Identification of Sites Within a Monomeric Red Fluorescent Protein that Tolerate Peptide Insertion and Testing of Corresponding Circular Permutations

Yankun Li, Aillette M. Sierra, Hui-wang Ai and Robert E. Campbell*

Department of Chemistry, University of Alberta, Edmonton, Alberta, Canada

ABSTRACT

In recent years the class of known fluorescent proteins (FPs) has dramatically expanded as an ever-increasing numbers of variants and homologs of the green fluorescent protein (GFP) from *Aequorea* jellyfish have been either engineered in the lab or discovered in other marine organisms. The red fluorescent protein (RFP) from *Discosoma* coral (also known as dsFP583 and DsRed) has proven to be a particularly fruitful progenitor of variants with biochemical and spectroscopic properties conducive to applications in live cell imaging. We have investigated the tolerance of an engineered monomeric descendent of *Discosoma* RFP, known as mCherry, towards peptide insertion and circular permutation. Starting from a random library of insertion variants, we identified six genetically distinct sites localized in three different loops where a sequence of five residues could be inserted without abolishing the ability of the protein to form its intrinsic red fluorescent chromophore. For each of these insertion variants, a corresponding circular permutation variant was created in which the original N- and C-termini were connected by a six-residue linker and new termini were introduced at the site of the insertion. All six circular permutation variants had significantly diminished brightness relative to the analogous insertion variants. The most promising circular permutation variant has termini at the position corresponding to residue 184 of mCherry and retains 37% of the intrinsic fluorescent brightness of mCherry. These circularly permuted variants may serve as the foundation for construction of genetically encoded Ca^{2+} sensors analogous to the previously reported camgaroo, pericam and G-CaMP sensors based on variants of *Aequorea* GFP.

INTRODUCTION

Engineered variants of fluorescent proteins (FPs) cloned from marine organisms of the phylum Cnidaria have provided biologists with a toolbox of genetically encoded fluorophores suitable for expression and imaging in living cells (1,2). The most common application of FPs is the creation of chimeras with particular proteins of interest (3,4). Cells transfected with the chimeric gene synthesize the protein of interest covalently linked to the FP via a suitable polypeptide linker. By employing fluorescence microscopy, a researcher can obtain high-resolution images of the transfected cells (or tissue) that reveal the subcellular localization of the protein of interest. For the researcher who wishes to extract a more comprehensive and detailed insight into the intracellular behavior and interaction partners of their protein of interest, there is a growing number of sophisticated imaging techniques at their disposal. For example, time-lapse imaging in combination with spatially confined photobleaching (5) or photoactivation (2) can provide insight into a protein's spatial and temporal dynamics. Furthermore, intermolecular protein–protein interactions and proximities of less than ~10 nm can be detected using fluorescence (or Förster) resonance energy transfer (FRET) between two different colors of FP using either a steady-state or time-resolved imaging modality (6).

Fluorescent protein chimeras combined with appropriate imaging techniques, such as those mentioned in the partial list above, are extremely useful for investigating where and when a particular protein is present in a cell. However, these "passive" FP techniques are not readily amenable to addressing less tractable questions regarding the intracellular environment such as the concentration and localization of small molecules or the functional state of specific enzymes. The most effective strategy for addressing such questions has been to engineer a series of new tools for inclusion in the FP toolbox: FP-based biosensors (4). In contrast to FPs acting as "passive" fluorophores in which environmentally insensitive fluorescence is a desirable trait, FP-based biosensors are "active" in the sense that they are designed such that their spectral properties (*i.e.* intensity or color) are dependent on a biochemical parameter (4). FP-based biosensors can be subdivided into two groups; those based on FRET between two different colors of FP and those based on a single FP (4). Some notable examples of each type of sensor include: single FP biosensors for intracellular Ca^{2+} (7–9), kinase activity (10) and beta-lactamase inhibitory protein (11); and FRET-type biosensors for Ca^{2+} (12), kinase activity (13,14), histone methylation (15) and phosphoinositide (16).

The challenge in developing single FP biosensors is to create a connection between a biochemical parameter of interest and the chemical microenvironment of the chromophore (17). Given that the chromophore is located in the heart of the FP and shielded from the outside environment by a formidable 11-stranded β-barrel (18,19), it is apparent that the structure of

*Corresponding author email: robert.e.campbell@ualberta.ca (Robert E. Campbell)

FPs has evolved to prevent just such connections from occurring. Fortunately, protein engineers have found means to circumvent the β-barrel shell of FPs and to create variants in which the fluorescence of the chromophore does depend on events or the environment outside the barrel. The two most effective approaches include insertion of a binding protein within the structure of the FP or circular permutation of the FP and fusing interacting proteins to the new termini (4). The key to both of these approaches is that the site of insertion, or the site of the new termini, are close in spatial proximity to the chromophore. Structural changes that occur upon analyte binding or chemical modification of a partner domain are communicated to the chromophore as changes in the chemical microenvironment. For example, a change in the chemical microenvironment could alter the pKa of the fluorophore (as proposed for the Ca^{2+} sensor camgaroo (7)) or the extent of fluorescence quenching (as appears to be the case for the hydrogen peroxide sensor HyPer (20)). It is self-evident that before one can begin to design a single FP biosensor, "permissive" sites within the FP that tolerate insertion or the introduction of new termini in a circularly permuted variant must be known. For *Aequorea* green fluorescent protein (GFP), these sites were identified by both screening of random libraries of circularly permuted variants of enhanced GFP (7) as well as by systematic structure guided circular permutation with new termini in each of the loops of the GFPuv variant (21). It has been possible to translate the permissive sites identified in these green GFP variants to a variety of *Aequorea* GFP-derived color shifted variants including yellow fluorescent protein (YFP) (7,22), cyan fluorescent protein (CFP) (7), and T-Sapphire (23).

In addition to their utility in single FP-based biosensors, circularly permuted FPs have also proven very useful for the optimization of FRET-based sensors. Replacement of either a donor or acceptor FP in a FRET construct with a circularly permuted FP of the same color can result in a sensor with improved dynamic range due to changes in the interchromophore dipole–dipole orientation (24–26). Optimization of FRET-based sensors by introduction of circularly permuted FP variants is a completely empirical process and is therefore critically dependent on having a large number of alternatives available.

In 1999 the spectrum of available hues of FP was greatly expanded when it was reported that a red fluorescent protein (RFP) had been cloned from *Discosoma* sp. coral (27). Unfortunately, the wild type RFP suffered from slow and incomplete chromophore maturation and obligate tetramerization which impeded its use in many applications (28). In 2002, a monomeric RFP (29), named mRFP1, was generated by multiple steps of random and site-directed mutagenesis. Further efforts to improve the properties of this monomeric version as well as diversify the color palette eventually produced the "mFruit" series of yellow to far red FPs (1,30). mCherry is the member of the mFruit series that provides the best compromise of red-shifted emission, photostability, maturation speed, brightness and pH resistance. There have been no reports to date describing either insertion variants or circularly permuted variants of a monomeric RFP, though there has been a report of a split variant of mRFP1 (31).

In this manuscript we describe the identification of permissive sites within mCherry at which five amino acids can be inserted without obliterating the intrinsic ability of the protein to form its red fluorescent chromophore. For each insertion variant we created a circularly permuted variant with new N- and C- termini at the site of the insertion.

MATERIALS AND METHODS

General methods and materials. Synthetic DNA oligonucleotides were purchased from Sigma-Genosys Canada or Integrated DNA Technologies. Unless otherwise noted, Pfu polymerase (Fermentas) was used for all polymerase chain reaction (PCR) amplifications in the buffer supplied by the manufacturer. All restriction enzymes were purchased from Invitrogen or New England Biolabs. T4 DNA ligase (Invitrogen) was used for all ligation reactions according to the supplier's recommended protocol. PCR products were routinely purified using the QIAquick PCR purification kit (Qiagen). DNA fragments resulting from preparative restriction digests were separated by agarose gel electrophoresis, excised from the gel using a razor blade and extracted from the gel using the QIAquick Gel extraction kit (Qiagen). Plasmid DNA was purified from bacteria culture or scraped colonies using the QIAprep spin miniprep kit (Qiagen). Luria–Bertani (LB) liquid media and LB/agar solid media used for growth of *Escherichia coli* transformed with pUC18 (Fermentas) or pBAD/His B (Invitrogen) was supplemented with 0.1 mg mL^{-1} ampicillin (amp). The integrity of all gene sequences was confirmed by dye terminator cycle sequencing using the DYEnamic ET kit (Amersham Biosciences). Sequencing reactions were analyzed at the University of Alberta Molecular Biology Service Unit. The sequences of all primers used for PCR amplifications and sequencing reactions are provided in supplementary Table S1.

Construction of a library of mCherry random insertion variants. The gene for mCherry in pRSET-B (1) was amplified by PCR. The 5′ primer (YLA1) contained *Xba*I and *Kpn*I restriction sites and encoded three amino acids (Gly-Gly-Ser) of an unstructured linker. The 3′ primer (YLB1) contained *Eco*RI and *Kpn*I restriction sites and encoded another three amino acids (Gly-Gly-Thr) of the unstructured linker. The PCR product was digested with *Xba*I and *Eco*RI and ligated into similarly digested pUC18. The ligation reaction was used for the transformation of electrocompetent *E. coli* DH10B (Invitrogen). An appropriate dilution of transformed bacteria was plated on LB/agar/amp. After overnight incubation at 37°C, a single colony was picked and used to inoculate 4 mL of LB/amp. Cultures were grown with shaking overnight at 37°C and the plasmid DNA, designated pUC-YL, was isolated.

A modified pBAD expression vector was prepared using the following procedure. The pBAD/His B vector (Invitrogen) was PCR amplified with a 5′ primer (YLC1) that annealed just after the multiple cloning site (MCS) and a 3′ primer (YLD1) that annealed just before the MCS. The resulting PCR product corresponded to a linear version of the complete pBAD plasmid that lacked the MCS but had new *Kpn*I and *Eco*RI sites appended to its 5′ end and new *Xba*I and *Kpn*I sites appended to its 3′ end. The PCR product was digested with *Kpn*I, purified by agarose gel electrophoresis, and self-ligated with T4 DNA ligase. *E. coli* was transformed with the ligation mixture and the plasmid DNA, designated pBAD-YL, isolated using a procedure identical to that described above.

The Mutation Generation System^{TM} Kit (Finnzymes) was used to construct the insertion mutation library of mCherry variants. Following the manufacturer's protocols, an *in vitro* transposition reaction was performed with 140 ng of pUC-YL and the M1-Kan^R Entranceposon that contains the kanamycin resistance gene. The reaction mixture was diluted 10-fold in deionized water and used to transform *E. coli* DH10B. The transformation reaction was plated on 5 LB/agar/amp plates supplemented with kanamycin (10 μg mL^{-1}) and incubated 14 h at 37°C. Bacterial colonies were scraped from all the plates and plasmid DNA was isolated. The plasmid DNA was digested with *Xba*I and *Eco*RI and the resulting fragments were separated by agarose gel electrophoresis with ethidium bromide staining. The four strongest bands on the gel had sizes of 3.8 kb (2.7 kilobase (kb) vector + 1.1 kb Entranceposon), 2.7 kb (empty vector only), 1.8 kb (0.7 kb mCherry gene + Entranceposon) and 0.7 kb (mCherry gene). The 1.8 kb band was excised, isolated and ligated with appropriately digested pBAD-YL. The ligation reaction was used to transform *E. coli* as described

above. The transformation reaction was diluted into LB/amp and incubated overnight at 37°C with shaking at 225 rpm. Plasmid DNA was isolated from this culture, digested with *Not*I to remove the Kan^R gene, and the resulting fragments separated by agarose gel electrophoresis. The larger 4.8 kb fragment (modified mCherry gene + p-BAD/His B) was excised and the DNA isolated as described above. The isolated DNA fragment was self-ligated to circularize the plasmid and thus provide the library of mCherry variants with random 15 base pair (five codon) insertions.

Library screening. The plasmid library of mCherry with random insertions was used to transform electrocompetent *E. coli* strain DH10B. The transformed bacteria were plated on LB/agar/amp supplemented with L-arabinose (0.02%). Plates were incubated for 14 h at 37°C prior to screening. The system for imaging the fluorescence of bacterial colonies grown on 10 cm Petri dishes is a custom built device that has been described in detail elsewhere (32,33). In brief, the light from a 175 W xenon arc lamp (Sutter) was passed through a 560/40 nm bandpass filter (Chroma) and used to illuminate a 10 cm diameter area. For screening of the library of insertion variants, colony fluorescence on illuminated plates was viewed through a pair of red wraparound goggles that block light of less than ∼580 nm (Lightning Powder Company, Inc.). Following screening of 10 plates with approximately 200 colonies per plate, 100 red fluorescing colonies were picked and used to inoculate separate 5 mL volumes of LB/amp. Cultures were grown overnight and plasmid DNA for each of the 100 clones was purified as described above.

Insertion mapping by PCR or restriction digest. Two different approaches were used to map the approximate location of the 15 bp insertion with the mCherry gene: a PCR method and a restriction digest method. The PCR method is based on the manufacturer's recommended protocol and involves two separate PCR reactions. One reaction amplifies the 5′ end of the gene up to the point of insertion and one reaction amplifies the 3′ end of the gene starting from the point of insertion. The PCR products are then analyzed by agarose gel electrophoresis to determine their relative sizes. We found that this method did not produce a detectable product for a significant fraction of the clones and thus we turned to a more reliable restriction digest method. For each clone three separate 10 μL double restriction digest reactions were performed: one with *Not*I and *Xba*I, one with *Not*I and *Eco*RI and one with *Xba*I and *Eco*RI. The digestion products were analyzed by agarose gel electrophoresis to determine the relative fragment sizes. Clones of interest were subjected to DNA sequencing to explicitly identify the location and sequence of the inserted residues.

Construction of circularly permuted mCherry variants. Circular permuted mCherry variants with new termini at the site of insertion were constructed by a PCR-based method (21). The pUC-YL plasmid containing mCherry was digested with *Kpn*I and the resulting fragments were separated by agarose gel electrophoresis. The fragment containing the mCherry gene was 0.7 kb and had a composition corresponding to *Kpn*I-Gly-Gly-Ser-mCherry2-Gly-*Kpn*I with sticky ends. This fragment was extracted from the gel and self-ligated to give a circular gene with a continuous reading frame. This circular gene served as the template for six individual PCR reactions using the primers with sequences shown in supplementary Table S1. The resulting PCR products encoding the permuted genes were digested sequentially with *Xba*I and *Eco*RI. The digested and purified PCR product was ligated into pBAD-YL that had been digested with the same two restriction enzymes. Plasmids were amplified in *E. coli* and the gene sequences were verified by DNA sequencing.

Protein production and purification. For all protein production, *E. coli* strain LMG194 (Invitrogen) was first transformed with the gene of interest in the pBAD-YL vector by electroporation. Single colonies were used to inoculate 4 mL of LB/amp that was cultured overnight (37°C, 225 rpm) before being diluted into 1 L of LB/amp supplemented with arabinose (0.2%). The culture was allowed to shake overnight (37°C, 225 rpm) before cells were harvested. Alternatively, the culture temperature was decreased to 18°C once an OD600 of 0.6 was reached and the culture allowed to shake for at least another 24 h. Bacteria were pelleted by centrifugation, resuspended in 50 mM Tris–HCl/300 mM NaCl and lysed by a French press. The bacterial lysates were centrifuged at 30 000 *g* for 30 min, and the proteins were purified from the supernatants by Ni-NTA chromatography (Amersham) following the manufacturer's protocol. Proteins were stored at 4°C for several weeks prior to spectral characterization.

Spectroscopy. Absorbance spectra were recorded on a DU-800 UV–visible spectrophotometer (Beckman). Fluorescence emission spectra were acquired on a QuantaMaster spectrofluorometer (Photon Technology International) and were corrected for the instrument response. Quantum yields (Φ) for all mCherry variants were measured using purified mCherry as the reference standard (1). Extinction coefficients (ε) were measured by UV–visible absorbance spectroscopy on purified proteins. To determine the ensemble ε of the total protein (inclusive of polypeptide chains that did not form a red chromophore), the intensity of the 587 nm absorbance peak for each protein was compared to that of a solution of mCherry with matched absorbance at 280 nm. The ε was calculated by multiplying the ratio of absorbance intensities at 587 nm by the ε of mCherry (72 000 $M^{-1}cm^{-1}$) (1). To determine the intrinsic ε for only those polypeptide chains that did form a red chromophore, the alkali denaturation method was employed (1).

Western blot. Single colonies of transformed bacteria expressing the insertion or circular permutation variants were used to inoculate 10 mL of LB/amp supplemented with arabinose (0.2%). The culture was grown overnight with shaking (37°C, 225 rpm) before cells were pelleted by centrifugation. To obtain the soluble fraction, the pellet was suspended in 1 mL B-PER II (Pierce) according to the manufacturer's protocol, repelleted and the supernatant removed (soluble fraction). The pellet was dissolved in 8 M urea (1 mL) to provide the "insoluble" fraction. Equal volumes of both the soluble and insoluble fraction (20 μL) were subject to electrophoresis on a sodium dodecyl sulfate (SDS)-polyacrylamide (12.5%) gel and then electro-blotted onto polyvinylidene difluoride (PVDF) membrane (Millipore). The membrane was blocked in Tris-buffered saline (TBS)-Tween 20 (0.1%) containing 5% nonfat dry milk (Carnation) overnight at 4°C, then probed with anti-His antibody conjugated with horseradish peroxidase (Roche) for 1 h at room temperature with orbital shaking. Detection was performed using the ECL chemiluminescence substrate (Pierce) and BioMax light film (Kodak). The exposed film was digitally scanned and the density of each band were calculated using the Image-Pro Plus 6.0 Software (Media Cybernetics).

RESULTS

mCherry variants with peptide insertions

Identification of variants with peptide insertions. Through the use of transposon-based mutagenesis, we constructed a library of mCherry gene variants that each contained a single consecutive sequence of 15 bps inserted at a random location. To ensure that each gene variant contained a single insertion (*versus* none or more than one), we employed the strategy outlined in Fig. 1. The mCherry gene in pUC18 was used as the template for a transposition reaction with a transposon that carried the kanamycin resistance (Kan^R) gene marker. Only plasmid products with at least one insertion of the transposon would confer resistance to kanamycin following *E. coli* transformation with the reaction mixture. However, due to the random nature of the insertion reaction ∼20% of the surviving clones have the insertion in the mCherry coding sequence while ∼80% have the insertion outside of the mCherry coding sequence. To isolate only those mCherry gene variants with a single insertion we digested the plasmid with restriction enzymes specific for sequences at either end of the mCherry gene. The resulting fragments were separated by agarose gel electrophoresis and the DNA fragment that corresponded to mCherry plus one transposon was isolated. This fragment was inserted back into the expression vector and the Kan^R gene removed to leave behind only a 15 bp insertion. This strategy was essential for the production of a high quality gene library in which the "background" (*i.e.* the plasmid with
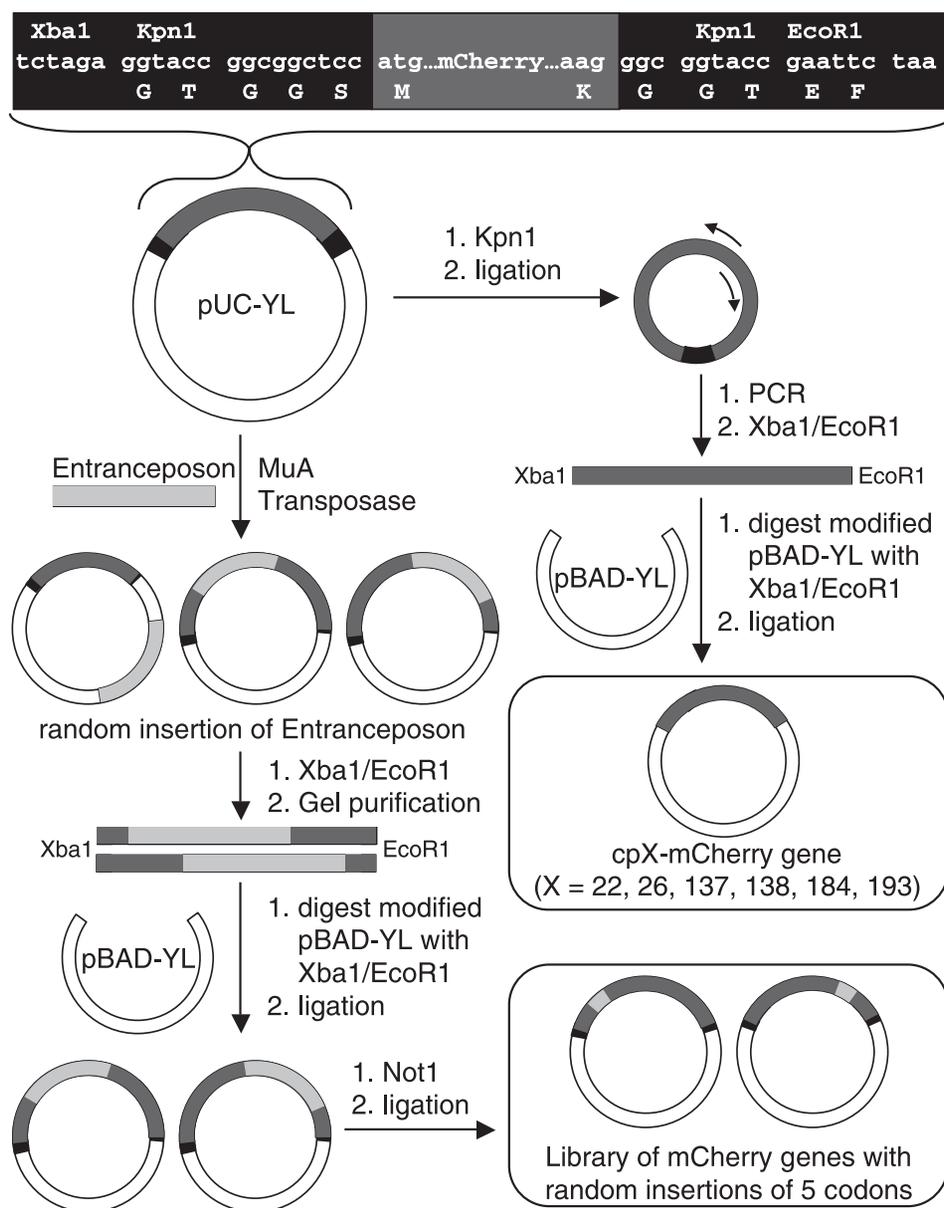
Figure 1. Strategy for generating the library of mCherry insertion variants and individual circularly permuted variants. "Kan[R]" is the gene encoding the kanamycin resistance marker, neomycin phosphotransferase (*npt*). "*Xba*I," "*Kpn*I," "*EcoR*I," and "*Not*I" represent either the DNA substrate for a restriction endonuclease or the endonuclease itself.

no insertion in the mCherry coding region) was minimized. The theoretical size of this library is ∼700 genetically distinct members.

The library of mCherry insertion variants was expressed in *E. coli* and several thousand individual clones, presented in the form of colonies on solid media, were screened for red fluorescence. Of those colonies that fluoresced red, intensities ranged from "bright" (readily observed through red goggles 12 h after plating, comparable to mCherry itself), to "dim" (<50% of the brightness of mCherry 12 h after plating), to "very dim" (barely perceptible 24 h after plating). The bright colonies were attributed to insertions within the 24 residues of unstructured protein sequence present at the N- and C-termini of this version of mCherry. These unstructured residues include eight residues at the

N-terminus (MVSKGEED) and eight residues at the C-terminus (GGMDELYK) that were not observed in the X-ray crystal structure (34), as well as eight extra residues (GTGGS at the N-terminus and GGT at the C-terminus) appended as a linker for creation of circularly permuted variants (Fig. 1). Unlike insertions within the β-barrel of mCherry, insertions in these unstructured portions are unlikely to have a detrimental effect on the ability of the protein to fold and efficiently form the intrinsic red chromophore. The goal of this research was to identify locations within the β-barrel of mCherry that would tolerate insertion or introduction of new termini to form circularly permuted variants, and thus insertions in the unstructured "tails" of the protein were not considered relevant. A total of 100 colonies that had dim red fluorescence were picked

for further investigation and the plasmid DNA purified for each individual clone.

To map the location of the insertion within each mCherry clone picked for further investigation, we employed both a PCR-based method and a restriction digest-based method. The latter proved to be more reliable and was used for the majority of the mapping experiments. Each clone was subject to a restriction digest that excised the portion of the gene between the start of the coding sequence and the point of insertion. A second digest excised the portion of the gene between the point of insertion and the end of the coding sequence. Comparison of the size of these two fragments by agarose gel electrophoresis allowed us to estimate the location of the insertion. All clones that were not background (*i.e.* could not be digested with *Not*I) were found to fall into one of four categories. The first of these categories accounted for the majority of the clones and was defined by one of the two fragments being within ~50 bp of the full-length mCherry gene as determined by agarose gel electrophoresis. These variants presumably contain insertions close to either the N- or C-termini that are partially disruptive with respect to protein folding or the development of the red chromophore. We speculate that these variants have insertions in the stretches of residues that are between the edge of the rigid β-barrel (Phe11 at the N-terminus and Arg220 at the C-terminus) and the beginning of the unstructured portion of termini (Asn4 at the N-terminus and Thr223 at the C-terminus). All other variants fell into one of three distinct categories as defined by the size of the fragments observed after digestion. The approximate fragment sizes were: 0.1 and 0.6 kb, 0.4 and 0.3 kb, and 0.6 and 0.1 kb, for the 5′ and 3′ portions of the gene, respectively. For each of these three categories, the two most brightly fluorescent representatives were subject to DNA sequencing to determine the specific site of insertion. For variants that gave 0.1 and 0.6 kb fragments, the insertions were AAASV between Val22 and Asn23 (i22-mCherry) and DAAAH between His25 and Glu26 (i26-mCherry). For variants that gave 0.4 and 0.3 kb fragments, the insertions were MRPQQ between Gln137 and Lys138 (i137-mCherry) and NAAAQ between Gln137 and Lys138 (i138-mCherry). For variants that gave 0.6 and 0.1 kb fragments, the insertions were NAAAA between Ala183 and Lys184 (i184-mCherry) and CGRTY between Tyr193 and Asn194 (i193-mCherry). The position of all insertions relative to mCherry's secondary structure and tertiary structure is shown in Fig. 2A,B, respectively.

Characterization of insertion variants. All six of the 15 bp insertion variants had fluorescence quantum yields (Table 1) and emission spectra (Fig. 3A) that were practically identical to that of mCherry. This result supports the conclusion that the immediate environment of the chromophore has not been significantly perturbed by the presence of the peptide insertions. One property in which the insertion variants are quite different from mCherry is in their "in colony" fluorescence (Table 1). By imaging of colonies of bacteria on identically treated plates, we could assess the cumulative efficiency of all steps required to produce a functional FP; transcription, translation, folding, and chromophore formation. It is reasonable to suspect that peptide insertion could have dramatic effects on the ability of the protein to fold or the chromophore to develop. Indeed, each of our six insertion
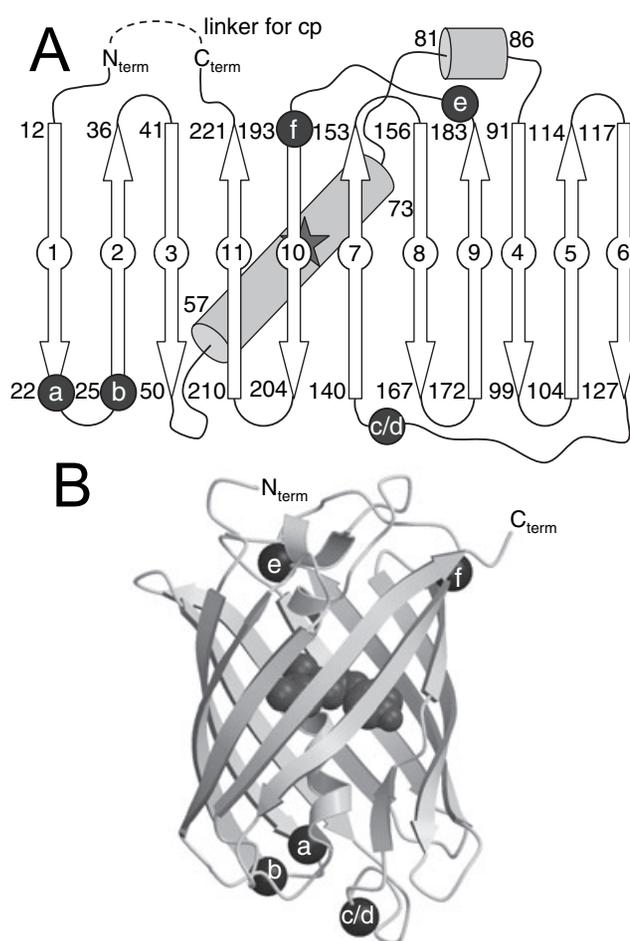

Figure 2. Locations of five amino acid insertions in mCherry variants. (A) Schematic representation of the secondary structure elements of mCherry with locations of insertions represented as labeled blue spheres. The approximate position of the chromophore is represented with a red star. Residue numbers for the start and end of secondary structure elements are taken from Protein Data Bank (PDB) entry 2H5Q (34). Helices are represented as grey cylinders and β-strands are represented as white arrows with consecutive numbering provided in a central white circle. Blue circles correspond to the point of insertion for: (a) i22-mCherry, (b) i26-mCherry, (c) i137-mCherry, (d) i138-mCherry, (e) i184-mCherry and (f) i193-mCherry. (B) Cartoon representation of the structure of mCherry (PDB ID 2H5Q) (34) with positions of peptide insertions labeled as in (A). The chromophore is shown in a red space-filling representation.

variants were decreased in "in colony" fluorescent brightness (to varying extents) relative to mCherry suggesting that the expression level, protein folding efficiency or chromophore maturation efficiency for each variant has been adversely affected (Table 1). To further investigate the reason for the decreased brightness of the insertion variants, we performed Western blots against the soluble and insoluble protein fractions isolated from bacteria (Fig. 4). This experiment revealed that all of the insertion variants were expressed at substantially decreased levels (7–23%) relative to mCherry itself. Furthermore, a large majority of i22-mCherry, and a significant fraction of i37-mCherry, is present in inclusion bodies.

In addition to the poor expression and folding of the insertion variants, the diminished ε values (Table 1) also

Table 1. Properties of mCherry insertion variants.

	Relative "in colony" fluorescence (%)*	Φ	ε (M^{-1} cm^{-1})	Brightness† (mM^{-1} cm^{-1})	(%)
mCherry	100	0.22	72 000	16	100
i22-mCherry	4	0.20	28 000	6	35
i26-mCherry	7	0.20	38 000	8	48
i137-mCherry	19	0.21	33 000	7	44
i138-mCherry	11	0.18	40 000	7	45
i184-mCherry	33	0.22	45 000	10	63
i193-mCherry	23	0.22	44 000	10	61

*All measurements are of red fluorescence. Fluorescence measured after 24 h at 37°C. †Brightness is defined as the product of Φ and ε as determined for the purified protein (38).
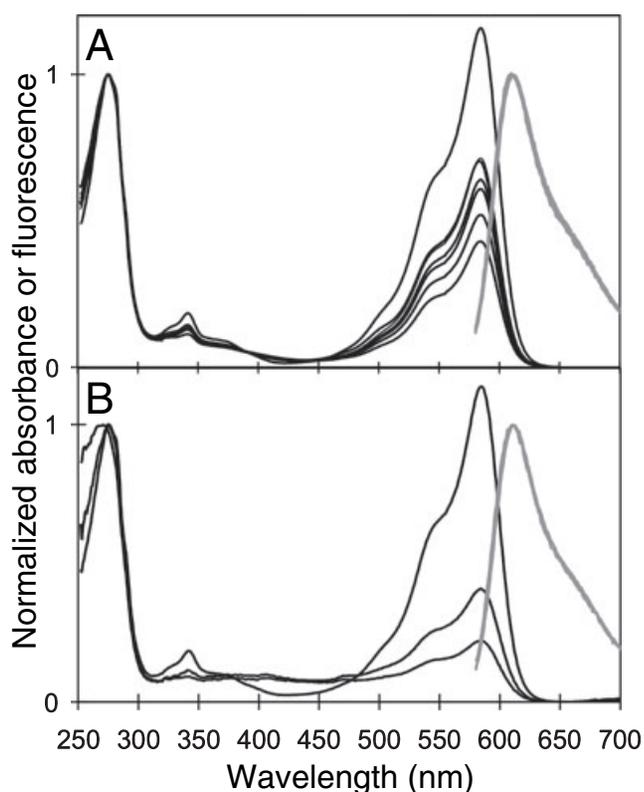

Figure 3. Absorbance and emission spectra for insertion (A) and circular permutation (B) variants. (A) Absorbance spectra (black lines) of insertion mCherry variants in order of highest to lowest intensity at 587 nm: mCherry (included for comparison), i184-mCherry, i193-mCherry, i138-mCherry, i26-mCherry, i137-mCherry and i22-mCherry. Absorbance spectra have been normalized at 280 nm. The emission spectra (grey lines) for all variants have been normalized to a maximum intensity of 1. (B) Absorbance spectra (black lines) of circular permutation mCherry variants that exhibited significant absorbance at wavelengths > 500 nm in order of highest to lowest intensity at 587 nm: mCherry (included for comparison), cp184-mCherry and cp22-mCherry. Emission spectra are represented as in (A). All other circular permutation variants showed no significant absorbance in the visible region, though weak red fluorescence was apparent for the purified protein. No significant shifts in either absorbance or emission maxima were observed.

contribute to the decreased "in colony" brightness. It is important to note that these ε values are an ensemble average and are based on the assumption that all protein molecules are equivalent. When we compared values of ε before and after
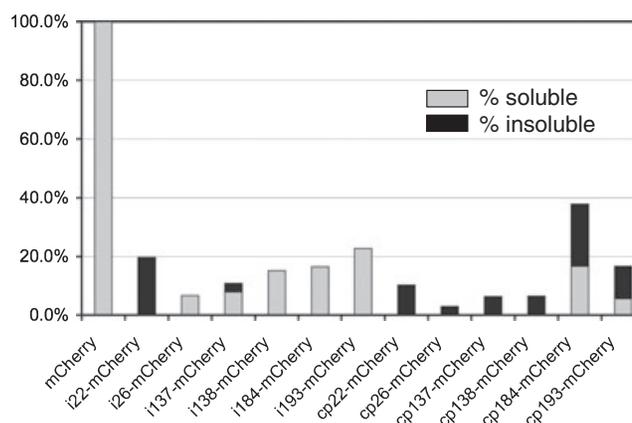

Figure 4. Assessment of the expression levels and amount of soluble *versus* insoluble protein for each variant. Western blots were performed against both the soluble and insoluble fractions from bacteria expressing each of the insertion and circular permutation constructs. Band intensities were integrated and plotted as percentage of total mCherry protein expressed under identical conditions (experimentally determined to be 99.97% soluble and 0.03% insoluble). It is not apparent from this graph, but i22- and cp22-mCherry had 0.5% and 0.04% soluble protein, respectively.

alkali denaturation (data not shown), we found that the ratio of $\varepsilon_{\text{native}}$ to $\varepsilon_{\text{denatured}}$ for all variants was identical to that of mCherry. We interpret this result as evidence that the "purified" soluble proteins are actually mixtures of a RFP (with brightness very similar to that of mCherry) and a non-fluorescent FP. This non-fluorescent FP may be a misfolded conformer that is incapable of promoting the series of post-translational modifications necessary for chromophore formation. Such a scenario is consistent with the observation that the quantum yields and emission peaks are also practically identical to mCherry. The corollary is that the relative concentration of these two components determines the ensemble ε.

Circularly permuted variants of mCherry

Design of circularly permuted variants. We hypothesized that sites within mCherry that were tolerant of peptide insertions would also be suitable for introduction of new termini in a circularly permuted variant. To test this hypothesis we used a previously reported method (21) to construct a series of circularly permuted variants in which new termini were introduced at each of the insertion locations identified by

library screening. A circular version of the mCherry gene with a continuous reading frame and a suitable linker joining the ends of the gene was constructed (Fig. 1). PCR amplification of the circular template using appropriately designed primers provided the target genes encoding circularly permuted variants designated cpX-mCherry where X = 22, 26, 137, 138, 184, 193 (Fig. 5). In each case, the primers were designed such that three residues of mCherry located at the site of insertion were duplicated at both the N- and C-termini of the circular permutation protein variant. For example, the circularly permuted variant based on the i22-mCherry insertion began with residue Ser21 at its N-terminus and ended with Asn23 at its C-terminus (Fig. 5). We speculated that conformational constraints would favor one of the two duplicated tripeptide sequences (*i.e.* residues 21, 22 and 23) to adopt a native-like conformation and thus it seemed prudent to include both possibilities.

Characterization of circularly permuted variants. The six circular permutation mCherry variants were constructed and expressed in *E. coli*. Bacterial colonies expressing each of these variants were subjected to fluorescence imaging to determine if the proteins retained the ability to fluoresce. Following incubation at 37°C for 24 h, no fluorescence above the background autofluorescence of *E. coli* colonies was detected for any of the six variants (Table 2). However, after incubating at 4°C for a further 72 h colonies expressing cp184-mCherry exhibited strong red fluorescence. Under the same conditions,
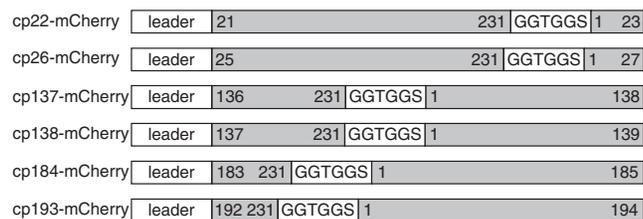
colonies expressing cp22- and cp193-mCherry were weakly fluorescent and colonies expressing cp26-, cp137- and cp138-mCherry were no more fluorescent than colonies of untransformed *E. coli*.

When we attempted to purify and further characterize the six circular permutation variants, we found that only the two purified variants that had exhibited the brightest "in colony" fluorescence, cp22- and cp184-mCherry, gave solutions with measurable absorbance at 587 nm. For the other four purified variants the intensity of the 587 absorbance peak was not significantly greater than the baseline noise of our spectrophotometer. However, it was possible to detect weak red fluorescence at 610 nm from the four variants that did not exhibit a significant absorbance peak. Due to the lack of a significant absorbance peak, we were unable to measure either Φ or ε for these weakly fluorescent variants. We expect that, as with the insertion variants, each of the circular permutation variants exists as a mixture of a bright RFP and a non-fluorescent FP. This conclusion is supported by the observation that Φ (Table 2) and the ratio of ε_{native} to $\varepsilon_{denatured}$ (data not shown) for cp22- and cp184-mCherry are essentially identical to that of mCherry.

Analysis of protein expression levels and solubility by Western blot revealed that for the circularly permuted variants, the majority of expressed protein was targeted to inclusion bodies in *E. coli* (Fig. 4). In fact, in the case of cp26-, cp137- and cp138-mCherry, no soluble protein could be detected by Western blot. It is apparent that the circular permutation variants are significantly more destabilized than the analogous insertion variants.

Figure 5. Primary structure of all circular permutation mCherry variants. The grey portion of the bar represents the primary sequence of mCherry (1) between the indicated residues. The "leader" includes the His$_6$ affinity purification tag and enterokinase cleavage site of pBAD/His B (MGGSHHHHHHGMASMTGGQQMGRDLY-DDDDKDPSSR). The "GGTGGS" region is the six residue linker between the original N- and C-termini (Fig. 2A).

DISCUSSION

We identified six variants of mCherry with genetically distinct insertions of five-residue peptides. The positions of the insertions localize to three distinct regions of mCherry: the loop between strands 1 and 2, the loop between strands 6 and 7, and the loop between strands 9 and 10 (Fig. 2). Of these three regions of the protein, the loop between strands 9 and 10 is the most tolerant of both insertions and the introduction of new termini to create a circularly permuted variant. Interestingly, previous efforts in screening of random permutations of *Aequorea* GFP (7) did not identify this loop as a position that tolerated introduction of new termini.

Table 2. Properties of circular permutation mCherry variants.

	Relative "in colony" fluorescence (%)*		Φ	ε (M^{-1} cm^{-1})	Brightness	
	24 h at 37°C	72 h at 4°C			(mM^{-1} cm^{-1})	(%)
mCherry	100	100	0.22	72 000	16	100
cp22-mCherry	<0.5†	3	0.21	15 000	3	20
cp26-mCherry	<0.5	<0.5	ND‡	ND	ND	ND
cp137-mCherry	<0.5	<0.5	ND	ND	ND	ND
cp138-mCherry	<0.5	<0.5	ND	ND	ND	ND
cp184-mCherry	<0.5	18	0.22	26 600	6	37
cp193-mCherry	<0.5	1.0	ND	ND	ND	ND

*All measurements are of red fluorescence. †This lower threshold is determined by the autofluorescence of colonies of *E. coli* when imaged under identical conditions. ‡Not determined. Weak red fluorescence was detected but no significant absorbance peak was observed for the purified protein at 587 nm.

The loops between strands 1 and 2 and strands 6 and 7 of mCherry were significantly less permissive than the loop between strands 9 and 10 with respect to peptide insertions and the introduction of new termini in circularly permuted variants. Insertion in the loops between either strands 1 and 2 or strands 6 and 7 caused the fluorescent brightness to decrease to less than 50% of mCherry. Circularly permuted variants in which new termini were introduced in the loops between either strands 1 and 2 or strands 6 and 7 had, for the most part, greatly diminished ability to fold and form the red chromophore. The one exception was the cp22-mCherry variant that retained 20% of the intrinsic brightness of mCherry and has new termini between strands 1 and 2. Previous work on screening of random circular permutations of *Aequorea* GFP did not identify this loop as a position that tolerates new termini (7). The loop between strands 6 and 7 and the N-terminal half of strand 7 is one of the regions of *Aequorea* GFP that best tolerates insertions (7–9) as well as new termini in circularly permuted variants (7,21). Indeed, screening of libraries of random circularly permuted variants resulted in the identification of a variant with new termini at Glu142 of GFP, a position only three residues removed from the site of insertion in the i137- and i138-mCherry variants in primary sequence alignments (7). Likewise, systematic screening of circularly permuted variants of GFP identified Asn144/Tyr145 as a location where new termini could be introduced with retention of bright green fluorescence (21). In contrast, we have found that while peptide insertions in the loop between strands 6 and 7 of mCherry are reasonably well tolerated, introduction of new termini completely abolishes the ability of protein to correctly fold and form the intrinsic red chromophore in *E. coli*. In addition, we did not identify any insertions in the β-strands of the protein. Taken together, our results provide strong support for the conclusion that the locations of new termini that are tolerated in circularly permuted *Aequorea* GFP variants cannot be directly translated to mCherry or other mFruit-type variants.

Our strategy for identifying positions within mCherry that can accept new termini in circularly permuted variants was a test of the hypothesis that a site that is permissive towards peptide insertion should also be permissive towards the introduction of new termini. We conclude that the ability of a particular site to accept a peptide insertion does not guarantee that the same site can accept new termini in a circularly permuted variant. Each of the six circularly permuted variants that were constructed in this work was significantly destabilized relative to the corresponding insertion variant. On the other hand, the fact that two of six attempted circularly permuted variants retained 20% or more of the brightness of mCherry shows that the sites that did accept peptide insertions are more likely to accept new termini than a site within the protein chosen at random. From this entirely pragmatic perspective, the strategy of using insertional mutagenesis to find candidate sites for new termini in circularly permuted variants was quite effective. There remains a good possibility that screening of libraries of circular permutation mCherry variants would result in the identification of additional sites within the protein that tolerated the new termini; though these same sites would probably not tolerate insertions. Another assumption in this work is that the linker between the original N- and C-termini did not significantly interfere with protein folding or chromophore maturation. The distance between the first (Asn4) and last residue (Thr223) resolved in the X-ray crystal structure of mCherry is 26 Å (34). In our circularly permuted variants Asn4 and Thr223 are connected by a 24 residue linker and based on previous work with circular permutation GFP variants (35), this linker should be compatible with normal protein folding and function. The fact that cp184-mCherry retains significant fluorescent brightness provides further support for this assumption.

In the context of new fluorescent probes for use in live cell imaging, what are the prospects for the mCherry variants reported in this work? We see the most direct and promising future application of these variants as being the construction of genetically encoded red fluorescent Ca^{2+} sensors. For example, insertion of calmodulin into one of the sites identified in this work could potentially result in Ca^{2+} sensor of the Camgaroo-type (7). For the circularly permuted variants that retained significant fluorescent brightness (cp184- and, to a lesser extent, cp22-mCherry), fusion of interacting protein pairs such as calmodulin and the M13 peptide to the new N- and C-termini may lead to the creation of Ca^{2+} sensors of the pericam (8) and G-CaMP (9) type. A particularly exciting possibility for single FP sensors based on mFruit variants is that they could in principle be intrinsically green/red ratiometric. Another possible application of these variants is in guiding the construction of "split" RFPs for use in bimolecular fragment complementation (BiFC) (31). In recent years BiFC with split versions of *Aequorea* GFP (and its variants) has emerged as a versatile technique for the detection of protein–protein interactions in live cells (36).

As discussed above, circular permutation FP variants have recently found utility as donor or acceptor FPs in efforts to empirically optimize the response of FRET-based sensor constructs. To date, the performance of the mFruit-type variants, such as mCherry, in FRET-based sensors has been disappointing. It seems as though the modest Φ of these variants result in sensitized emission that is less than typically observed with the well-established CFP/YFP FRET pair. For these reasons, even the brightest of the circularly permuted variants described in this work, which suffer from decreased brightness due to poor folding of chromophore maturation affinity relative to mCherry, are not recommended for use in FRET-type constructs. However, we are optimistic that future mFruit variants with improved Φ and further improvements in folding efficiency will have better tolerance of insertion and circular permutation. It has previously been established that FPs engineered for higher folding efficiency are more tolerant of circular permutation (23,37). Alternatively, continued directed evolution of the circularly permuted variants described herein may lead to brighter and more stable versions that will serve as the basis for red fluorescent sensors and as donors or acceptors in FRET constructs that are spectrally orthogonal to the CFP and YFP pair.

REFERENCES

1. Shaner, N. C., R. E. Campbell, P. A. Steinbach, B. N. Giepmans, A. E. Palmer and R. Y. Tsien (2004) Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. *Nat. Biotechnol.* **22**, 1567–1572.
2. Lukyanov, K. A., D. M. Chudakov, S. Lukyanov and V. V. Verkhusha (2005) Innovation: Photoactivatable fluorescent proteins. *Nat. Rev. Mol. Cell. Biol.* **6**, 885–891.
3. Tsien, R. Y. (1998) The green fluorescent protein. *Annu. Rev. Biochem.* **67**, 509–544.
4. Zhang, J., R. E. Campbell, A. Y. Ting and R. Y. Tsien (2002) Creating new fluorescent probes for cell biology. *Nat. Rev. Mol. Cell Biol.* **3**, 906–918.
5. Lippincott-Schwartz, J. and G. H. Patterson (2003) Development and use of fluorescent protein markers in living cells. *Science* **300**, 87–91.
6. Wallrabe, H. and A. Periasamy (2005) Imaging protein molecules using FRET and FLIM microscopy. *Curr. Opin. Biotechnol.* **16**, 19–27.
7. Baird, G. S., D. A. Zacharias and R. Y. Tsien (1999) Circular permutation and receptor insertion within green fluorescent proteins. *Proc. Natl Acad. Sci. USA* **96**, 11241–11246.
8. Nagai, T., A. Sawano, E. S. Park and A. Miyawaki (2001) Circularly permuted green fluorescent proteins engineered to sense Ca^{2+}. *Proc. Natl Acad. Sci. USA* **98**, 3197–3202.
9. Nakai, J., M. Ohkura and K. Imoto (2001) A high signal-to-noise Ca(2+) probe composed of a single green fluorescent protein. *Nat. Biotechnol.* **19**, 137–141.
10. Kawai, Y., M. Sato and Y. Umezawa (2004) Single color fluorescent indicators of protein phosphorylation for multicolor imaging of intracellular signal flow dynamics. *Anal. Chem.* **76**, 6144–6149.
11. Doi, N. and H. Yanagawa (1999) Design of generic biosensors based on green fluorescent proteins with allosteric sites by directed evolution. *FEBS Lett.* **453**, 305–307.
12. Miyawaki, A., J. Llopis, R. Heim, J. M. McCaffery, J. A. Adams, M. Ikura and R. Y. Tsien (1997) Fluorescent indicators for Ca^{2+} based on green fluorescent proteins and calmodulin. *Nature* **388**, 882–887.
13. Ting, A. Y., K. H. Kain, R. L. Klemke and R. Y. Tsien (2001) Genetically encoded fluorescent reporters of protein tyrosine kinase activities in living cells. *Proc. Natl Acad. Sci. USA* **98**, 15003–15008.
14. Zhang, J., Y. Ma, S. S. Taylor and R. Y. Tsien (2001) Genetically encoded reporters of protein kinase A activity reveal impact of substrate tethering. *Proc. Natl Acad. Sci. USA* **98**, 14997–15002.
15. Lin, C. W., C. Y. Jao and A. Y. Ting (2004) Genetically encoded fluorescent reporters of histone methylation in living cells. *J. Am. Chem. Soc.* **126**, 5982–5983.
16. Ananthanarayanan, B., Q. Ni and J. Zhang (2005) Signal propagation from membrane messengers to nuclear effectors revealed by reporters of phosphoinositide dynamics and Akt activity. *Proc. Natl Acad. Sci. USA* **102**, 15081–15086.
17. Ostermeier, M. (2005) Engineering allosteric protein switches by domain insertion. *Protein Eng. Des. Sel.* **18**, 359–364.
18. Ormo, M., A. B. Cubitt, K. Kallio, L. A. Gross, R. Y. Tsien and S. J. Remington (1996) Crystal structure of the *Aequorea victoria* green fluorescent protein. *Science* **273**, 1392–1395.
19. Yang, F., L. G. Moss and G. N. Phillips Jr. (1996) The molecular structure of green fluorescent protein. *Nat. Biotechnol.* **14**, 1246–1251.
20. Belousov, V. V., A. F. Fradkov, K. A. Lukyanov, D. B. Staroverov, K. S. Shakhbazov, A. V. Terskikh and S. Lukyanov (2006) Genetically encoded fluorescent indicator for intracellular hydrogen peroxide. *Nat. Methods* **3**, 281–286.
21. Topell, S., J. Hennecke and R. Glockshuber (1999) Circularly permuted variants of the green fluorescent protein. *FEBS Lett.* **457**, 283–289.
22. Chiang, J. J., I. Li and K. Truong (2006) Creation of circularly permutated yellow fluorescent proteins using fluorescence screening and a tandem fusion template. *Biotechnol. Lett.* **28**, 471–475.
23. Zapata-Hommer, O. and O. Griesbeck (2003) Efficiently folding and circularly permuted variants of the Sapphire mutant of GFP. *BMC Biotechnol.* **3**, 5.
24. Nagai, T., S. Yamada, T. Tominaga, M. Ichikawa and A. Miyawaki (2004) Expanded dynamic range of fluorescent indicators for Ca(2+) by circularly permuted yellow fluorescent proteins. *Proc. Natl Acad. Sci. USA* **101**, 10554–10559.
25. Allen, M. D. and J. Zhang (2006) Subcellular dynamics of protein kinase A activity visualized by FRET-based reporters. *Biochem. Biophys. Res. Commun.* **348**, 716–721.
26. Mank, M., D. F. Reiff, N. Heim, M. W. Friedrich, A. Borst and O. Griesbeck (2006) A FRET-based calcium biosensor with fast signal kinetics and high fluorescence change. *Biophys. J.* **90**, 1790–1796.
27. Matz, M. V., A. F. Fradkov, Y. A. Labas, A. P. Savitsky, A. G. Zaraisky, M. L. Markelov and S. A. Lukyanov (1999) Fluorescent proteins from nonbioluminescent Anthozoa species. *Nat. Biotechnol.* **17**, 969–973.
28. Baird, G. S., D. A. Zacharias and R. Y. Tsien (2000) Biochemistry, mutagenesis, and oligomerization of DsRed, a red fluorescent protein from coral. *Proc. Natl Acad. Sci. USA* **97**, 11984–11989.
29. Campbell, R. E., O. Tour, A. E. Palmer, P. A. Steinbach, G. S. Baird, D. A. Zacharias and R. Y. Tsien (2002) A monomeric red fluorescent protein. *Proc. Natl Acad. Sci. USA* **99**, 7877–7882.
30. Wang, L., W. C. Jackson, P. A. Steinbach and R. Y. Tsien (2004) Evolution of new nonantibody proteins via iterative somatic hypermutation. *Proc. Natl Acad. Sci. USA* **101**, 16745–16749.
31. Jach, G., M. Pesch, K. Richter, S. Frings and J. F. Uhrig (2006) An improved mRFP1 adds red to bimolecular fluorescence complementation. *Nat. Methods* **3**, 597–600.
32. Ai, H. W., J. N. Henderson, S. J. Remington and R. E. Campbell (2006) Directed evolution of a monomeric, bright and photostable version of *Clavularia* cyan fluorescent protein: Structural characterization and applications in fluorescence imaging. *Biochem. J.* **400**, 531–540.
33. Cheng, Z. and R. E. Campbell (2006) Assessing the structural stability of designed beta-hairpin peptides in the cytoplasm of live cells. *Chembiochem* **7**, 1147–1150.
34. Shu, X., N. C. Shaner, C. A. Yarbrough, R. Y. Tsien and S. J. Remington (2006) Novel chromophores and buried charges control color in mFruits. *Biochemistry* **45**, 9639–9647.
35. Akemann, W., C. D. Raj and T. Knopfel (2001) Functional characterization of permuted enhanced green fluorescent proteins comprising varying linker peptides. *Photochem. Photobiol.* **74**, 356–363.
36. Kerppola, T. K. (2006) Visualization of molecular interactions by fluorescence complementation. *Nat. Rev. Mol. Cell. Biol.* **7**, 449–456.
37. Pedelacq, J. D., S. Cabantous, T. Tran, T. C. Terwilliger and G. S. Waldo (2006) Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* **24**, 79–88.
38. Shaner, N. C., P. A. Steinbach and R. Y. Tsien (2005) A guide to choosing fluorescent proteins. *Nat. Methods* **2**, 905–909.

SUPPLEMENTAL MATERIALS

The following supplemental material is available for this article:

Table S1. Sequences of all primers used for molecular biology procedures.

This material is available as part of the online article from: http://www.blackwell-synergy.com/doi/full/10.1111/j.1751-1097.2007.00206.x