Supporting Information

Color Hues in Red Fluorescent Proteins Are Due to Internal Quadratic Stark Effect

Mikhail Drobizhev[1], Shane Tillo[2], Nikolay S. Makarov[1], Thomas E. Hughes[2], Aleksander Rebane[1,3]

[1]Department of Physics and [2]Department of Cell Biology and Neuroscience, Montana State University, Bozeman, MT 59717, USA

[3] National Institute of Chemical Physics and Biophysics, Tallinn, EE 12618, Estonia

Methods

The DNA encoding the mRFP[1], fruit fluorescent proteins[2] (generous gifts from Roger Tsien, UC San Diego), and tagRFP[3] were positioned in the pRSET plasmid (Invitrogen, Carlsbad CA) such that a fusion protein is produced with an N-terminal 6xHis tag. Transformed E. Coli (T7 express pLysY, New England Biolabs) were grown at 37^0C to an optical density of 0.5 at 600 nm, followed by IPTG induction and growth for eight hours at room temperature. Cells were lysed (Bugbuster, Novagen) and affinity purified with Ni-NTA His Bind resin (Novagen). The proteins were eluted with 250 mM imidazole (pH 8). All samples were studied at pH 8, while mCherry was studied at both pH 8 and pH 11.2.

The methods of measurement of chromophore concentration, extinction, 2PA spectra, and cross sections are described in detail in (*1*). Briefly, to measure chromophore extinction in solution we use the Strickler–Berg (S-B) equation (*1*). This approach makes it possible to find the peak extinction coefficient ε_{max}(S-B) of only

mature chromophore in properly folded protein. The chromophore concentration was then obtained using Lambert-Beer law.

Two-photon absorption measurements were carried out by using a relative fluorescence method with femtosecond excitation (*1*,*2*). Rhodamine B in methanol was used as a standard (*2*).

Evaluation of $\Delta\mu$ was based on the formula relating 2PA cross section of the lowest $S_0 \rightarrow S_1$ 0 - 0 transition, $\sigma_2(0 - 0)$, with $\Delta\mu$ and maximum 1PA extinction coefficient $\varepsilon(0\text{-}0)$ (*3*). If all intermediate electronic states are lying much higher in energy than S_1, and also have small transition dipole moments connecting them to S_0, one can restrict the sum-over-states description of two-photon tensor to only two intermediate states, ground *0* and final two-photon excited state *1* (*3-11*). This two-level approximation will result in the appearance of the change of permanent dipole moments between S_1 and S_0, $\Delta\boldsymbol{\mu} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_0$, in the expression for σ_2 (in cm^4 s) (*3,7-11*):

$$\sigma_2(2\nu) = 2\frac{(2\pi)^4}{(hc)^2}\frac{(1+2\cos^2\gamma)}{15}\frac{f_{opt}^4}{n^2}\left|\Delta\mu\right|^2\left|\mu\right|^2 g(2\nu)\,. \qquad (1)$$

Here *h* is the Planck's constant, *c* is the speed of light, N_A is the Avogadro number, *n* is the refractive index of the medium (*n* = 1.33). Equation (1) implies linear polarization of two laser photons of the same frequency ν (in Hz) and takes into account the space-averaging of 2PA tensor. In (1), $\boldsymbol{\mu}$ is the transition dipole moment (in vacuum) between states $S_{0,0}$ and $S_{1,0}$, (i.e. electronic transition dipole moment times the square root of the Franck-Condon factor of the 0-0 transition), γ is the angle between vectors $\Delta\boldsymbol{\mu}$ and $\boldsymbol{\mu}$, f_{opt} is the local field factor at optical frequency, $g(2\nu)$ is the 2PA line shape function in

(Hz^{-1}), normalized such that $\int_{-\infty}^{\infty} g(2\nu)d(2\nu) = 1$. The integration in the last formula is carried over the 0-0, $S_0 \rightarrow S_1$, transition. Function $g(2\nu)$ describes the line broadening, which we assume to be the same for one- and two-photon absorption.

To obtain $|\Delta\mu|$, we can re-write (1) in terms of the maximum 2PA cross section in the 0-0 peak σ_2(0-0), maximum 1PA extinction coefficient $\varepsilon(0-0)$, and the central frequency of this transition $\bar{\nu}_{0-0}$ (in cm^{-1}). By using a straightforward relation between $|\mu|^2$ and $\varepsilon(0-0)$, which takes into account the local field correction (see e.g. (*12*)):

$$|\mu|^2 g(\nu_{0-0}) = \frac{3\times 10^3 \ln 10\, h}{(2\pi)^3 N_A} \frac{n}{f_{opt}^2} \frac{\varepsilon(0-0)}{\bar{\nu}_{0-0}}, \qquad (2)$$

substituting (2) into (1), and solving (1) with respect to $|\Delta\mu|$ we obtain (cf. (*3*)):

$$|\Delta\mu| = \left(\frac{5}{4(1+2\cos^2\gamma)} \frac{hc^2 N_A}{\pi 10^3 \ln 10} \frac{n}{f_{opt}^2} \frac{\bar{\nu}_{0-0}}{\varepsilon(0-0)} \sigma_2(0-0) \right)^{1/2}, \qquad (3)$$

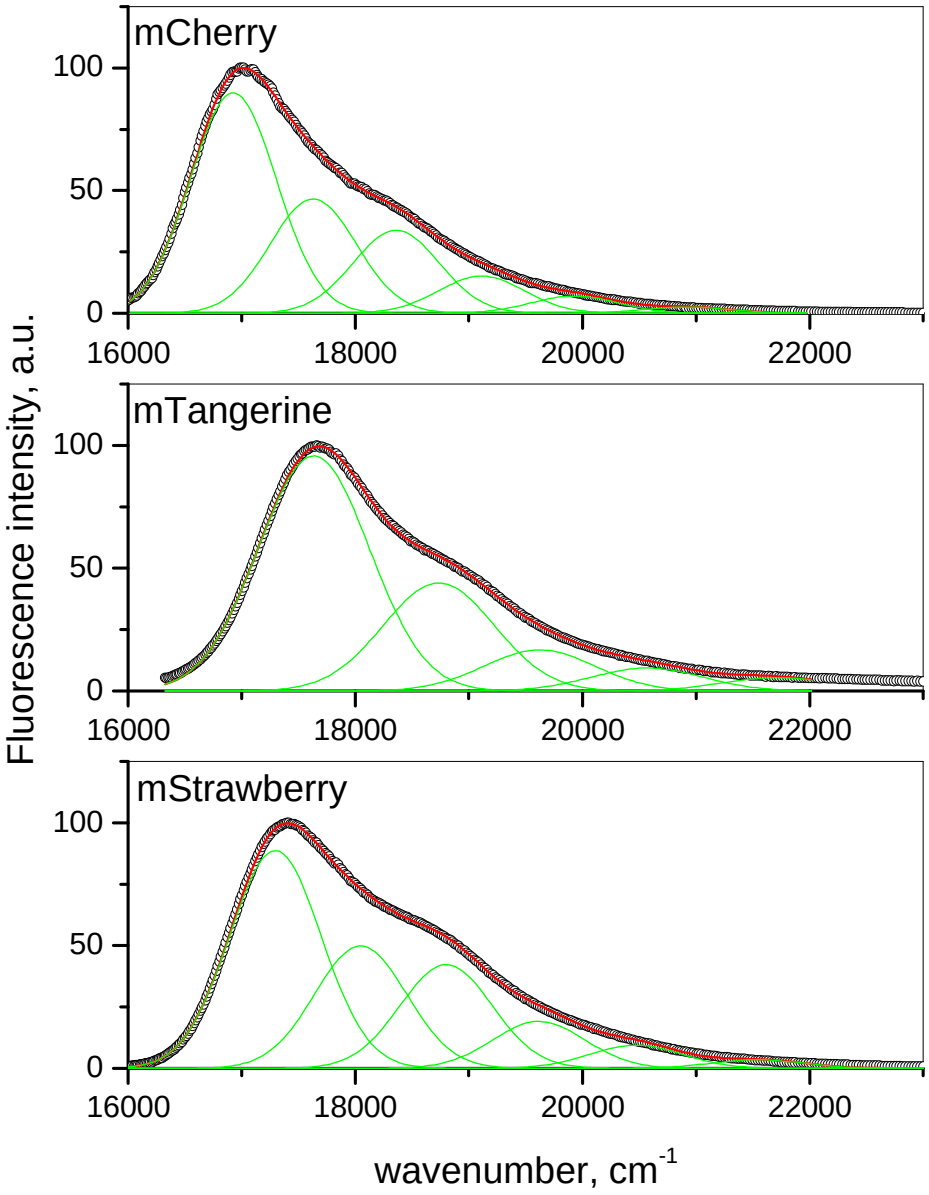

Fig 1S (a). One-photon fluorescence excitation profiles with multi-Gaussian fits for mCherry, mTangerine, and mStrawberry.

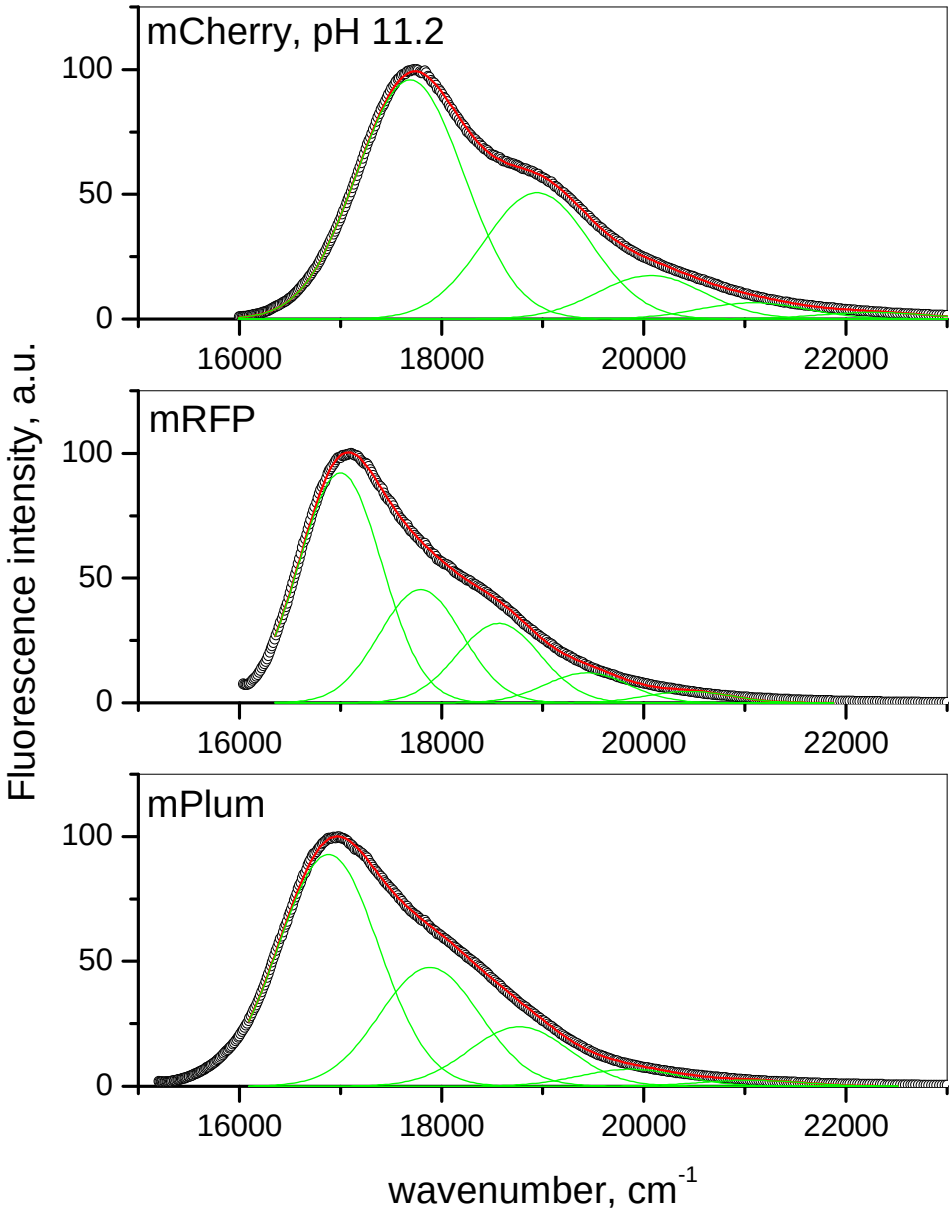

Fig 1S (b). One-photon fluorescence excitation profiles with multi-Gaussian fits for mCherry at pH11.2, mRFP, and mPlum.

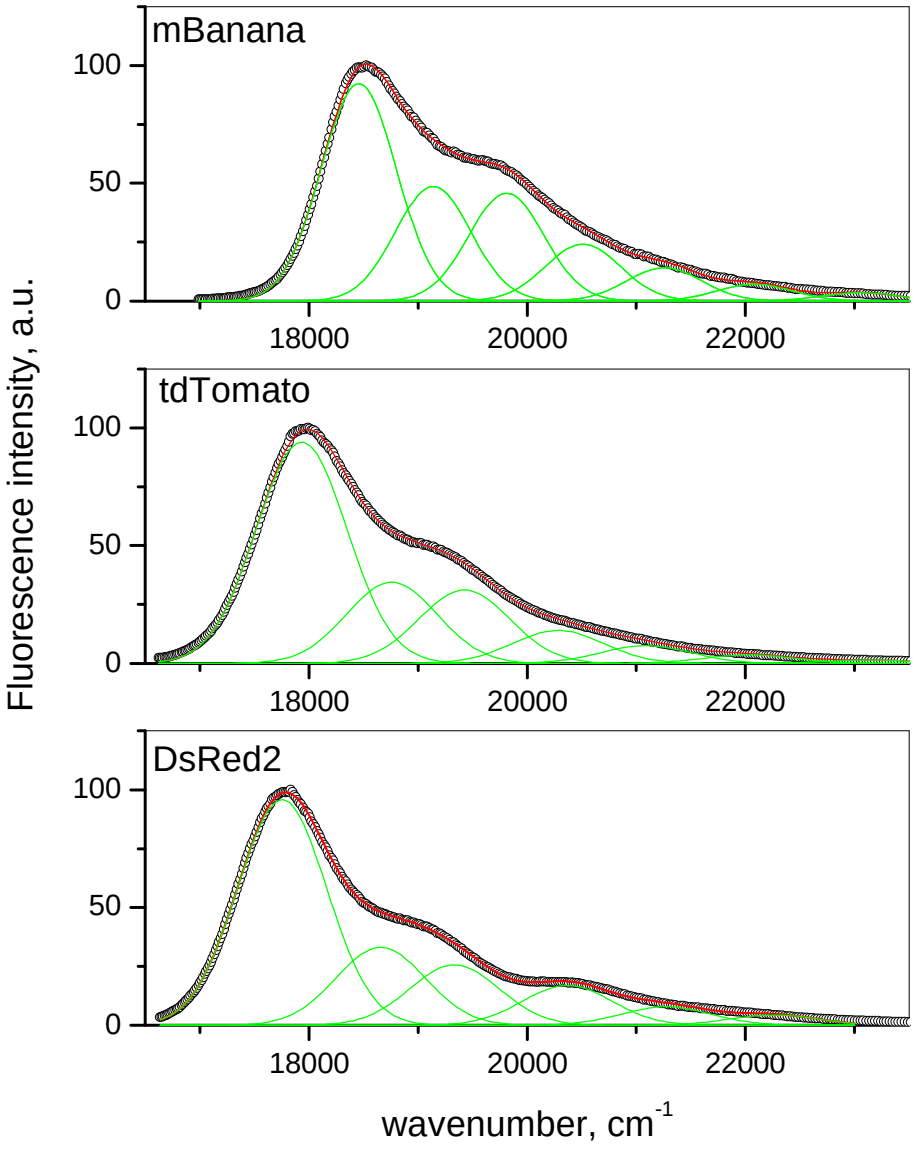

Fig 1S (c). One-photon fluorescence excitation profiles with multi-Gaussian fits for mBanana, tdTomato, and DsRed2.

Figs 1S (a)-(c) show the corrected (to the spectral intensity distribution of excitation monochromator) and calibrated (to the spectral position of excitation monochromator) fluorescence excitation spectra. To obtain the $\bar{\nu}_{0-0}$ and $\varepsilon(0-0)$ values, these spectra were fitted to a sum of 5-7 Gaussians. The attempts to fit the spectra with 4 Gaussians gave unsatisfactory results (with large systematic deviations). In each case, we first tried to fit the spectrum with 5 Gaussians. This was done by fixing the Gaussian widths w_i (i = 1, 2, ..5) of all components equal to one particular value w. Then the w value was systematically varied to get the minimum χ^2 of the total multi-Gaussian fit. If after this procedure the systematic deviations of the fit (especially in the high-frequency region) still remained, the number of Gaussians was increased subsequently to 6 or 7 and the routine was repeated to obtain the minimum χ^2. The final multi-Gaussian fits (where all $w_i = w$) are shown in Figs. S1 (a)-(c) with the red line. Note that in most cases the obtained distribution of peaks (shown with green lines) corresponds well to two Franck-Condon progressions, built on two main characteristic modes with similar (for all proteins) Franck-Condon factors and similar frequencies. Also note that the 0-0 (lowest-frequency) transition is the dominant transition in all cases and therefore its frequency and amplitude are well defined. The assignment of this transition to 0-0 (in terms of vibronic excitation) is further supported by hole burning spectroscopy of DsRed presented in (*13*). In other proteins, this assignment also holds because of their spectra are quite similar in shape. We estimate an error in position of $\bar{\nu}_{0-0}$ to be ± 50 cm^{-1} (~ 2 nm). The $\varepsilon(0-0)$ value was obtained from the above fits by comparing the amplitude of the lowest frequency Gaussian peak with the known value ε_{max} at the

spectral maximum. In all cases $\varepsilon(0-0)$ constitutes 0.89 – 0.96 of ε_{max}. The main error of measurement of $\varepsilon(0-0)$ comes from determination of ε_{max} and constitutes ~15%. Note, however, that the measurement of $\Delta\mu$ is not sensitive to the errors in ε_{max} because the experimentally determined σ_2(0-0) depends linearly on the measured $\varepsilon(0-0)$, but the formula (1) for $\Delta\mu$ involves the fraction $\sigma_2(0\text{-}0)/\varepsilon(0-0)$. Also note that $\Delta\mu$ is calculated as square root of a combination of measured parameters, including σ_2(0-0), see (1). Out of these parameters, the largest error is contributed by σ_2(0-0), which is ~ 20%. Therefore our estimation of experimental error in $\Delta\mu$, due to a random error of σ_2(0-0) is of the order of 10%.

The choice of local field factor and refractive index inside the protein particle, entering (3), is a little bit arbitrary. We select $n^2 = 1.8$ ($n = 1.33$), giving for the Lorentz local field factor $f_{opt} = (n^2+2)/3 = 1.27$. (Simple analysis of function $f_{opt}{}^2/n$ shows that the systematic error in $|\Delta\mu_{10}|$, which can arise as a result of error in n is less than 3% when n is allowed to vary in a maximum reasonable range, i.e. between 1.33 and 1.5).

The angle γ between $\Delta\boldsymbol{\mu}$ and the $S_0 \rightarrow S_1$ transition dipole moment $\boldsymbol{\mu}$ was obtained by measuring the 2PA isotropic polarization ratio $\Omega = \sigma_{2,cir}/\sigma_{2,lin}$, where $\sigma_{2,lin}$ is the 2PA cross section measured with linear (vertical) polarization of excitation and fluorescence analyser set 54.7^0 with respect to the polarization of the excitation light; $\sigma_{2,cir}$ is the 2PA cross section obtained with circular polarization of excitation and fluorescence analyser set 35.3^0 with respect to the polarization plane of the excitation light (*14*). In the two-level approximation employed here, Ω and γ are related through the following equation (*15*):

$$\Omega = \frac{\cos^2 \gamma + 3}{4\cos^2 \gamma + 2}. \quad (7)$$

For all the proteins we obtained $\gamma \approx 0^0$. If experimental errors were considered, the maximum possible γcould be ~ 20^0, which contributed less than 10% to the final $\Delta\mu$ value. These estimations are in good agreement with the small angle $\gamma = 13^0$ obtained for wt-DsRed using Stark effect spectroscopy (*16*).

An additional, systematic, error in determining σ_2(0-0), and, consequently, $\Delta\mu$, can emerge from selecting a particular method of fitting the 2PA spectrum. We have attempted several models (methods of fitting) to describe the 2PA spectra. In all these methods the frequency of the 0-0 transition was not allowed to vary and was fixed equal to the $\bar{\nu}_{0-0}$ value obtained from 1PA data (see above).

Method 1 (Fig. 1 of the main text). The main part of the 2PA spectrum was fitted with two Gaussians with fixed $\bar{\nu}_{0-0}$ and all other parameters free. In the case of mCherry at pH11.2, an ambiguity in the σ_2(0-0) value was rather large because σ_2(0-0) depended strongly on how to select the upper frequency limit in the fitting procedure. In this case, we have chosen an average σ_2(0-0), obtained from a set of values, emerging from a variation of the upper frequency limit. The resulting spread in σ_2(0-0) values was not larger than 30%, translating into 15% variation in $\Delta\mu$.

Method 2. The whole 2PA band, corresponding to the S0 → S1 transition was fitted with 3 Gaussians. Only $\bar{\nu}_{0-0}$ was fixed. The results are shown in Fig. 2S.

Method 3. The main part of the 2PA spectrum was fitted with two Gaussians with both $\bar{\nu}_{0-0}$ and transition width w_1 fixed and equal to the numbers obtained from 1PA fits. Other parameters were free.

Method 4. The whole 2PA band, corresponding to the $S_0 \rightarrow S_1$ transition was fitted with 3 Gaussians. Both $\bar{\nu}_{0-0}$ and w_1 were fixed (as in Method 3). Other parameters were free.

Method 5. The whole 2PA band, corresponding to the $S_0 \rightarrow S_1$ transition was fitted to 3-7 Gaussians, all with the same widths, w_i, fixed, and equal to the width w obtained from 1PA spectra. $\bar{\nu}_{0-0}$ was also fixed as in methods 1 - 4.

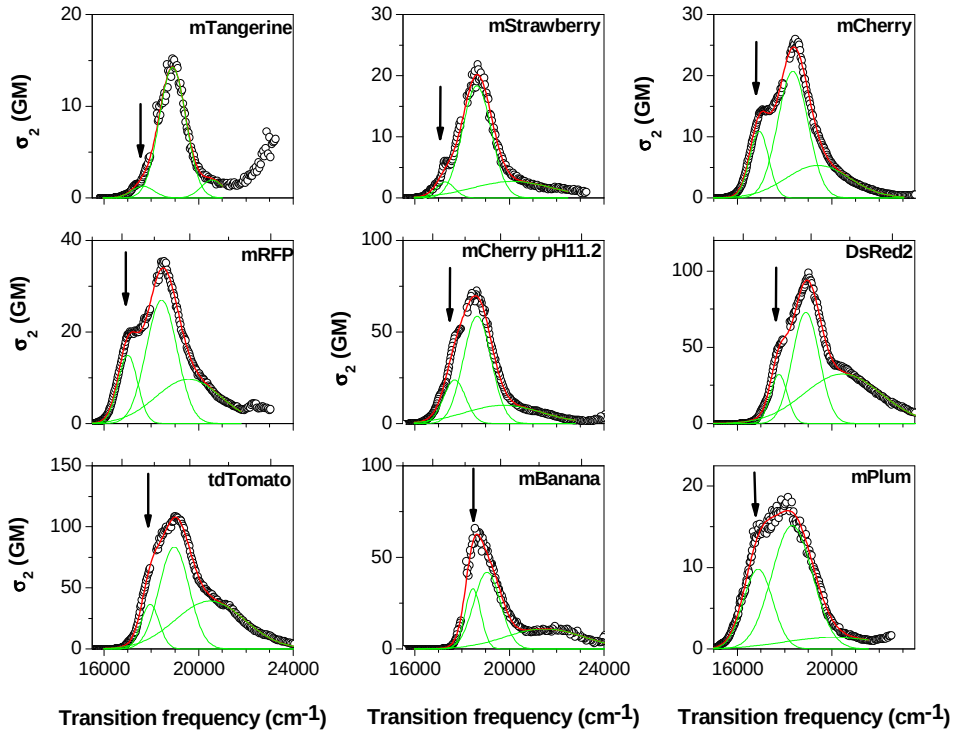

Fig. 2S. 3-Gaussian fits to the 2PA spectra with fixed position of the 0-0 transition (lowest-frequency peak), corresponding to Method 2.

Using methods 1 - 5, the $\sigma_2(0\text{-}0)$ values were extracted and $\Delta\mu$ values calculated, according to (1). Finally, the plots, similar to Fig. 2 were produced and presented in Fig 3S. The second order polynomial fits are shown for each method.

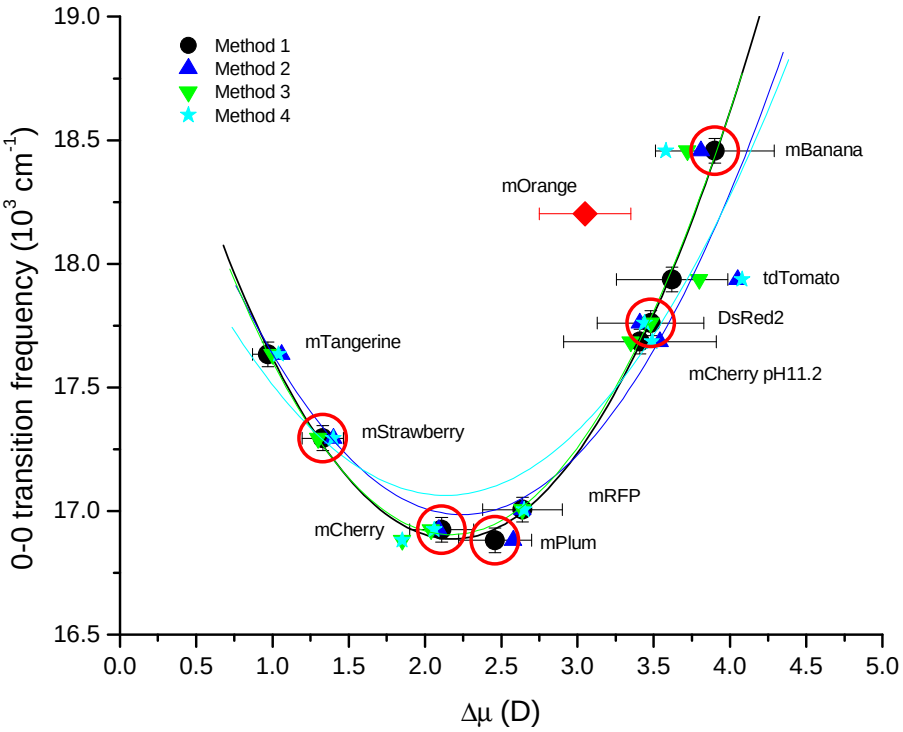

Fig. 3S. The dependence of the 0-0 transition energy (obtained from 1PA spectra) on Δμ (obtained from 2PA cross sections in the 0-0 transition maximum). Different symbols correspond to different methods of fitting 2PA spectra. The proteins with crystallographically solved structures are highlighted in red circles. All of the Fruits proteins with the same chromophore structure are well described by the model, while mOrange, whose chromophore structure is different (*17*), is not.

As one can see, in all cases the data points for 9 proteins group around the parabola obtained with method 1 (black line) and discussed in the main text. The largest deviations from the fits are observed for tdTomato when using Methods 2 and 4 (i.e. 3-Gaussians fits). Table 1S presents the parameters of the second order polynomial fits for all 5 methods with the corresponding statistical analysis. The results of the fits, obtained

with methods 2 and 4, but excluding tdTomato (total number of points = 8), are also presented and provide an improvement of correlation coefficients, compared to the fits which use all 9 points. As one can seen from the Table, the estimated correlation coefficient of the parabolic dependence is always larger than 0.8 and P-value smaller than 4% and, in most cases, P < 1%. For each protein, the different fitting methods produce $\Delta\mu$ values that vary by less than 30%.

Table S1. Statistical analysis of the data obtained with second-order polynomial fits of the $\bar{\nu}_{0-0}$ versus $\Delta\mu$ dependences, shown in Fig. 3S.

Method #	Number of points	R	SD	P	A = ν_0 (cm^{-1})	B	C	$\Delta\mu_0$ (D)	$\Delta\alpha_0$ ($Å^3$)
1	9	0.998	32.9	<0.0001	**19383** ± 77	-2287 ± 69	524 ± 14	**4.36 ± 0.18**	**-19** ± 0.5
2	9	0.894	272	0.0080	19107 ± 667	-1891 ±572	422 ±110	4.5 ± 1.8	-24 ± 6
2	8	0.983	115	0.0002	19789 ± 308	-2626 ± 277	590 ± 56	4.45 ± 0.63	-17 ±1.6
3	9	0.949	193	0.0010	19311 ± 496	-2217 ± 457	511 ± 92	4.3 ± 1.2	-19.6 ± 3.5
4	9	0.810	357	0.0405	18648 ± 904	-1484 ± 794	348 ± 154	4 ± 3	-29 ± 13
4	8	0.945	208	0.0038	19882 ± 630	-2783 ± 589	639 ±121	4.4 ± 1.2	-16 ±3
5	9	0.900	265	0.0069	18840 ± 623	-1379 ± 444	265 ± 71	5 ± 2	-38 ± 10
5	8	0.975	141	0.0005	19331 ±352	-1832 ±261	355 ±44	5 ±1	-28 ±3.5

R is the correlation coefficient, P is the P-value showing the probability of the absence of correlation with the given correlation coefficient, SD – standard deviation, A, B, and C are the coefficients of parabola, ν_0, $\Delta\mu_0$ and $\Delta\alpha_0$ are the chromophore parameters defined in the text.

References

1. M. Drobizhev, S. Tillo, N. S. Makarov, T. E. Hughes, A. Rebane, *J. Phys. Chem. B* **113**, 855-859 (2009).
2. N. S. Makarov, M. Drobizhev, A. Rebane, *Opt. Express* **16**, 4029-4047 (2008).
3. M. Drobizhev, N. S. Makarov, T. Hughes, A. Rebane, *J. Phys. Chem. B* **111**, 14051-14054 (2007).
4. P. R. Callis, T. W. Scott, A. C. Albrecht, *J. Chem. Phys.* **75**, 5640-5646 (1981).
5. B. Dick, G. Hohlneicher, *J. Chem. Phys.* **76**, 5755-5760 (1982)
6. R. Birge, C.-F. Zhang, *J. Chem. Phys.* **92**, 7178-7195 (1990).
7. P. Feneyrou, *et al. Opt. Lett.* **22**, 1132-1134 (1997).
8. A. Miniewicz, S. Delysse, J.-M. Nunzi, F. Kajzar, *Chem. Phys. Lett.* **287**, 17-21 (1998).
9. M. Drobizhev, *et al. J. Phys. Chem. B* **110**, 9802-9814 (2006).
10. M. Drobizhev, N. S. Makarov, A. Rebane, G. de la Torre, T. Torres, J. Phys. Chem. C **112**, 848-859 (2008).
11. A. Rebane, *et al. J. Phys. Chem C* **112**, 7997-8004 (2008).
12. D. Toptygin, *J. Fluorescence* **13**, 201-219 (2003).
13. S. Bonsma, *et al.*, *ChemPhysChem* **6**, 838-849 (2005).
14. C. Z. Wan, C. K. Johnson, *Chem. Phys.* **179**, 513-531 (1994).
15. W. J. Meath, E. A. Power, *J. Phys. B – Atom. Mol. Opt. Phys.* **17**, 763-781 (1984).

16. B. Lounis, J. Deich, F. I. Rosell, S. G. Boxer, W. E. Moerner, *J. Phys. Chem. B* **105**, 5048-5054 (2001).

17. X. Shu, N. C. Shaner, C. A. Yarbrough, R. Y. Tsien, S. J. Remington, *Biochem.* **45,** 9639-9647 (2006).