*I sites by adding *Pst*I linkers and deleting the internal 382-bp fragment, containing the polylinker region. The resulting plasmid was designated pTRgreen. The neomycin resistance gene (*neo*) cassette, driven by the herpes simplex virus thymidine kinase gene promoter and the enhancer from polyomavirus, was obtained from plasmid pMClneo (Stratagene) by cutting the plasmid with *Xho*I, filling in the ends with Klenow DNA polymerase, adding *Sal*I linkers, and digesting with *Sal*I. The DNA fragment containing the *neo* cassette was gel purified and subcloned into the *Sal*I site of pTRgreen. The resulting construct, pTR_{BS}-UF (UF for user friendly), is depicted in Fig. 2. To construct pTR_{BS}-UF1, pTR_{BS}-UF2, or pTR_{BS}-UFB, we substituted the *Not*I fragment of pBS-GFP_H1 (wild type), pBS-GFP_H2 (Thr-65), or pBS-GFP_HB (His-66), respectively, for the *Not*I fragment of pTR_{BS}-UF (Fig. 2). Any DNA fragment that had undergone PCR amplification was sequenced to confirm the identity of the original sequence.

To construct pTR_{BS}-UF3, the *Eco*RI site of plasmid pSBC-1 (7) was converted to an *Not*I site after digestion with *Eco*RI, filling in of the 5′ overhang with Klenow polymerase, and ligation of *Not*I linkers. The 680-bp *Not*I fragment, consisting of a polylinker and the internal ribosome entry site (IRES) element of poliovirus type 1, was then subcloned into one of the *Not*I sites of pTR_{BS}-UF2 (Fig. 2).

Construction of a recombinant Ad shuttle plasmid. To construct the Ad shuttle vector pΔE1GFP, the parent plasmid pTR-UF3 was partially digested with *Sal*I and then digested to completion with *Bgl*II. The transcription cassette consisting of the CMV promoter, intron, IRES element, GFP_H cDNA, and poly(A) site was isolated from an agarose gel. This fragment was subcloned into pΔE1sp1A (3) which had been digested with *Bam*HI and *Sal*I.

Production of rAAV. To generate rAAV, 293 cells were transfected with pTR_{BS}-UF, pTR_{BS}-UF1, pTR_{BS}-UF2, or pTR_{BS}-UF3 and cotransfected with the helper plasmid pIM45, carrying the wild-type AAV genome without the terminal repeats (24). The same cells were also infected with Ad5 at a multiplicity of infection (MOI) of 10. rAAV was harvested after 60 h by freezing and thawing the cells three times, heat inactivating Ad for 30 min at 56°C, spinning down the cell debris, and purifying the virus through a CsCl gradient (1.40 g/ml) formed in an SW41 rotor for 48 h at 200,000 × *g*. The gradient was fractionated, and the density was determined by refractometry. Fractions with densities of between 1.38 and 1.4 g/cm^3 were pooled and dialyzed against Dulbecco modified Eagle medium for 4 h. The AAV titer was determined by the infectious center assay (26).

Construction of recombinant Ad. To generate recombinant Ad, the shuttle vector pΔE1GFP (3) and the Ad vector pJM17 (25) were cotransfected into 293 cells, using the procedure recommended by the supplier (Microbix Biosystems Inc.). Plaques containing recombinant Ad were screened by visual selection under epifluorescence for a group of bright green cells displaying typical cytopathic effect. The recombinant Ad was designated AdΔE1GFP and propagated by using standard techniques.

Monitoring of fluorescence. 293 cells were infected with CsCl-purified rAAV-GFP_H2 at an MOI of 10. At 36 h postinfection, the cells were photographed in a fluorescence microscope, using a CHROMA no. 41014 GFP-HQ filter cube (excitation at 450 ± 25 nm). Alternatively, after infection at an MOI of 1 and selection with G418 for 2 weeks, three independent observers scored the number of green fluorescent cells within the G418 colonies by fluorescence microscopy. The mean of the frequencies obtained by the three observers was calculated. At least 11,500 cells in 150 separate colonies were scored for each viral preparation, rAAV-GFP_J, rAAV-GFP_H1, and rAAV-GFP_H2.

To monitor the fluorescence of the blue mutant, 293 cells on a 6-cm-diameter plate were cotransfected with plasmids pTR_{BS}-UF2 and pTR_{BS}-UFB, using Lipofectamine (GIBCO, Life Technologies). The DNA-liposome complex was formed separately for each plasmid and added to the same plate of cells. After 4 days, cells were photographed in a fluorescence microscope using a Nikon V-2B filter cube.

FACS analysis of cells expressing GFP. 293 cells, transfected with the pTR_{BS}-UF series plasmids or infected with rAAV-GFP, were harvested and analyzed on a flow cytometer (Becton Dickinson) equipped for fluorescein isothiocyanate detection at an excitation wavelength of 488 nm. Alternatively, 293 cells were stably transduced with rAAV-GFP and selected for G418 resistance (200 μg/ml) for 2 weeks. Resistant colonies were trypsinized, pooled (at least 1,000 colonies each for rAAV-GFP_J, rAAV-GFP_H1, and rAAV-GFP_H2), resuspended in OPTI-MEM, and analyzed by fluorescence-activated cell sorting (FACS) as described above.

Infection of the guinea pig retina. Guinea pigs were anesthetized by intramuscular injection of a 1:1 mixture of ketamine (35 mg/kg of body weight) and xylazine (8 mg/kg). Each eye was dilated with 2.5% phenylephrine (Neo-Synephrine) and 0.5% tropicamide, and a topical anesthetic (proparacaine HCl) was administered to the cornea. The eye pressure was lowered by paracentesis of the anterior chamber. Then a 30-gauge needle was inserted at the pars plana into the vitreous under visual guidance of an indirect ophthaloscope, and 25 μl of rAAV-GFP_H2 (2.5×10^7 infectious particles) was delivered. The eyes were examined by ophthalmoscopy for fluorescence and sites of inflammation; none were seen. At 28 days postinjection, animals were anesthetized and euthanized with an intramuscular injection of ketamine HCl followed by an intraperitoneal pentobarbital sodium overdose. The animals were then perfused by using 4% paraformaldehyde in 0.1 M phosphate-buffered saline. The eyes were dissected out, and the lens and cornea were removed. The retina and eye cup were additionally fixed overnight at 4°C. The retina was then infiltrated with 7.5, 15, and 30% sucrose and cryosectioned at a thickness of 20 to 25 μm. Tissue specimens were visualized by using fluorescein excitation/emission filters on a Bio-Rad confocal microscope.

Nucleotide sequence accession number. The gfp_h cDNA sequence has been submitted to GenBank; its accession number is U50963.

RESULTS

Construction of rAAV-GFP. To test the possibility of using the *gfp* gene as a reporter in the context of an AAV vector, we constructed the rAAV vector plasmid pTR_{BS}-UF (Fig. 2), which carried both the jellyfish *gfp10* gene and the *neo* gene (see Materials and Methods). Plasmid TU#65 (35) was used as the source of the *gfp10* coding sequence, and the gene was placed under the control of the immediate-early CMV promoter. When pTR_{BS}-UF plasmid DNA was transfected into 293 cells, the average number of cells expressing GFP was disappointingly low, usually less than 5% (Fig. 3). Furthermore, 293 cells infected with rAAV carrying the same GFP expression cassette were repeatedly scored as GFP negative (not shown). The only difference between these two experiments was apparently the number of GFP cDNAs delivered into each cell. During DNA transfection, hundreds or even thousands of plasmid copies were being delivered to the cell,

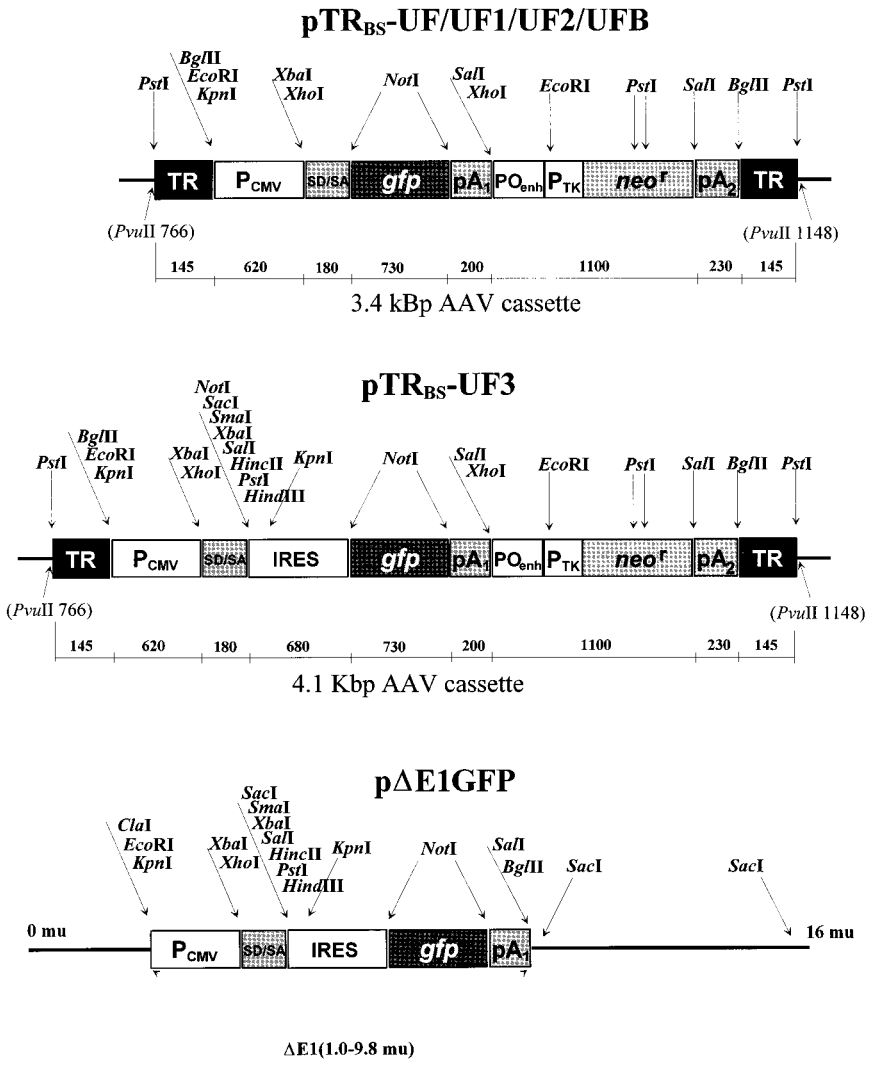


FIG. 2. Restriction maps of the AAV and Ad vector plasmids. Only those restriction sites relevant for the construction of the rAAV plasmids are shown. The sizes of removable elements and reporter gene cassettes are shown in base pairs. The origins of the genes and transcription elements are as follows: TR (terminal repeat) is the *Pst*I-*Bgl*II fragment [145 bp plus oligo(dG)-oligo(dC), 160 bp in total] from *dl*3-94 (26); P_{CMV} is the CMV immediate-early promoter/enhancer; SD/SA represents the SV40 late viral protein gene 16S/19S splice donor and acceptor signals; *gfp* is the *A. victoria gfp* cDNA in pTR$_{BS}$-UF, the chemically synthesized wild-type *gfp*$_h$ cDNA in pTR$_{BS}$-UF1, the Thr-65 *gfp*$_h$ in pTR$_{BS}$-UF2, or the His-66 *gfp*$_h$ in pTR$_{BS}$-UFB; pA_1 is the SV40 polyadenylation signal from the SV40 genome; PO_{enh} is a tandem repeat of the enhancer from the polyomavirus mutant PYF441; P_{TK} is the thymidine kinase promoter of herpes simplex virus; *neo*^{r} is the *neo* gene from Tn5; pA_2 is the bovine growth hormone polyadenylation signal from pRc/CMV (Invitrogen); and IRES is the IRES of poliovirus type 1 from pSBC-1 (7).

whereas viral infection under conditions of low MOI (less than 1) delivered only a single copy of a gene. Having found the *gfp10* cDNA, as originally described by Chalfie et al. (4), to be a poor reporter when expressed in primate and human cells, we attempted to enhance the expression of *gfp10*.

Design of a humanized GFP. There are several ways to boost the amount of desired gene product which is under control of a given promoter. One could increase the stability of the mRNA by introducing an intron sequence which directs the pre-mRNA into the processing/splicing pathway through protein-RNA interactions and transport. In this regard, our GFP expression cassette already contained the sequence of the SV40 late gene 16S/19S splice donor/splice acceptor signal (Fig. 2). Another way to increase the protein yield is to maximize the translation efficiency by introducing sequences that facilitate translation of eukaryotic mRNA. One such sequence, immediately preceding the AUG initiator codon (underlined), is the Kozak consensus sequence GCCGCC(A/G)CCATG (18). Additionally, an optimally positioned stem-loop hairpin structure, located about 14 nucleotides downstream of the AUG codon, can be used (19). However, other laboratories have placed a Kozak sequence upstream of *gfp10* and found that it did not significantly change the expression efficiency. We therefore hypothesized that one of the reasons for the low expression of GFP was the poor translation efficiency of the mRNA in the human cell environment, which is characterized by a set of isoacceptor tRNAs different from that in the jellyfish.

It is known that the choice of synonymous codons in both prokaryotic and eukaryotic genes is strongly biased. There exist clear differences in codon usage between taxonomically distant organisms, even among genes encoding cognate proteins (8, 11, 12). The differences in codon choice have been attributed to differences in the populations of isoacceptor tRNAs and to

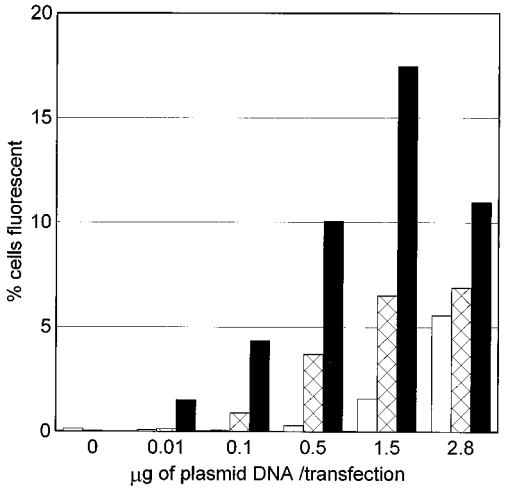


FIG. 3. FACS analysis of 293 cells transfected with the pTR_{BS}-UF series of recombinant plasmids. 293 cells (six-well dish) were transfected with a total of 2.8 μg of DNA, consisting of different ratios of *gfp*-containing plasmid and sonicated salmon sperm carrier DNA, using the conventional calcium phosphate transfection protocol. Cells were harvested 36 h posttransfection and analyzed on a flow cytometer as described in Materials and Methods. Cells scored as positive were plotted on the graph as a function of the amount of *gfp*-carrying plasmid transfected. Open bars, pTR_{BS}-UF; crosshatched bars, pTR_{BS}-UF1; black bars, pTR_{BS}-UF2.

differences in modified nucleotides at the anticodon wobble position (13, 14). Synonymous codon choices do not affect the nature of the protein synthesized but may affect the level of gene expression (2, 12–14). An analysis of the sequence of *gfp10* cDNA showed that the codon usage frequencies of this jellyfish gene are quite different from those prevalent in the human genome. For example, Leu amino acid residues at positions 18, 53, 125, 178, 195, and 236, Ser at 208, and Val at 93, 150, and 224 (Fig. 1) of the jellyfish GFP are encoded by triplets which are almost never used in human genes (34). The remaining amino acids also display a codon bias different from that in human genes, though not as dramatic.

To test the possibility that inappropriate codon usage was responsible for the poor expression of the *gfp* gene, we designed a synthetic version of the *gfp10* gene in which we substituted codons preferentially used in the human genome for rare or less frequently used codons in the original jellyfish gene. A total of 92 base substitutions were made in 88 codons without changing the amino acid sequence (Fig. 1). This construct was called pTR_{BS}-UF1 (see Materials and Methods). In addition, the sequence immediately preceding the start codon for the GFP protein in pTR_{BS}-UF1 was modified to produce a Kozak consensus sequence. We also took note of the recent report by Heim et al. (9), who described a Ser-65-to-Thr substitution which increased the fluorescence yield in the context of the original jellyfish codon sequence. Reasoning that this mutation might be even more effective in the context of the humanized pTR_{BS}-UF1 sequence, we reproduced this point mutation in the pTR_{BS}-UF1 background to produce plasmid pTR_{BS}-UF2. Another point mutation, Tyr-66 to His, which resulted in blue fluorescence (10), also was built into the humanized background of pTR_{BS}-UF1 to produce the vector pTR_{BS}-UFB. Finally, although it is not believed to affect expression, in all three versions of GFP_H we reverted codon 80 to a wild-type glutamine residue (29) from the arginine missense codon described by Chalfie et al. (4).

Comparison of jellyfish and humanized *gfp* genes. To compare the expression efficiency of the gfp_h constructs with the original jellyfish sequence, we transfected 293 cells with pTR_{BS}-UF, pTR_{BS}-UF1, or pTR_{BS}-UF2 plasmid DNA at various DNA concentrations. The transfected cells were then analyzed by FACS 36 h after transfection (Fig. 3). pTR_{BS}-UF1 carrying the gfp_h sequence consistently produced 5 to 10 times more cells scored as positive for green fluorescence than the jellyfish sequence. The point mutation in pTR_{BS}-UF2 increased the number of fluorescent cells an additional 5- to 10-fold over pTR_{BS}-UF1. At relatively low plasmid DNA

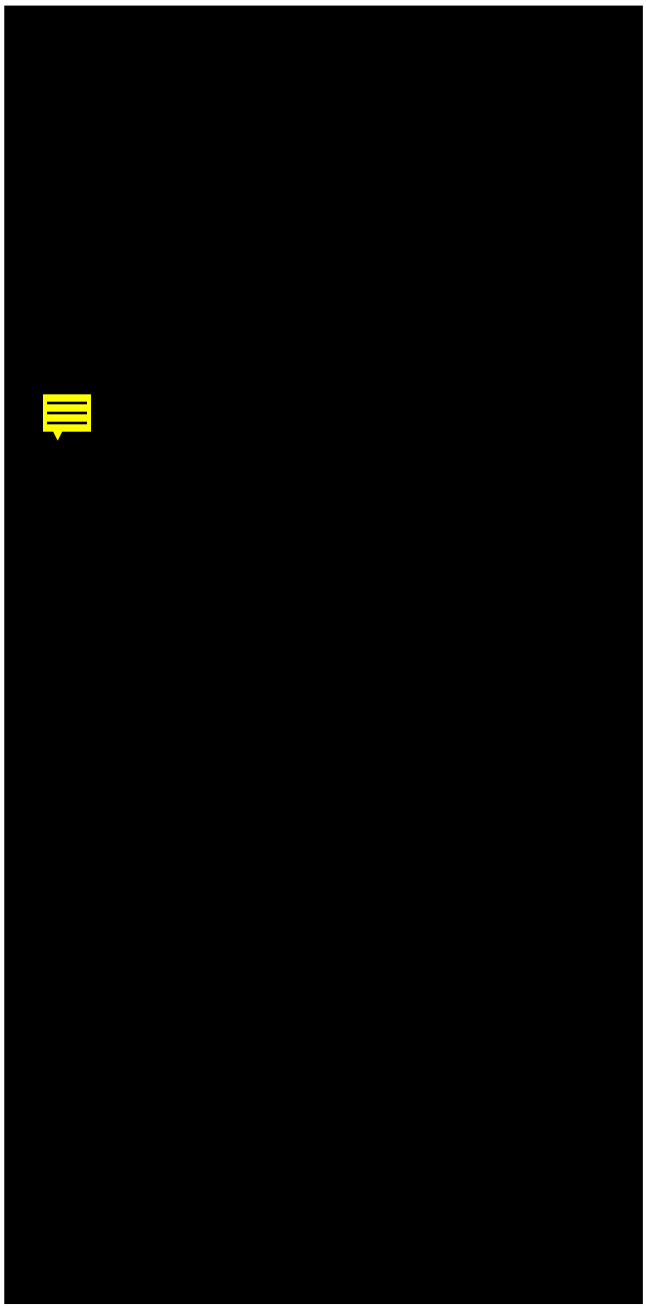


FIG. 4. Expression of rAAV-GFP_H2 in 293 cells. 293 cells were infected with CsCl-purified rAAV-GFP_H2 at an MOI of 10; 36 h postinfection, cells were photographed in a fluorescence microscope, using a CHROMA no. 41014 GFP-HQ filter cube (excitation at 450 ± 25 nm). (A) Cells under phase contrast, light field; (B) same field, epifluorescence.

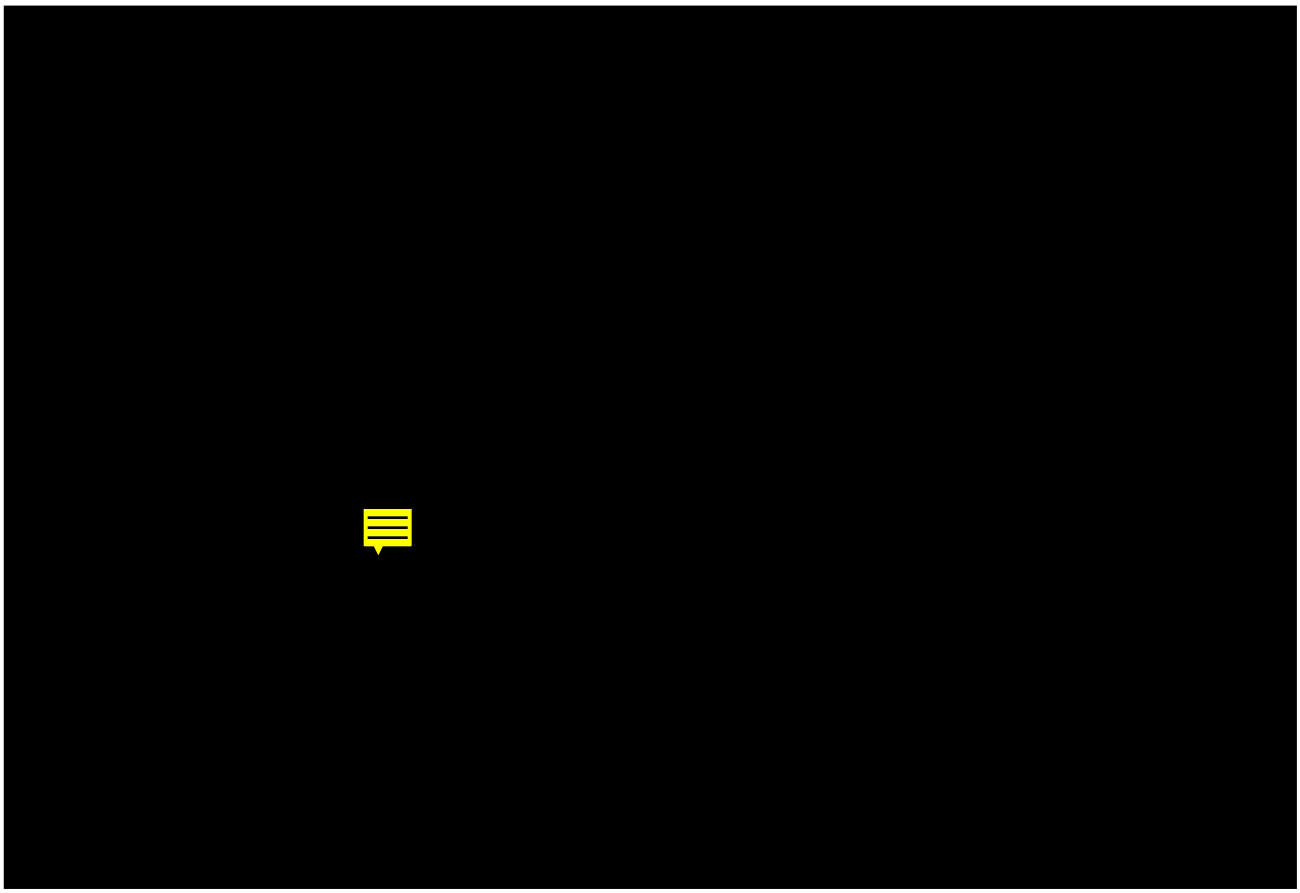


FIG. 5. Fluorescence of G418-resistant clones containing rAAV-GFP_H2 provirus. 293 cells were infected with CsCl-purified rAAV-GFP_H2 at an MOI of 1; 48 h postinfection, cells were split and plated at a low (less than 10%) confluency; 18 h later, G418 was added at a final concentration of 200 μg/ml. The medium was changed every 4 days, and G418-resistant colonies were photographed after 14 days of selection. (A and C) G418-resistant colonies under phase contrast, light field; (B and D) same fields as in panels A and C, epifluorescence.

concentrations, the difference between pTR_{BS}-UF2 and pTR_{BS}-UF was greater than 70-fold. At higher concentrations of transfected plasmid DNA, the difference in the number of cells expressing GFP was reduced. This result was consistent with the idea that the inability to translate the jellyfish *gfp* sequence could be overcome in part by increasing the gene copy number.

To determine whether the modified *gfp* cDNA was sufficient now to detect the marker gene at a low gene copy number, we isolated rAAV by packaging the three *gfp* constructs (UF, UF1, and UF2) and used them to transduce the *gfp* marker into 293 cells by virus infection. While there was almost no detectable GFP expression from a virus carrying the *gfp10* cDNA (rAAV-GFP_J), cells infected with a virus carrying the gfp_h gene (rAAV-GFP_H1 or rAAV-GFP_H2) were readily detected either visually (Fig. 4) or by FACS analysis. At a high MOI (approximately 20), the ratio of infected cells, scored by FACS as fluorescence positive, reached 70% for rAAV-GFP_H2 (not shown).

To determine more accurately the relative efficiency of the different *gfp* constructs, 293 cells infected at a low multiplicity (MOI of 1) with rAAV-GFP_J, rAAV-GFP_H1, or rAAV-GFP_H2 were first selected for the expression of the second reporter gene, *neo*. G418-resistant colonies that have been transduced by an AAV-Neo^r recombinant virus have been shown by us and others (5, 20, 26, 31) to contain an average of two to three copies of the recombinant viral genome integrated into host DNA. After 2 weeks of selection, approximately 11% of the UF1-transduced cells and 23% of the UF2-transduced cells that were G418 resistant were also found to express GFP, as judged by fluorescence microscopy. The visual pattern of GFP expression was patchy, with the number of green cells per colony ranging from 1% to about 100% (Fig. 5). In contrast, only 0.5% of the G418-resistant cells containing the jellyfish GFP-coding rAAV-GFP_J provirus were fluorescent. Thus, optimization of the codon usage within the *gfp* gene increased the level of detection at a low copy number approximately 22-fold, and the Ser-65→Thr mutation increased the level of detection an additional 2-fold, for a total of 45-fold. Analysis of the G418-resistant cells by FACS, which may be inherently more sensitive to the level of expression, revealed similar differences in the level of detection between GFP_J, GFP_H1, and GFP_H2 (Fig. 6). No difference in the number of fluorescent cells was detected between GFP_J and uninfected parental 293 cells. In contrast, approximately 1.6% of the GFP_H1 and 10% of the GFP_H2 cells were scored positive for green fluorescence. Since no positive cells were detected with GFP_J, it was difficult to judge accurately the difference in the frequency of detection between GFP_J and GFP_H2. However, the frequency of detecting a green fluorescent cell in the GFP_H2 population was at least 190-fold higher than the background frequency found for

FIG. 6. FACS analysis of 293 cells stably transduced with rAAV-GFP_J, rAAV-GFP_H1, or rAAV-GFP_H2 and selected for 2 weeks with G418. The 293 panel shows the FACS histogram plot of the parental 293 cell line, the UF panel shows 293 cells transduced with rAAV-GFP_J, the UF1 panel shows rAAV-GFP_H1, and the UF2 panel shows rAAV-GFP_H2. In each case, 20,000 cells were sorted. The uncorrected frequencies of cells scored positive for each cell population were as follows: uninfected 293 cells, 0.05%; GFP_J, 0.05%; GFP_H1, 1.67%; and GFP_H2, 9.6%.

GFP_J and uninfected parental 293 cells (see the legend to Fig. 6).

As expected, the blue GFP mutant, pTR_{BS}-UFB, when reproduced in a humanized background, induced 293 cells to fluoresce in a true blue color. However, the intensity of the fluorescence was considerably less than that induced by GFP_H2. For example, Fig. 7 shows 293 cells cotransfected with pTR_{BS}-UF2 and pTR_{BS}-UFB and viewed under conditions favoring the blue fluorescence. We also noticed a rather fast (10 to 15 s) bleaching of the blue fluorescence, when observing without a neutral density filter, which rarely happened with GFP_H2.

Construction of the IRES-GFP cassette AAV vector. Expression of the transduced gene of interest is often hard to monitor for various technical reasons. On these occasions, the monitoring of a marker gene delivered by the same vector is of a little help, since it is usually transcribed from a separate promoter. However, coordinate expression of both the reporter gene and the gene under study can be achieved if these genes are placed within one dicistronic transcription unit. The cap-independent translational initiation of the second cistron in this array is mediated by an untranslated RNA sequence which functions as an IRES (15, 16, 23).

To incorporate this feature into our AAV vectors, we constructed plasmid pTR_{BS}-UF3, in which translation of GFP is controlled by a poliovirus type 1 IRES element (7). A restriction site polylinker sequence also was inserted upstream of the IRES element to facilitate the insertion of the gene of interest, and the dicistronic mRNA was under the control of the CMV promoter.

The level of IRES-driven GFP expression with the pTR_{BS}-UF3 vector, as judged by the fluorescence intensity, was lower

FIG. 7. Fluorescence of the blue His-66 mutant of GFP_H. 293 cells were cotransfected with pTR_{BS}-UF2 and pTR_{BS}-UFB as described in Materials and Methods and photographed 4 days posttransfection in a fluorescence microscope, using a Nikon V-2B filter cube.

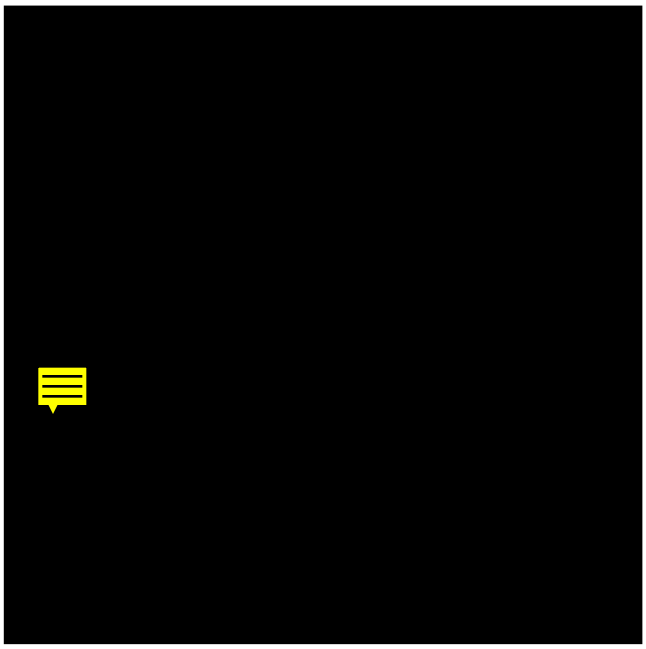


FIG. 8. A single plaque of recombinant AdΔE1GFP as seen in a fluorescence microscope. The plaque was photographed at 40 h postinfection.

than that seen with pTR_{BS}-UF2, the parental plasmid, and was comparable to that of the pTR_{BS}-UF1 vector (not shown). However, when another open reading frame (human B-chain insulin cDNA) was inserted upstream from the IRES element, the expression of GFP increased and was indistinguishable from that of the parent vector, pTR_{BS}-UF2 (not shown).

Construction of recombinant Ad-GFP. To test the utility of the gfp_h gene as a reporter in a different vector system, we constructed an Ad shuttle plasmid carrying the dicistronic expression cassette from pTR_{BS}-UF3. The GFP cassette was inserted into the Ad vector plasmid pΔE1sp1A (3), which contains the left end of Ad5, to produce plasmid pΔE1GFP (Fig. 2). When pΔE1GFP was recombined in vivo with a plasmid containing the remainder of the Ad genome, pJM17 (33), a recombinant Ad which carried and expressed GFP was produced (Fig. 8). The GFP reporter gene allowed an easy selection of recombinant Ad plaques. As examined by fluorescence microscopy, a true recombinant plaque consisted of a compact group of bright green cells displaying typical Ad cytopathic effect, whereas a false recombinant plaque contained no green cells. The ratio of true to false plaques was about 1:2 when the combination of the pΔE1GFP shuttle plasmid and the pJM17 donor plasmid was used. Thus, use of the GFP selection significantly simplified the screening process.

In vivo transduction of cells in the neurosensory retina of the guinea pig eye with rAAV-GFP_H2. To test the utility of gfp_h cDNA as a reporter gene in an in vivo system, we injected rAAV-GFP_H2 into the right-eye vitreous body in two strain 13 guinea pigs. Tissue sections of the eye revealed weak GFP_H2 fluorescence predominantly in cells of the ganglion cell layer (the layer closest to the vitreous injection). In addition, a few horizontal cells exhibited GFP_H2 fluorescence. The greatest intensity of GFP_H2 was seen in cells of the retinal pigment epithelium (RPE) (Fig. 9). With rAAV-GFP_H2, every tissue section examined had RPE cells that fluoresced. This preference for CMV promoter-driven expression in RPE cells has been previously noted (1, 22). Examination of tissue specimens from the control left eyes revealed no cell-specific emission except for autofluorescence within pigment granules of the RPE. The fact that inoculation of AAV into the guinea pig vitreous cavity led to GFP expression in RPE cells demonstrated that AAV can traverse the neural retina, 100 to 200 μm thick. This property may be related to the small diameter of AAV particles.

DISCUSSION

In this study we designed, synthesized, and cloned a humanized mutant of *gfp10* cDNA, adapted for high-level expression in mammalian cells. For fluorescence microscopy, we increased the sensitivity of this reporter gene system approximately 22-fold for gfp_h1 and at least 45-fold for gfp_h2. For FACS analysis, the gfp_h2 gene was at least 190-fold more detectable than the original jellyfish gene. Because the amino acid sequence of the GFP produced from the humanized cDNA is virtually identical to that produced from the original jellyfish gene, it is unlikely that protein stability played a role in the improved expression of the humanized gene. However, it is possible that optimization of the codon usage of the *gfp* gene affected the stability of the *gfp* mRNA.

Our data allowed us to estimate the minimum number of gfp_h genes per cell necessary to detect the fluorescence. When gfp_h is stably integrated as part of the *gfp-neo* cassette of the rAAV provirus in G418-resistant cell lines, a considerable portion of the cells express a visually detectable GFP1 or GFP2. According to previously published data, rAAV integrates as a tandem repeat, with the number of genome copies per cell ranging from 1 to 10 (5, 20, 26, 31). Therefore, this range of 1 to 10 copies of gfp_h per cell under the control of a strong promoter can be detected. For the mutant Thr-65–GFP2 construct, this number could be as low as one. We noted earlier that the G418-resistant proviral clones displayed patchy green fluorescence. We believe that this was due to the incomplete penetrance of the *gfp* marker. However, we cannot rule out the possibility that the nonuniform fluorescence was due to selective loss of *gfp* expression.

We also constructed a blue GFP mutant version of the gene in a humanized background. This mutant induced true blue fluorescence, but the fluorescence intensity was too low for the mutant to be useful as a marker gene. Work is in progress to insert a nuclear localization signal within the blue mutant gene, which may localize the GFP within the much smaller space of the nucleus and enhance the fluorescence intensity.

In addition to the newly designed *gfp* gene, we have described a versatile rAAV vector. The design of the pTR_{BS}-UF series of vectors (Fig. 2) provides convenience and flexibility in the construction of rAAV vectors. To use the maximum cloning capacity of 5 kbp, the whole reporter gene cassette can be deleted by digestion with *Bgl*II, thus leaving the two terminal repeats of AAV, which are the only sequences required for replication and packaging of AAV DNA. The pTR_{BS}-UF series contains two reporter gene cassettes, *gfp* and *neo*, each with its own promoter and polyadenylation signal. These two transcription units can be independently deleted (*Kpn*I-*Not*I digest for *gfp* and *Sal*I digest for *neo*), which increases the cloning space for the gene of interest. Even if used as is, the vector can accommodate another transcription unit of up to 1.6 kbp. Furthermore, the efficiency of a particular promoter in any given cell type or tissue could also be tested by substituting it for the CMV promoter upstream of the *gfp* gene after digesting the vector DNA with *Kpn*I and *Xba*I. Finally, the design of the pTR_{BS}-UF3 vector allows for the coordinate expression

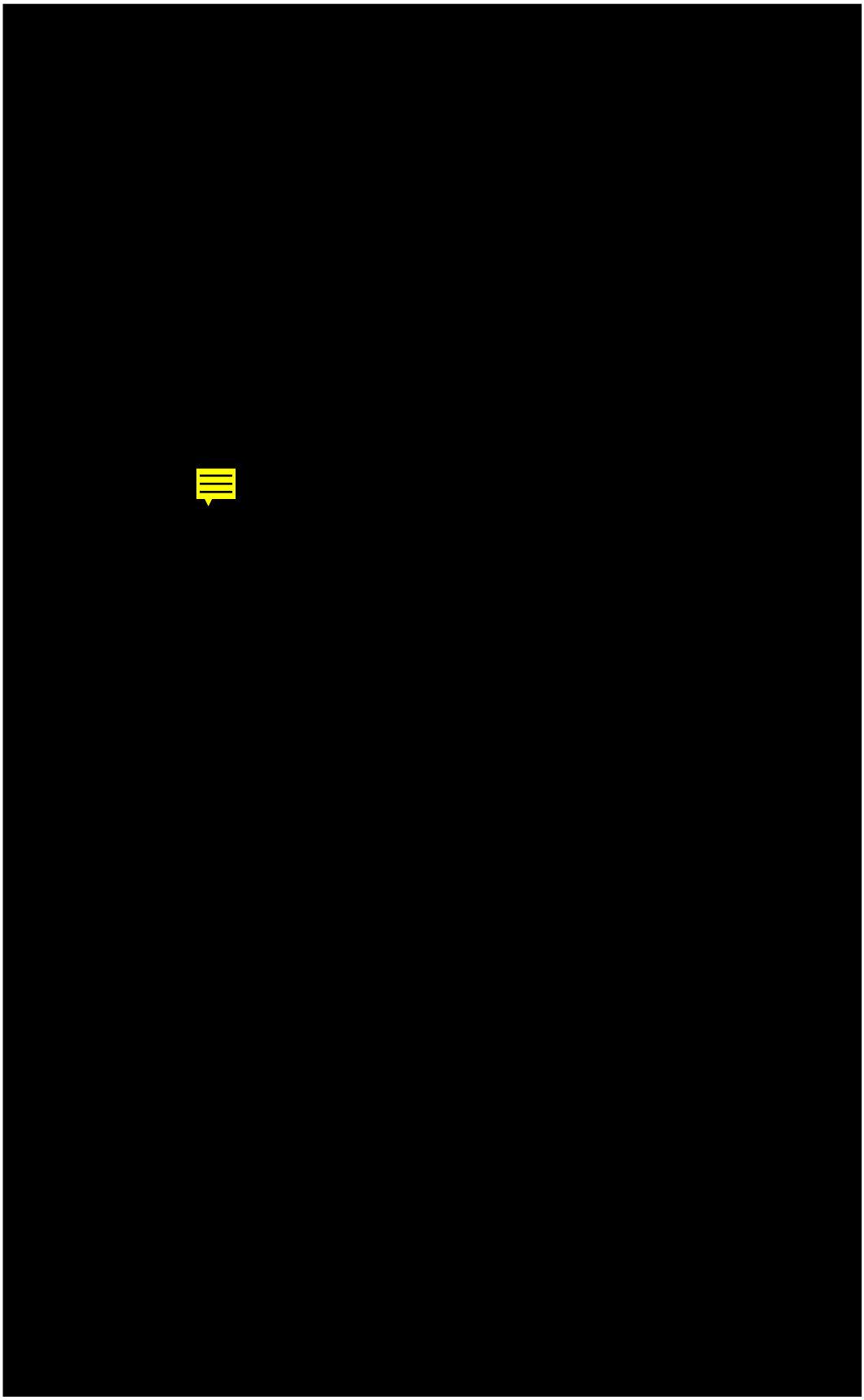


FIG. 9. GFP fluorescence in a segment of guinea pig RPE infected with rAAV-GFP_H2. (A) Differential interference contrast image of retina from an infected eye near the region shown in panel B. The darkly pigmented layer of cells near the top of the retina shown is the RPE layer in a slightly oblique section. The photoreceptor cell layer and other neuroretinal layers can be seen below the RPE. (B) RPE layer from an rAAV-GFP_H2-inoculated eye near the injection site viewed under short-wavelength excitation and fluorescein emission optics by confocal microscopy. (C) Fluorescence of the RPE layer from the same eye as in panel B at a site distal to the injection site. (D) Fluorescence of the RPE layer from the uninjected eye of the same animal as in panels A to C.

of the reporter *gfp* gene and the gene of interest from the same promoter by the use of an IRES element.

In addition, we described the construction of an Ad shuttle vector, carrying the gfp_h reporter gene under the control of the IRES element. 293 cells infected with recombinant Ad displayed typical cytopathic effect and bright green fluorescence. Expression of GFP allowed for quick and easy selection of true recombinant Ad clones, discriminating them from false plaques. In principle, gfp_h can be incorporated into other viral and nonviral vector and expression systems. For example, Levy et al. (21) have recently used the gfp_h cDNA described here in the context of a retrovirus vector.

Finally, the system described here could be used for efficient transduction and expression of genes into cells of mammalian origin. rAAV-GFP_H transduces a reporter gene sensitive enough to allow detection by a simple sorting of infected cells by FACS, thereby eliminating selection of transduced cells with drugs such as G418 or manipulation of cells for the visualization of enzymatic activities such as β-galactosidase. Since AAV and Ad have very broad host ranges, the described vectors could be used in many gene delivery technologies, including human gene therapy.

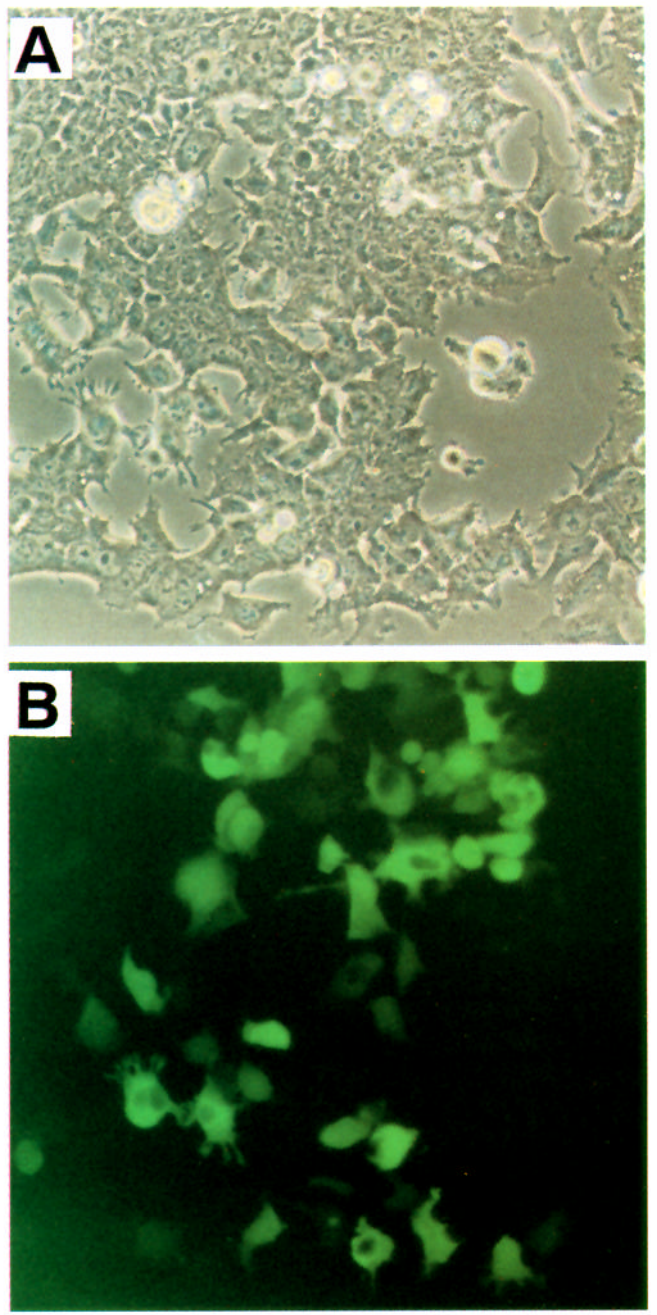
ACKNOWLEDGMENTS

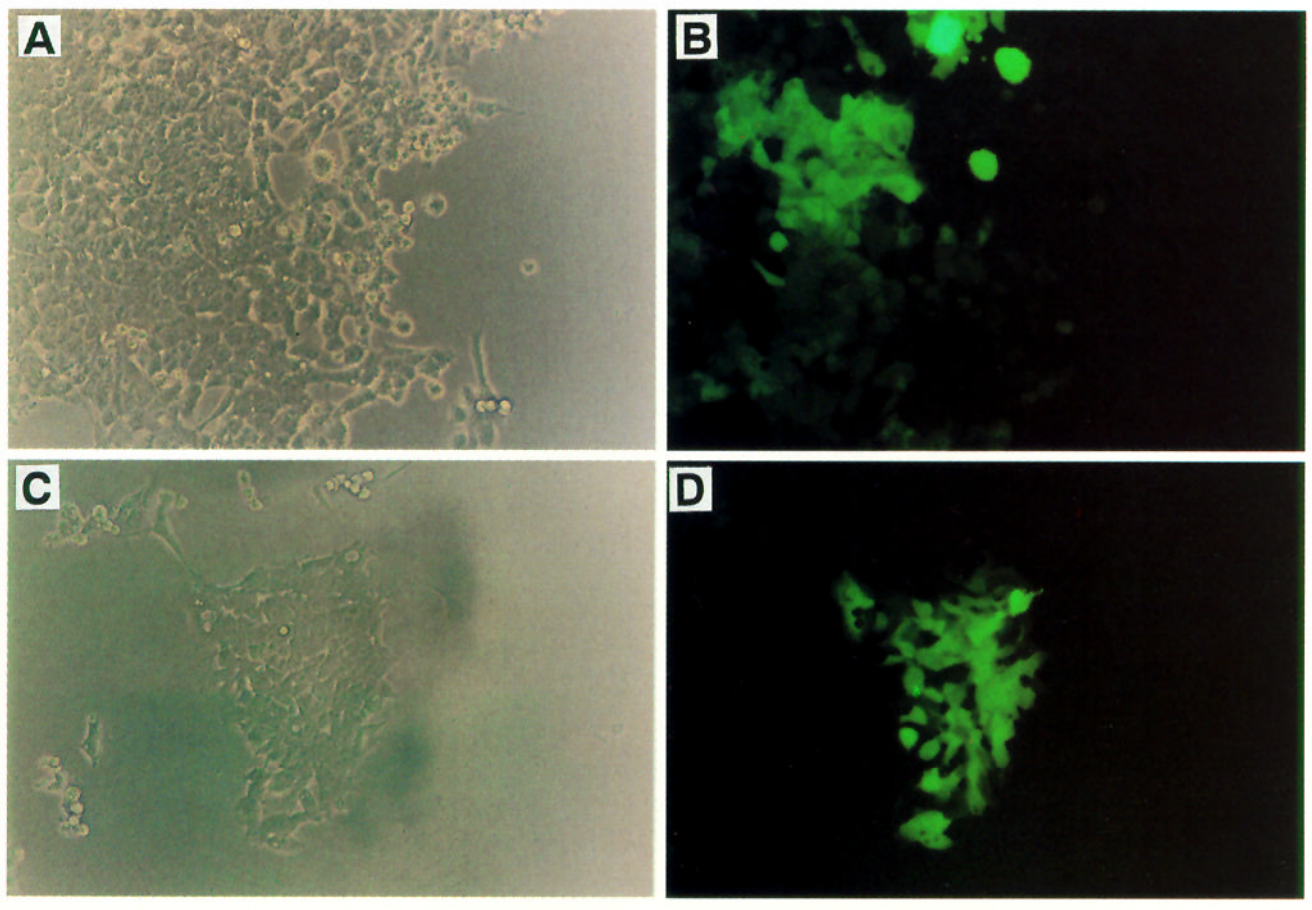
We thank R. J. Samulski (University of North Carolina), J. Resnick (University of Florida), and J. P. Levy (Central Iowa Health Systems) for sharing unpublished observations with us.

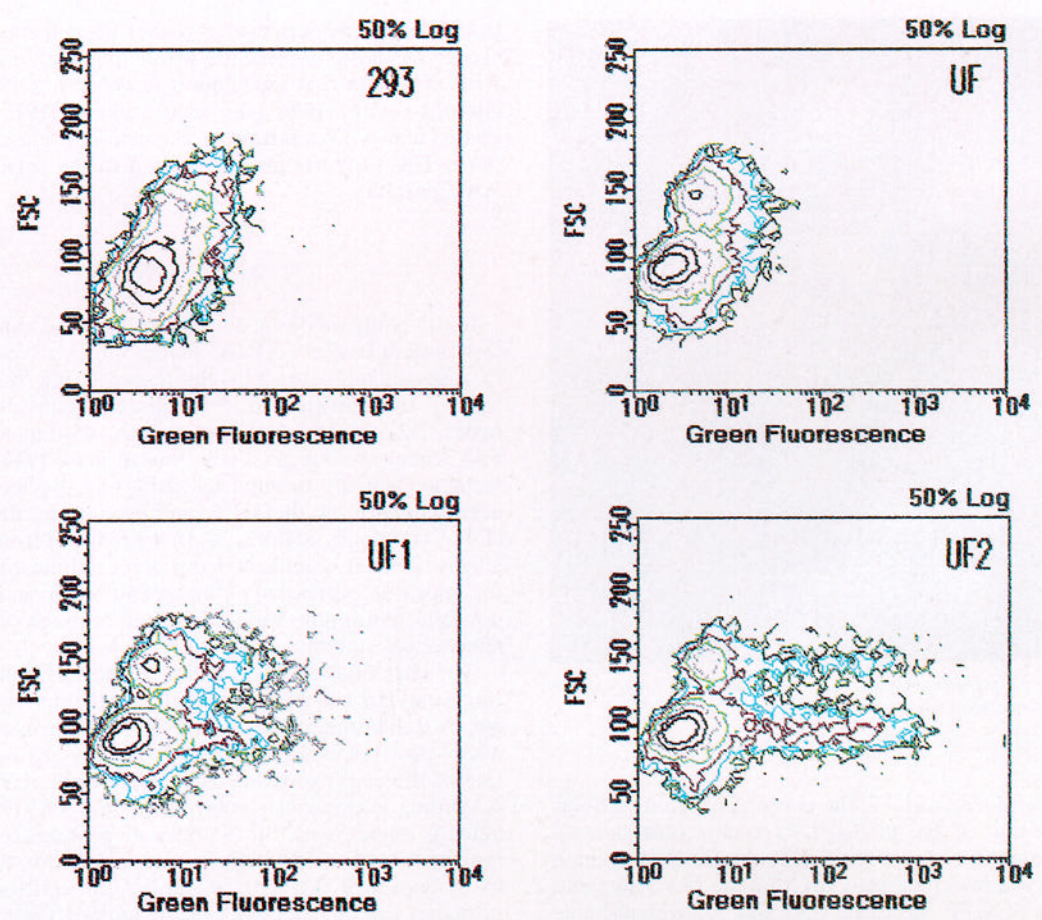
This work was supported by the University of Florida Gene Therapy Center.

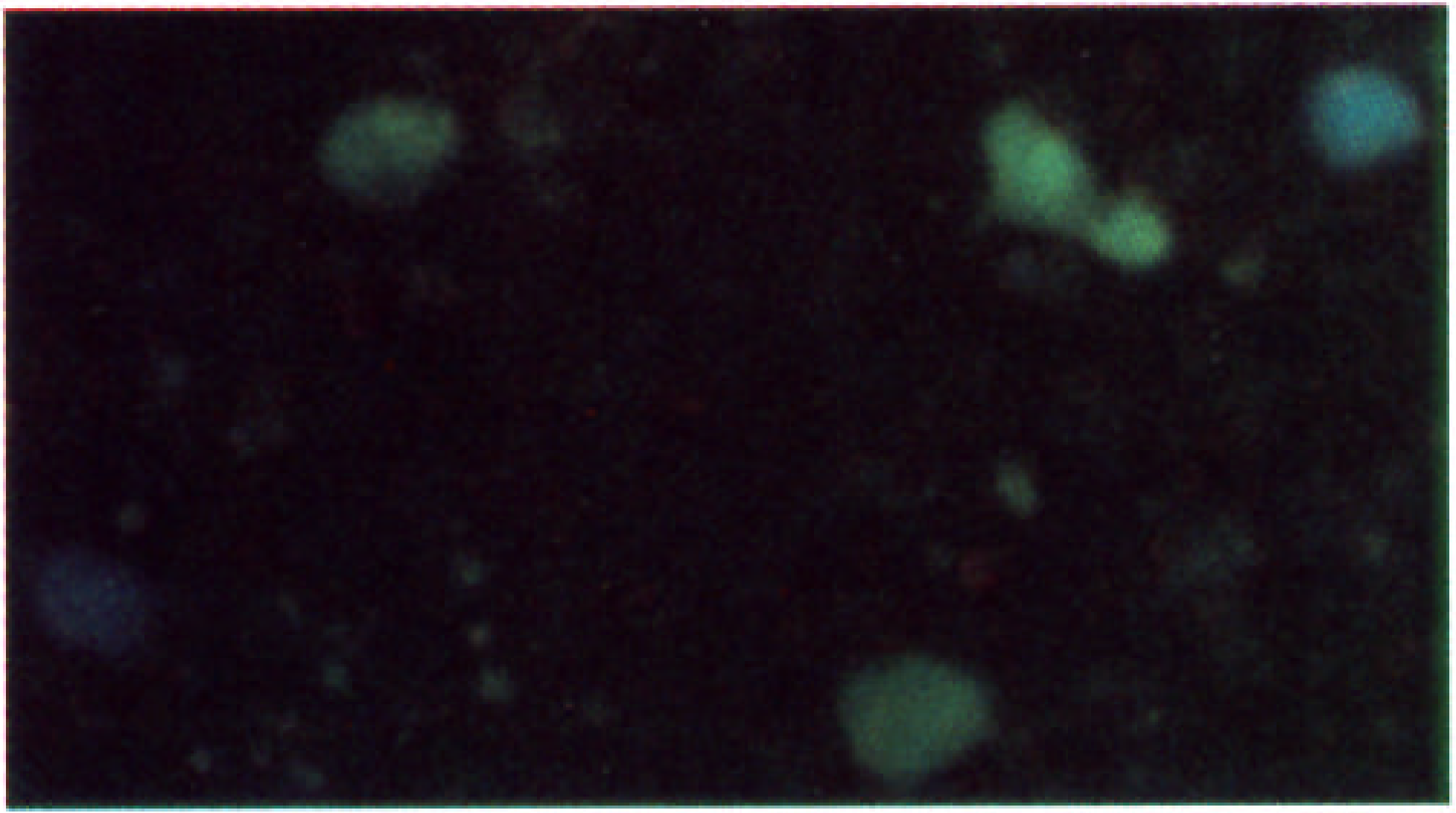
REFERENCES

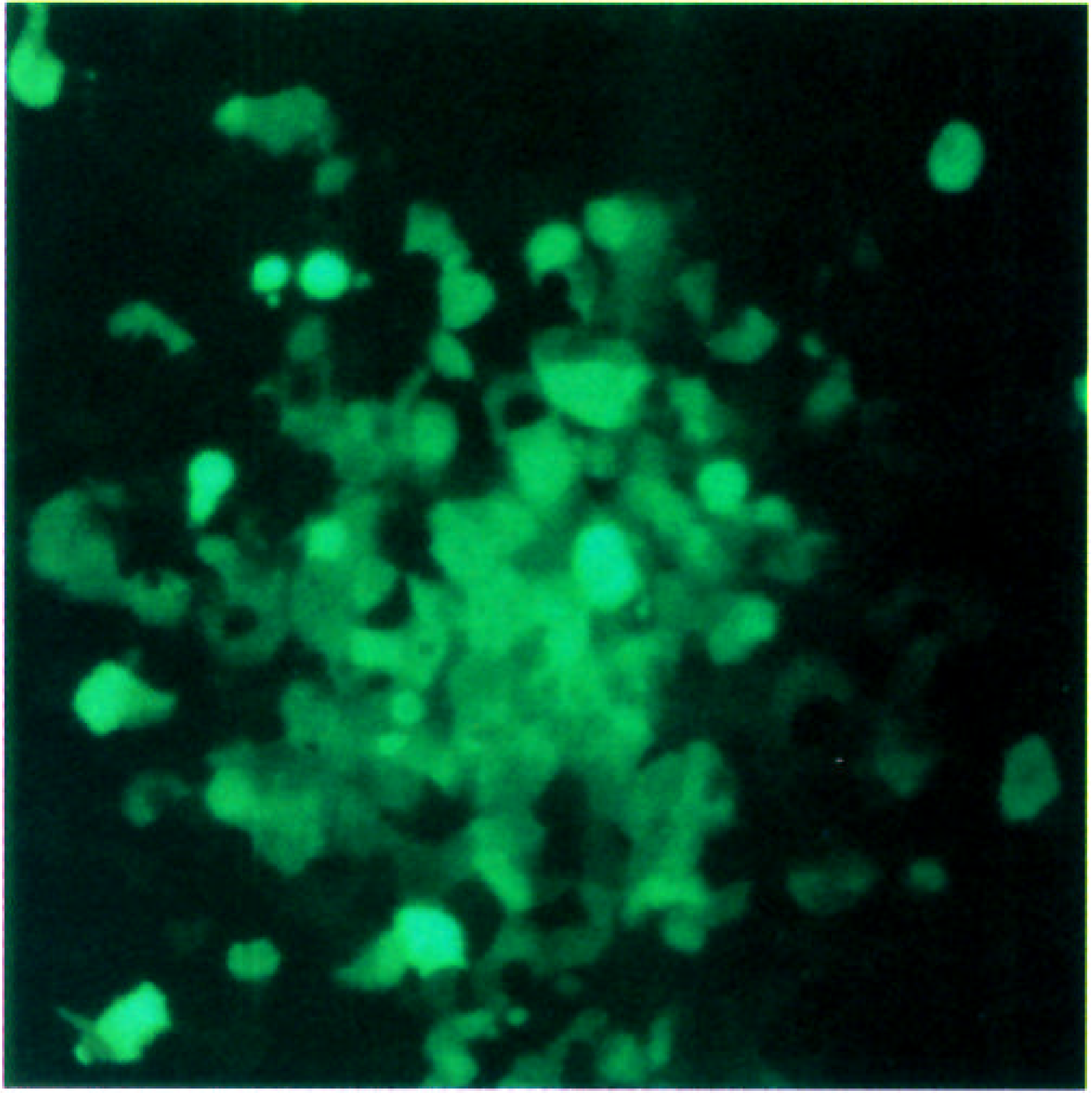
1. **Bennett, J., J. Wilson, D. Sun, B. Forbes, and A. Maguire.** 1994. Adenovirus vector-mediated in vivo gene transfer into adult murine retina. Invest. Ophthalmol. Visual Sci. **35:**2535–2542.
2. **Bennetzen, J. L., and B. D. Hall.** 1982. Codon selection in yeast. J. Biol. Chem. **257:**3026–3031.
3. **Bett, A. J., W. Haddara, L. Prevec, and F. L. Graham.** 1994. An efficient and flexible system for construction of adenovirus vectors with insertions or deletions in early regions 1 and 3. Proc. Natl. Acad. Sci. USA **91:**8802–8806.
4. **Chalfie, M., Y. Tu, G. Euskirchen, W. W. Ward, and D. C. Prasher.** 1994. Green fluorescent protein as a marker for gene expression. Science **263:**802–805.
5. **Cheung, A. K., M. D. Hoggan, W. W. Hauswirth, and K. I. Berns.** 1980. Integration of the adeno-associated virus genome into cellular DNA in latently infected human Detroit 6 cells. J. Virol. **33:**739–748.
6. **Cody, C. W., D. C. Prasher, W. M. Westler, F. G. Prendergast, and W. W. Ward.** 1993. Chemical structure of the hexapeptide chromophore of the Aequorea green-fluorescent protein. Biochemistry **32:**1212–1218.
7. **Dirks, W., M. Wirth, and H. Hauser.** 1993. Dicistronic transcription units for gene expression in mammalian cells. Gene **128:**247–249.
8. **Grantham, R., C. Gautier, M. Gouy, M. Jacobzone, and R. Mercier.** 1981. Codon catalog usage is a genome strategy modulated for gene expressivity. Nucleic Acids Res. **9:**r43–r74.
9. **Heim, R., A. B. Cubitt, and R. Y. Tsien.** 1995. Improved green fluorescence. Nature (London) **373:**663–664. (Letter.)
10. **Heim, R., D. C. Prasher, and R. Y. Tsien.** 1994. Wavelength mutations and posttranslational autoxidation of green fluorescent protein. Proc. Natl. Acad. Sci. USA **91:**12501–12504.
11. **Ikemura, T.** 1980. The frequency of codon usage in E. coli genes: correlation with abundance of cognate tRNA, p. 519–534. *In* S. Osawa et al. (ed.), Genetics and evolution of RNA polymerase, tRNA, and ribosomes. University of Tokyo Press, Tokyo, and Elsevier/North Holland, Amsterdam.
12. **Ikemura, T.** 1981. Correlation between the abundance of Escherichia coli transfer RNAs and the occurrence of the respective codons in its protein genes. J. Mol. Biol. **146:**1–21.
13. **Ikemura, T.** 1981. Correlation between the abundance of Escherichia coli transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the E. coli translational system. J. Mol. Biol. **151:**389–409.
14. **Ikemura, T.** 1982. Correlation between the abundance of yeast transfer RNAs and the occurrence of the respective codons in protein genes. Differences in synonymous codon choice patterns of yeast and Escherichia coli with reference to the abundance of isoaccepting transfer RNAs. J. Mol. Biol. **158:**573–597.
15. **Jackson, R. J., M. T. Howell, and A. Kaminski.** 1990. The novel mechanism of initiation of picornavirus RNA translation. Trends Biochem. Sci. **15:**477–483.
16. **Jang, S. K., H. G. Krausslich, M. J. Nicklin, G. M. Duke, A. C. Palmenberg, and E. Wimmer.** 1988. A segment of the 5′ nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during in vitro translation. J. Virol. **62:**2636–2643.
17. **Kotin, R. M.** 1996. Prospects for the use of adeno-associated virus as a vector for human gene therapy. Hum. Gene Ther. **5:**793–801.
18. **Kozak, M.** 1987. At least six nucleotides preceding the AUG initiator codon enhance translation in mammalian cells. J. Mol. Biol. **196:**947–950.
19. **Kozak, M.** 1990. Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. Proc. Natl. Acad. Sci. USA **87:**8301–8305.
20. **Laughlin, C. A., C. B. Cardellichio, and H. C. Coon.** 1986. Latent infection of KB cells with adeno-associated virus type 2. J. Virol. **60:**515–524.
21. **Levy, J. P., R. R. Muldoon, S. Zolotukhin, and C. J. Link, Jr.** 1996. Retroviral transfer and expression of a humanized, red shifted green fluorescent protein gene into human tumor cells. Nat. Biotechnol. **14:**610–614.
22. **Li, T., M. Adamian, D. J. Roof, E. L. Berson, T. P. Dryja, B. J. Roessler, and B. L. Davidson.** 1994. In vivo transfer of a reporter gene to the retina mediated by an adenoviral vector. Invest. Ophthalmol. Visual Sci. **35:**2543–2549.
23. **Macejak, D. G., and P. Sarnow.** 1991. Internal initiation of translation mediated by the 5′ leader of a cellular mRNA. Nature (London) **353:**90–94.
24. **McCarty, D. M., M. Christensen, and N. Muzyczka.** 1991. Sequences required for coordinate induction of adeno-associated virus p19 and p40 promoters by Rep protein. J. Virol. **65:**2936–2945.
25. **McGrory, W. J., D. S. Bautista, and F. L. Graham.** 1988. A simple technique for the rescue of early region I mutations into infectious human adenovirus type 5. Virology **163:**614–617.
26. **McLaughlin, S. K., P. Collis, P. L. Hermonat, and N. Muzyczka.** 1988. Adeno-associated virus general transduction vectors: analysis of proviral structures. J. Virol. **62:**1963–1973.
27. **Morin, J. G., and J. W. Hastings.** 1971. Energy transfer in a bioluminescent system. J. Cell Physiol. **77:**313–318.
28. **Muzyczka, N.** 1992. Use of adeno-associated virus as a general transduction vector for mammalian cells. Curr. Top. Microbiol. Immunol. **158:**97–129.
29. **Prasher, D. C., V. K. Eckenrode, W. W. Ward, F. G. Prendergast, and M. J. Cormier.** 1992. Primary structure of the Aequorea victoria green-fluorescent protein. Gene **111:**229–233.
30. **Ryan, J. H., S. Zolotukhin, and N. Muzyczka.** 1996. Sequence requirements for binding of Rep68 to the adeno-associated virus terminal repeats. J. Virol. **70:**1542–1553.
31. **Samulski, R. J., L. S. Chang, and T. Shenk.** 1989. Helper-free stocks of recombinant adeno-associated viruses: normal integration does not require viral gene expression. J. Virol. **63:**3822–3828.
32. **Shimomura, O.** 1979. Structure of the chromophore of the Aequorea green fluorescent protein. FEBS Lett. **104:**220–222.
33. **Snyder, R. O., D. S. Im, T. Ni, X. Xiao, R. J. Samulski, and N. Muzyczka.** 1993. Features of the adeno-associated virus origin involved in substrate recognition by the viral Rep protein. J. Virol. **67:**6096–6104.
34. **Wada, K., S. Aota, R. Tsuchiya, F. Ishibashi, T. Gojobori, and T. Ikemura.** 1990. Codon usage tabulated from the GenBank genetic sequence data. Nucleic Acids Res. **18**(Suppl.)**:**2367–2411.
35. **Ward, P., E. Urcelay, R. Kotin, B. Safer, and K. I. Berns.** 1994. Adeno-associated virus DNA replication in vitro: activation by a maltose-binding protein/Rep 68 fusion protein. J. Virol. **68:**6029–6037.
36. **Ward, W. W., C. W. Cody, R. C. Hart, and M. J. Cormier.** 1980. Spectrophotometric identity of the energy-transfer chromophores in Renilla and Aequorea green fluorescent proteins. Photochem. Photobiol. **31:**611–615.

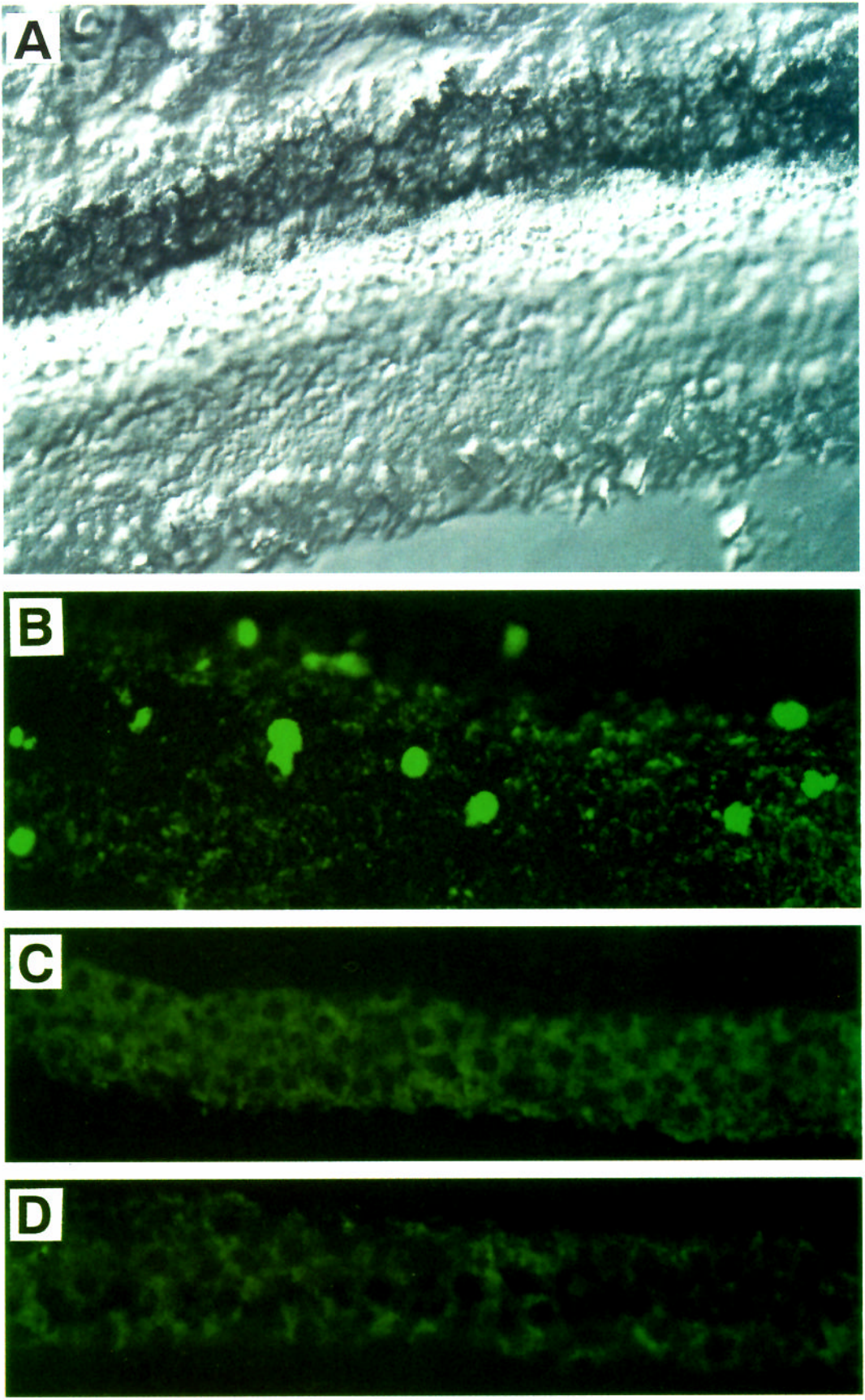

A
B


A
B
C
D


50% Log
293
FSC
0
50
100
150
200
250
10^0
10^1
10^2
10^3
10^4
Green Fluorescence
50% Log
UF
FSC
Green Fluorescence
50% Log
UF1
FSC
Green Fluorescence
50% Log
UF2
FSC
Green Fluorescence






A
B
C
D