Mammalian Expression of Infrared Fluorescent Proteins Engineered from a Bacterial Phytochrome
Xiaokun Shu, et al.
Science 324, 804 (2009);
DOI: 10.1126/science.1168683
Mammalian Expression of Infrared Fluorescent Proteins Engineered from a Bacterial Phytocrome

Xiaokun Shu,1,2 Antoine Royant,3 Michael Z. Lin,2 Todd A. Aguilera,2 Varda Lev-Ram,2 Paul A. Steinbach,1,2 Roger Y. Tsien1,2,4*

Visibly fluorescent proteins (FPs) from jellyfish and corals have revolutionized many areas of molecular and cell biology, but the use of FPs in intact animals, such as mice, has been handicapped by poor penetration of excitation light. We now show that a bacteriophytocrome from Deinococcus radiodurans, incorporating biliverdin as the chromophore, can be engineered into monomeric, infrared-fluorescent proteins (IFPs), with excitation and emission maxima of 684 and 708 nm, respectively; extinction coefficient >90,000 M−1 cm−1, and quantum yield of 0.07. IFPs express well in mammalian cells and mice and spontaneously incorporate biliverdin, which is ubiquitous as the initial intermediate in heme catabolism but has negligible fluorescence. However, excitation and emission maxima of FPs have not yet exceeded 598 and 655 nm, respectively (5–7). Somewhat longer wavelengths (644-nm excitation, 672-nm emission) have been observed in a phytocrome-based FP that incorporates phycocyanobilin (PCB) as the chromophore (8). However, neither incorporation of exogenous PCB nor transfer of its biosynthetic pathway into animal cells has yet been demonstrated. Bacterial phytocromes are more promising because they incorporate biliverdin IXα (BV) instead of PCB (9), and BV is the initial intermediate in heme catabolism by heme oxygenase (HO-1) in all aerobic organisms, including animals. For example, normal adult humans endogenously generate and metabolize 300 to 500 mg BV each day simply from routine heme breakdown (10). Recently, a full-length bacteriophytocrome (DrBphP) from Deinococcus radiodurans with a single mutation (D207H) (11) was reported to be red fluorescent at 622 nm upon excitation of the Soret band near 416 nm (12). Excitation of the Q band absorbing at 699 nm gave no fluorescence (12), which contradicted Kasha’s rule that fluorescence occurs from the lowest excited state. Emission peaks at 710 to 725 nm have been observed from various forms of Rhodopsseudomonas palustris (13) and Pseudomonas aeruginosa (14) bacteriophytochromes expressed in Escherichia coli, but fluorescence efficiencies have not been quantified, and reconstitution in nonbacterial systems has not yet been demonstrated.

To minimize the probability of nonradiative decay, we chose to limit DrBphP to its chromophore-binding domain (CBD), consisting of the PAS and GAF domains, which are supported by hemoglobin, water, and other macromolecules.

Fig. 1. Infrared fluorescent proteins created by structure-based engineering of a bacteriophytocrome. (A) Fourteen residues surrounding the biliverdin in DrCBD [Protein Data Bank (PDB) ID: 1ztu] (16) were divided into seven groups (shown in different colors) and targeted for mutagenesis. (B) Normalized excitation (blue) and emission (red) spectra of IFP1.4.

Supporting Online Material
www.sciencemag.org/cgi/content/full/1171583/DC1
Materials and Methods
Figs. S1 to S4
Tables S1 to S6
References
Movies S1 and S2

Published online 19 March 2009;
10.1126/science.1171583
Include this information when citing this paper.
necessary and sufficient for covalent incorporation of BV \((15, 16)\). We discarded the PHY domain and the C-terminal histidine kinase–related domain (HKRD) \((9)\), which transduce excited-state energy into conformational change and biochemical signaling \((16)\). A gene encoding DrCBD (321 amino acids) with the D207H mutation (in which His replaces Asp at codon 207) was synthesized with codons optimized for \(E. coli\) \((17)\). When coexpressed with cyanobacterial HO-1 in \(E. coli\) and excited near 700 nm, the truncated mutant fluoresced in the infrared with emission maximum of 722 nm. However, this mutant, dubbed IFP1.0, is weakly fluorescent (table S1), reversibly photofatiguable (fig. S1), and dimeric (fig. S2). The dimerization of IFP1.0 is due to at least four residues (Y307/L311/L314/V318) through hydrophobic interactions (fig. S3A).

On the basis of the crystal structure of DrCBD \((16)\), nonradiative decay of the excited chromophore is probably promoted by rotation of the D pyrrole ring because of relatively sparse packing of surrounding residues. Multiple sequence alignment of >100 phytochromes revealed conserved residues, some of which may contribute to photoisomerization [see supporting online material (SOM) text]. In order to increase the brightness of IFP1.0, 14 residues near the D ring were chosen and divided into seven groups for saturation mutagenesis (Fig. 1A), followed by DNA shuffling, which generated IFP1.1, with excitation and emission maxima of 686 and 713 nm, respectively, and brightness greater than that of IFP1.0 by a factor of ~2.6 (table S1). Several more rounds of directed evolution of IFP1.1 led to IFP1.4 (fig. S4).

IFP1.4 is about four times brighter than IFP1.0 (table S1), and its fluorescence is stable over a wide pH range from 5 to 9 (fig. S5). IFP1.4 is monomeric (fig. S2) and no longer shows significant reversible photofatigue (fig. S1). At an excitation rate that initially produces 1000 emitted photons per s per molecule of IFP1.1 or 1.4, the time to photobleach by 50% \((t_{1/2})\) is 8.5 or 8.4 s, respectively. For comparison, \(t_{1/2}\) of the popular yellow fluorescent protein Venus is 15 s \((18)\). A rationally introduced mutation L311K replaced a hydrophobic group by a charged amino acid and disrupted the dimer interface in IFP1.2 (fig. S3B). Mutation A288V likely eliminated the residual photoconversion of IFP1.0, because two additional methyl groups of Val288 may limit the D-ring rotation. However, the excitation and emission maxima of IFP1.4, 684 and 708 nm, respectively, are slightly blue-shifted compared with IFP1.0 (Fig. 1B). The blue shift may have resulted from using 676-nm excitation, the longest-wavelength laser line available to us, to select for higher brightness during fluorescence-activated cell sorting.

Expression of IFP1.4 alone without exogenous BV leads to bright and homogeneous infrared fluorescence in human embryonic kidney cells (HEK293A) (Fig. 2A). Furthermore, exogenously added BV further increased infrared fluorescence of transfected cells including neurons (fig. S6), which demonstrated that BV is membrane-permeant and adds rapidly to fill

![Fig. 2. Imaging of IFP1.4 and IFP1.4-PHAKT1 in HEK293A cells. (A) Fluorescence image of IFP1.4 taken with Cy5.5 filter set (665 ± 22.5 nm excitation, 725 ± 25 nm emission) with and without differential interference contrast (DIC). (B) Confocal laser-scanning microscopy of IFP1.4-PHAKT1 before and after insulin stimulation (excitation by 635-nm laser, emission by 650-nm long-pass filter).](https://www.sciencemag.org)

![Fig. 3. Imaging of GFP, mKate, and IFP1.1 in living mice. (A) Liver fluorescence of living mice injected with Ad5I (top row) and Ad5K (bottom row), imaged in the IFP excitation and emission channel before and after BV administration, mKate channel before and after BV, and GFP channel \((17)\). Images through the mKate channel have been 5× brightened in software to render them visible, so the relative gains of the IFP channel, mKate channel, and GFP channel were 1, 5, and 1, respectively. Arrows point to the liver. Note that the GFP images are dominated by autofluorescence, which renders the livers invisible. (B) Time course of averaged and normalized Ad5I liver fluorescence before and after BV injection. (C) Images of IFP-expressing mouse showing spectrally deconvoluted liver fluorescence (left, red) and its overlay (right) with autofluorescence (gray).](https://www.sciencemag.org)
IFP1.4 when endogenous BV had not saturated the protein. The half-life of IFP1.4 is about 4 hours in HEK293A cells (figs. S7, S8).

As a simple demonstration that IFP1.4 fusions can be functional, IFP1.4 was fused to the pleckstrin homology (PH) domain of human AKT1 (19). This PH domain is known to bind to phosphatidylinositol 3,4,5-trisphosphate formed at the plasma membrane after growth factor stimulation. Serum-starved HEK293 cells expressing the IFP1.4-PHAKT1 fusion showed IR fluorescence diffusely distributed in the cytosol, but this signal translocated to the plasma membrane within 10 min after insulin stimulation (Fig. 2B), which illustrated that IFP1.4 can highlight the trafficking of fusion proteins.

Expression of IFPs in intact mice via adenovirus serotype 5 (Ad5) also produced infrared fluorescence. Ad5 is well known to infect mouse liver specifically (20). Two modified versions of Ad5 were generated: Ad5I and Ad5K. Ad5I contains the genes for IFP1.1 and green fluorescent protein (GFP), the latter controlled by an internal ribosome entry sequence (IRES) (17). Ad5K encodes mKate, a red fluorescent protein advocated for in vivo imaging (5), and IRES-GFP. Weak infrared fluorescence of liver was detected 5 days after intravenous injection of Ad5I through tail vein (Fig. 3A). The whole liver was easily detected after intravenous injection of 250 nmol (~7 mg/kg) BV (Fig. 3A). The increase in liver fluorescence was half-maximal in ~10 min and maximal (about fivefold) 1 hour after BV injection (Fig. 3B). Resolution of IFP fluorescence from background autofluorescence can be enhanced by spectral deconvolution (Fig. 3C).

The three-dimensional distribution of IFP fluorescence in the mouse liver can be reconstructed tomographically (fig. S9). BV injection did not cause observable toxicity in mice (21). Furthermore, higher doses of BV (35 to 50 mg/kg) have been reported to give beneficial protection in vivo against reactive oxygen species (22) and transplantation-induced injury (23). As a control, intravenous injection of 250 nmol BV did not generate infrared fluorescence in either Ad5K-infected mice (Fig. 3A) or uninfected mice (not shown). The far-red fluorescence of mKate was observed in Ad5K-infected liver and was unaffected by BV. Neither Ad5I- nor Ad5K-infected intact mice displayed GFP fluorescence in their livers (Fig. 3A). Removal of the overlying skin, followed by complete exposure of the liver, increased the mKate fluorescence by a much greater factor than for the IFP signal (Fig. 4 and fig. S10), which showed how overlying tissues attenuate mKate’s excitation and emission wavelengths to a greater extent than those for IFP.

The entire dissected liver was fluorescent for mice expressing IFP and mKate (fig. S11), which suggested virus infection of the whole liver. GFP fluorescence became visible only after complete extraction of the liver and was similar for Ad5I and Ad5K (fig. S11), which indicated similar efficiencies of viral infection. Fluorescence microscopically of frozen sections showed fluorescence increasing in the order IFP1.1 < mKate < GFP fluorescence (Fig. 4C), which confirmed that IFP remains detectable in histology and that its improved visibility in vivo is due not to higher expression levels but rather to superior penetration of longer excitation and emission wavelengths through bulk pigmented tissue.

IFPs can be imaged over spatial scales from subcellular resolution up to strongly pigmented organs within intact whole mammals, whereas luciferase-based bioluminescence is useful mainly for whole-body imaging (24). The wavelengths of IFPs are particularly well-suited to optical tomographic reconstruction (fig. S9) (25). Even for microscopic imaging where existing FPs are highly effective, IFPs should reduce the contribution of cellular autofluorescence; enable excitation by cheap laser diodes; add new wavelengths for multicolor labeling; and allow resonance energy transfer from other dyes, FPs, or bioluminescent proteins. BV is uniquely advantageous as a cofactor because it is spontaneously and irreversibly incorporated into bacteriophytochromes (26), nontoxic at appropriate doses (21–23), nonfluorescent by itself, endogenously produced, and can be further supplemented either by expression of HO-1 or by direct administration of commercially available material. HO-1 is an important enzyme in its own right and is involved in various diseases (27). Its cumulative activity could be monitored by IFP fluorescence if apoprotein expression were in excess over BV. More than 1500 bacteriophytochrome-like sequences are already available in the National Center for Biotechnology Information (NIH) and CAMERA databases (28). These genes should provide raw material for selection and directed evolution of photochemical transducers based on a scaffold completely independent of the 11-stranded β barrel of coelenterate FPs.

References and Notes
High-Throughput Sequencing of the Zebrafish Antibody Repertoire

Joshua A. Weinstein,1* Ning Jiang,2* Richard A. White III,3 Daniel S. Fisher1,4,5 Stephen R. Quake2,3,4,†

Despite tremendous progress in understanding the nature of the immune system, the full diversity of an organism’s antibody repertoire is unknown. We used high-throughput sequencing of the variable domain of the antibody heavy chain from 14 zebrafish to analyze VDJ usage and antibody sequence. Zebrafish were found to use between 50 and 86% of all possible VDJ combinations and shared a similar frequency distribution, with some correlation of VDJ patterns between individuals. Zebrafish antibodies retained a few thousand unique heavy chains that also exhibited a shared frequency distribution. We found evidence of convergence, suggesting that despite tremendous progress in understanding the nature of the immune system, the full repertoire is highly plastic and can be directed to create antibodies with broad chemical diversity and high selectivity (1, 2). There is also a good understanding of the potential diversity available and the mechanistic aspects of how this diversity is generated. Antibodies are composed of two types of chains (heavy and light), each containing a highly diversified antigen-binding domain (variable). The V, D, and J gene segments of the antibody heavy-chain variable region is generated by a series of recombination events to generate a new heavy-chain gene (Fig. 1). Antibodies are formed by a mixture of recombination among gene segments, sequence diversification at the junctions of these segments, and point mutations throughout the gene (3). Estimates of immune diversity for antibodies or the related T cell receptors have attempted to extrapolate from small samples to entire systems or have been limited by coarse resolution of immune receptor genes (4). However, certain very elementary questions have remained open more than a half-century after being posed (1, 5, 6). It is still unclear what fraction of the potential repertoire is expressed in an individual at any point in time and how similar repertoires are between individuals who have lived in similar environments. Moreover, because each individual’s immune system is an independent experiment in evolution by natural selection, questions about repertoire similarity also inform our understanding of evolutionary diversity and convergence.

Zebrafish are an ideal model system for studying the adaptive immune system because they have the earliest recognition of an adaptive immune system whose features match the essential human elements (7, 8). Like humans, zebrafish have a recombination activating gene (RAG) and a combinatorial rearrangement of V, D, and J gene segments to create antibodies. They also have junctional diversity during recombination and somatic hypermutation of antibodies to improve specificity, and the organization of their immunoglobulin (Ig) gene loci approximates that of human (9). In addition, the zebrafish immune system has only ~300,000 antibody-producing B cells, making it three orders of magnitude simpler than mouse and five orders simpler than human in this regard.

We developed an approach to characterize the antibody repertoire of zebrafish by analyzing complementarity-determining region 3 (CDR3) of the heavy chain, which contains the vast majority of immunoglobulin diversity (10, 11) and can be captured in a single sequencing read (Fig. 1). Using the 454 GS FLX high-throughput pyrosequencing technology allowed sequencing of 640 million bases of zebrafish antibody cDNA from 14 zebrafish in four families (Fig. 1B). Zebrafish were raised in separate aquaria for each family and were allowed to have normal interactions with the environment, including the development of internal flora. We chose to investigate the quiescent state of the immune system, a state where the zebrafish had sampled a complex but fairly innocuous environment and had established an equilibrium of normal immune function. mRNA was prepared from whole fish, and we synthesized cDNA using primers designed to capture the entire variable region.

Between 28,000 and 112,000 useful sequencing reads were obtained per fish, and we focused our analysis on CDR3 sequences. Each read was assigned V and J by alignment to a reference with a 99.6% success rate (table S3); failures were due to similarity in some of the V gene segments. D was determined for each read by applying a clustering algorithm to all of the reads within a given