THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 283, NO. 43, pp. 28980–28987, October 24, 2008
Printed in the U.S.A.

# A Crystallographic Study of Bright Far-Red Fluorescent Protein mKate Reveals pH-induced *cis-trans* Isomerization of the Chromophore*[S]

Received for publication, January 23, 2008, and in revised form, June 17, 2008 Published, JBC Papers in Press, August 4, 2008, DOI 10.1074/jbc.M800599200

**Sergei Pletnev[‡§1], Dmitry Shcherbo[¶], Dmitry M. Chudakov[¶], Nadezhda Pletneva[¶], Ekaterina M. Merzlyak[‖], Alexander Wlodawer**, Zbigniew Dauter[‡], and Vladimir Pletnev[¶]**

*From the [‡]Macromolecular Crystallography Laboratory, NCI, National Institutes of Health, Argonne, Illinois 60439, the [§]Basic Research Program, SAIC-Frederick Inc., Argonne, Illinois 60439, the [¶]Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Science, 117997 Moscow, Miklukho-Maklaya 16/10, Russia, [‖]Evrogen JSC, 117997 Moscow, Miklukho-Maklaya 16/10, Russia, and the **Macromolecular Crystallography Laboratory, NCI-Frederick, National Institutes of Health, Frederick, Maryland 21702*

**The far-red fluorescent protein mKate ($\lambda_{ex}$, 588 nm; $\lambda_{em}$, 635 nm; chromophore-forming triad Met[63]-Tyr[64]-Gly[65]), originating from wild-type red fluorescent progenitor eqFP578 (sea anemone *Entacmaea quadricolor*), is monomeric and characterized by the pronounced pH dependence of fluorescence, relatively high brightness, and high photostability. The protein has been crystallized at a pH ranging from 2 to 9 in three space groups, and four structures have been determined by x-ray crystallography at the resolution of 1.75–2.6 Å. The pH-dependent fluorescence of mKate has been shown to be due to reversible *cis-trans* isomerization of the chromophore phenolic ring. In the non-fluorescent state at pH 2.0, the chromophore of mKate is in the *trans*-isomeric form. The weakly fluorescent state of the protein at pH 4.2 is characterized by a mixture of *trans* and *cis* isomers. The chromophore in a highly fluorescent state at pH 7.0/9.0 adopts the *cis* form. Three key residues, Ser[143], Leu[174], and Arg[197] residing in the vicinity of the chromophore, have been identified as being primarily responsible for the far-red shift in the spectra. A group of residues consisting of Val[93], Arg[122], Glu[155], Arg[157], Asp[159], His[169], Ile[171], Asn[173], Val[192], Tyr[194], and Val[216], are most likely responsible for the observed monomeric state of the protein in solution.**

Green fluorescent proteins (GFP)[2] and GFP-like proteins (FP) have become important noninvasive tools for visualization and monitoring of the internal processes within cells or whole organisms, such as gene expression, monitoring the cellular pH, ion concentration, embryogenesis, inflammatory processes, tracking protein trafficking, the migration of parasites within a host, *etc* (1–13). Fluorescent proteins can be used to visualize many types of cancer processes, including primary tumor growth, tumor cell motility and invasion, metastatic seeding and colonization, angiogenesis, and interactions between the tumor and its host microenvironment (14–16). FPs might be very useful in real-time testing of the efficacy of cancer drugs in animal models of human cancer.

The extensive spectral diversity of fluorescent proteins arises mostly from variations in the chemical structure of the mature chromophore and in the stereochemistry of its adjacent environment. The FP chromophore forms autocatalytically *in vivo* and *in vitro* from three residues, Xxx-Tyr-Gly, without need for any cofactors or enzymes, except for molecular oxygen (17). In most cases, the post-translational modification results in a blue/green emitting state, characterized by formation of an imidazolinone heterocycle with a *p*-hydroxybenzylidene substituent. Often, the reaction chain propagates further with formation of an additional *N*-acylimine double bond, which extends the conjugation of the chromophore $\pi$ electronic system and results in a bathochromic shift in spectra (18–22).

Proteins that emit red, and especially far-red light, are of particular interest (13). The longer wavelength light extends the range of fluorescence resonance energy transfer (FRET)-based applications and causes fewer damaging events to proteins and DNA because of its lower energy. The most favorable "optical window" for the visualization in living tissues is ~650–1100 nm (23). Light with wavelength longer than 1100 nm is absorbed by water. Detection of fluorescence from proteins with emission peaks much shorter than 650 nm encounters the problem of interfering cellular autofluorescence. At present the brightest red fluorescent proteins have emission maxima too far from the preferred "optical window." Besides, their excitation maxima are located in a range 550–560 nm, where living tissues are almost opaque and fluorescence of these proteins cannot be effectively excited (see Table 1 in Ref. 13). Recently, far-red fluorescent variants, HcRed, mPlum, and AQ143, reaching the 650 nm barrier, have been developed (24–26).

* This work was supported, in whole or in part, by the National Institutes of Health Intramural Research Program of the Center for Cancer Research, NCI, and with Federal funds from NCI Contract No. NO1-CO-12400. This work was also supported by the Russian Foundation for Basic Research (Grants 07-04-00054 and 07-04-12189), Russian Federal Agency for Science and Innovations (Grant 02.513.12.3013). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

*The atomic coordinates and structure factors (codes 3BX9, 3BXA, and 3BXB) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).*

[S] The on-line version of this article (available at http://www.jbc.org) contains supplemental "Experimental Procedures" and Fig. S1.

[1] To whom correspondence should be addressed: SAIC-Frederick Inc., Basic Research Program, Argonne, IL 60439. E-mail: svp@ncifcrf.gov.

[2] The abbreviations used are: GFP, green fluorescent protein; PEG, polyethylene glycol, PDB, Protein Data Bank; R.M.S.D., root mean-squared deviation.

However, these proteins are characterized by low brightness, strongly limiting their practical application.

A majority of the wild-type GFP-like proteins form tetramers, complicating their practical use. To be the most useful tools for practical applications, the designed biomarkers should preferably exist in a monomeric form, emit in the far-red fluorescent spectral range, have high brightness, be photostable, and exhibit a high rate of chromophore maturation. It has always been difficult to develop variants meeting all of these criteria simultaneously. Recently, however, the bright far-red dimeric variant Katushka and its monomeric version mKate, both characterized by a chromophore-forming sequence, $Met^{63}$-$Tyr^{64}$-$Gly^{65}$, have been successfully designed (13). Both proteins, characterized by similar spectral properties ($\lambda_{ex}$, 588 nm; $\lambda_{em}$, 635 nm), are derived from wild-type red fluorescent progenitor eqFP578 (*Entacmaea quadricolor*), the latter having spectral maxima $\lambda_{ex}$, 552 nm; $\lambda_{em}$, 578 nm. Both Katushka and mKate are significantly brighter than the spectrally close HcRed or mPlum (24, 25) and display fast maturation, as well as high pH- and photo-stability. The fluorescence of Katushka and mKate is pH-dependent, showing maximum emission at pH ~8, which gradually diminishes to zero at pH ~4. Compared with other far-red FPs, Katushka exhibits evident superiority for visualization in living tissues. The monomeric mKate is an excellent fluorescent label for monitoring fused proteins in whole organisms, multicolor labeling, and FRET applications.

We present here the results of crystallographic studies of three-dimensional structures of the far-red variant mKate in different fluorescent states, corresponding to pH 2.0, 4.2, and 7.0. These structures were solved at resolution ranging from 1.75 to 2.6 Å. We have analyzed in detail the stereochemical features in the chromophore area that are responsible for the outstanding spectral characteristics of the protein and its relatively high pH dependence of fluorescence, as well as the nature of surface residues responsible for the predominantly monomeric state of the protein.

## EXPERIMENTAL PROCEDURES

The details of cloning, purification, and characterization of the studied proteins are presented in supplemental "Experimental Procedures."

*Crystallization, Structure Solution, and Crystallographic Refinement*—Crystals have been obtained by the hanging drop vapor diffusion method in four different conditions. The mKate_pH2.0 crystals have appeared from 20% w/v PEG 3350, 0.2 M ammonium citrate tribasic, 0.4 M citric acid, pH 2.0, initial protein concentration 12 mg/ml. The mKate_pH4.2 was crystallized from 17.5% (w/v) PEG 3350, 0.07 M citric acid, pH 4.2, initial protein concentration 15 mg/ml. Crystals of mKate_pH7.0 and mKate_pH9.0 were grown from 20% w/v PEG6000, 1 M LiCl, 0.1 M HEPES pH 7.0, and 20% w/v PEG6000, 1 M LiCl, 0.1 M Bicine pH 9.0, respectively. Initial protein concentration for both conditions was 18 mg/ml.

X-ray diffraction data were collected from single crystals flash-cooled in a 100 K nitrogen stream. Prior to cooling, the crystals were transferred to a cryo-protecting solution containing 20% glycerol and 80% reservoir solution. Data were collected with a MAR300 CCD detector at the SER-CAT beamline 22ID (Advanced Photon Source, Argonne National Laboratory, Argonne, IL) and were processed with *HKL2000* (27).

Crystal structure of mKate at pH 4.2 was solved by the molecular replacement method with *MOLREP* (28, 29), using the coordinates of the eqFP611 monomer without the chromophore (71% sequence identity, PDB ID: 1UIS; (19)). The refined coordinates of mKate at pH 4.2 were used to solve the other mKate structures at pH 2.0, 7.0, and 9.0. Structure refinement was performed with *REFMAC5* (30) and *PHENIX* (31), alternating with manual revision of the model using *COOT* (32). Water molecules were located with *ARP/wARP* (33). Non-crystallographic symmetry restraints were applied in refinement of mKate_pH7.0 and mKate_pH9.0 structures with eight subunits in asymmetric unit. The occupancy of each chromophore state was set to reach the best possible agreement between the model and difference electron density map. Crystallographic data and refinement statistics are presented in Table 1. Although the values of $R_{merge}$ were relatively high in the outermost shells of all data sets, the corresponding values of $I/\sigma(I)$ indicated that these data were still significant.

Structure validation was performed with *PROCHECK* (34). The coordinates and structure factors were deposited in the Protein Data Bank under accession codes 3BX9 (mKate_pH2.0), 3BXA (mKate_pH4.2), 3BXB (mKate_pH7.0). The structure at pH 9.0 (accession code 3BXC) was found almost identical to the structure at pH 7.0 and was skipped from discussion.

## RESULTS AND DISCUSSION

*Electron Density Interpretation*—The asymmetric unit in mKate_pH2.0 and mKate_pH4.2 crystals contains one dimer. Crystallographic symmetry operations transform the dimers to the corresponding tetramers. The asymmetric unit in mKate_pH7.0 possesses two tetramers. The electron density for all structures allowed unambiguous fitting of residues 2/3/4–228 for all monomers in the asymmetric unit. No density was observed for the N-terminal His tag fragment introduced into the expressed construct and used for protein purification. The relatively high resolution ~1.8 Å of the low pH structures enabled us to detect alternative stable conformations for a number of side chains. We located between 259 and 351 hydrogen-bonded water molecules in the asymmetric unit of each crystal. Several citric acid and glycerol molecules (the components of the crystallization and cryoprotectant solutions) were located in the mKate_pH2.0 and mKate_pH4.2 structures.

*Monomer Structure*—The principal structural fold of the mKate is an 11-stranded $\beta$-barrel, closed from both sides by loop caps, with a chromophore (matured from the sequence $Met^{63}$-$Tyr^{64}$-$Gly^{65}$) embedded in the middle of an internal $\alpha$-helix that is wound along the $\beta$-barrel axis. The C-terminal tail 222–228 has irregular conformation and goes away from the $\beta$-barrel body. The R.M.S.D. values from pairwise superposition of the mKate monomer structures corresponding to different pH are within the range of 0.32–0.41 Å for all equivalent $C^{\alpha}$ atoms, indicating a very similar fold of the monomers. Two *cis* peptide bonds preceding $Pro^{50}$ and $Pro^{85}$ in the loop area have been detected. Similarly to TurboGFP and the FPs from *Zoanthus* (22, 35), the $\beta$-barrel frame of mKate shows the presence of a pore, formed by the backbone of $Trp^{140}$, $Glu1^{41}$,

**TABLE 1**
**Crystallographic data and refinement statistics**

| Protein | mKate pH 2.0 (PDB_ID: 3BX9) | mKate pH 4.2 (PDB_ID: 3BXA) | mKate pH 7.0 (PDB_ID: 3BXB) |
|---|---|---|---|
| **Crystallographic data** | | | |
| Space group | $P6_522$ | I4 | $P2_1$ |
| Cell dimensions (Å, °) | $a, b = 67.9, c = 413.0$ | $a, b = 98.2, c = 106.5$ | $a = 74.2, b = 105.0, c = 123.0$ $\beta = 105.8$ |
| Z/(Z′) | 24 (2) | 16 (2) | 16 (8) |
| Estimated solvent content (%) | 51 | 47 | 41 |
| Temperature (K) | 100 | 100 | 100 |
| Wavelength (Å) | 1.00 | 1.00 | 1.00 |
| Resolution range (Å) | 30.0–1.80 (1.86–1.80)[a] | 30.0–1.75 (1.81–1.75)[a] | 30.0–2.60 (2.69–2.60)[a] |
| Total reflections measured | 762,295 | 387,816 | 210,733 |
| Unique reflections observed (F > 0) | 53,955 | 50,567 | 55,386 |
| Redundancy | 14.1 (14.4)[a] | 7.7 (7.6)[a] | 3.8 (3.4)[a] |
| I/σ (I) | 46.1 (5.6)[a] | 43.9 (3.7)[a] | 11.6 (2.1)[a] |
| $R_{merge}$ | 0.061 (0.556)[a] | 0.041 (0.522)[a] | 0.122 (0.506)[a] |
| Completeness | 99.9 (99.7)[a] | 99.6 (99.1)[a] | 99.9 (99.4)[a] |
| **Refinement statistics** | | | |
| Non-H atoms in model | | | |
| Protein | 3,586 [2 x (2–228) res] | 3,574 [2 x (3–228) res] | 14,224 [8 x (4–228) res] |
| Water | 259 | 302 | 351 |
| Citrate | 78 (6 mol) | 13 (1mol) | |
| Glycerol | 12 (2 mol) | | |
| $R_{work}$ | 0.204 (97.9%)[b] | 0.180 (97.9%)[b] | 0.170 (94.8%)[b] |
| $R_{free}$ | 0.247 (2.1%)[b] | 0.232 (2.1%)[b] | 0.263 (5.2%)[b] |
| **Mean B factor/(R.M.S.D.) (Å²)** | | | |
| Protein atoms | | | |
| Main chain | 25.5 (0.8) | 26.4 (0.7) | 39.1 (0.5) |
| Side chain | 28.3 (1.8) | 28.9 (1.8) | 39.4 (1.0) |
| Chromophore | 23.4 (2.8) | 23.4 (2.2) | 37.1 (1.5) |
| **Ramachandran statistics (%)(for non-Gly/Pro residues)** | | | |
| Most favorable/additional allowed | 91.1/8.9 | 91.3/8.7 | 87.4/12.6 |
| Generously allowed/disallowed regions | 0.0/0.0 | 0.0/0.0 | 0.0/0.0 |

[a] Values for the data in the highest resolution shell.
[b] Percent of the data reserved for working and free sets.

Ala1[42], Arg[197], Arg[198], and Leu[199], leading to the hydroxyphenyl moiety of the chromophore. A chain of hydrogen-bonded water molecules, going through the pore from the outside, could be identified in the mKate_pH2.0 and mKate_pH4.2 structures. Evdokimov *et al.* (35) suggested that this pore is essential for chromophore maturation, providing access for molecular oxygen.

*Monomer Association*—According to gel filtration data, mKate exists in solution in the monomeric state at concentration as high as 10 mg/ml (13). However, in the crystalline state, which corresponds to a much higher protein concentration, mKate adopts at all pH values tetrameric arrangement with 222 symmetry, typically seen in GFP-like proteins. The interacting surfaces of the subunits create two types of interfaces. Interface IF1 is located between two antiparallel monomers that form an "antiparallel" dimer (A-dimer), whereas the IF2 interface is found between two monomers belonging to adjacent A-dimers. Those monomers positioned at ~75° with respect to each other form a "crossed" dimer (C-dimer) (Table 2). The irregular C-terminal tail, consisting of residues 222–228, goes away from the β-barrel and sticks to cylindrical surface of the interacting counterpart, contributing to the IF2-contacting surface.

The tetrameric assemblies of mKate in the crystal forms grown at different pH have similar topology but exhibit significant packing differences. The pairwise three-dimensional superposition of the tetramers for all equivalent $C^{\alpha}$ atoms gives the following values of R.M.S.D. for corresponding pairs: mKate_pH7.0 and mKate_pH2.0; 1.85 Å, mKate_pH4.2 and mKate_pH2.0; 2.61 Å, mKate_pH4.2 and mKate_pH7.0; 4.30 Å.

**TABLE 2**
**Characteristics of the tetramer interfaces**

| mKate_pH2.0 | | mKate_pH4.2 | | mKate_pH7.0 | |
|---|---|---|---|---|---|
| **Intratetramer contact area (Å²)[a]** | | | | | |
| IF1 | IF2 | IF1 | IF2 | IF1 | IF2 |
| 690 | 1500 | 300 | 1490 | 930 | 1490 |
| **Interface-stabilizing H-bonds/salt-bridges (within 3.3 Å)** | | | | | |
| 4 | 17 | 10 | 23 | 13 | 25 |
| **Angle between β-barrel axes in an A/C-dimer[b] (°)** | | | | | |
| A | C | A | C | A | C |
| 31 | 75 | 41 | 74 | 23 | 72 |

[a] Half of the buried surface area at the dimer interface corresponding to a radius of the probe solvent molecule 1.4 Å.
[b] The axis goes through $C^{\alpha}$ atoms of residues 60 and 67 of the central α-helix.

The IF1 interface within the A-dimer is noticeably weaker, compared with the IF2 interface within the C-dimer. In the structures reported here, IF1 exhibits significant variation in its contact area, the number and composition of stabilizing interactions, and the angle between the antiparallel β-barrel axes (Table 2). In all crystal forms, the IF2 interfaces are more extensive and more uniform than the IF1 interfaces.

Although the monomeric mKate and the dimeric Katushka have similar spectral characteristics ($\lambda_{ex}$, 588 nm, $\lambda_{em}$, 635 nm; (13)), a comparison of their primary structures shows 18 differences (Fig. 1). Three of them, corresponding to positions 93, 122, and 155 are situated at the interface IF1, whereas eight differences at positions 157, 159, 169, 171, 173, 192, 194, and 216 are found at the IF2 interface. At least some of these differences must be responsible for the observed variation of the oligomeric states of these proteins in solu-

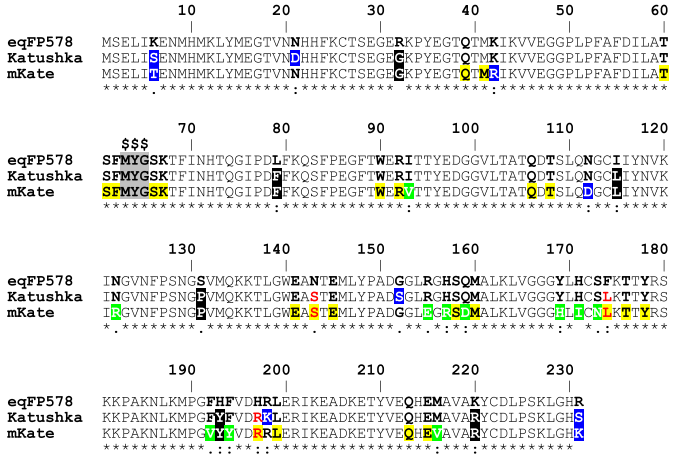


FIGURE 1. **Sequence alignment of the wild-type eqFP578 and its mutant far-red variants, Katushka and mKate.** Residue positions in three-dimensional structure: $, chromophore forming residues. *Black* and *red* font (highlighted in *yellow*), chromophore nearest environment (see Figs. 4 and 5). *Red*, chromophore environment; apparently responsible for the far-red shift. *White* font (highlighted in *black*), β-barrel caps area; presumably silent mutations or covariant with those in *red*. *White* font (highlighted in *green*), crystal intratetramer interfaces; apparently responsible for monomerization in solution. *White* font (highlighted in *blue*), presumably silent mutations in random mutagenesis process.

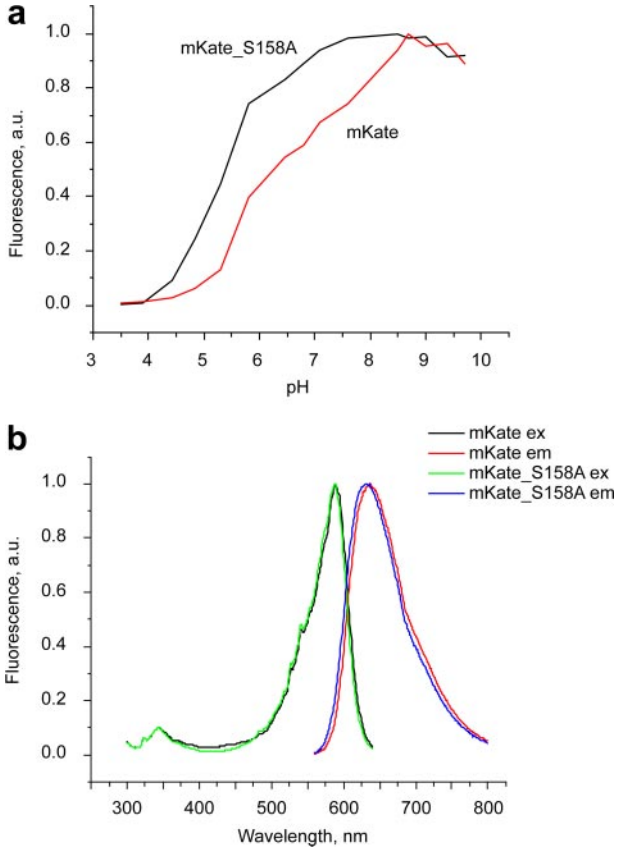


FIGURE 2. **mKate and rationally designed structure-based variant mKate_S158A.** *a*, pH dependence of the fluorescence; *b*, fluorescence spectra.

tion. These 11 interface positions (highlighted in *green* for mKate in Fig. 1) are occupied by identical amino acids in Katushka and its wild-type progenitor eqFP578, both of which form dimers in solution (13, 36).

*pH-induced cis-trans Isomerization of the Chromophore*—The spectral properties of the far-red fluorescent protein mKate have been comprehensively investigated by Shcherbo *et al.* (13). The protein exhibits maximum emission at pH ~8, gradually disappearing at pH ~4 (Fig. 2*a*). In mKate the post-translational modification of the chromophore-forming sequence $Met^{63}$-$Tyr^{64}$-$Gly^{65}$ results in a conventional GFP two ring-conjugated core consisting of a five-membered imidazolinone heterocycle with a *p*-hydroxybenzylidene substituent. Similar to other red and far-red fluorescent proteins (18–20, 37, 38), the first chromophore residue $Met^{63}$ in mKate is characterized by formation of an *N*-acylimine partially double bond, $N{=}C^{\alpha}$, the $sp^2$ hybridization of the corresponding $C^{\alpha}$ atom, and the *cis* configuration of the preceding peptide bond. An additional *N*-acylimine bond apparently extends the chromophore-conjugated π electronic system, resulting in a bathochromic shift in spectra.

The unique feature of mKate, revealed by this study, is the observed pH-induced cis-*trans* isomerization of the chromophore $Tyr^{64}$ phenolic ring with respect to the $C^{\alpha}$-N bond. The predominant *trans* conformation of the phenolic ring (Fig. 3*A*) was detected in the mKate_pH2.0 crystal structure that corresponds to the non-fluorescent (dark) state (Fig. 2). The difference electron density indicates the presence of ~10% of the *cis* isomer. In contrast, the highly fluorescent (bright) state of the mKate_pH7.0 structure is characterized by a mostly *cis* conformation of the phenolic ring, with ~10% contamination by *trans* isomer in four out of eight independent subunits (Fig. 3*C*). The structure of mKate_pH4.2, exhibiting a low level of fluorescence, shows the presence of both the *trans* and *cis* isomers in a ratio ~60% to ~40% in subunit *A* (Fig. 3*B*) and ~80% to ~20% in subunit *B*.

Both the *trans* and *cis* forms of the chromophore, representing the dark and the bright states, respectively, exhibit noticeable distortion from coplanarity of the imidazolinone and phenolic rings. In two subunits of the mKate_pH2.0 crystal structure with the *trans* chromophore, the values of $\chi^1$ and $\chi^2$ torsion angles around the $C^{\alpha}{=}C^{\beta}$ and $C^{\beta}$-$C^{\gamma}$ bonds of the tyrosine are relatively low, ~173° (7° deviation from the ideal planar form) and ~15° respectively. In eight subunits of the mKate_pH7.0 structure with a *cis* chromophore, these angles are ~1° and 25 ± 5°, respectively. In FPs, both chromophore rings generally are more coplanar in the *cis* than in the *trans* arrangement (Table 3*B* in Ref. 39). Non-coplanar ring arrangement was mostly associated with the non-fluorescent state. However, the observed coplanarity of the *trans* chromophore in eqFP611 (19) and the non-coplanarity of the *cis* chromophore in the fluorescent variant of Rtms5 (40) show that this is not always the case. In mKate, the non-coplanar *cis* chromophore (with a relatively large value of $\chi^2$), exhibits high fluorescence. We suggest that the energy difference between nonplanar and planar chromophore conformations is small and could be overcome by light excitation. In other words, the nonplanar conformation of the *cis* chromophore observed in the crystals presumably corresponds to the resting state, which at small energy expense may be transformed to planar conformation in the excited fluorescent state. The bond angle $C^{\alpha}$-$C^{\beta}$-$C^{\gamma}$ of $Tyr^{64}$

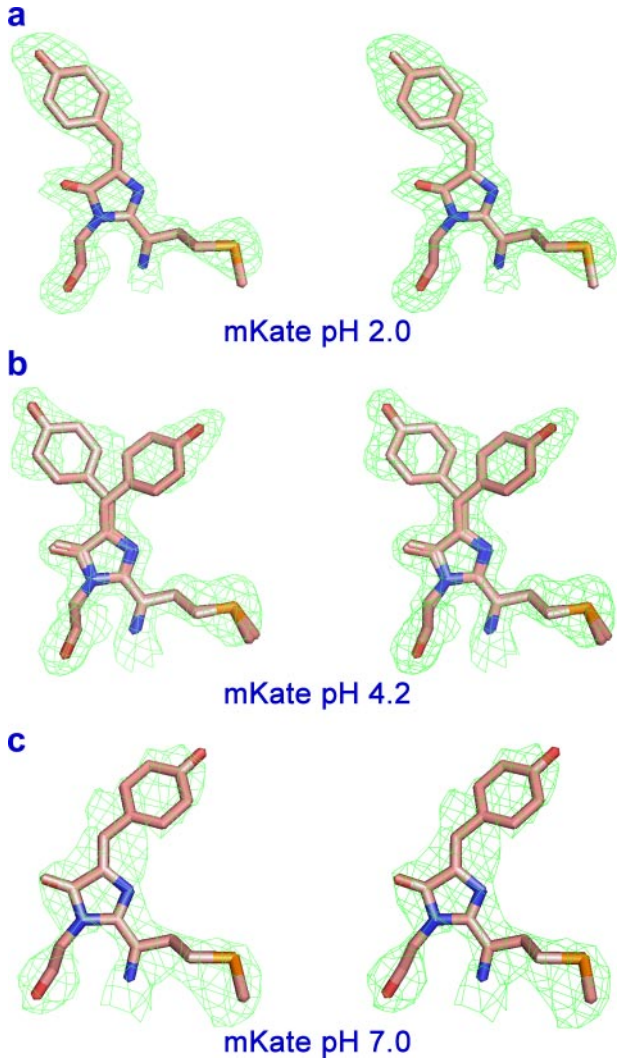


FIGURE 3. **A stereoview of the chromophores in a difference Fo-Fc electron density (cutoff $\rho = 3.0\sigma$).** This was calculated with the phases omitting the contribution of the chromophore in the mKate crystal structures at pH 2.0 (*a*), pH 4.2 (subunit A) (*b*), pH 7.0 (*c*). This figure was produced with *PYMOL* (44).

observed in the *trans* mKate chromophore is 5° larger than that in the *cis* chromophore (~135° *versus* ~130°). This difference, as well as the deviation from planarity of the *trans* chromophore, apparently arises from the steric repulsion between $C^{\delta}$ and carbonyl O atoms of $Tyr^{64}$.

Different geometric restraint schemes were tested to determine the optimal geometry of the group (63)$C^{\alpha}$=N-C(O)-$C^{\alpha}$(62) bridging the $C^{\alpha}$ atoms of the first chromophore residue $Met^{63}$ and the preceding $Phe^{62}$. Similar to HcRed (38), it exhibits, at optimal fit to electron density, considerable deviation from planarity with $\omega$ torsion angle around the quasi-peptide N-C(O) bond in a range 20–35°. Moreover, similarly to other red and far-red fluorescent proteins (18, 19, 22, 38), the C(O)-N-$C^{\alpha}$ bond angle of the linkage in both chromophore isomers is strongly linearized, in the range of 140–160°. The nature of such unusual geometry of the linkage is not clear. It may be assumed that steric tension in the central $\alpha$-helix that, according to a hypothesis of Barondeau *et al.* (17), drives chromophore formation, might leave in the mature structure some remnant strain, partially responsible for the observed effect. The equilibrium strain relaxation is presumably achieved at the expense of the compromised distortion of the linkage preceding the chromophore.

The chromophore in the *trans* and *cis* conformations makes three direct H-bonds (≤3.3 Å) with the side chains of the protein, three H-bonds with water molecules (each water molecule mediates H-bonding with two residues), and the respective 106 and 93 van der Waals contacts (≤3.9 Å) (Fig. 4). Two of the three direct H-bonds are formed by the carbonyls of $Tyr^{64}$ and $Gly^{65}$ of the chromophore, interacting with the side chains of $Arg^{92}$ and $Trp^{90}$, respectively. The third H-bond is formed between the hydroxyl of $Tyr^{64}$ and the side chains of either $Ser^{158}$ or $Ser^{143}$ in the *trans* or *cis* conformational states, respectively.

The consensus part of the chromophore nearest shell in the mKate_pH2.0 and mKate_pH7.0 structures is composed of 17 residues, most of which are involved in an extensive H-bond network formed by the side chain and backbone interactions (Fig. 5). Among them are the catalytic residues, $Glu^{215}$, $Arg^{92}$, and $Thr^{60}$. Five proximal waters (presumed reaction products of the maturation process) are actively involved in forming the network-mediating residue interactions. The hydrogen-bonded network interacts with the chromophore and is apparently functionally important, creating a potential proton wire in the maturation process.

Surprisingly, the positions of the imidazolinone rings, as well as the three-dimensional arrangements of the side chains in the chromophore environment, are practically identical in mKate_pH2.0 and mKate_pH7.0. In both the *trans* and *cis* states the phenolic ring almost

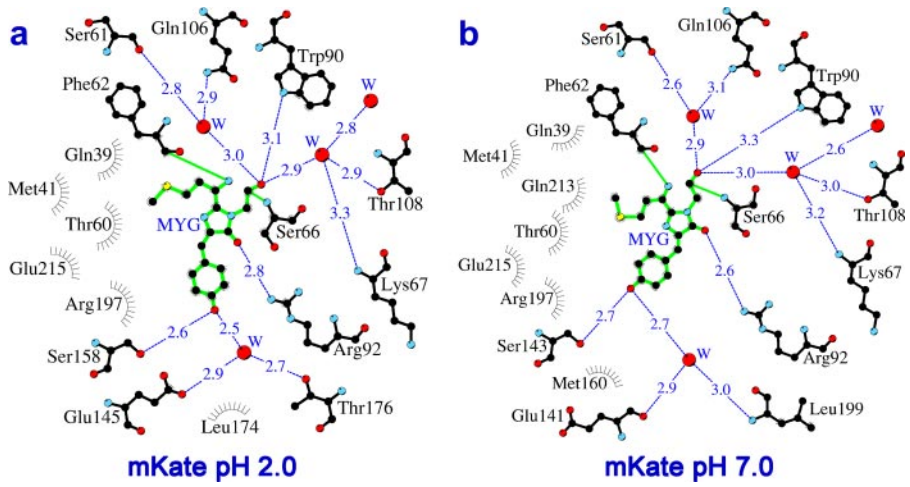


FIGURE 4. **A schematic diagram illustrating the environment of the *trans* chromophore in the mKate_pH2.0 structure (*a*) and the *cis* chromophore in the mKate_pH7.0 structure (*b*).** Hydrogen bonds (≤3.3 Å) are shown as *blue dashed lines*, waters (*W*) as *red spheres*, and van der Waals contacts (≤3.9 Å) as black "eyelashes." This figure was produced with *LIGPLOT/HBPLUS* (45, 46).

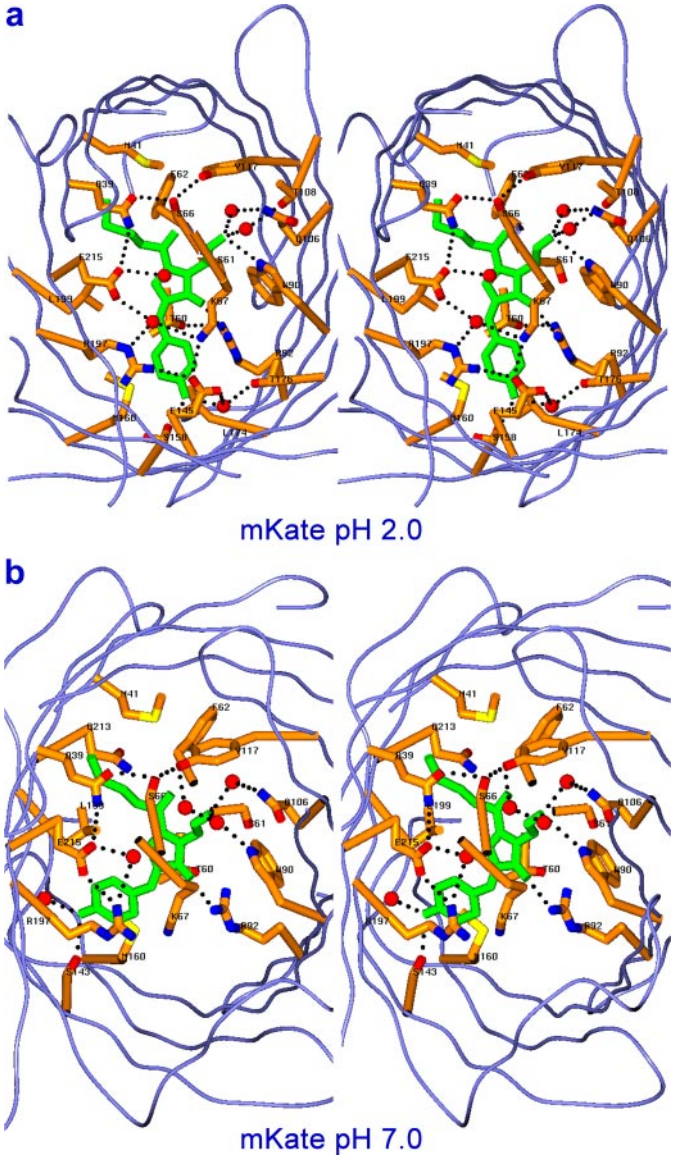


FIGURE 5. **Stereoview of the H-bond network in the vicinity of the *trans* chromophore (shown in *green*) in the mKate_pH2.0 structure (*a*) and the *cis* chromophore in the mKate_pH7.0 structure (*b*).** Mediating waters are shown by *red spheres*. This figure was produced with *SETOR* (47).

does not disturb the chromophore environment. The chromophore isomerization mostly affects the conformational state of Arg$^{197}$. Two subunits from each tetramer in mKate_pH7.0 crystal structure with the *cis* chromophore present two alternative orientations of the Arg$^{197}$ side chain. The first orientation is generally similar to the one found in mKate_pH2.0 with the *trans* chromophore and mediates H-bonding between the side chains of Ser$^{158}$ and Ser$^{143}$. Those residues were found to be important in stabilization of respective *trans* and *cis* conformational states of the phenolic ring. The second orientation of the Arg$^{197}$ side chain is completely different and mediates the connection between the Glu$^{145}$ and Glu$^{215}$ side chains by two salt bridges. The observed alternative orientations of the Arg$^{197}$ side chain presumably correspond to its different protonation states. In the mKate_pH2.0 the side chain of Ser$^{158}$ also adopts two alternative orientations. The first orientation is identical to that in mKate_pH7.0 and the second one provides H-bonding with the hydroxyl of the chromophore phenolic ring in the *trans* isomeric form.

As expected, the hydroxyphenyl moiety in the *trans* and *cis* isomeric forms exhibits different interactions within the interior of the β-barrel. The *trans-cis* isomerization results in the replacement of the H-bonding of the chromophore tyrosine hydroxyl with the Ser$^{158}$ side chains and via water with the Glu$^{145}$ and Thr$^{176}$ side chains by H-bonding with Ser$^{143}$ side chain and via water with the backbone of Glu$^{141}$ and Leu$^{199}$ (Fig. 4).

In the bright fluorescent state of mKate the phenolic ring of the *cis* chromophore occupies a local pocket bordered by the side chains of six residues: Thr$^{60}$, Ser$^{143}$, Met$^{160}$, Arg$^{197}$, Leu$^{199}$, and Glu$^{215}$ (Fig. 5*B*). The positions occupied by Ser$^{143}$ and Arg$^{197}$ in mKate are filled in wild-type progenitor eqFP578 by Asn and His, respectively (36), whereas the other residues are invariant. Ser$^{143}$ appears to be very important for stabilizing the *cis* conformation of the chromophore.

In the dark state, the phenolic ring of the *trans* chromophore moves to an adjacent pocket composed of the side chains of nine residues: Thr$^{60}$, Lys$^{67}$, Arg$^{92}$, Glu$^{145}$, Ser$^{158}$, Met$^{160}$, Leu$^{174}$, Tyr$^{178}$, and Arg$^{197}$ (Fig. 5*A*). Two of these residues, Leu$^{174}$ and Arg$^{197}$, are Phe and His, respectively, in wt eqFP578. Besides stabilizing H-bonds with Ser$^{158}$ and the nearest water molecule, the *trans* phenolic ring makes stacking interactions with pH-susceptible guanidinium group of Arg$^{197}$. The position of Arg$^{197}$ is fixed by H-bonds with Ser$^{143}$ and with another pH-susceptible residue, Glu$^{145}$. Ser$^{158}$ is critically responsible for the observed pH dependence of the fluorescence.

An NMR study of the relationship between *cis-trans* isomerization and the protonation state of a synthetic chromophore model in solution demonstrated that the free energy gain of the *cis* form over *trans* form is on the order of a single hydrogen bond (41). In the protein, the immediate chromophore environment provides an additional contribution that influences the *cis-trans* equilibrium and selectively stabilizes one form or the other. The change of the protonation state of mKate upon pH variation influences the electrostatic field in the β-barrel interior, apparently affecting the hydrogen bond system by gain or loss of H-bonds, which might trigger chromophore isomerization.

As mentioned above, the far-red fluorescent proteins, monomeric mKate and dimeric Katushka, have similar spectral characteristics (13, 36). Our structural results suggest that three key residues, Ser$^{143}$, Leu$^{174}$, and Arg$^{197}$ (shown in *red* in the sequence alignment, Fig. 1) residing in the vicinity of the chromophore in mKate (Figs. 4 and 5) and apparently in Katushka and differing from those their progenitor eqFP578, are primarily responsible for the far-red shift. The other six positions in mKate and Katushka that differ from those in eqFP578 (highlighted in *blue* in Fig. 1) reside in the cap area of the β-barrel. The corresponding mutations are presumably silent or covariant with those directly responsible the far-red spectral shift.

*Improved Variant mKate_S158A*—Our structural results suggested that replacement of Ser$^{158}$ by a hydrophobic residue of an appropriate size would cause partial destabilization of the dark *trans* state of the chromophore, shifting the equilibrium toward the bright *cis* state and thus increasing the fluorescence

**TABLE 3**
**Spectral characteristics of mKate and mKate_S158A**

| Protein | Excitation maximum | Emission maximum | QY | EC, at ex max | Relative brightness[a] | p$K_a$ |
|---|---|---|---|---|---|---|
| | | | | $M^{-1}\ cm^{-1}$ | | |
| mKate | 588 | 635 | 0.33 | 45,000 | 0.45 | 6.2 |
| mKate_S158A | 588 | 633 | 0.40 | 70,000 | 0.85 | 5.3 |

[a] Brightness is calculated as a product of the molar extinction coefficient and the fluorescence quantum yield and is given by comparison to the brightness of EGFP.

power at low pH. The mKate_S158A variant was prepared, and we compared its spectral and biochemical characteristics with those of mKate (Fig. 2 and Table 3). mKate_S158A is a bright fluorescent protein with $\lambda_{ex}$, 588 nm; $\lambda_{em}$, 633 nm; *i.e.* emission spectra is slightly blue-shifted, compared with mKate (Fig. 2*b*). As expected, this variant is characterized by substantially higher pH stability (Fig. 2*a*), with p$K_a$ = 5.3, compared with 6.2 of mKate. The mKate_S158A is characterized by an essentially higher molar extinction and also by a higher fluorescence quantum yield, resulting in almost 2-fold brighter fluorescence, as well as by a higher rate of chromophore maturation. In contrast to mKate, at physiological pH 7.0–7.5 it demonstrates no kindling effect upon irradiation by excitation light, in agreement with the *trans-cis* isomerization hypothesis of kindling effect (42, 43). At this pH, characterized by relatively low level of cumulative protonation, the equilibrium state of the mKate chromophore has approximately ~10% fraction of the *trans*-form. Photoactivation of mKate by light irradiation causes switching of the remaining dark *trans*- form to the fluorescent *cis*-form, resulting in gradual increase of the intensity at fluorescence maximum by an additional ~10%. Reverse transformation was found to be less favorable. At physiological pH the chromophore of mKate_158A is, presumably, in ~100% *cis*-form, thus irradiation does not cause the kindling effect. In addition, the photostability of mKate_S158A is comparable to that of the highly photostable mKate, making it an excellent fluorescent tag for labeling of fusion proteins in the far-red part of the visible spectrum.

*Acknowledgments—We acknowledge the use of beamline 22-ID of the Southeast Regional Collaborative Access Team (SER-CAT), located at the Advanced Photon Source, Argonne National Laboratory. Use of the APS was supported by the United States Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. W-31-109-Eng-38. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does the mention of trade names, commercial products or organizations imply endorsement by the U.S. Government.*

## REFERENCES


1. (1998) *Green Fluorescent Protein: Properties, Applications, and Protocols*, pp. 120–171, Wiley-Liss, New York
2. Tsien, R. Y. (1998) *Annu. Rev. Biochem.* **67,** 509–544
3. Haugwitz, M., Dery, O., Turpin, P., and Fang, Y. (2003) *Genet. Eng. News* **23,** 36–39
4. Zubova, N. N., and Savitsky, A. P. (2005) *Uspekhi Biol. Khim. (Rus)* **45,** 1–66
5. Verkhusha, V. V., and Lukyanov, K. A. (2004) *Nat. Biotechnol.* **22,** 289–296
6. Chudakov, D. M., Lukyanov, S., and Lukyanov, K. A. (2005) *Trends Biotechnol.* **23,** 605–613
7. Tallini, Y. N., Ohkura, M., Choi, B. R., Ji, G., Imoto, K., Doran, R., Lee, J., Plan, P., Wilson, J., Xin, H. B., Sanbe, A., Gulick, J., Mathai, J., Robbins, J., Salama, G., Nakai, J., and Kotlikoff, M. I. (2006) *Proc. Natl. Acad. Sci. U. S. A.* **103,** 4753–4758
8. Remington, S. J. (2006) *Curr. Opin. Struct. Biol.* **16,** 714–721
9. Passamaneck, Y. J., Di, G. A., Papaioannou, V. E., and Hadjantonakis, A. K. (2006) *Microsc. Res. Tech.* **69,** 160–167
10. Stewart, C. N., Jr. (2006) *Trends Biotechnol.* **24,** 155–162
11. Seitz, G., Warmann, S. W., Fuchs, J., Mau-Holzmann, U. A., Ruck, P., Heitmann, H., Hoffman, R. M., Mahrt, J., Muller, G. A., and Wessels, J. T. (2006) *J. Pediatr. Surg.* **41,** 1369–1376
12. Wacker, S. A., Oswald, F., Wiedenmann, J., and Knochel, W. (2007) *Dev. Dyn.* **236,** 473–480
13. Shcherbo, D., Merzlyak, E. M., Chepurnykh, T. V., Fradkov, A. F., Ermakova, G. V., Solovieva, E. A., Lukyanov, K. A., Bogdanova, E. A., Zaraisky, A. G., Lukyanov, S., and Chudakov, D. M. (2007) *Nat. Methods* **4,** 741–746
14. Hoffman, R. M. (2005) *J. Biomed. Opt.* **10,** 41202
15. Hoffman, R. M. (2005) *Curr. Top. Dev. Biol.* **70,** 121–144
16. Hoffman, R. M. (2005) *Nat. Rev. Cancer* **5,** 796–806
17. Barondeau, D. P., Putnam, C. D., Kassmann, C. J., Tainer, J. A., and Getzoff, E. D. (2003) *Proc. Natl. Acad. Sci. U. S. A.* **100,** 12111–12116
18. Yarbrough, D., Wachter, R. M., Kallio, K., Matz, M. V., and Remington, S. J. (2001) *Proc. Natl. Acad. Sci. U. S. A.* **98,** 462–467
19. Petersen, J., Wilmann, P. G., Beddoe, T., Oakley, A. J., Devenish, R. J., Prescott, M., and Rossjohn, J. (2003) *J. Biol. Chem.* **278,** 44626–44631
20. Tubbs, J. L., Tainer, J. A., and Getzoff, E. D. (2005) *Biochemistry* **44,** 9833–9840
21. Pakhomov, A. A., Pletneva, N. V., Balashova, T. A., and Martynov, V. I. (2006) *Biochemistry* **45,** 7256–7264
22. Pletneva, N., Pletnev, V., Tikhonova, T., Pakhomov, A. A., Popov, V., Martynov, V. I., Wlodawer, A., Dauter, Z., and Pletnev, S. (2007) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **63,** 1082–1093
23. Konig, K. (2000) *J. Microsc.* **200,** 83–104
24. Gurskaya, N. G., Fradkov, A. F., Terskikh, A., Matz, M. V., Labas, Y. A., Martynov, V. I., Yanushevich, Y. G., Lukyanov, K. A., and Lukyanov, S. A. (2001) *FEBS Lett.* **507,** 16–20
25. Wang, L., Jackson, W. C., Steinbach, P. A., and Tsien, R. Y. (2004) *Proc. Natl. Acad. Sci. U. S. A.* **101,** 16745–16749
26. Shkrob, M. A., Yanushevich, Y. G., Chudakov, D. M., Gurskaya, N. G., Labas, Y. A., Poponov, S. Y., Mudrik, N. N., Lukyanov, S., and Lukyanov, K. A. (2005) *Biochem. J.* **392,** 649–654
27. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* **276,** 307–326
28. Collaborative Computational Project No. 4 (1994) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **50,** 760–763
29. Vagin, A., and Teplyakov, A. (1997) *J. Appl. Crystallogr.* **30,** 1022–1025
30. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **53,** 240–255
31. Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K., and Terwilliger, T. C. (2002) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **58,** 1948–1954
32. Emsley, P., and Cowtan, K. (2004) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60,** 2126–2132
33. Perrakis, A., Sixma, T. K., Wilson, K. S., and Lamzin, V. S. (1997) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **53,** 448–455
34. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr.* **26,** 283–291
35. Evdokimov, A. G., Pokross, M. E., Egorov, N. S., Zaraisky, A. G., Yampolsky, I. V., Merzlyak, E. M., Shkoporov, A. N., Sander, I., Lukyanov, K. A., and Chudakov, D. M. (2006) *EMBO Rep.* **7,** 1006–1012
36. Merzlyak, E. M., Goedhart, J., Shcherbo, D., Bulina, M. E., Shcheglov, A. S., Fradkov, A. F., Gaintzeva, A., Lukyanov, K. A., Lukyanov, S., Gadella, T. W., and Chudakov, D. M. (2007) *Nat. Methods* **4,** 555–557
37. Wall, M. A., Socolich, M., and Ranganathan, R. (2000) *Nat. Struct. Biol.* **7,** 1133–1138
38. Wilmann, P. G., Petersen, J., Pettikiriarachchi, A., Buckle, A. M., Smith, S. C., Olsen, S., Perugini, M. A., Devenish, R. J., Prescott, M., and Rossjohn, J. (2005) *J. Mol. Biol.* **349,** 223–237

39. Quillin, M. L., Anstrom, D. M., Shu, X., O'Leary, S., Kallio, K., Chudakov, D. M., and Remington, S. J. (2005) *Biochemistry* **44,** 5774–5787
40. Prescott, M., Ling, M., Beddoe, T., Oakley, A. J., Dove, S., Hoegh-Guldberg, O., Devenish, R. J., and Rossjohn, J. (2003) *Structure.* **11,** 275–284
41. He, X., Bell, A. F., and Tonge, P. J. (2003) *FEBS Lett.* **549,** 35–38
42. Chudakov, D. M., Feofanov, A. V., Mudrik, N. N., Lukyanov, S., and Lukyanov, K. A. (2003) *J. Biol. Chem.* **278,** 7215–7219
43. Andresen, M., Wahl, M. C., Stiel, A. C., Grater, F., Schafer, L. V., Trowitzsch, S., Weber, G., Eggeling, C., Grubmuller, H., Hell, S. W., and Jakobs, S. (2005) *Proc. Natl. Acad. Sci. U. S. A.* **102,** 13070–13074
44. DeLano, W. L. (2002) The PyMOL Molecular Graphics System
45. Wallace, A. C., Laskowski, R. A., and Thornton, J. M. (1995) *Protein Eng.* **8,** 127–134
46. McDonald, I. K., and Thornton, J. M. (1994) *J. Mol. Biol.* **238,** 777–793
47. Evans, S. V. (1993) *J. Mol. Graph.* **11,** 134–138
48. Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., and Pease, L. R. (1989) *Gene* **77,** 51–59

## SUPPLEMENTARY DATA

*Cloning, Purification, and Characterization*- For bacterial expression, a PCR-amplified BamHI/HindIII fragment encoding a fluorescent protein was cloned into the pQE30 vector (Qiagen). Site-directed mutagenesis was performed by overlap-extension PCR1, with primers containing the appropriate target substitutions (48). Protein fused to the N-terminal polyhistidine tag was expressed in *E. coli* XL1 Blue strain (Invitrogen). The bacterial culture was grown overnight at 37 °C and additionally incubated for 12 hr at 25 °C. No induction by IPTG was applied since promoter leakage was sufficient for effective expression. Culture was centrifuged, the cell pellet re-suspended in 20 mM Tris-HCl, 100 mM NaCl, pH 7.4 buffer, and lysed by sonication. The recombinant protein was purified using TALON metal-affinity resin (Clontech), followed by gel filtration on a Superdex-200 (16/60) size exclusion column (Amersham). Absorption spectra were recorded with a Beckman DU520 UV/VIS spectrophotometer. A Varian Cary Eclipse fluorescence spectrophotometer was used for measuring excitation-emission spectra. The molar extinction coefficient for the native state was calculated based on the absorption of the native and alkali-denatured proteins, the latter characterized by an extinction coefficient 44,000 $M^{-1}cm^{-1}$ at 446 nm. For determination of the quantum yield, fluorescence spectra of the mutant variants were compared with equally absorbing mPlum (quantum yield 0.10). The pH titrations were performed by using a series of buffers in the range from 3.5 to 9.7. For each pH value an aliquot of purified protein was diluted in an equal volume of the corresponding buffer solution and the fluorescence brightness was estimated as a product of molar extinction coefficient and a fluorescence quantum yield after 12 hr incubation at room temperature.

Supplementary Figure 1S