Mammalian Expression of Infrared Fluorescent Proteins Engineered from a Bacterial Phytochrome

Xiaokun Shu, *et al.*

Science **324**, 804 (2009);
DOI: 10.1126/science.1168683

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005.

6. U. H. Manjunatha *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 431 (2006).
7. M. Matsumoto *et al.*, *PLoS Med.* **3**, e466 (2006).
8. R. Singh *et al.*, *Science* **322**, 1392 (2008).
9. T. S. Balganesh, P. M. Alzari, S. T. Cole, *Trends Pharmacol. Sci.* **29**, 576 (2008).
10. Global Alliance for TB, *Tuberculosis (Edinb.)* **81**, (suppl. 1), 1 (2001).
11. D. B. Young, M. D. Perkins, K. Duncan, C. E. Barry 3rd, *J. Clin. Invest.* **118**, 1255 (2008).
12. V. Makarov, U. Möllmann, S. T. Cole, Eurasian Patent Application EP2029583 (2007).
13. V. Makarov *et al.*, *J. Antimicrob. Chemother.* **57**, 1134 (2006).
14. Y. Zhang, C. Vilcheze, W. R. Jacobs Jr., in *Tuberculosis and the Tubercle Bacillus*, S. T. Cole, K. D. Eisenach, D. N. McMurray, W. R. Jacobs Jr., Eds. (American Society for Microbiology Press, Washington, DC, 2005), pp. 115–140.
15. N. Q. Balaban, J. Merrin, R. Chait, L. Kowalik, S. Leibler, *Science* **305**, 1622 (2004).
16. H. I. Boshoff *et al.*, *J. Biol. Chem.* **279**, 40174 (2004).
17. S. J. Waddell, P. D. Butcher, *Curr. Mol. Med.* **7**, 287 (2007).
18. V. C. Abraham, D. L. Taylor, J. R. Haskins, *Trends Biotechnol.* **22**, 15 (2004).
19. D. Fenistein, B. Lenseigne, T. Christophe, P. Brodin, A. Genovesio, *Cytometry A* **73**, 958 (2008).
20. K. Mikusova *et al.*, *J. Bacteriol.* **187**, 8020 (2005).
21. C. M. Sassetti, D. H. Boyd, E. J. Rubin, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 12712 (2001).
22. B. A. Wolucka, *FEBS J.* **275**, 2691 (2008).
23. C. E. Barry, D. C. Crick, M. R. McNeil, *Infect. Disord. Drug Targets* **7**, 1 (2007).
24. K. Mikusova, R. A. Slayden, G. S. Besra, P. J. Brennan, *Antimicrob. Agents Chemother.* **39**, 2484 (1995).
25. L. J. Alderwick *et al.*, *J. Biol. Chem.* **280**, 32362 (2005).
26. Materials and methods are available as supporting material on *Science* Online.

27. We thank A. Deshpande, I. Heinemann, P. Højrup, K. Johnsson, M. K. N. Kumar, N. Kumar, L. Pagani, P. Marone, M. R. McNeil, I. Old, J. Reddy, S. Schmitt, P. Vachaspati, and C. Weigel for their help and support. Patents related to this work have been filed (WO/2007/134625, WO/2009/010163, and PCT/EP2008/001088). The NM4TB Consortium is funded by the European Commission (LHSP-CT-2005-018923), and microarray work at St George's Hospital, University of London, is supported by the Wellcome Trust (grant 062511). Microarray data are Minimum Information About a Microarray Experiment (MIAME)–compliant and deposited under accession number E-BUGS-80 at http://bugs.sgul.ac.uk/E-BUGS-80.

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

29 January 2009; accepted 13 March 2009
Published online 19 March 2009;
10.1126/science.1171583
Include this information when citing this paper.

Mammalian Expression of Infrared Fluorescent Proteins Engineered from a Bacterial Phytochrome

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. Tsien[1,2,4]*

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 bacteriophytochrome 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 by itself. Because their wavelengths penetrate tissue well, IFPs are suitable for whole-body imaging. The IFPs developed here provide a scaffold for further engineering.

In vivo optical imaging of deep tissues in animals is most feasible between 650 and 900 nm because such wavelengths minimize the absorbance by hemoglobin, water, and lipids, as well as light-scattering (*1*, *2*). Thus, genetically encoded IFPs would be particularly valuable for whole-body imaging in cancer and stem cell biology (*3*, *4*), gene therapy, and so on. 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 phytochrome-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 phytochromes 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 bacteriophytochrome (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 *Rhodopseudomonas 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

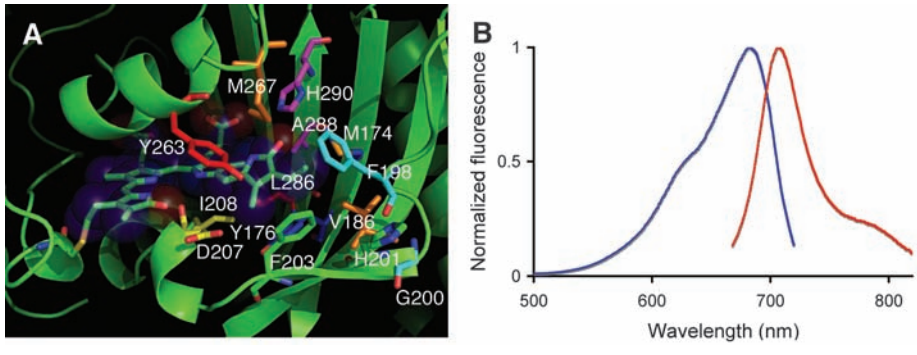

Fig. 1. Infrared fluorescent proteins created by structure-based engineering of a bacteriophytochrome. (**A**) Fourteen residues surrounding the biliverdin in DrCBD [Protein Data Bank (PDB) ID: 1ztu] (*16*) were divided into seven groups (shown in different colors) and targeted for mutagenesis. (**B**)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

[1]Howard Hughes Medical Institute, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093–0647, USA. [2]Department of Pharmacology, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093–0647, USA. [3]Institut de Biologie Structurale 41, rue Jules Horowitz, 38027 Grenoble CEDEX 1, France. [4]Departments of Chemistry and Biochemistry, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093–0647, USA.

*To whom correspondence should be addressed. E-mail: rtsien@ucsd.edu

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 Val^{288} 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

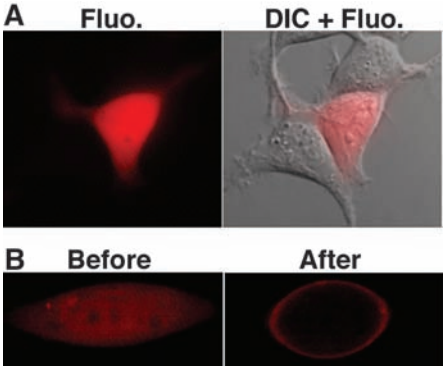

Fig. 2. Imaging of IFP1.4 and IFP1.4-PH^{AKT1} in HEK293A cells. (**A**) Fluorescence image of IFP1.4 taken with Cy5.5 filter set (665 ± 22.5 nm excitation, 725 ± 25 nm emission) with and without differential interference contrast (DIC). (**B**) Confocal laser-scanning microscopy of IFP1.4-PH^{AKT1} before and after insulin stimulation (excitation by 635-nm laser, emission by 650-nm long-pass filter).

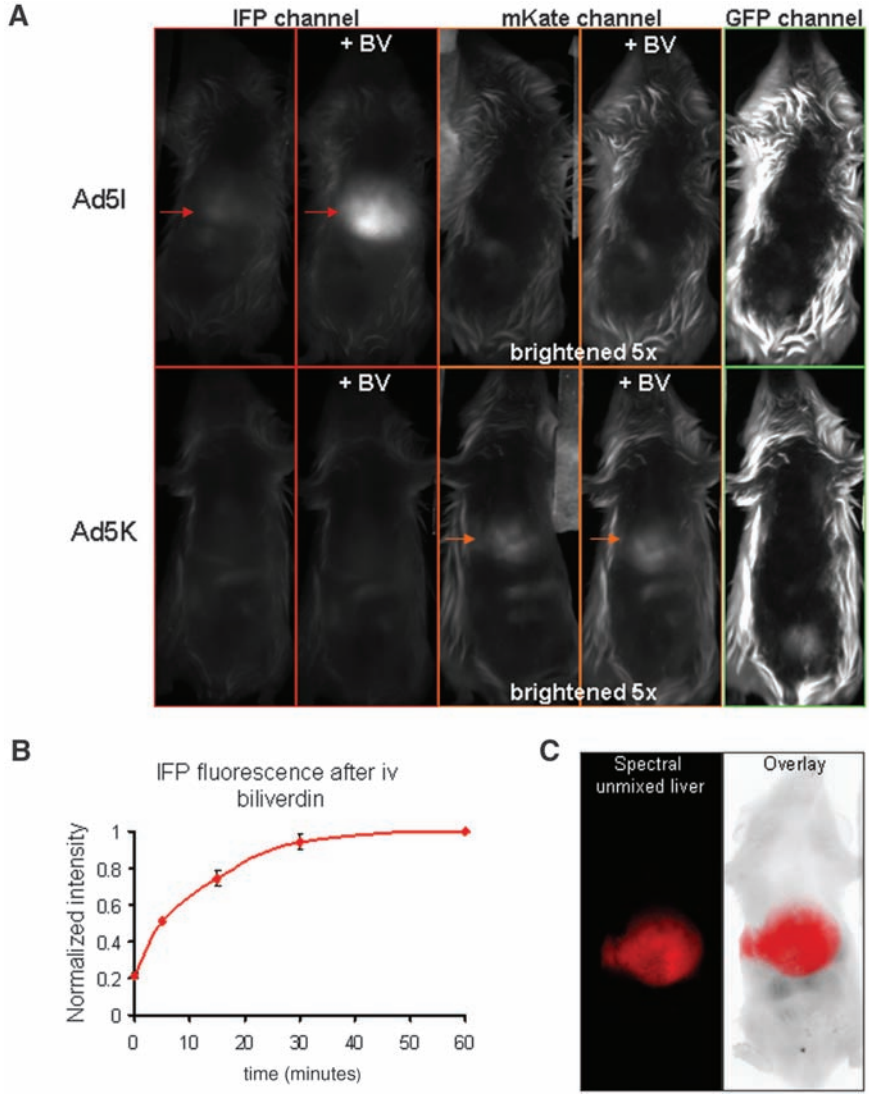

Fig. 3. Imaging of GFP, mKate, and IFP1.1 in living mice. (**A**) Liver fluorescence of living mice injected with Ad5I (top row) and Ad5K (bottom row), imaged in the IFP excitation and emission channel before and after BV administration, mKate channel before and after BV, and GFP channel (*17*). Images 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).

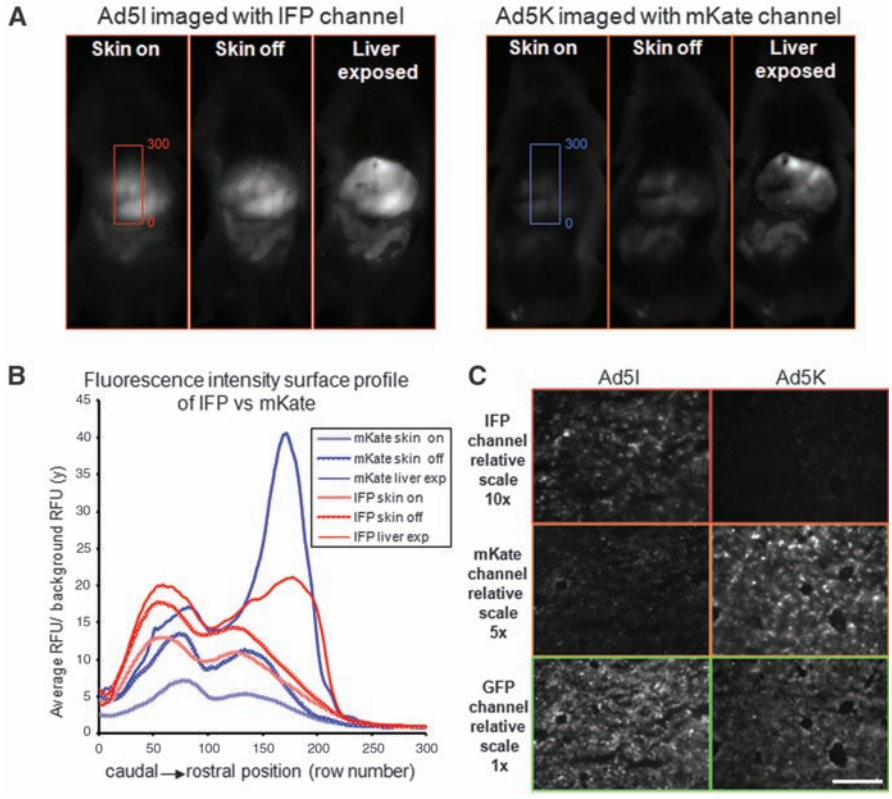

Fig. 4. Analysis of mKate and IFP1.1 visibility and expression levels in livers of mice infected with Ad5I and Ad5K. (**A**) IFP and mKate fluorescence images before dissection (skin on), after removal of skin (skin off), and after removal of overlying peritoneum and rib cage (liver exposed). mKate images were 2.5× brightened relative to IFP images. Note that the Ad5I-infected mouse was imaged after 250 nmol intravenous injection of BV. See bright-field images in fig. S9. (**B**) Fluorescence intensity (signal-to-noise ratio) analysis of livers from (**A**). Average of 80 pixels in the horizontal *x* axis divided by the average of the 30 most rostral horizontal lines from each of the images starting below the liver (i.e., signal-to-noise ratio), moving rostrally for 300 lines. (**C**) Frozen sections were imaged to show IFP, mKate, and GFP expression with a fluorescence stereomicroscope (Lumar, Zeiss) (*17*), displayed with relative intensity gains of 10, 5, and 1, respectively. Scale bar, 200 μm.

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-PH^{AKT1} 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 microscopy 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 accept 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

1. F. F. Jöbsis, *Science* **198**, 1264 (1977).
2. R. Weissleder, V. Ntziachristos, *Nat. Med.* **9**, 123 (2003).
3. T. Schroeder, *Nature* **453**, 345 (2008).
4. R. Weissleder, M. J. Pittet, *Nature* **452**, 580 (2008).
5. D. Shcherbo *et al.*, *Nat. Methods* **4**, 741 (2007).
6. M. A. Shkrob *et al.*, *Biochem. J.* **392**, 649 (2005).

7. L. Wang, W. C. Jackson, P. A. Steinbach, R. Y. Tsien, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 16745 (2004).
8. A. J. Fischer, J. C. Lagarias, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 17334 (2004).
9. S. J. Davis, A. V. Vener, R. D. Vierstra, *Science* **286**, 2517 (1999).
10. J. W. Harris, R. W. Kellermeyer, *The Red Cell* (Harvard Univ. Press, Cambridge, MA, 1970).
11. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
12. J. R. Wagner *et al.*, *J. Biol. Chem.* **283**, 12212 (2008).
13. E. Giraud *et al.*, *J. Biol. Chem.* **280**, 32389 (2005).
14. X. Yang, J. Kuk, K. Moffat, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 14715 (2008).
15. N. C. Rockwell, Y. S. Su, J. C. Lagarias, *Annu. Rev. Plant Biol.* **57**, 837 (2006).
16. J. R. Wagner, J. S. Brunzelle, K. T. Forest, R. D. Vierstra, *Nature* **438**, 325 (2005).
17. Materials and methods are available as supporting material on *Science* Online.
18. N. C. Shaner, P. A. Steinbach, R. Y. Tsien, *Nat. Methods* **2**, 905 (2005).
19. A. Bellacosa, J. R. Testa, S. P. Staal, P. N. Tsichlis, *Science* **254**, 274 (1991).
20. S. N. Waddington *et al.*, *Cell* **132**, 397 (2008).
21. Six mice after BV injection were observed for 3 days (the maximum time that we could hold mice for imaging according to our university-approved animal protocol).
22. R. Ollinger *et al.*, *Antioxid. Redox Signal.* **9**, 2175 (2007).
23. A. Nakao *et al.*, *Gastroenterology* **127**, 595 (2004).
24. C. H. Contag, M. H. Bachmann, *Annu. Rev. Biomed. Eng.* **4**, 235 (2002).
25. V. Ntziachristos *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 12294 (2004).
26. Electrospray mass spectrometry of bacterially expressed IFP1.4 holoprotein in 5% acetonitrile, 0.05% trifluoroacetic acid reported an average relative molecular mass (M_r) of 36,342 daltons, within experimental error of the M_r (36,332 daltons) expected from covalent incorporation of BV (582.6 daltons) into the apoprotein (35,749.6 daltons) from Ala^2 to the C terminus.
27. N. G. Abraham, A. Kappas, *Pharmacol. Rev.* **60**, 79 (2008).
28. D. B. Rusch *et al.*, *PLoS Biol.* **5**, e77 (2007).

29. We thank J. C. Lagarias and S. Field for donation of cDNAs encoding HO-1 and AKT1's PH domain, respectively; J. M. Saathoff and G. Tran for help with plasmid purification; M. Timmers for help with tissue culture; S. Adams for help with light scattering measurements; L. Gross for mass spectrometry; and Q. Xiong for flow cytometry. This work was supported by NIGMS grant R01 GM086197 and the Howard Hughes Medical Institute.

Supporting Online Material

www.sciencemag.org/cgi/content/full/324/5928/804/DC1
Materials and Methods
SOM Text
Figs. S1 to S11
Table S1
References

18 November 2008; accepted 10 March 2009
10.1126/science.1168683

High-Throughput Sequencing of the Zebrafish Antibody Repertoire

Joshua A. Weinstein,[1]* Ning Jiang,[2]* Richard A. White III,[3] Daniel S. Fisher,[1,4,5] Stephen R. Quake[1,2,3,4]†

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, in which different individuals made the same antibody. This approach provides insight into the breadth of the expressed antibody repertoire and immunological diversity at the level of an individual organism.

The nature of the immune system's antibody repertoire has been a subject of fascination for more than a century. This 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 genes go through 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 either 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, these 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 in evolutionary terms they have the earliest recognizable 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 complimentarity-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 natural 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

[1]Biophysics Program, Stanford University, Stanford, CA 94305, USA. [2]Department of Bioengineering, Stanford University, Stanford, CA 94305, USA. [3]Howard Hughes Medical Institute, Stanford University, Stanford, CA 94305, USA. [4]Department of Applied Physics, Stanford University, Stanford, CA 94305, USA. [5]Department of Biology, Stanford University, Stanford, CA 94305, USA.

*These authors contributed equally to this work.
†To whom correspondence may be addressed. E-mail: quake@stanford.edu