Hidden Photoinduced Reactivity of the Blue Fluorescent Protein mKalama1

Russell B. Vegh, Dmitry A. Bloch, Andreas S. Bommarius, Michael Verkhovsky, Sergei Pletnev, Hideo Iwai, Anastasia V. Bochenkova, and Kyril M. Solntsev

ELECTRONIC SUPPLEMENTARY INFORMATION

MATERIALS AND METHODS

(a) Protein expression and purification

The gene encoding mKalama1 was purchased from AddGene (Cambridge, MA). The gene was inserted between the *Xho*I and *Eco*RI restrictions sites in pBAD-His B and contained an N-terminal hexahistidine tag. The plasmid was transformed into *E. coli* DH5 αPro. A 5 mL culture was inoculated into 1 L of LB (Luria-Bertani broth, pH 7) containing 50 μg/mL of ampicillin, grown to an OD600 of 0.5 at 37 °C (2-3 hrs), and protein expression was induced with 0.2% arabinose. We overexpressed mKalama1 to ~20% of total cell protein and purified the protein via Ni^{2+}-NTA-immobilized metal affinity chromatography (IMAC). All measurements were performed in phosphate buffer saline (PBS, 50 mM sodium phosphate, 250 mM NaCl, pH 7.5), unless otherwise stated.

(b) Crystallization, data collection, structure solution and refinement

The crystallization conditions were screened at the Biocenter Finland crystallization facility (Institute of Biotechnology, University of Helsinki) using solution from Index HT screen (Hampton Research), by sitting drop vapor diffusion in Innovadyne SD2 96 well plates with drops containing 100 nL protein solution (30 mg/ml in H_2O) and 100 nL reservoir solution, incubated against 80 μl of the latter for 35 days at 293 K. A single crystal was picked from the drop containing 1.8 M ammonium sulfate, 0.01 M cobalt chloride, 0.1 M MES at pH 6.5. The crystal growth was visually controlled using Minstrel DT UV robotic imaging system (Rigaku, Japan).

The X-ray diffraction data were collected at I24 at the Diamond Light Source, UK. The diffraction intensities were registered with Pilatus (DECTRIS Ltd). Prior to data acquisition, a single crystal was washed in Paratone-N and flash frozen in a 100 K nitrogen stream. During the experiment cryogenic temperature was maintained. Diffraction images were processed with Pointless, XDS, Xia2.[1]

The structure of mKalama1 was solved by a molecular replacement method with Phaser,[2] using the coordinates of a single monomer of the GFP S205V mutant (PDB ID: 2QLE)[3] as a template. The structure was refined with PHENIX.REFINE[4] and COOT.[5] The latter program was also used for manual rebuilding of protein chains and addition of ordered solvent molecules. The structure was validated with COOT and PROCHECK.[6] The data processing and structure refinement statistics are given in Table S1. The coordinates and the structure factors were deposited in the Protein Data Bank under accession code 4ORN.

(c) Transient absorption spectroscopy measurements

1. µs-to-second time domain

In the present study, a custom-built kinetic setup based on ANDOR and Basler Vision microarray imaging cameras was used, allowing broadband TA measurements with 1 µs time resolution after a single pulse excitation. This set-up, previously used to study multiheme electron-transfer enzymes such as terminal respiratory oxidases,[7-,9] is located at the University of Helsinki (Helsinki, Finland), where the measurements were performed. A pulsed 150-W xenon arc lamp (Applied Photophysics, Surrey, U.K.) was used as the probe light source. After passing through the sample, the probe light was directed using a fiber optic cable (300 µm, UV-VIS XSR solarization-resistant, Ocean Optics, Dunedin, FL) to either of the two detectors incorporated into one setup. A Triax-180 compact imaging spectrometer (HORIBA JobinYvon, Edison, NJ) delivered spectral imaging over a fast kinetic CCD matrix (DV420-UV-FK; Andor Technology, Belfast City, Ireland) to enable of

recording absorption change surfaces with a time resolution of 1–16 μs between spectra. A slower response on a time scale from tens of microseconds to minutes was measured with an in-house-built spectrometer based on Sprint spL 2,048–140 km linear scan camera (Basler vision technologies, time resolution of 12 μs between spectra, Basler AG, Ahrensburg, Germany). The wavelength-time-absorption surfaces from these two detectors were combined into a single surface containing absorbance changes developed after laser flash. The setup was operated by the software written by Nikolai Belevich (University of Helsinki, Helsinki, Finland).

The photolysis of mKalama1 was achieved with single-flash laser excitation at 355 nm (maximum energy of 100 mJ, 50 mm^2 spot area, 4 ns per pulse) provided by a 3rd harmonic generator attached to a Q-switched Nd-YAG laser (Brilliant B, Quantel, Les Ulis, France). The laser beam was focused, using a pair of cylindrical quartz lenses, on a larger side of a quartz fluorescence microcuvette (10 × 2 mm optical path, Hellma Analytics, Müllheim, Germany) and covered a 10 × 2 mm^2 area of the sample. The laser-illuminated spot was slightly larger than the sample area exposed to the probe light, lessening the possibility of slow, diffusion-related artifacts. Each surface was obtained as an average over the data following 20-25 laser flashes. A fresh, dark-adapted sample (OD_{390} = 1.0, unless otherwise stated) was taken every time, so that each sample was illuminated only once. The sample was exposed to the probe light ~0.5 ms prior to flash excitation using a mechanical shutter (Uniblitz LS2, 2 mm aperture, 300 μs opening time, Vincent Associates, Rochester, NY) to minimize the actinic effect of white light. After each flash, less than 2% of the protein population underwent irreversible photolysis, as checked by photobleaching at 390 nm. Typically, focused laser flashes provided ~500 mJ/cm^2 energy per shot (unless otherwise stated), which caused saturation of most kinetic phases at the level ranging between 40 and 60%, all depending hyperbolically (n=1) on laser flash energy with a half-saturation energy ranging from 325 to 700 mJ/cm^2. All measurements were performed at +21 °C. The wavelength-time-absorption

surfaces of TA changes in the 1 μs – 1 min time window were decomposed into a multiexponential decay function using a global fit algorithm:

$$\Delta A(t,\lambda) = \sum_i \varepsilon_i(\lambda) \exp(-t/\tau_i) \quad \text{(S1)}$$

where $\Delta A(t,\lambda)$ is the optical absorption change at time t (t = 0 at the instance of laser flash) and at wavelength λ, and $\varepsilon_i(\lambda)$ and τ_i are the characteristic spectrum and the time constant for the i-th exponential decay component, respectively. The global data fit with a sequential reaction model[10] (a reaction that consists of several sequential, irreversible steps) was performed using the algorithm described previously.[11] All data analysis was performed in MATLAB (version 7.8, MathWorks, Natick, MA).

2. picosecond-to-nanosecond time domain

Subpicosecond transient absorption measurements were performed in Dr. Joseph Perry's lab (Georgia Institute of Technology, Atlanta, GA). Spectra were acquired using a commercial transient absorption spectroscopy system (Newport, Helios). This system accepts two input laser beams, one of variable wavelength used as the pump beam and one of a fixed wavelength used to generate the probe beam in a proprietary nonlinear optical crystal. All experiments were performed using the TOPAS output at 355 nm with a pulse width of 100 fs. A small portion (~5%) of a Ti:Sapphire regenerative amplifier (Newport, Spitfire, 800 nm, 1 kHz, Spectra-Physics, Santa Clara, CA) provided the probe pulse to generate the white-light continuum (WLC, 420-850 nm, 850-1650 nm). At these specifications, the instrument response function (IRF) had a full width at half maximum (FWHM) of approximately 200 fs. At each time point, data were averaged for 2 s. The Helios pump beam was chopped at 500 Hz to obtain pumped (signal) and non-pumped (reference) absorption spectra of the sample. A correction factor for chirp in the probe beam was generated using the ultrafast response of CCl_4, and was applied to all data sets. The data were stored as wavelength-time-absorbance surfaces that were exported for use with the fitting software. The protein solutions

had an OD_{390} of ~0.5 in 2 mm path-length cuvettes, and were stirred continually throughout the data acquisition. The pump energy was 145 mW/cm^2 per pulse.

(d) Time-correlated single photon counting measurements

Time-resolved fluorescence was acquired using the time-correlated single photon counting method (TCSPC). Picosecond pulsed diode lasers were used to excite the samples. The detector was a high speed microchannel plate photomultiplier tube (Hamamatsu R3809U-50) cooled to –20 °C to reduce noise. The pulses were communicated to a hardware controller by a constant fraction discriminator (CFD). A second CFD was used to obtain a timing reference pulse from the light source. The time-amplitude-converter output voltage was sent through a biased amplifier with a variable gain and a variable offset and the amplified signal was fed to the analog-to-digital converter. A multichannel analyzer software (Edinburgh Instruments F900, Chorley, U.K.) was used to process the signal and convert it to a PC format. Fluorescence lifetime data were fitted using the program FFIT provided by professor Nikolai Tkachenko at the Tampere University of Technology (Tampere, Finland). Excitation of both neutral and anionic Cro was achieved using 372 nm and 476 nm diodes, respectively. The IRF with a FWHM of 100 ps was measured for both wavelengths and used to deconvolute the fluorescence decay.

(e) H/D kinetic isotope effect measurements

For the kinetic measurements in D_2O, the concentrated protein samples were diluted in 99.9% D_2O-based PBS (pD 7.5); the final H_2O concentration in the samples did not exceed 2%. pD was adjusted with DCl or KOD, taking into account a typical pH correction for the pH glass electrode: $pD = pH^{app}(D_2O) - 0.4$.[12] The samples were incubated for 14 hours in the D_2O medium before measurements.

(f) Aerobic and anaerobic measurements

Anaerobic samples were prepared using a gas/vacuum line of local design. A 4 × 10 × 30 mm fluorescence quartz cell was sealed to a Kimble-Kontes high-vacuum stopcock (Kimble-Chase, Vineland, NJ), which had a vacuum O-ring connection to the vacuum line. As a typical degassing procedure, 20 cycles of exchanging vacuum (10^{-5} bar) and pure Ar were followed by shaking the cell filled with Ar for 15 min on an orbital shaker; then the procedure was repeated 2 more times. 99.99% Ar was additionally purified using the Agilent Technologies BOT-2 and IOT oxygen scrubbers (Santa Clara, CA). Aerobic samples were bubbled with oxygen for 10 minutes.

(g) Protein structure simulations

The model systems of the BFP mKalama1 were constructed based on the crystal structure of the GFP S205V mutant (PDB ID: 2QLE, chain A)[3] and compared to our newly determined structure of mKalama1 (PDB ID: 4ORN, this study). The initial system was obtained by introducing 23 point mutations relative to GFP S205V according to the sequence of mKalama1[13]. Missing hydrogen atoms were added to heavy nuclei. All histidine residues were treated as neutral and δ-protonated, except for H39 forming a salt bridge with D36, which was treated as charged and doubly protonated. Upon substitution, the overall net charge of the protein changed from -6 to -1. It was set to zero by performing an additional point mutation D117N at the solvent-exposed site. The structure was then extensively solvated by adding 14632 water molecules (Figure S12a). Afterwards, it was exposed to the preliminary geometry optimization (10000 steps) followed by the constant temperature NVT molecular dynamics (MD) simulations with the CHARMM force field parameters.[14] The periodic boundary conditions were applied. The equilibration temperature was set to 300 K using a Langevin thermostat. The long-range electrostatic interactions were treated with the particle mesh Ewald method. The simulations were carried out with a 1 fs integration step for 2

ns. After that, the temperature of the system was gradually reduced down to 50 K in steps of 50 K during 1.2 ns. The final MD geometry was obtained by performing energy minimization during another 10000 steps. The atomistic model of mKalama1 was constructed using the program VMD,[15] while all MD simulations were run using NAMD, version 2.7b2.[16]

The final annealed MD atomistic model was reduced for a subsequent quantum mechanical/molecular mechanical (QM/MM) treatment by removing outer water molecules. The reduced system contained 750 water molecules, while the total number of atoms, including the entire protein, was 5830 (Figure S12b). The Cro as well as three charged residues (R96, E222, R224) together with the other two residues (Q94, Q183) and water molecules that formed a hydrogen bonding network inside the Cro binding pocket were included in the QM part (Figure S12c). In total, 136 atoms, including nine capping hydrogen atoms, were assigned to this subsystem treated using 1335 basis functions at the PBE0/(aug)-cc-pVDZ level of theory, while performing the ground-state geometry optimization of the neutral chromophore inside the protein. For this, the mechanical embedding QM/MM scheme was employed together with the CHARMM[14] force field parameters. The additional constraints were imposed on the outer part of the protein located more than 7 Å away from the Cro, which was held fixed during the QM/MM optimization procedure.

The equilibrium geometry parameters of the Cro radical forms were found using the same QM/MM methodology, with the partial charges used for calculating the long-range electrostatic interactions of the QM and MM parts recalculated at the PBE0/ROHF/(aug)-cc-pVDZ level of theory, applying the natural bond orbital analysis (NBO).[17] The differences in the calculated NBO charges between a parental species and its radical were used to redefine the corresponding force field parameters.

The modified PC GAMESS/Firefly version 6.4[18] was used for the QM/MM ground-state optimization procedure.

(h) Excited-state electronic structure calculations

The vertical excitation energies were calculated using the QM/EFP approach (EFP)[19], enabling the QM electron density to be polarized in the electrostatic field of the MM part. The CHARMM partial point charges were used. Here, the same QM parts were employed as those used for the ground-state optimization procedure (Figure S12c).

The QM parts were treated at the extended multi-configuration quasi-degenerate perturbation theory (XMCQDPT2[20]) level. The reference wave functions of the ground state and the target excited states were constructed within the complete active space self-consistent field (CASSCF) method, with the active spaces comprising 14 electrons distributed over 13 orbitals of a π type. A state-averaging (SA) procedure over two and four states was applied for anionic and neutral species, respectively. In the latter case, the number of states included in the state-averaging procedure was larger due to a deficient CASSCF description of the order of excited states, which were subsequently mixed within a perturbative treatment through the effective Hamiltonian. At the XMCQDPT2 level, the correct order of states restored, and the target bright state appeared to be the first excited state for the neutral forms. The XMCQDPT2 effective Hamiltonians were constructed in the frame of reference spaces spanned by seven CASCI zeroth-order wave functions.

The Cro neutral radical and radical cation species required special attention, since several excited states were of interest here. A substantial mixing of the states occurred at the XMCQDPT2 level, also accompanied by their flipping. Here, the active spaces were enlarged to include all valence π and π^* orbitals. The occupation numbers of the CASSCF natural orbitals indicated the importance of enlarged dimensions of the active spaces. The XMCQDPT2/SA(7)-CASSCF(15,14)/(aug)-cc-pVDZ approach resulted in 4294290 configuration state functions (CSFs) used to expand the reference CASSCF functions. A selection procedure for the most important CSFs was applied while constructing effective Hamiltonians with a threshold of 10^{-7} for the weights

of retaining CSFs. The number of the reference CSFs selected for the perturbative treatment was *ca.* 500000.

Finally, the energies of all semi-canonical orbitals used in the perturbation theory were defined as eigenvalues of the corresponding blocks of a DFT-based effective Fock operator, diagonalized in the basis of the computed CASSCF molecular orbitals. The hybrid PBE0 functional was used in constructing the DFT-based Fock matrix.

For all excited-state electronic structure calculations, the Firefly package, version 8.0,[18] partially based on the GAMESS (US) source code,[21] was used.

SUPPLEMENTARY RESULTS, DISCUSSION, FIGURES, AND TABLES

The supplementary figures and tables are grouped into sections according to particular measurements and calculations performed in the present study. The data presented in the ESI figures and tables are also supplemented by the details of their acquiring, as well as by the detailed discussion when necessary. Below, we provide a list of sections with figures and tables numbered sequentially, in accordance to the order they are referenced in the main text:

1. Time-correlated single photon counting measurements

Time-resolved fluorescence was acquired using the time-correlated single photon counting method (TCSPC). Picosecond pulsed diode lasers were used to excite the samples. The detector was a high speed microchannel plate photomultiplier tube (Hamamatsu R3809U-50) cooled to -20 °C to reduce noise. The pulses were communicated to a hardware controller by a constant fraction discriminator (CFD). A second CFD was used to obtain a timing reference pulse from the light source. The time-amplitude-converter output voltage was sent through a biased amplifier with a variable gain and a variable offset and the amplified signal was fed to the analog-to-digital converter. A multichannel analyzer software (Edinburgh Instruments F900, Chorely, U.K.) was used to process the signal and convert it to a PC format. Fluorescence lifetime data was fit using the program FFIT provided by professor Nikolai Tkachenko at the Tampere University of Technology (Tampere, Finland). Excitation of both the neutral and anionic Cro was achieved using 372 nm and 476 nm diodes, respectively. The instrument response functions (IRF) with a FWHM of 100 ps was measured for both wavelengths and used to deconvolute the fluorescence decay. At pH 7.5 ($\lambda_{ex/em}$=372/460 nm), the data was fit with a bi-exponential decay with lifetimes of 0.8 ns and 1.9 ns and an amplitude ratio of 0.25 (Figure S1). The same lifetimes were observed by exciting at 372 nm and monitoring the emission at 510 nm (data not shown). At pH 11.25 ($\lambda_{ex/em}$=476/510 nm), the data was fit with a bi-exponential decay and had lifetimes of 78.0 ns and 3.9 ns with an amplitude ratio of 0.005; therefore, the 78 ns component is negligible (Figure S2). The absence of both a rise time component at the red edge of the spectrum and an increase of the decay time (Figure S2) indicates there is no anion emission and hence no ESPT.

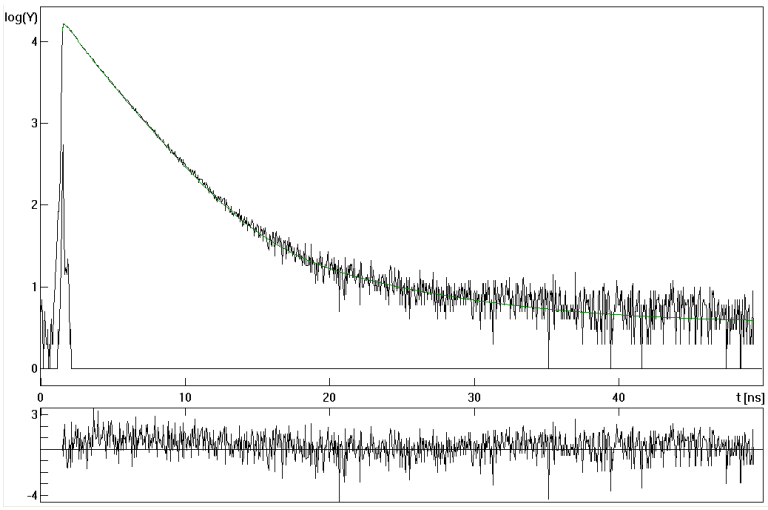

Figure S1. Fluorescence decay of mKalama1 at 460 nm excited with a 372 nm diode at pH 7.5 (upper panel) and multiexponential fit residuals (lower panel).

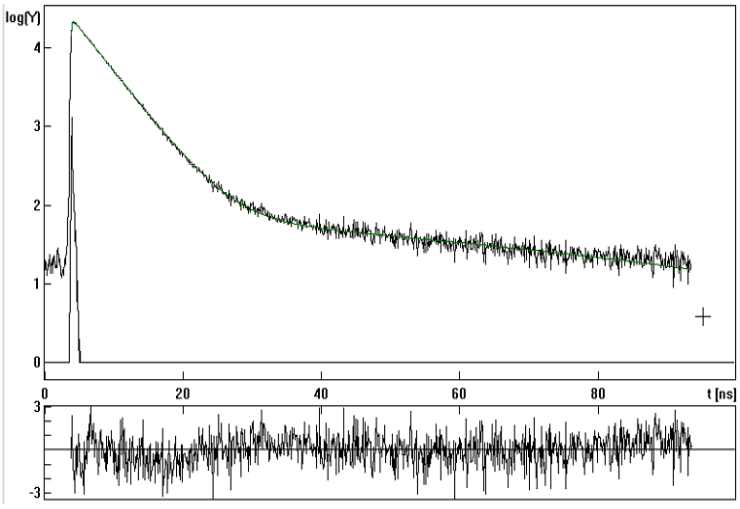

Figure S2. Fluorescence decay of mKalama1 at 510 nm excited with a 476 nm diode at pH 11.25 (upper panel) and multiexponential fit residuals (lower panel).

2. pH-dependence of mKalama1 absorption and excitation spectra

mKalama1 (Buffer: 20 mM *N*-cyclohexyl-3-aminopropanesulfonic acid (CAPS), 20 mM 4-(Cyclohexylamino)-1-butanesulfonic acid (CABS), 20 mMKPi, 200 mMNaCl) was titrated to pH 12.0 using KOH. The protein was then titrated back to pH 7.5 using HCl (Figure S3). The anionic form of mKalama1 chromophore seen at alkaline pH with a broad maximum at *ca*. 590 nm was also fluorescent ($\lambda_{ex/em}$=470/509 nm) and the excitation spectra at various pH are shown in Figure S4.

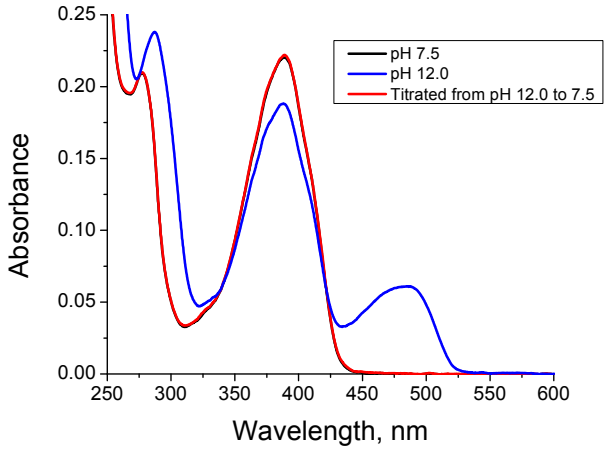

Figure S3. Absorbance spectra of mKalama1 at pH 7.5 (black), at pH 12.0 (blue), and after titration back to pH 7.5 (red).

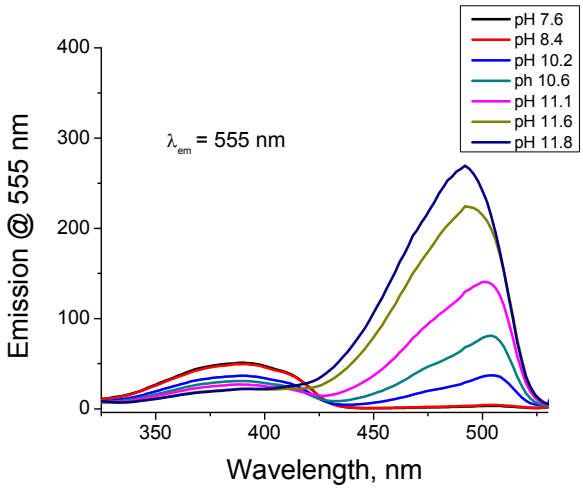

Figure S4. Excitation spectra of mKalama1 at various pH as monitored through emission at 555 nm.

3. Laser flash-triggered, transient absorption kinetics of mKalama1

The transient absorption data of mKalama1 are shown in Figure 1 in the main text. Figure S5 illustrates the expanded absorption area at λ > 525 nm.

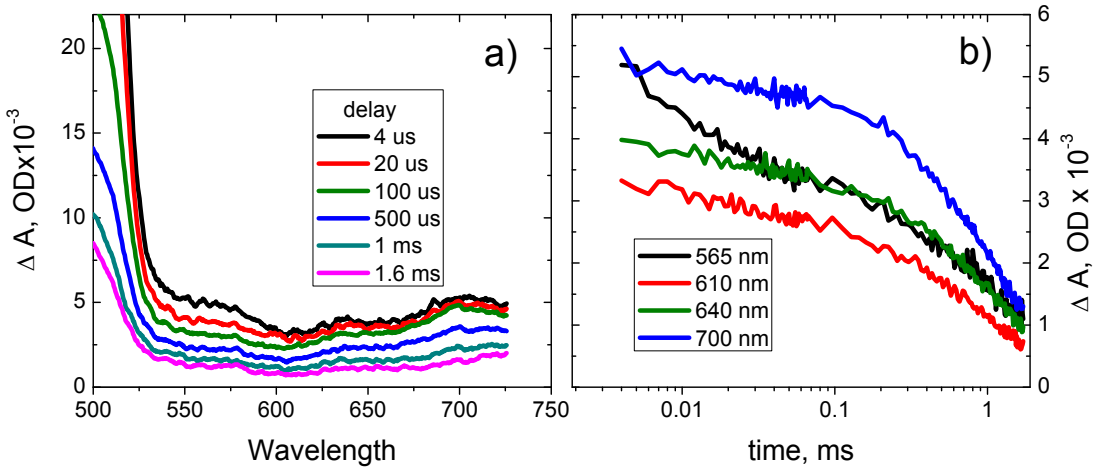

Figure S5. TA kinetics in mKalama1 illustrating weak absorption at λ > 525 nm. (a) TA spectra at selected time points. (b) Kinetic traces of the TA signal at selected wavelengths.

The best least-square multiexponential fit of the experimental time-wavelength-absorption surfaces at λ=380-540 nm is shown in Figure S6a. The data were also fit using the algorithm for a sequential reaction model. The algorithm used for the kinetic analysis of sequential reactions is described in Bloch, *et al*.[10] In a general case of *n* consecutive irreversible steps with reduced concentrations an analytical solution can easily be found for any given *n* using symbolic algebra software. Note that the solution for *n*=3 was first published by Rakowski in 1906.[22] The best least-square sequential reaction fit, with and without the slower components subtracted out, is shown in Figure S6b,c.

For the TA studies at various pH values, a fresh sample was used for each pH measurement and was titrated with NaOH. The analysis of the TA data at different pH is shown in Figure S7.

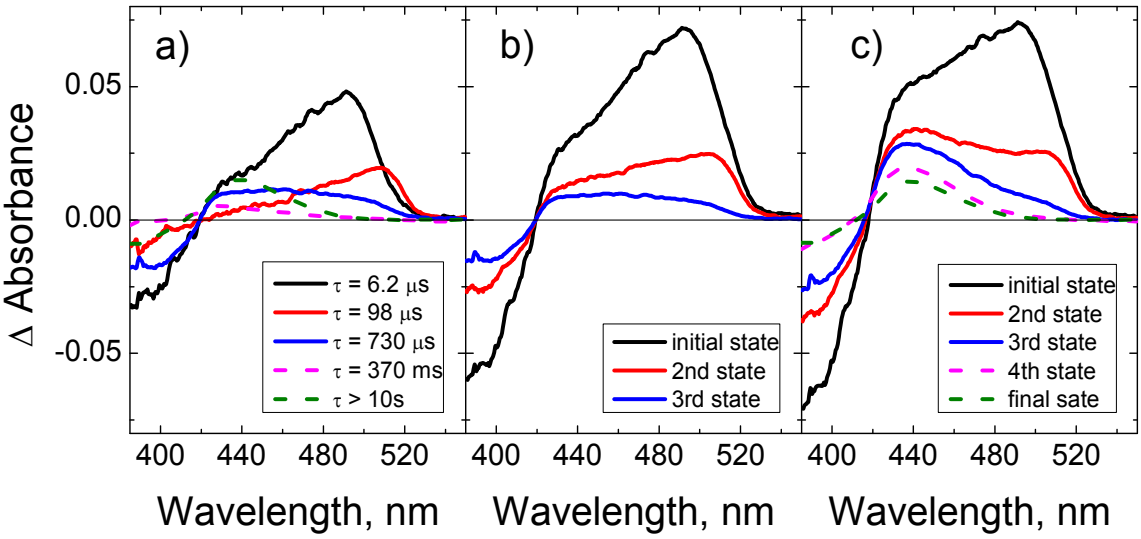

Figure S6. (a) The best least-square multiexponential fit results: the three faster (sub-ms) exponential phases (solid lines) and the two slower components (dashed lines); the irreversible component is subtracted. (b) The best least-square Rakowski (sequential reaction) fit results. The two slower exponential components (with τ = 370 ms and 9.5 s were subtracted from the original data matrix. (c) The best least-square Rakowski (sequential reaction) fit results. The original data matrix was used as such for the fit.

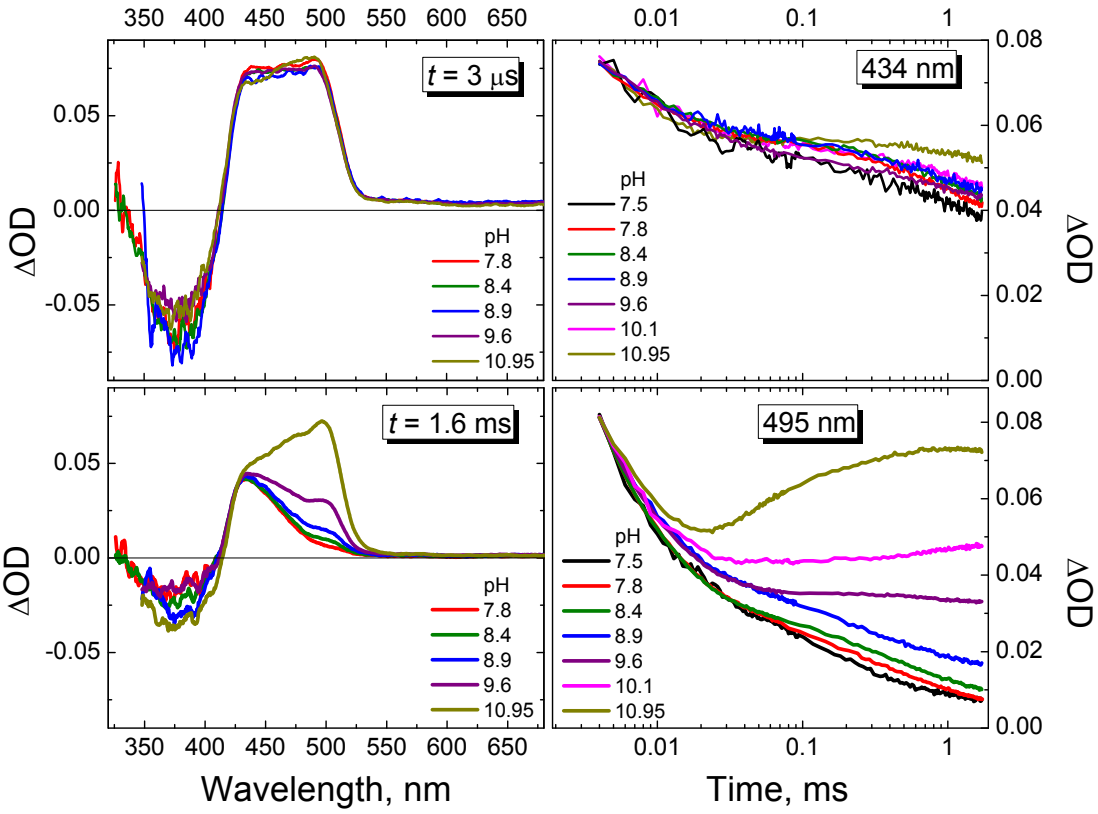

Figure S7. Effect of pH on the TA kinetics of mKalama1. Left column: TA spectra at various pH after excitation with 4 ns 355 nm laser pulse at the time delays of 3 μs and 1.6 ms. Right column: TA kinetics at various pH at 434 and 495 nm.

4. Photodecomposition of mKalama1

The photodecomposition experiments were performed using a flash energy of *ca.* 500 mJ/cm^2, a spot size of 20 mm^2, and an optical path of 1 cm. Figure S8a shows the absorption spectra of mKalama1 before illumination and 5 min after each of the 25 sequential flashes. The accumulation of the "irreversible" spectral changes following photolysis is shown in Figure S8b. The TA spectra after each flash at *ca.* 2 ms are shown in Figure S8c. To determine whether the accumulation of this photoproduct is reversible, a separate experiment was performed, where the sample was exposed to 30 sequential flashes upon stirring (2 min in between flashes). The absorption spectra were recorded at 0, 16, 40, 85, and 170 minutes following the 30 flashes (Figure S9). From these spectra, it is shown that the photoproduct slowly disappears concomitantly with an increase in the ground state

mKalama1 absorption. This is further illustrated in Figure S10, where the difference spectrum is plotted as a function of time.

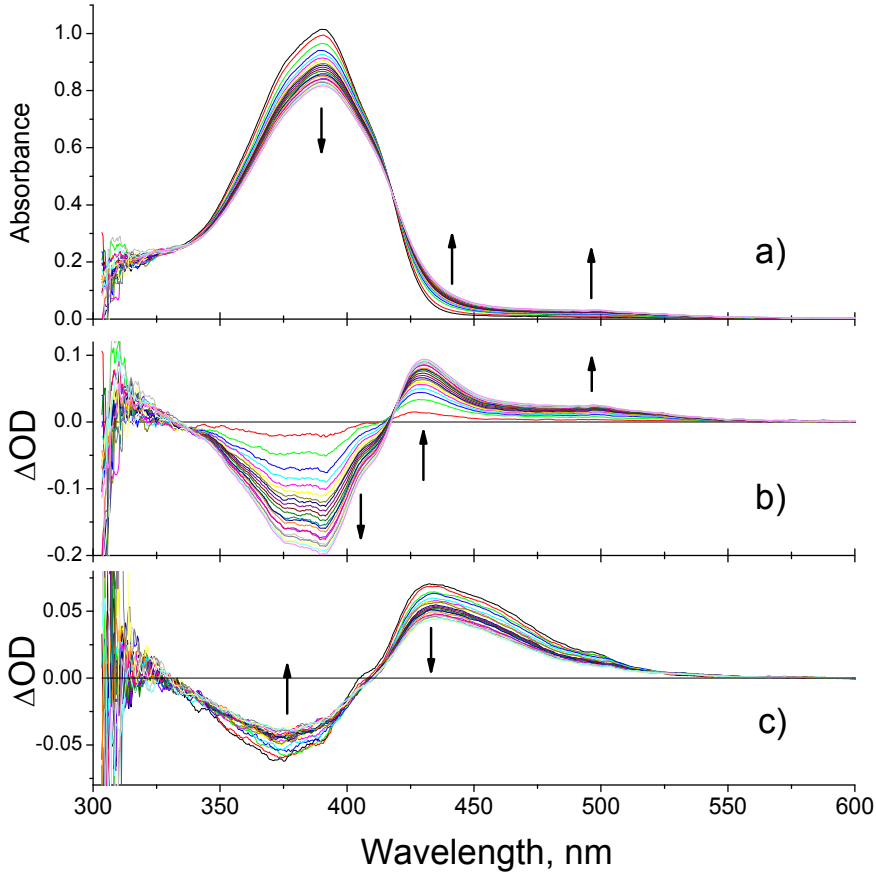

Figure S8. Irreversible photodecomposition of mKalama1 in response to sequential flash illumination at 355 nm. (a) Absolute absorption spectra before illumination and 5 min after each of the 25 sequential flashes. (b) Accumulation of the irreversible spectral changes as the effect of 25 sequential flashes. (c) Light-*minus*-dark difference spectra measured at *ca.* 2 ms following each flash. Arrows show the direction of the flash-induced, irreversible spectral changes after 25 sequential flashes. The time delay between flashes was 5 min.

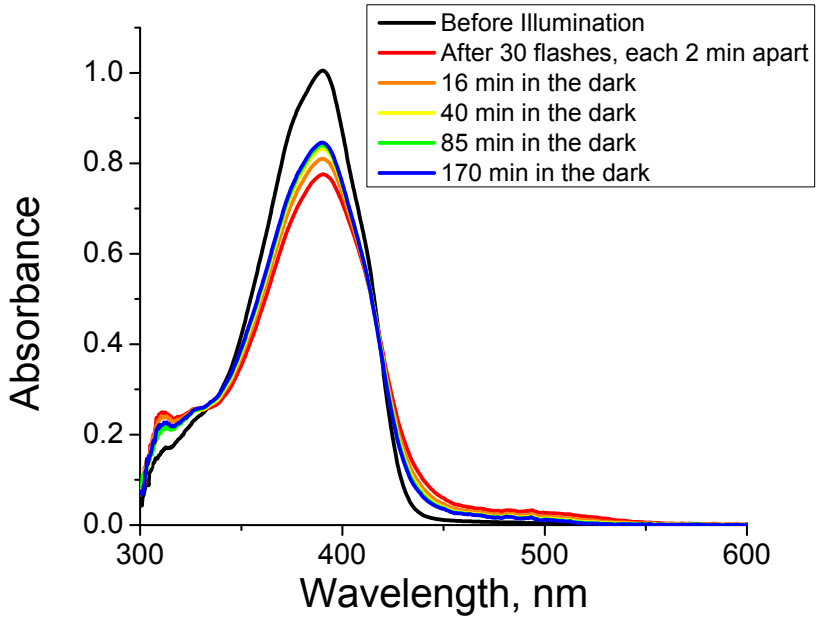

Figure S9. Partial spectral recovery of the dark-adapted state of mKalama1 after sequential flash illumination at 355 nm. Shown are the absolute absorption spectrum before illumination and the transient absorption spectra at different time delays after 30 flashes.

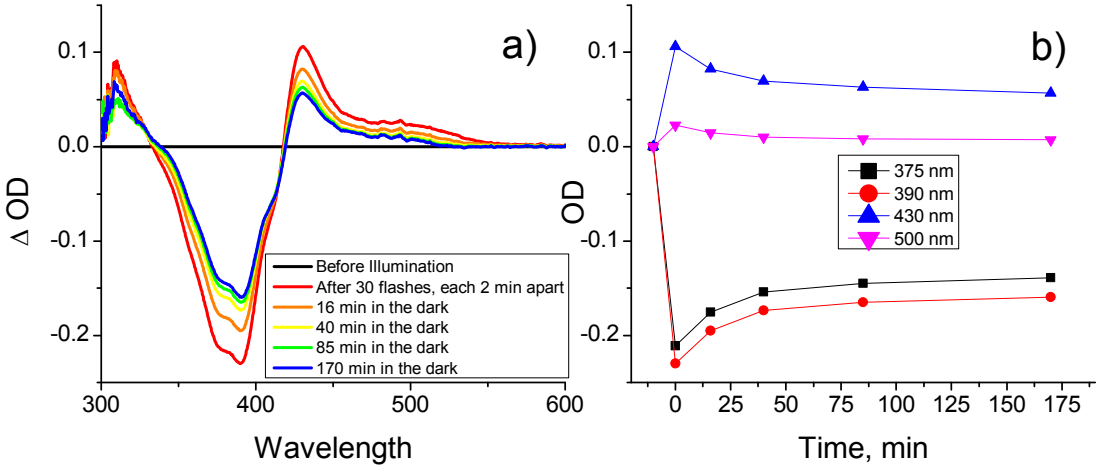

Figure S10. (a) Difference spectra (light-*minus*-dark) taken at different time points following illumination at 355 nm. (b) The changes in absorption at selected wavelengths are plotted as a function of time, where 0 min is the absorption recorded immediately following the 30 flashes. The same data set as in Figure S9 was used.

5. X-ray structural analysis of the initial state in the photocycle

To explore the structural basis for the observed photoinduced processes in mKalama1, we have determined the crystal structure of this protein at 1.71 Å resolution. mKalama1 is crystallized in a space group $P2_12_12_1$. The asymmetric unit of the crystal contains two monomers positioned at a ~70° angle between their β-barrel axes. The electron density for the residues 2-232 is well defined with the exception of the side chains of several lysines, arginines, glutamines, and glutamic acids located at the protein surface, that are disordered. Multiple conformations of side chains are observed for the total of nine (3.8%) and thirteen (5.5%) residues in crystal monomers A and B, respectively. The final atomic model is of a high quality, as indicated by low deviations of bonds, angles, dihedrals and planes from their library values, and by the absence of residues in the disallowed regions of Ramachandran plot (Table S1). The superposition of mKalama1 and wtGFP by Cα atoms results in a root-mean-square deviation of 0.6 Å, indicating close similarity of the two structures.

Table S1. Data collection and refinement statistics of mKalama1

Data Statistics	
Space group	$P2_12_12_1$
Unit cell parameters (Å)	a = 80.5, b = 83.5, c = 92.2
Temperature (K)	100
Wavelength (Å)	1.00
Resolution (Å)	80.5 - 1.71
Total reflections	886,720
Unique reflections	67,686
Completeness (%)	99.8
I/σ<I>	21.5
R-merge	0.07
Multiplicity	13.1
Refinement Statistics	
No. of protein atoms	3,816

No. of solvent atoms	402
Resolution range (Å)	60.7 - 1.71
R-work	0.158
R-free	0.183
R.m.s.d. bond lengths (Å)	0.012
R.m.s.d. angles (°)	1.67
R.m.s.d. chirality ($Å^3$)	0.12
R.m.s.d. planarity (Å)	0.007
R.m.s.d. dihedral (°)	16.1
Ramachandran statistics (%), for non-Gly/Pro residues	
most favorable	98.7
additional allowed	1.3

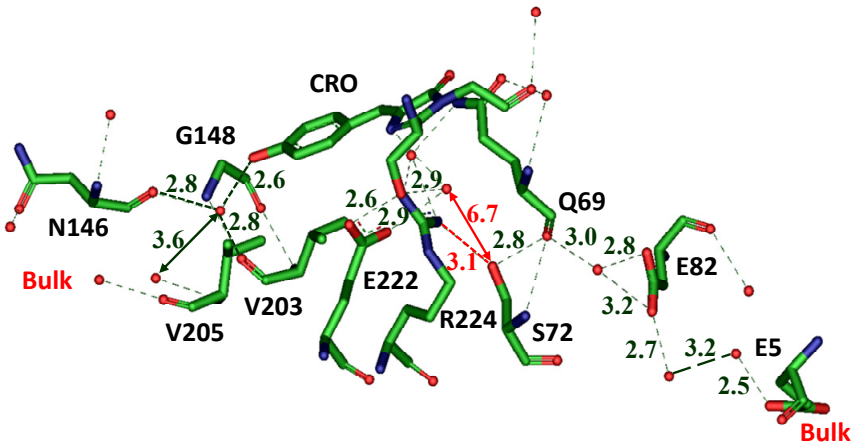

Figure S11. Extended hydrogen-bonding network inside mKalama1 according to the X-ray crystal structure obtained in this work.

6. Electronic structure calculations

Table S2. Dark-adapted and transient states of mKalama1 analyzed in the text.*

Cro protonation and redox state	Cro configuration		
	Z		E
	Protonation state of E222 and R224		
	$E^{-}\cdots R^{+}$	EH R^{+}	EH R^{+}
PhOH-Im	$Z^{N}{}_{390}$ (376 nm/3.30 eV, $S_0 \rightarrow S_1$)		
PhO^{-}-Im	$Z^{A}{}_{430}$ (440 nm/2.82 eV, $S_0 \rightarrow S_1$)	$Z^{A}{}_{489}$ (488 nm/2.54 eV, $S_0 \rightarrow S_1$)	$E^{A}{}_{489}$ (490 nm/2.53 eV, $S_0 \rightarrow S_1$)
$PhO^{\bullet}$-Im	$Z^{\bullet}{}_{430:700}$ (442 nm/2.81 eV, $D_0 \rightarrow D_3$; 700 nm/1.77 eV, $D_0 \rightarrow D_1$)		
$PhOH^{+\bullet}$-Im	$Z^{+\bullet}$ (528 nm/2.35 eV, $D_0 \rightarrow D_3$; 780 nm/1.59 eV, $D_0 \rightarrow D_1$)		
PhO^{-}-ImH^{+}	$Z^{ZW}{}_{505}$ (508 nm/2.44 eV, $S_0 \rightarrow S_1$)		

* Cells are highlighted according to the photocycle branches (Figure 9): the dark-adapted state (blue), branch I (green), branch II (yellow), branch III (pink). The unlikely states are denoted in grey. The subscripts show the experimentally observed absorption maxima (when known). The calculated vertical transition energies for the major absorption bands are shown in brackets (see also Figure 5). PhOH and Im are the phenol and imidazolone moieties of the chromophore.

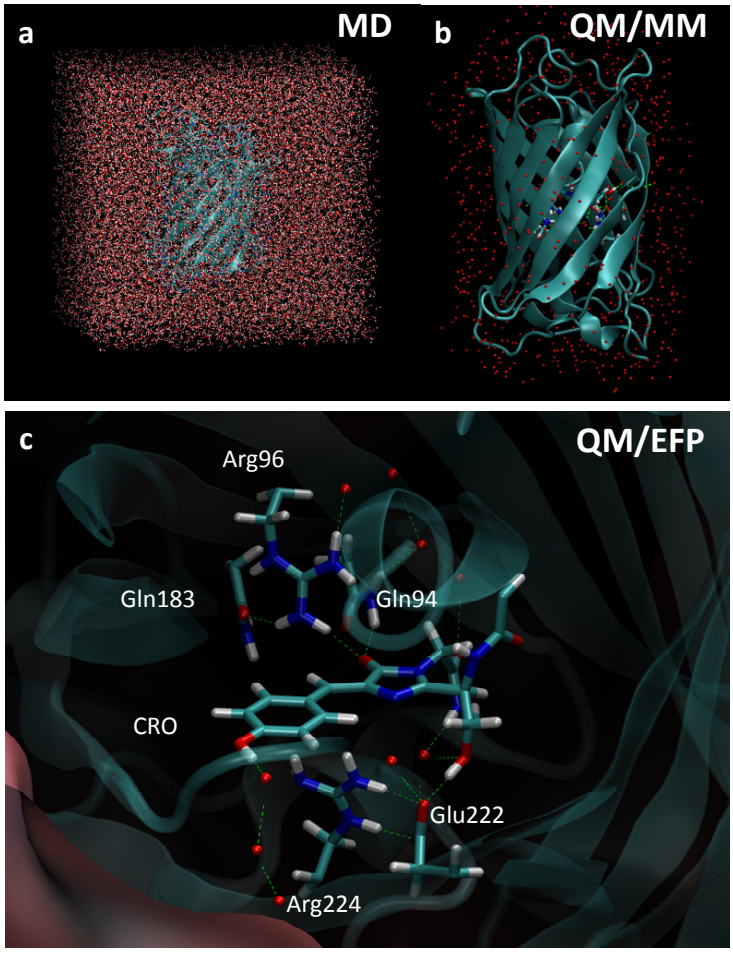

Figure S12. Model systems used for (a) MD and (b,c) QM/MM calculations. The QM part is shown in the stick representation, where the color code denotes the C, O, N, and H atoms depicted in cyan (green), red, blue, and white, respectively. Note that the same QM parts are employed for the ground and excited state calculations.

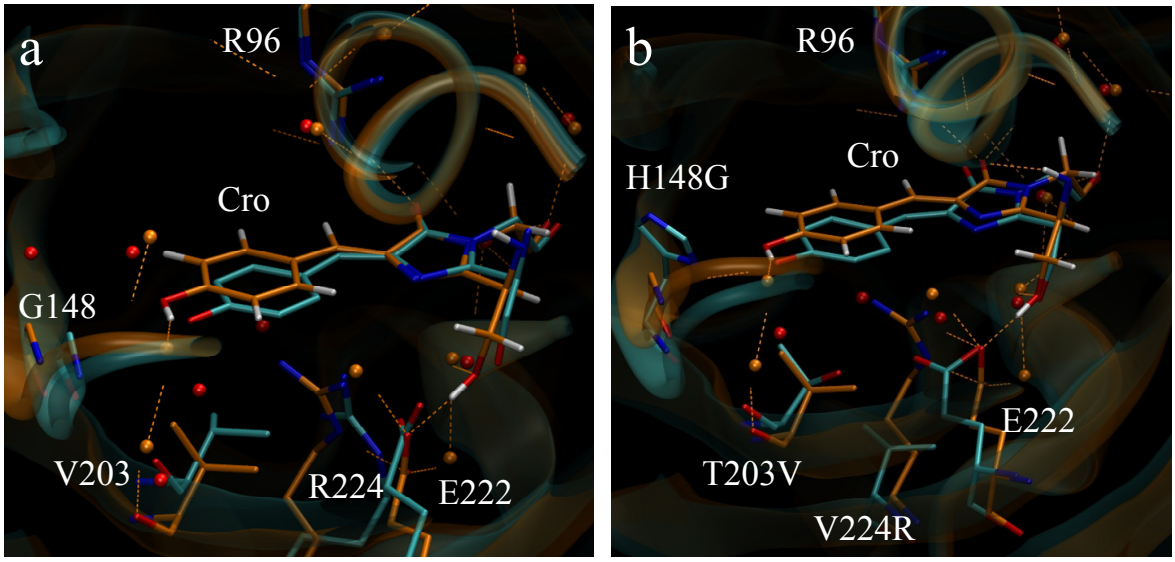

Figure S13. Superposition of the calculated structure (orange) of the chromophore-protein pocket of mKalama1 with those obtained through the X-ray analysis (cyan) of (a) mKalama1 and (b) GFP S205V (PDB ID, 2QLE). Note that the different conformations of the side chain of R224 do not affect the calculated vertical excitation energies. The calculations also reveal a conformational flexibility of this residue.

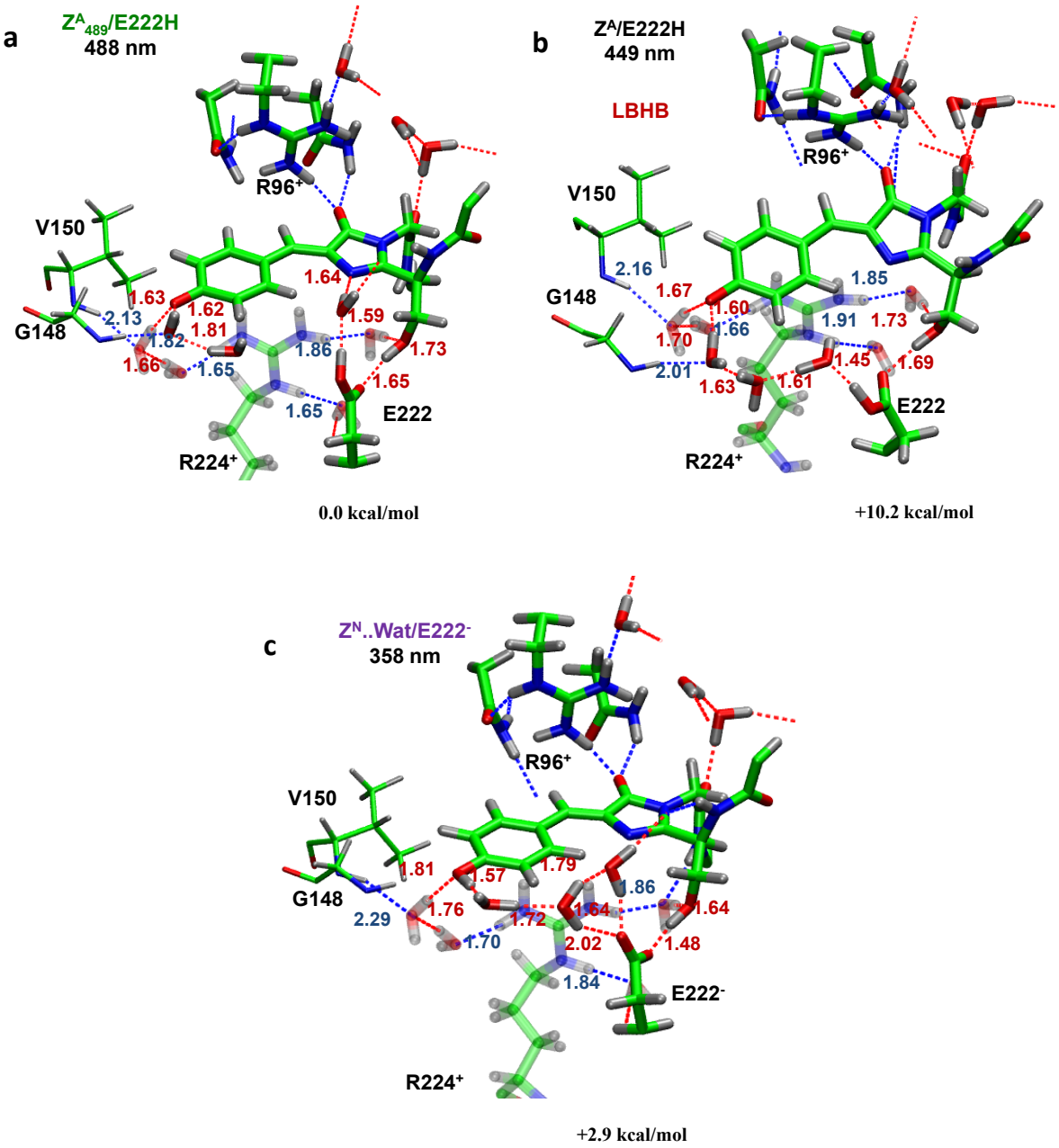

Figure S14. The QM/MM ground-state equilibrium structures of the *Z*-transients formed along the intraprotein, proton back-transfer pathway of branch I in the photocycle of mKalama1: (a,b) anionic and (c) neutral forms of the Cro inside the protein. Shown are the QM parts (sticks) and the MM residues, G148 and V150, (bold lines). The structures are labeled according to the charge state of E222 and to the assigned experimental absorption maxima. The calculated vertical excitation energies (nm) and the relative total energies (kcal/mol) are shown in black. All distances are in Å. Note that the structure (a) can be referred to that of the anionic intermediate B in the photocycle of wtGFP, while the structure (b) refers to the anionic intermediate with a low barrier hydrogen bond (LBHB), similar to that of I_2 in wtGFP, but for the proton of E222. This transient is hence characterized by a blue-shifted absorption.

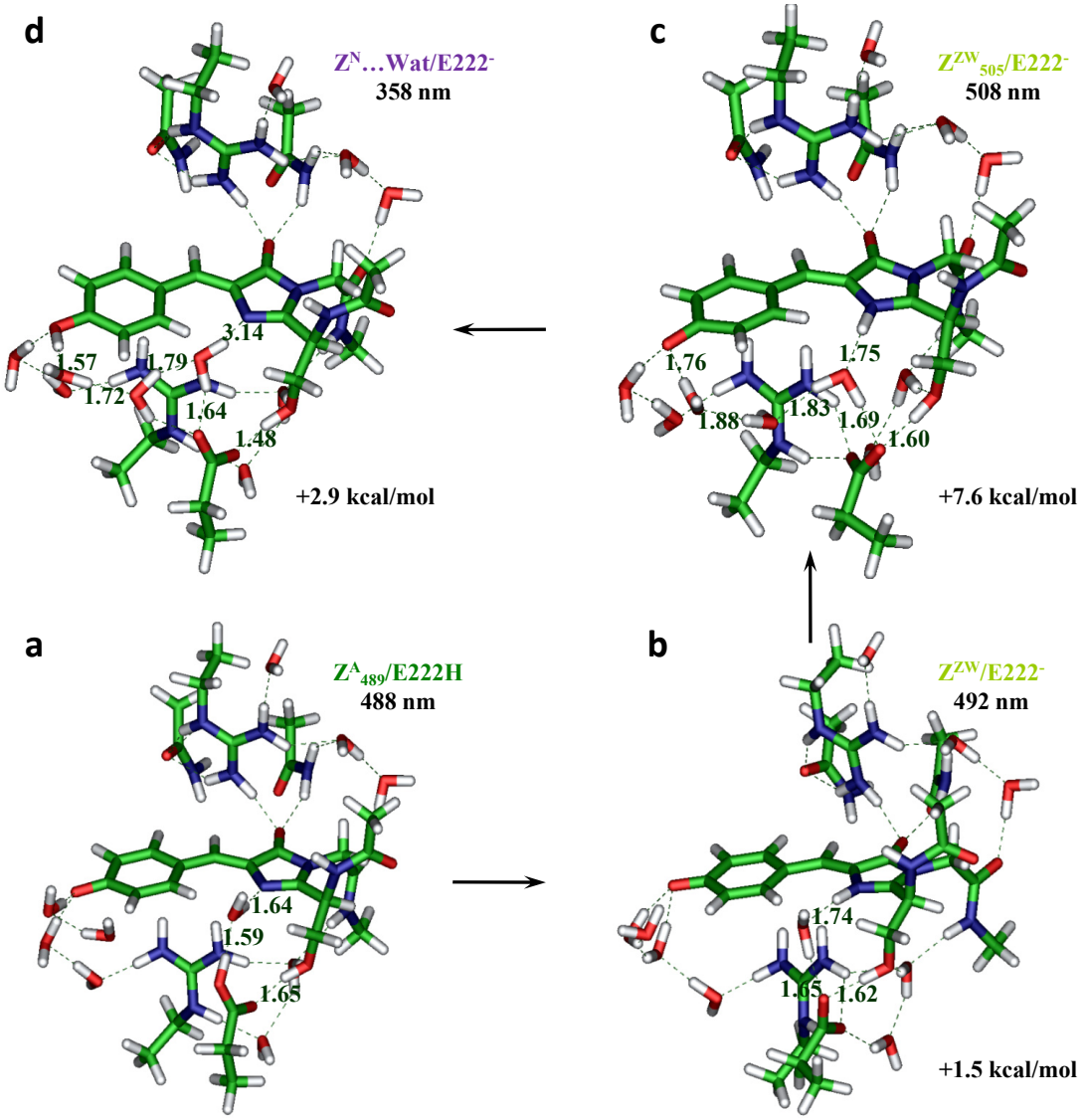

Figure S15. The QM/MM ground-state equilibrium structures of the transients involved in branch II of the photocycle of mKalama1: (a) anionic, (b,c) zwitterionic and (d) neutral forms of the Cro inside the protein formed upon the intraprotein proton transfer. Shown are the QM parts only. The structures are labeled according to the charge state of E222 and to the assigned experimental absorption maxima. The calculated vertical excitation energies are shown in black. All distances are in Å.

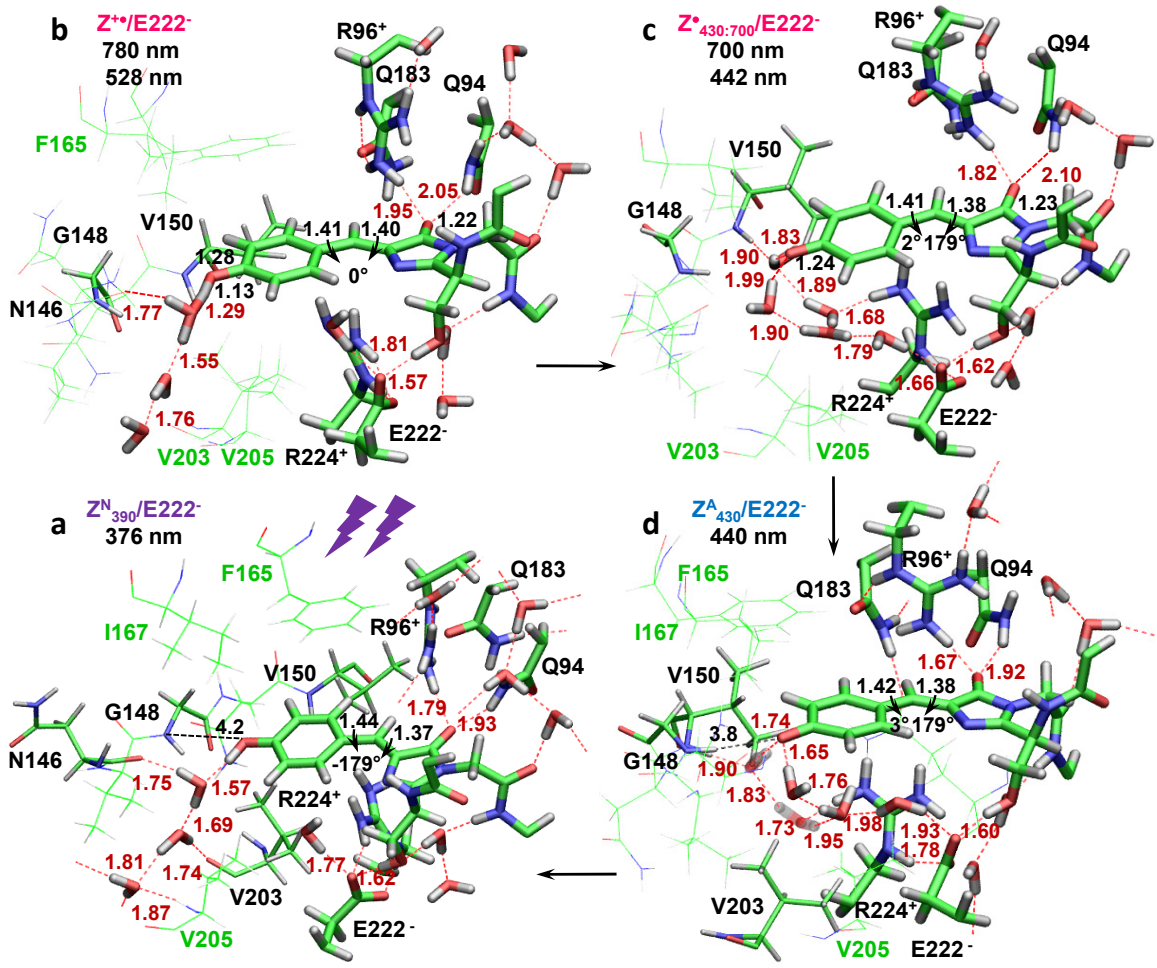

Figure S16. The QM/MM ground-state equilibrium structures of (a) the dark-adapted ground state, (b) the radical cation, (c) the neutral radical, and (d) the long-lived transient anionic state, involved in branch III of the photocycle of mKalama1. This radical branch is initiated by a sequential absorption of two photons and coupled to the proton release to the bulk. Shown are the QM parts (sticks) and the nearby MM residues (lines and bold lines). The structures are labeled according to the charge state of E222 and to the assigned experimental absorption maxima. Shown also are the calculated vertical excitation energies. All distances are in Å. Note that the structure of the radical cation (b) exhibits only a partially protonated Cro, with a low barrier hydrogen bond between the Y66 hydroxyl group and the adjacent water molecule. The regeneration of the dark-adapted ground state (a) from the neutral radical state (c) is thought to occur step-wise through a radical recombination with an electron (d), which is followed by slow reprotonation from the bulk.

7. H/D kinetic isotope effect measurements

For the kinetic measurements in D_2O, the concentrated protein samples were diluted in 99.9% D_2O-based PBS buffer (pD 7.5); the final H_2O concentration in the samples did not exceed 2%. pD was adjusted with a pH-meter, taking into account the typical pH correction for the pH glass electrode: $pD = pH^{app}(D_2O) - 0.4$.[23] The samples were incubated for 14 hours in the D_2O medium before measurements. The comparison of the kinetics at 505 nm in H_2O and D_2O are shown in Figure S17.

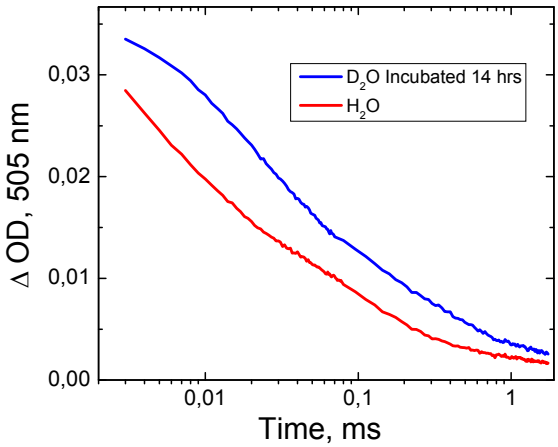

Figure S17. Kinetic traces of the transient absorption signal at 505 nm for mKalama1 in H_2O (red) and in D_2O (blue). The kinetics in H_2O was scaled to match the intensity of that in D_2O.

8. Laser flash-triggered, transient absorption of GFP T203V/S205V

The transient absorption of GFP T203V/S205V was recorded following excitation at 355 nm.

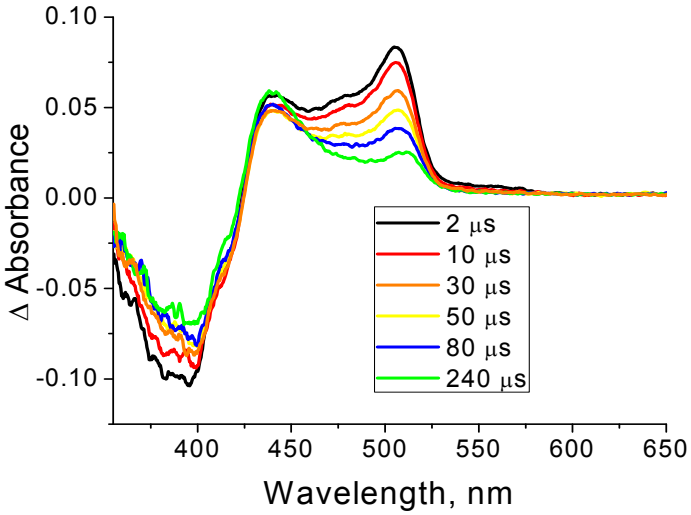

Figure S18. Transient absorption spectra of GFP S203V/T205V.

9. Power dependence of transient absorption changes in mKalama1

The power dependence measurements of the transient signal of mKalama1 (after 355 nm excitation) were performed by adjusting a flash lamp Q-switch time. Figure S19 shows the TA spectra of mKalama1 at different flash energies at both 3 μs and 100 μs. The spectral decomposition of the TA data from Figure S19 gives the amplitude of each component as a function of flash energy (Figure S20). The data were fitted using a hyperbolic function:

$$A = A_{\max} \times \left(1 - \frac{1}{1 + \left(\frac{E}{E_{1/2}} \right)^{n}} \right) \qquad \text{Eq. S2}$$

where A is the observed amplitude, E is the ambient laser energy, $A_{\max}$ and $E_{1/2}$ are the two fitting variables. After fitting, the data and the fit function were scaled for $A_{\max}$ taken as 100%. All major kinetic phases were adequately fitted with n=1 hyperbolic titration curves; the two faster phases (τ, *ca*. 6 μs and 100 μs) were titrated with half-maximum energy of 325 mJ/cm^2, whereas the slower phase, showing the decay of the "430 nm" form, was titrated with half-maximum energy of 700 mJ/cm^2. For comparison, when a quadratic hyperbolic function (n=2) was also used to fit the same data sets, there was no adequate fit (Figure S20).

Due to the relatively long pulse duration of 4 ns, not only excitation but also deactivation contributed to the observed energy sensitivity, hence giving rise to the hyperbolic fit in equation S2. Despite the fact that all major kinetic phases at 490 nm, 505 nm, 430 nm, and 700 nm were adequately fitted with the n=1 hyperbolic curve, indicating one-photon absorption conditions (Figure S20), a two-photon process (n=2) can also result in approximately linear power dependence within a certain range of flash energies; in particular, when two photons are absorbed sequentially via the first excited state.[24]

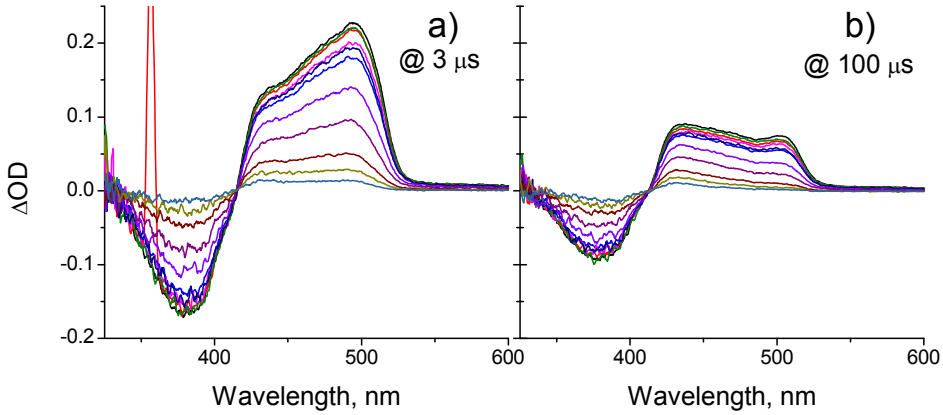

Figure S19. Flash-photolysis of mKalama1 at 355 nm at different flash energies. (a) TA spectra at 3 μs after the flash. (b) TA spectra at 100 μs after the flash. The maximum flash energy was *ca.* 500 mJ/cm^2. Laser pulse FWHM, 4 ns. Fresh samples were used for each shot.

The present results indicate the existence of multiple deexcitation pathways, most likely induced by both one- and two-photon excitation processes. At least a partial spectral overlap in the absorption of transients is expected in the region of 400-505 nm within a sub-ms time window. When amplitudes of spectral components as a function of flash energy exhibit a mixed behavior, the power dependence should appear linear at relatively low flash energies, dominated by one-photon absorption processes. Therefore, a clear signature for the two-photon excitation mechanism should be revealed for the components in the red part of the spectrum, ascribed to the absorption of radical transients formed through multiphoton ionization only. For the component at 700 nm, the observed linear (n=1) power dependence at flash energies larger than 20 mJ/cm^2 cannot rule out a resonantly-enhanced two-photon absorption mechanism[24]. In this case, the quadratic behavior can only be possible to observe at very low flash energies down to 1 mJ/cm^2.[25] Unfortunately, the signal-to-noise ratio does not allow us to analyze the power dependence at very low flash energies due to a very weak absorption in this part of the spectrum. Nevertheless, based on all available data, the consecutive two-photon absorption mechanism is considered to be likely in photoionization of mKalama1.

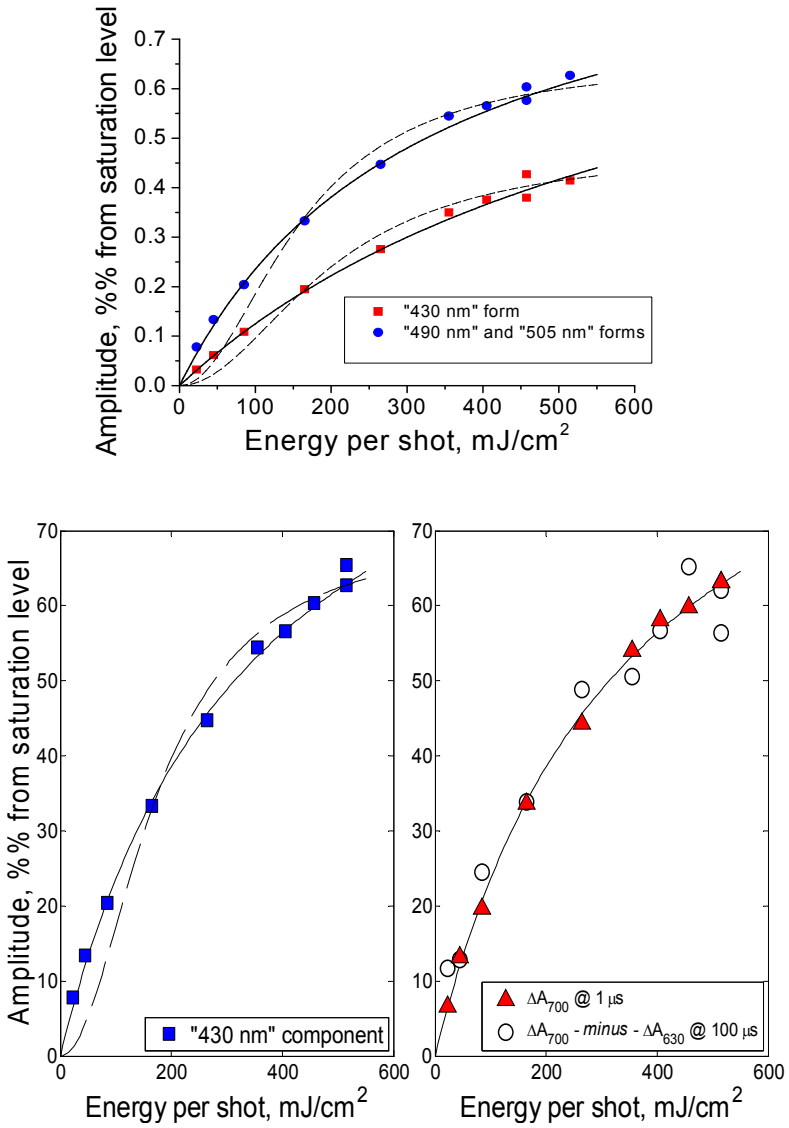

Figure S20. Flash-photolysis of mKalama1 at 355 nm as a function of flash energy. Upper panel: contributions of the fast-relaxing forms ("490 nm" and "505 nm", red symbols) and of the slow-relaxing, reversible form ("430 nm", blue symbols) as determined at each energy by a spectral decomposition of the wavelength-time-absorption matrices. The spectra are shown in Figure S19. The half-saturation energies are found by fitting the data with a single-quantum (n=1) hyperbolic function (solid lines). The dashed lines show the best fit results of the same data sets with a two-quanta (quadratic, n=2) hyperbolic function. Lower panel: saturation curves in the power dependence of the TA for 430 and 700 nm spectral components. The contribution of the radical form at its pure line around 700-710 nm is determined either as the light-*minus*-dark absorption change (ΔA) at 700 nm (at 1 μs following the laser flash, red symbols) or the respective ΔA difference at 700 and 630 nm (at 100 μs following the laser flash, open symbols). The spectra are shown in Figure S6. Same line with $E_{½}$ = 345 mJ/cm2, n=1, is shown for both spectral components at 430 and 700 nm.

10. Transient absorption kinetics of mKalama1 in aerobic *vs* anaerobic samples

Anaerobic samples were prepared using a gas/vacuum line of local design. A 4 × 10 × 30 mm fluorescence quartz cell was sealed to a Kimble-Kontes high-vacuum stopcock, which had a vacuum O-ring connection to the vacuum line. As a typical degassing procedure, 20 cycles of exchanging vacuum (10^{-3} bar) and pure Ar were followed by shaking the cell filled with Ar for 15 min on an orbital shaker; then the procedure was repeated 2 more times. 99.99% Ar was additionally purified using the Agilent Technologies BOT-2 and IOT oxygen scrubbers. Aerobic samples were bubbled with oxygen for 10 minutes. For the aerobic and anaerobic measurements, the data was fit using only the 1 µs - 1.6 ms data. The kinetic traces were decomposed into the corresponding spectra and lifetimes (Figure S21). The lifetime of the 730 µs component (Figure 1e, main text) decreased in the oxygen saturated sample (426 µs, Figure S21) and increased in the sample depleted of oxygen (1.02 ms, Figure S21). The kinetic traces of the 625 nm signal under anaerobic and aerobic conditions are shown in Figure S22.

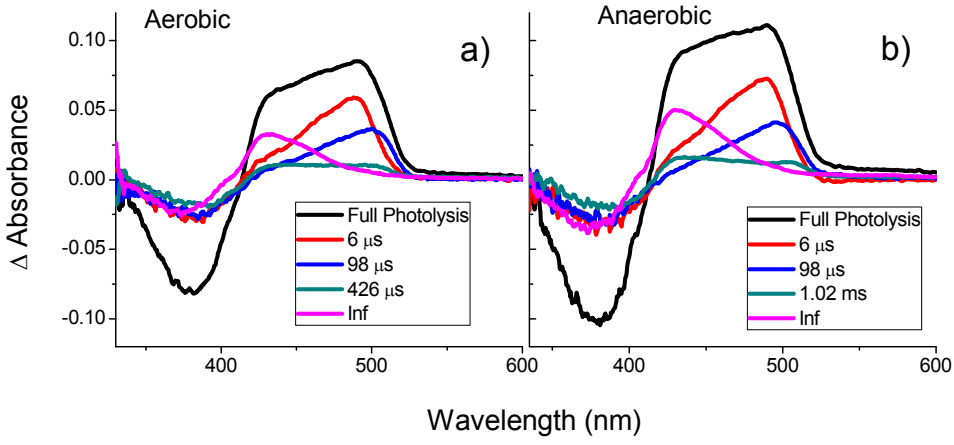

Figure S21: TA spectra of mKalama1 as a function of time under (a) aerobic conditions and (b) anaerobic conditions.

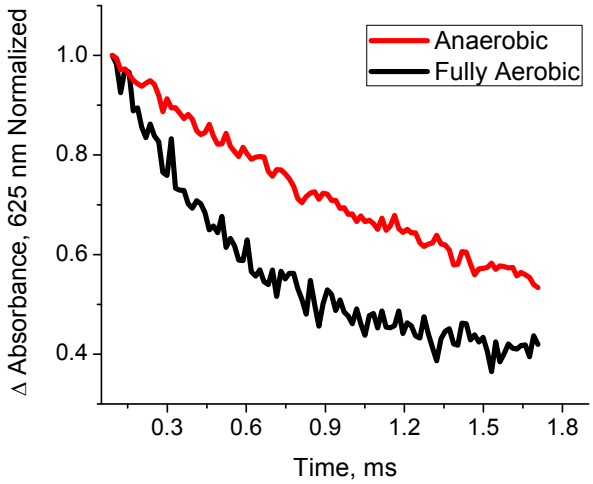

Figure S22: Kinetics of the TA signal at 625 nm under anaerobic conditions (red) and fully aerobic conditions (black).

11. Laser flash-triggered, transient absorption of wtGFP

The transient absorption of wtGFP was recorded following excitation at 355 nm.

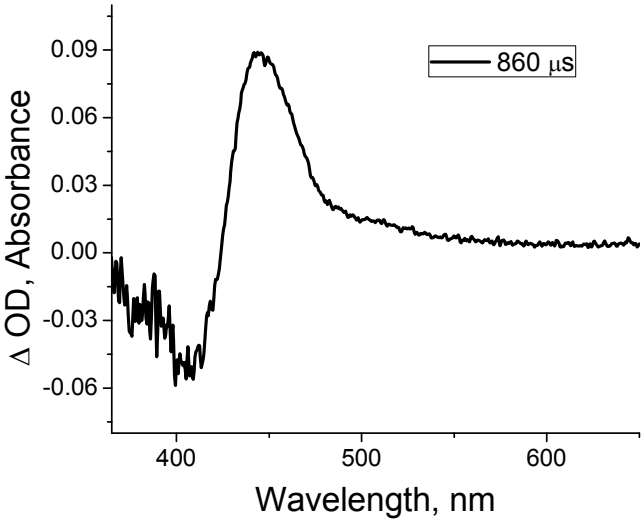

Figure S23. Transient absorption spectrum at 860 μs for wtGFP (pH 7.5).

References

1 (a) Evans, P. *Acta Crystallogr. D* **2006**, *62*, 72. (b) Kabsch, W. *Acta Crystallogr. D* **2010**, *66*, 125. (c) Winter, G. *J. Appl. Cryst*. **2010**, *43*, 186.

2 McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. *J. Appl. Cryst*. **2007**, *40*, 658.

3 Shu, X.; Leiderman, P.; Gepshtein, R.; Smith, N. R.; Kallio, K.; Huppert, D.; Remington, S. J.; *Protein Sci.* **2007**, *16*, 2703.

4 Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L. W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H. *Acta Crystallogr. D* **2010**, *66*, 213.

5 Emsley, P.; Cowtan, K. *Acta Crystallogr. D*. **2004**, *60*, 2126.

6 Laskowski, R. A.; Rullmannn, J. A.; MacArthur, M. W.; Kaptein, R.; Thornton, J. M. *J. Biomol. NMR* **1996**, *8*, 477.

7 Belevich, I.; Gorbikova, E.; Belevich, N. P.; Rauhamäki, V.; Wikström, M.; Verkhovsky, M. I. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 18469.

8 Borisov, V. B.; Belevich, I.; Bloch, D. A.; Mogi, T.; Verkhovsky, M. I. *Biochemistry* **2008**, *47*, 7907.

9 Belevich, I.; Bloch, D. A.; Belevich, N.; Wikström, M.; Verkhovsky, M. I. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 2685.

10 Bloch, D. A.; Jasaitis, A.; Verkhovsky, M. I. *Biophys. J.* **2009**, *96*, 4733.

11 Rakowski, A. *Z. Phys.Chem. Stochiom.Verwandtschafts*. **1906**, *57*, 321.

12 Bell, R. P. *J. Mol. Struct.* **1976**, *33*, 173.

13 Ai, H.-W.; Shaner, N. C.; Cheng, Z.; Tsien, R. Y.; Campbell, R. E. *Biochemistry* **2007**, *46*, 5904.

14 MacKerell Jr, A.; Bashford, D.; Bellott, M.; Dunbrack Jr, R.; Evanseck, J.; Field, M.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F.; Mattos, C.;

Michnick, S.; Ngo, T.; Nguyen, D.; Prodhom, B.; Reiher III, W.; Roux, B.; Schlenkrich, M.; Smith, J.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586.

15 Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33.

16 Phillips, J.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R.; Kale, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781.

17 Glendening, E. D.; Badenhoop, J. K.; Reed, A. E.; Carpenter, J. E.; Bohmann, J. A.; Morales, C. M.; Weinhold, F. NBO 5.G., Theoretical Chemistry Institute, University of Wisconsin, Madison, WI, https://www.chem.wisc.edu/~nbo5/index.htm.

18 Granovsky, A. A. Firefly, http://classic.chem.msu.su/gran/firefly/index.html.

19 Day, P.; Jensen, J.; Gordon, M.; Webb, S.; Stevens, W.; Krauss, M.; Garmer, D.; Basch, H.; Cohen, D. *J. Chem. Phys.* **1996**, *105*, 1968.

20 Granovsky, A. A. *J. Chem. Phys.* **2011**, *134*, 214113.

21 Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347.

22 Rakowski, A. *Z. Phys. Chem. Stochiom. Verwandtschafts*. **1906**, *57*, 321.

23 Bell, R. P. *J. Mol. Struct.***1976**, *33*, 173

24 Larsen, D. S.; Vengris, M.; van Stokkum, I. H. M.; van der Horst, M. A.; de Weerd, F. L.; Hellingwerf, K. J.; van Grondelle, R. *Biophys. J.* **2004**, *86*, 2538.

25 Zhu, J.; Paparelli, L.; Hospes, M.; Arents, J.; Kennis, J. T. M.; van Stokkum, I. H. M.; Hellingwerf, K. J.; Louise Groot, M. *J. Phys. Chem. B* **2013**, *117*, 11042.