Supplemental Information: Observation of robust energy transfer in the photosynthetic protein allophycocyanin using single-molecule pump-probe spectroscopy

Raymundo Moya[1], Audrey C. Norris[1], Toru Kondo[2], Gabriela S. Schlau-Cohen[1*]

[1]Department of Chemistry, Massachusetts Institute of Technology,
77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA

[*]To whom correspondence should be addressed; E-mail: gssc@mit.edu

[2]Department of Life Science and Technology, Tokyo Institute of Technology

1 Fit function used for parameter estimation

For the single-molecule pump-probe (SM2P) experiments, the fluorescence intensity was modeled as a function of the delay time as previously reported [1–3]. A complete derivation is shown in Ref. [4], but a brief explanation is illustrated below.

The modulated fluorescence intensity as a function of time delay, $I(\tau)$, is given by the convolution of the autocorrelation of the Gaussian pulse with the δ-pulse excitation solution as shown below: [5–7].

$$I(\tau) \propto \int AC(t-\tau)N_2(t)dt \tag{1}$$

where

$$AC(\tau) = \int f(t)f(t-\tau)d\tau \tag{2}$$

$AC(\tau)$ is the intensity autocorrelation assuming Gaussian pulses, $f(t) = \frac{\sqrt{2}}{d\sqrt{\pi}}e^{\frac{-2t^2}{d^2}}$, where d is the adjusted pulse FWHM given by $d = \frac{d_{fwhm}}{\sqrt{2ln(2)}}$, τ is the time elapsed before the second δ-pulse arrives, t is a dummy variable for integration, and N_2 is the population in the second, emissive state.

Equation 1 leads to the equation derived and used by the van Grondelle group [1]:

$$I(\tau) \propto I_\infty \left(1 - \frac{p}{2-p}\frac{1}{2}e^{\frac{k_{er}^2 d^2 \tau}{4}}\left(e^{-k_{er}\tau}erfc\left(\frac{d^2k_{er}-2\tau}{2d}\right) + e^{k_{er}\tau}erfc(\frac{d^2k_{er}+2\tau}{2d}\right)\right)\right) \tag{3}$$

There are three fit parameters in Equation 3: the baseline intensity, I_∞, the probability of excitation by a single pulse, p, and the energy transfer rate, $k_{er} = \frac{1}{\tau_{er}}$. The data was fit to this function through a custom Matlab program that minimized the negative log-likelihood function (LL):

$$-LL = -2\sum_i (C_i - M_i) + C_i\ log\left(\frac{M_i}{C_i}\right) \tag{4}$$

where M_i are the counts predicted by the intensity model (Equation 3) and C_i are the experimental counts detected [8].

2 Characterization of pulse duration for SM2P

The optical pulses were characterized with an interferometric autocorrelation (IAC) at the sample plane of the single-molecule pump-probe setup. A schematic of the setup is shown in Fig. S1. A GaP photodiode (Marktech, MTPD3650D-1.4), which has a bandgap above the single photon energy, was placed at the sample position as in previous work [9–11]. Sample IAC's are shown in Fig. S2 for 610nm (a,b; compressed 118 fs, c,d; uncompressed 180 fs) and 645 nm (e,f; uncompressed 303 fs). The intensity autocorrelations (b, d, f) were extracted by Fourier transforming the IAC and filtering out the coherent oscillations. The intensity autocorrelations were fit assuming a Gaussian envelope [11]. All data manipulation was completed in Matlab.

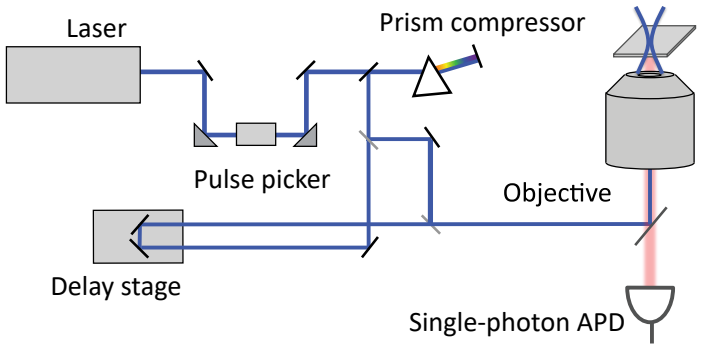

Figure S1: Single-molecule pump-probe spectroscopy. Simplified schematic of the apparatus in which a femtosecond laser proceeds through an acousto-otpic pulse picker, a single-prism pulse compressor [12], and then through a set of beam'splitters and a delay stage in a Mach-Zehnder configuration. The laser pulse is focused onto the sample into a diffraction-limited spot by an objective and emission is collected through the same objective.

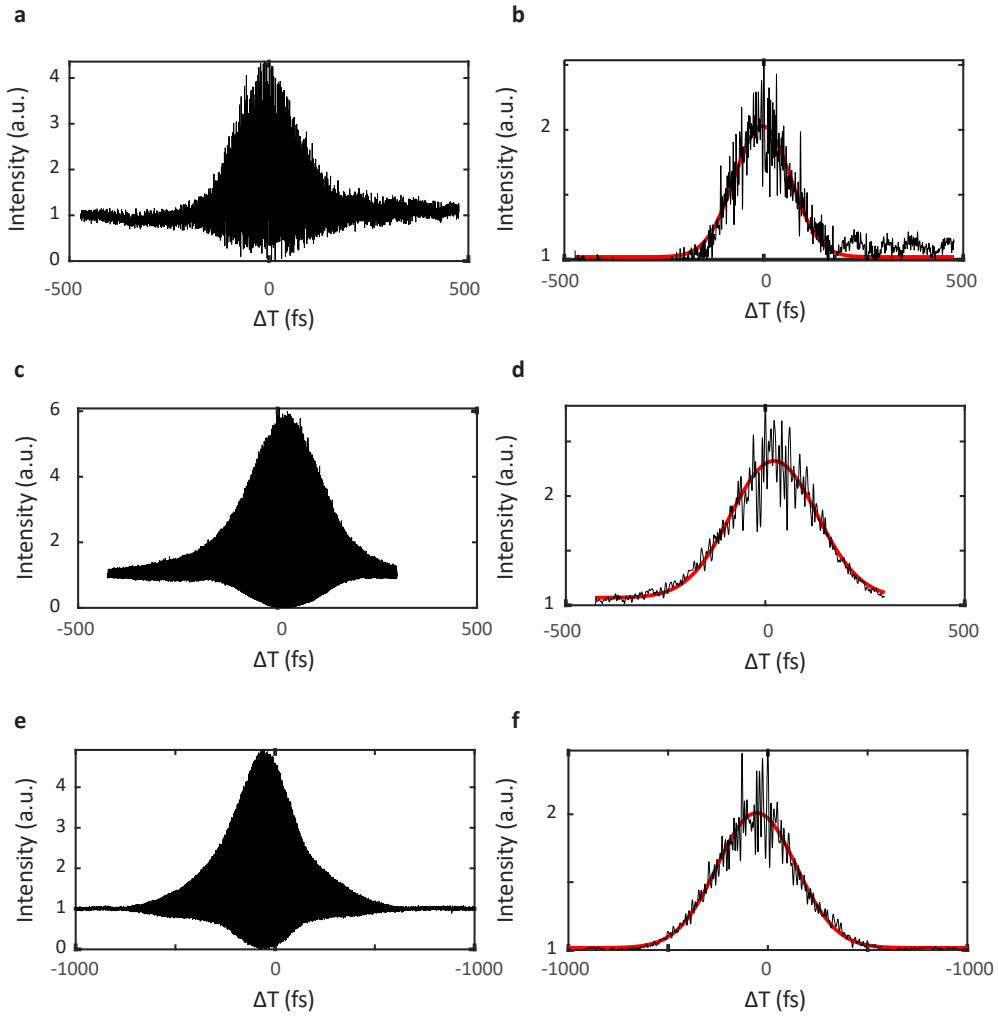

Figure S2: Inteferometric autocorrelations of excitation pulses. Interferometric autocorrelations are shown in (**a**, **c**, **e**) for compressed 610 nm excitation, uncompressed 610 nm excitation, and uncompressed 645 nm excitation, respectively. Corresponding intensity autocorrelations with fits are shown in (**b**, **d**, **f**).

3 Characterization and SM2P experiments on C-phycocyanin

3.1 Verification of aggregation state

Fluorescence correlation spectroscopy (FCS) was used to confirm the aggregation state of CPC and cross-linked APC. Fig. S3 shows the normalized autocorrelation curve calculated from the fluorescence photon stream for our samples at ∼5 nM concentration.

The autocorrelation curves ($G(\tau)$) were fit to:

$$G(\tau) = C + \left(\frac{1}{N}\right)\left(\frac{1}{1+\frac{\tau}{|\tau_D|}}\right)\left(\frac{1}{\sqrt{1+\frac{\tau}{V^2|\tau_D|}}}\right) \tag{5}$$

where C is a constant, N is the number of particles in the focal volume, τ is the delay time, τ_D is the diffusion time, and V is the detection volume defined as Z_0/w_0, where Z_0 and w_0 are the distances at which the 3D Gaussian volume has decayed to 1/e in the axial and radial directions, respectively. The fit is shown in black. Diffusion constants of $D_{APC} = 46\frac{\mu m^2}{s}$ and $D_{CPC} = 77\frac{\mu m^2}{s}$ were calculated for APC and CPC, respectively. These numbers indicate that the aggregation state of APC is trimeric [13] and the aggregation state of CPC is monomeric [14]. The mismatch at long delay times for CPC is likely due to a small trimer population at the nM concentrations used, which is expected to disassociate at the pM concentrations of the SM2P measurements assuming an equilibrium.

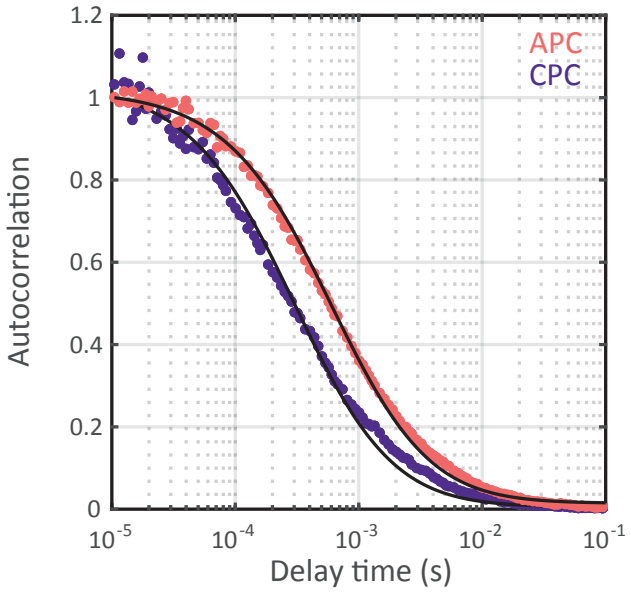

Figure S3: Fluorescence autocorrelation curves of light-harvesting sub-units. FCS curves for APC (red) and CPC (purple) with least squares fits (solid black) were used to extract diffusion times of $D_{APC} = 46\frac{\mu m^2}{s}$ and $D_{CPC} = 77\frac{\mu m^2}{s}$. These values confirm that APC exists in a trimer aggregation state while CPC dissociates into monomers.

3.2 SM2P representative traces

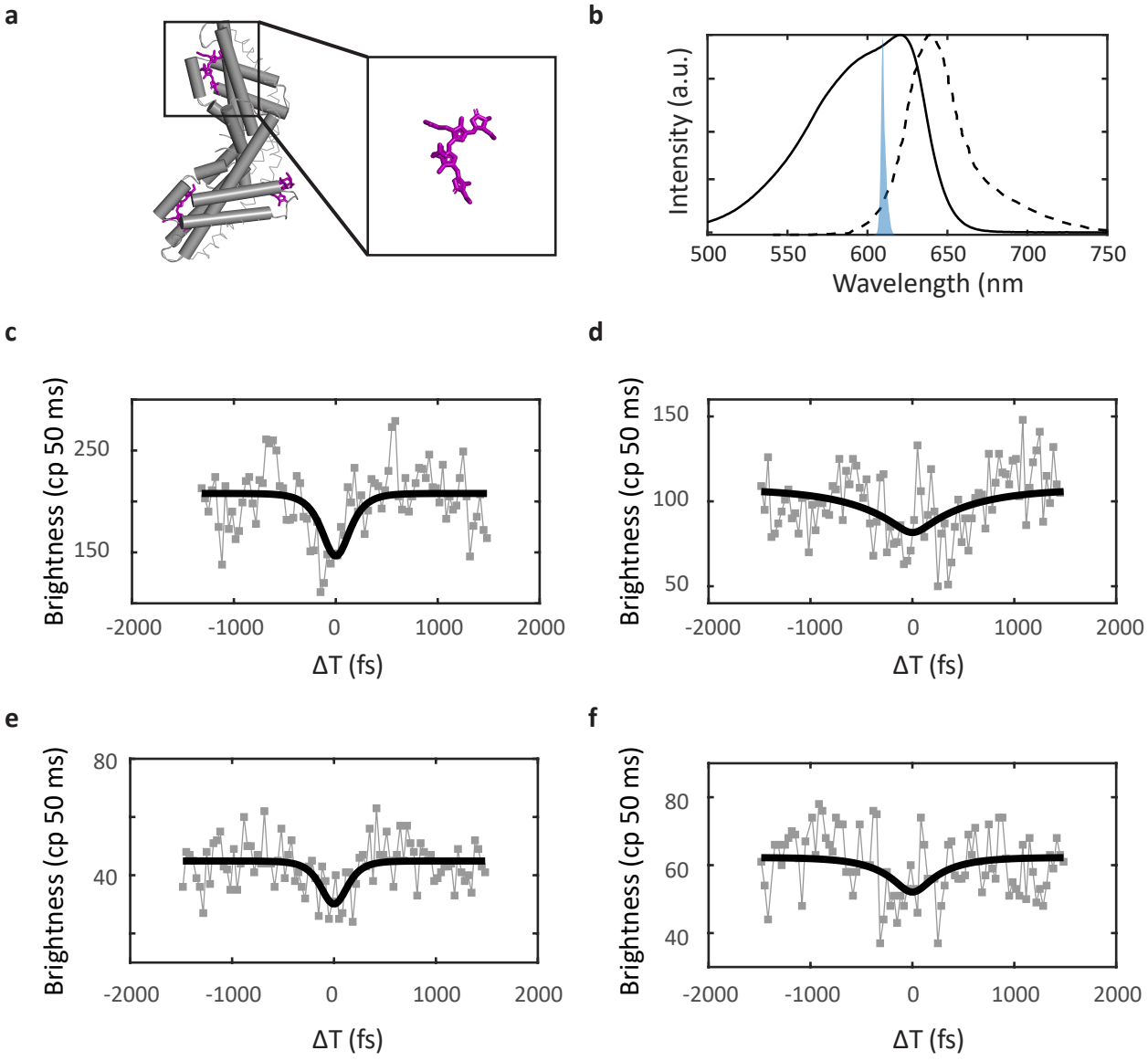

Figure S4: Single-molecule pump probe experiments on C-phycocyanin:. (**a**) The structure of C-phycocyanin (Protein Data Bank ID Code 1GH0) is shown with a callout of a tetrapyrole chromophore (purple). (**b**) The corresponding absorption (solid) and emission (dashed) spectra are shown with the 610 nm excitation shown in blue. (**c-f**) Representative traces for C-phycocyanin with 610 nm excitation are with values of 125±2, 503±124, 113±29, and 270±51 fs, respectively.

4 Concomitant measurement of fluorescence Lifetime

Fluorescence lifetime data was simultaneously recorded during the SM2P experiments and used to characterize fluorescence states. Representative particles are shown in Fig. S5 with both the SM2P modulation of the fluorescence intensity and the corresponding fluorescence lifetime. Lifetimes were extracted from the decay traces with a custom MATLAB program that minimized the negative log-likelihood function (Equation 4) with the exponential model used by Maus et al. [15]. The model function was a single exponential convolved with an experimentally measured instrument response function (FWHM $\sim$ 500 ps, solid gray line) plus the background. The model function (M_{exp}) is shown below in Equation 5 where IRF is normalized by the sum of photon counts, BG is an experimentally measured background trace at similar laser intensity, and γ weighs the BG component to signal intensity. In our case, γ was a tightly constrained fit parameter allowing for variation between samples. The exponential function is given by G where N is the total photon counts, A_0 is an offset, A_1 is the amplitude, and τ_1 is the time constant.

$$M_{exp} = IRF \otimes G + \gamma\, BG \tag{6}$$

where G is the normalized exponential decay given by

$$G = N\, \frac{(A_0 + A_1 e^{-t/\tau_1})}{\int A_1 e^{-t/\tau_1}\, dt} \tag{7}$$

Fluorescence lifetimes were measured for all successful SM2P experiments and are histogrammed in Fig. S6. The fluorescence lifetime distributions for APC with 610 nm and 645 nm excitation are similarly peaked around 1.5 ns as compared to previously reported experiments [16, 17]. However, the distributions cannot be directly compared because rapid photobleaching will bias the distributions to short timescales.

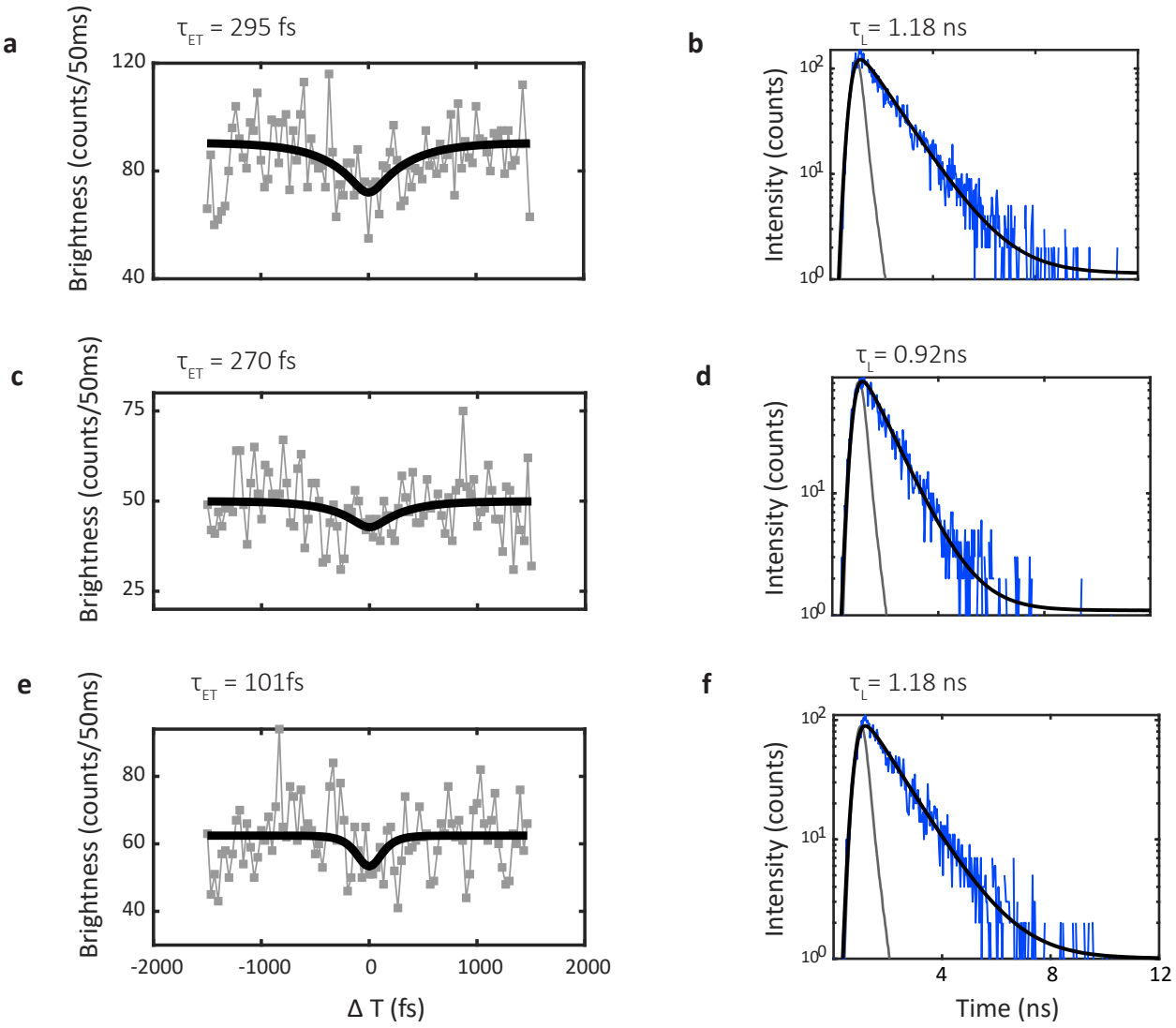

Figure S5: SM2P experimental traces of APC with corresponding fluorescence lifetimes. SM2P data from sequential measurements on a single APC. Energy transfer fits and SM2P fits are shown for the experiments pairs of (**a**, **b**), (**c**, **d**), and (**e**, **f**).

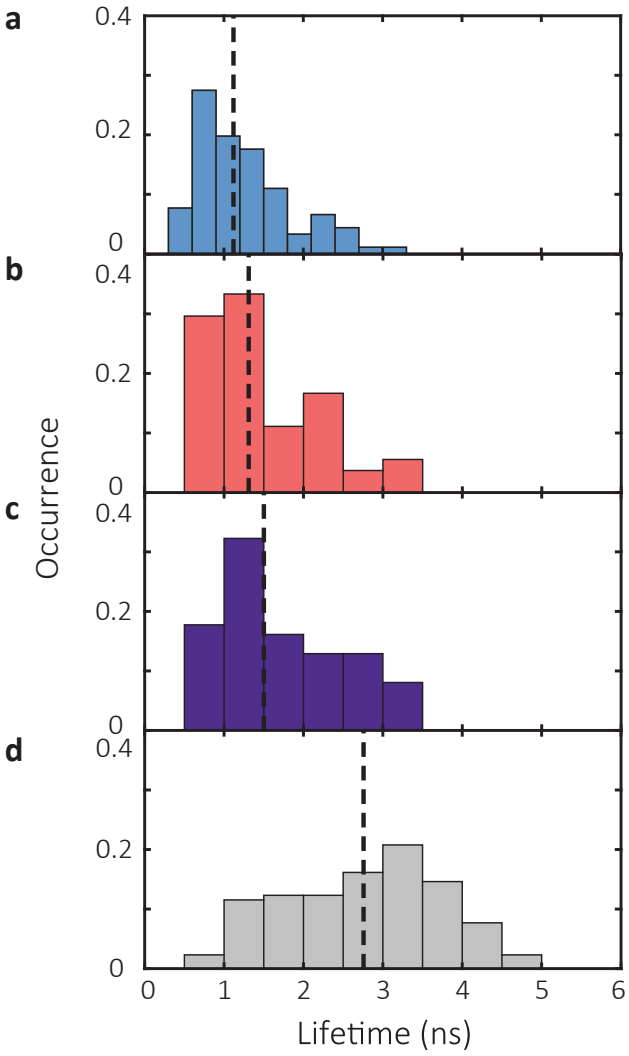

Figure S6: Fluorescence lifetime distributions. Histograms of fluorescence lifetimes for APC (610 nm, **a**), APC (645 nm, **b**), CPC (**c**), and Atto647N (**d**). Median values are marked with a dashed line at 1.12 ns (**a**), 1.31 ns (**b**), 1.50 ns (**c**), and 2.76 ns (**d**).

5 Statistical analysis of the distribution of energetic relaxation timescales

5.1 Dependence of energy relaxation histograms on excitation wavelength and pulse length

To ensure that the extracted microscopic dynamics were independent, *i.e.,* not due to, the pulse duration, we compared energy relaxation distributions for APC with 610 nm excitation with compressed and uncompressed pulses. This is shown in Fig. S7a These distributions were not significantly different from one another (two sample t-test gave a p-value ~ 0.6) and had similar qualitative distributions. The two data sets were combined for further analysis as described in the main text. Fig. S7b compares the energy relaxation distributions for Atto647N with 610 nm excitation (FWHM $\sim$ 180 fs) and 645 nm excitation (FWHM $\sim$ 303 fs). Similarly, these two data sets were found to derive from the same distribution (two sample t-test gave a p-value ~ 0.14). Further analysis of the Atto647N distributions is provided in [4] and these distributions were combined for analysis within this text. These results are in agreement with previous experimental and theoretical work in which the pulse duration was found to not affect the accuracy of values from SM2P [1–3, 18–20].

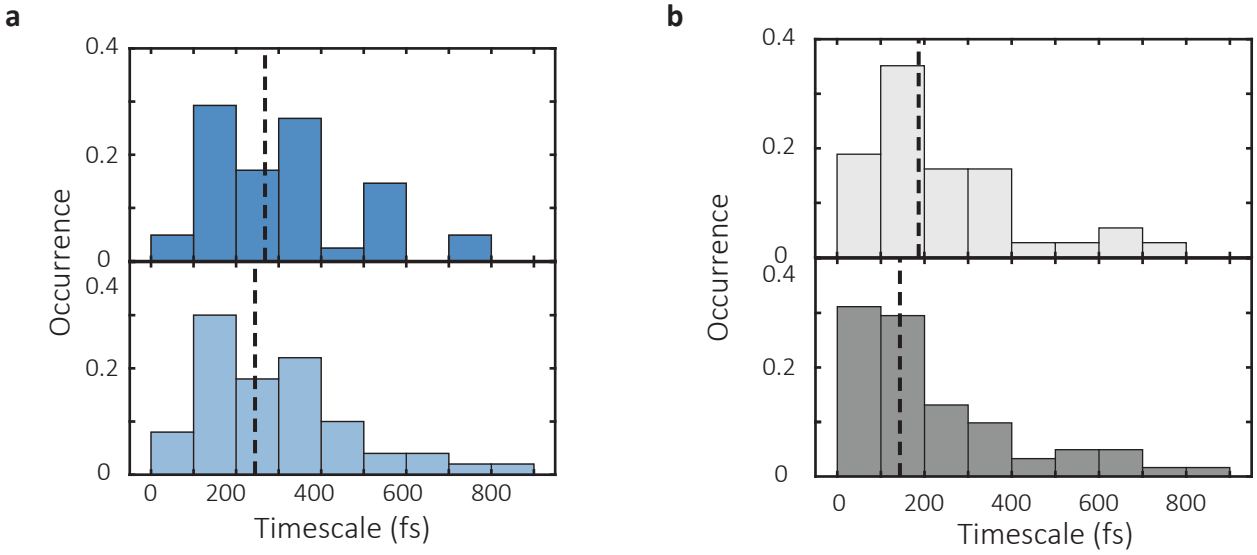

Figure S7: Histograms of APC data. (**a**) SM2P data for APC with 610 nm excitation. Dark blue represents the longer pulse length, (180 fs), and light blue represents the shorter pulse length (118 fs). (**b**) SM2P data for Atto647N with 610 nm excitation and 180 fs FWHM (top) and 645 nm excitation with 303 fs FWHM (bottom). Dashed lines indicate the medians of the distributions and are listed in Table S1.

Table S1: Statistical parameters for samples with different exication wavlength and pulse length

Sample	median (fs)	mean (fs)	σ (fs)	IQR (fs)	N
Atto647N (610 nm, 180 fs)	185	247	189	201	52
Atto647N (645 nm, 303 fs)	124	210	189	192	78
APC (610 nm, 118 fs)	245	294	181	228	51
APC (610 nm, 180 fs)	268	308	183	209	51

5.2 Comparison of energy-relaxation distributions for APC, CPC, and Atto647N

A combination of permutation tests and bootstrapping methods were used to investigate the timescale distributions presented in this work [21]. Bootstrap methods were used to estimate standard errors to compare parameters that describe the distributions. Fig. S8a shows the median values with the standard error determined by bootstrapping 10,000 times. Atto647N was found to have a significantly shorter median timescale than the phycobiliproteins. Fig. S8b shows the results of a permutation test between APC with 610 nm excitation and the combined Atto647N dataset. A permutation test pools the data and randomly assigns data points into two groups with the same size as the initial pooled data sets. The difference in statistical values between the resulting two groups (median) is calculated and this process is repeated many times. This tests the null hypothesis that the new data sets come from the same distribution and takes into account the fluctuations due to finite data samples. After repeating the process 10,000 times, we see that on average the difference between the two groups is zero, as expected for the null hypothesis, and that the chance of observing the experimental value for the median difference of 105 fs is 0.01 % as indicated by the red line in Fig. S8b. P-values of 0.0001, 0.001 and 0.001 were measured for APC with 610 nm excitation, APC with 645 nm excitation, and CPC when compared with Atto647N. The Benjamini–Hochberg procedure was used to take into account multiple testing, but the results stand as indicating that the phycobiliproteins are significantly different than Atto647N. Table S2 contains the statistical paramaters for all datasets.

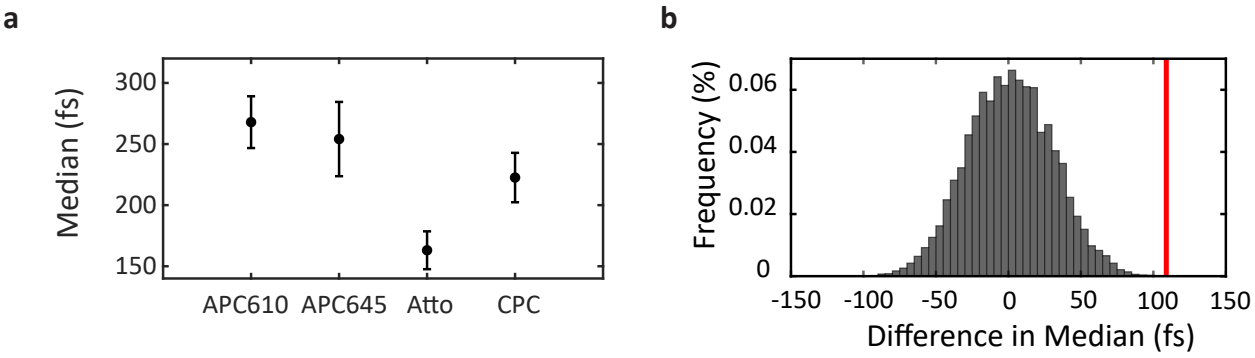

Figure S8: Statistical analysis of the medians. (**a**) The median values of the various samples with error bars given by the standard error obtained by bootstrapping. (**b**) The results of 10,000 iterations of a permutation test for Atto647N and APC with 610 nm excitation. The red line indicates the observed experimental value for the difference in medians. The location in the extreme tail of the distribution (p-value = 0.0001) shows that these distributions are significantly different

Table S2: Statistical values for SM2P experiments.

Sample	median (fs)	mean (fs)	σ (fs)	IQR (fs)	N
Combined Atto647N (610 nm, 645 nm)	148	224	187	205	130
Combined APC (610 nm)	257	300	180	223	93
APC (645 nm)	258	284	196	269	55
CPC	216	267	187	225	67

5.3 Comparison of the widths of the energy-relaxation distributions

To compare distribution widths, the data was smoothed using the kernel density estimation function in MATLAB and then the FWHM of the main peak was calculated. One example of the kernel density estimation (KDE) for APC with 610 nm excitation is shown in Fig. S9a. The resulting smoothed distribution curves for all data sets are shown in Fig. S9b. To determine whether the FWHM of Atto is significantly different from the phycobiliproteins, all of the data was pulled together and randomly assigned to four groups with sample sizes equal to the true data sets. The narrowest FWHM was chosen as the test statistic and is shown in Fig. S9c. This tests the probability of pulling a FWHM value as small as the experimentally observed Atto647N data set. Although the distribution is not normal, we can still determine that the chance of pulling a FWHM = 179 fs (the FWHM of Atto, red line in Fig. S9c) is less than 1.65%. This meets the p=0.05 standard, confirming that the narrower distribution for Atto647N is statistically significant. The measured FWHM values with standard errors are shown in Fig. S9d.

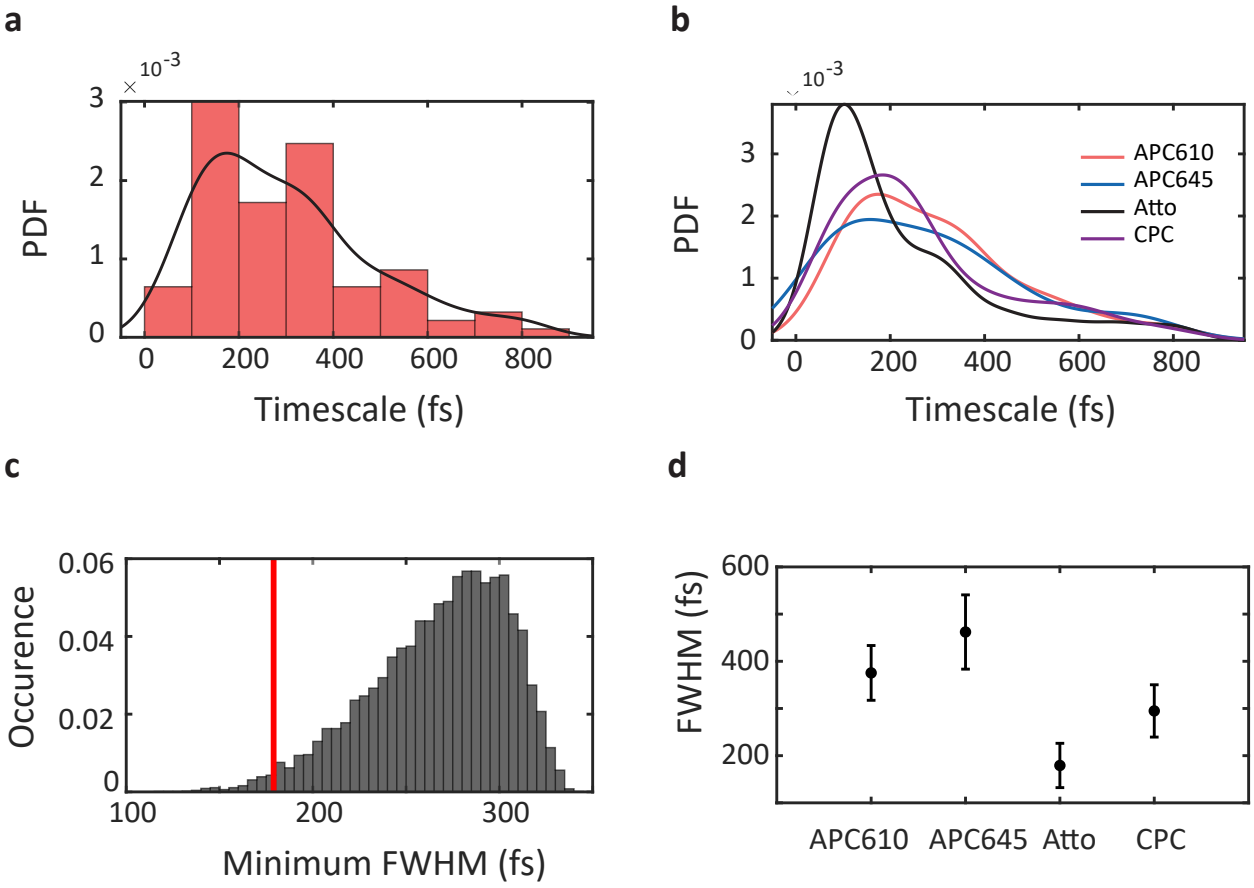

Figure S9: Analysis of the width of the distributions. (**a**) An example of a kernel density estimation (KDE) for APC with 610 nm excitation. (**b**) The KDEs for the four samples. (**c**) The results of a permutation test where the data was randomly divided into four groups and the smallest FWHM was determined. The chance of retrieving the experimentally observed value (FWHM_{Atto} = 179 fs, indicated by the solid red line) has a p-value of 0.0165. (**d**) The FWHM of the KDEs for the four samples with the error bars given by the standard error obtained from bootstrapping.

5.4 Comparison of the bright and quenched populations of APC with 610 nm excitation.

A permutation test with 10,000 iterations was used to further evaluate the statistical significance of the differences between bright and quenched states of APC with 610 nm excitation. The results of the permutation test are shown in Fig. S10. This distribution shows that the probability of randomly drawing two data sets with the appropriate sizes and seeing the observed difference in medians of 168 fs is only 0.05 %. Table S3 contains the statistical values for the different states.

To further illustrate the difference between the bright and quenched states, cumulative density plots are shown in Fig. S11 for APC with 610 nm excitation (a) and 645 nm excitation (b). These plots highlight the increased accumulation of data points at faster timescales for the bright versus quenched states for the 610 nm excitation compared to 645 nm excitation. $\sim 30\%$ of the quenched timescales occur before 200 fs, $\sim 50\%$ of the bright timescales arrive before 200 fs for the 610 nm excitation, while the two curves are almost identical for APC with 645 nm excitation.

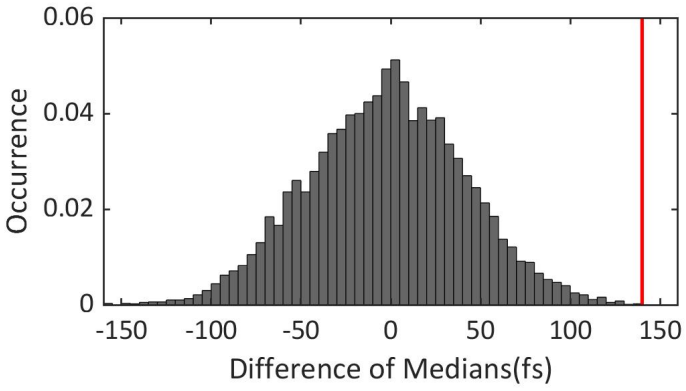

Figure S10: Permutation test results on APC with 610nm excitation. 33 timescales were randomly drawn without replacement from the APC with 610 nm excitation data-set to simulate the bright population and the median was compared to unselected values. This process was repeated 10,000 times. The red line indicates the observed experimental value of 168 fs that occurs less than 0.05% of the time.

Table S3: Statistical values filtered by lifetime.

	median (fs)	mean (fs)	σ (fs)	N
APC (610 nm, bright)	168	258	192	33
APC (610 nm, quenched)	308	324	171	60
APC (645 nm, bright)	276	300	217	24
APC (645 nm, quenched)	257	272	180	31

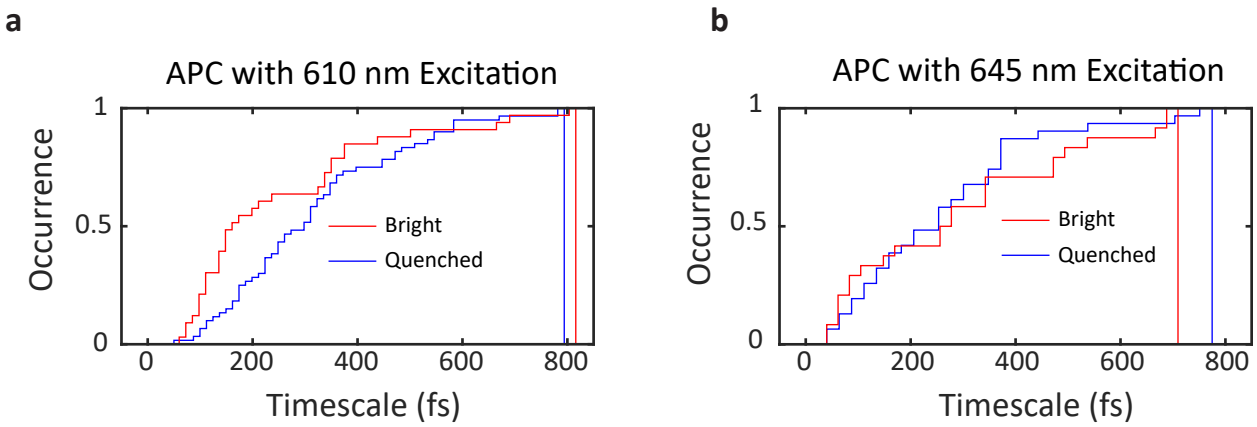

Figure S11: Bright and quenched cumulative probability histograms.. (**a**) For 610 nm excitation, a steep rise in the cumulative density function is shown for bright (red) versus quenched (blue). This is most easily seen at 200 fs, where a greater than 20% difference occurs. (**b**) For 645 nm excitation, the two histograms are effectively the same.

5.5 Analysis of the tails of energy transfer distribution

To further understand the tailed distribution of the bright population of APC with 610 nm excitation (APC610B), representative distributions were simulated. For the simulations, 50 values were randomly selected from a Gaussian distribution with center values (τ_{sim}) of 150 fs and 50 fs and a width of the same value. These parameters were then used to simulate SM2P traces, Poissonian noise was added, and the traces were fit to extract energy relaxation timescales. The pulse length was set to 120 fs (FWHM) for simulations. Fig. S12 shows histograms of the simulated data as well as the APC610B distribution. To quantify the spread of the energy relaxation timescales, the top of the histograms display the percent of values between 0 – 100 fs, 100 – 400 fs, and 400 – 850 fs. The simulations show that the bulk of the values fall between 100 – 400 fs, and that $\sim$ 10–15 % of the values are greater than 400 fs due to Poissonian noise, as discussed in more detail in Ref. [4]. Fig. S12C illustrates the distribution for τ_{sim} = 50 fs, which includes an increase of population of $\sim$ 10 % in the 0 – 100 fs range. This increase in timescales faster than the pulse length (150 fs), suggests that the fast timescales (50 fs) skews the distribution to short timescales. The lack of fast timescales for APC610B and τ_{sim} = 150 fs indicates that there is no significant component faster than pulse length in APC610B.

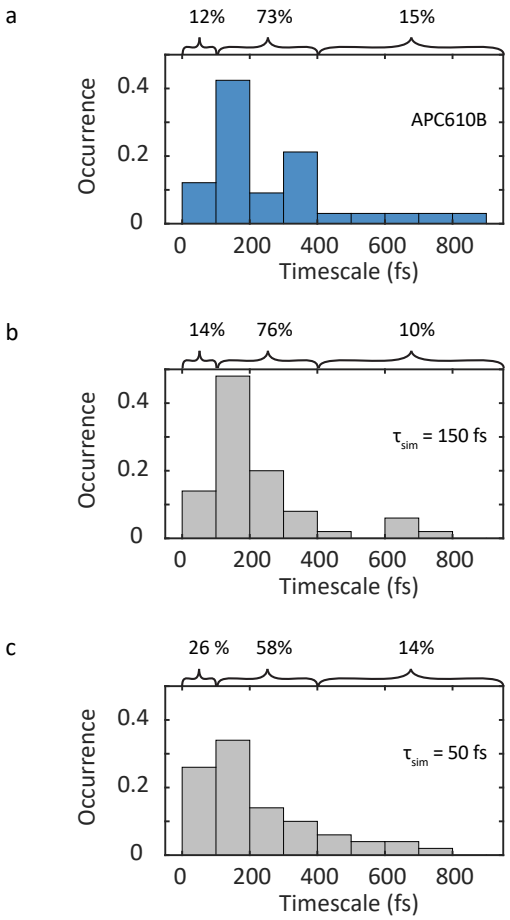

Figure S12: Comparison of energy transfer distribution with simulated timescales. Histograms of selected histograms with percentages shown above for population between 0–100 fs, 100 –400 fs, and 400–850 fs. Selected data sets are (**a**) APC bright population with 610 nm excitation, (**b**) Simulation data for τ = 150 fs, (**c**) τ = 50 fs

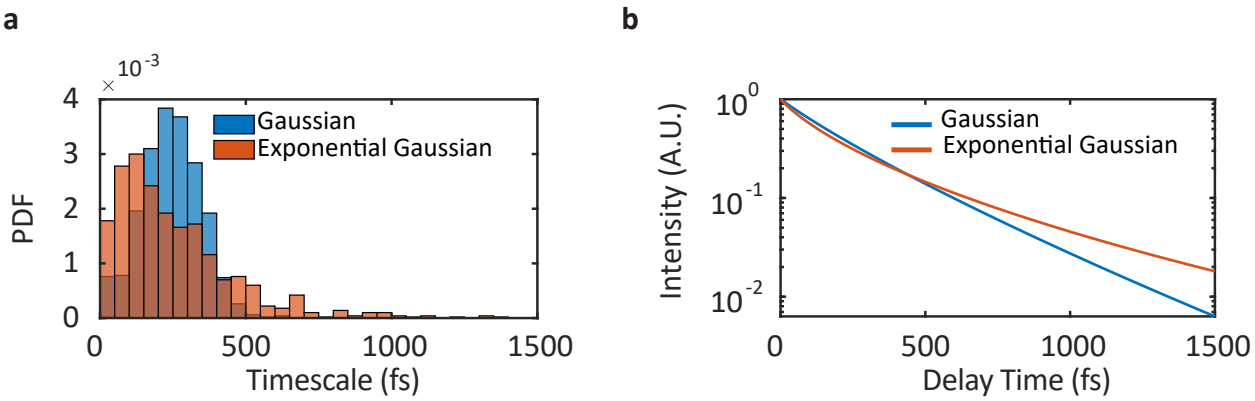

Figure S13: Effect of distribution on ensemble dynamics. Rate constants were randomly sampled from a Gaussian and exponential Gaussian distribution as histogrammed in (**a**). The rate constants were then used to create a set of decay curves that are averaged in and plotted in a semi-log plot in (**b**) to highlight a turnover from mono-exponential to bi-exponential behavior.

5.6 Effect of a skewed distribution on ensemble kinetics

The non-Gaussian distribution of energy transfer timescales observed may give rise to an ensemble averaged decay may have a more complex functional form than a simple mono-exponential. Previous ensemble ultrafast experiments generally modeled non-mono-exponential behavior as a sum of exponential terms where each time constant is assigned to a different physical interpretation. To better quantify how non-Gaussian distributions influence ensemble behavior, we sampled rate constants from a Gaussian distribution and an exponential Gaussian distribution (Fig. S13). The two distributions (Fig. S13) have the same mean value and similar spread. These distributions were sampled 1000 times and the rate constants were used to make simple exponential decay curves as a function of delay time and then averaged; these averaged decay curves are shown in Fig. S13b and simulate an ensemble measurement. We found that the Gaussian sampled data is well described by a single exponential decay, as shown by the straight line in the semi-log plot. However, the data sampled from the exponential Gaussian distribution is not well described by a mono-exponential decay, but instead requires a bi-exponential or other model. This brief analysis suggests that a non-Gaussian distribution of decay rates may lead to multi-exponential dynamics observed in ensemble experiments. Further investigation is required to firmly establish how non-Gaussian distributions can affect ensemble dynamics, and is underway.

References

1. Malý, P., Gruber, J. M., Cogdell, R. J., Mančal, T. & van Grondelle, R. Ultrafast energy relaxation in single light-harvesting complexes. *Proc. Natl. Acad. Sci.* **113,** 2934–2939 (2016).
2. Van Dijk, E., Hernando, J., Garcıa-Parajó, M. & Van Hulst, N. Single-molecule pump-probe experiments reveal variations in ultrafast energy redistribution. *J. Chem. Phys.* **123,** 064703 (2005).
3. Van Dijk, E. M., Hernando, J., Garcıa-López, J.-J., Crego-Calama, M., Reinhoudt, D. N., Kuipers, L., Garcıa-Parajó, M. F. & van Hulst, N. F. Single-molecule pump-probe detection resolves ultrafast pathways in individual and coupled quantum systems. *Phys. Rev. Lett.* **94,** 078302 (2005).
4. Moya, R., Kondo, T., Norris, A. & Schlau-Cohen, G. S. Spectrally-tunable femtosecond single-molecule pump-probe spectroscopy. *In Preparation.*
5. Tang, C. & Erskine, D. Femtosecond relaxation of photoexcited nonequilibrium carriers in Al x Ga 1- x As. *Physical Review Letters* **51,** 840 (1983).
6. Taylor, A., Erskine, D. & Tang, C. Equal-pulse correlation technique for measuring femtosecond excited state relaxation times. *Applied physics letters* **43,** 989–991 (1983).
7. Taylor, A., Erskine, D. & Tang, C. Femtosecond vibrational relaxation of large organic molecules. *Chemical physics letters* **103,** 430–435 (1984).
8. Turton, D. A., Reid, G. D. & Beddard, G. S. Accurate analysis of fluorescence decays from single molecules in photon counting experiments. *Analytical chemistry* **75,** 4182–4187 (2003).
9. Millard, A., Fittinghoff, D., Squier, J., Müller, M. & Gaeta, A. Using GaAsP photodiodes to characterize ultrashort pulses under high numerical aperture focusing in microscopy. *Journal of Microscopy* **193,** 179–181 (1999).
10. Diels, J.-C. M., Fontaine, J. J., McMichael, I. C. & Simoni, F. Control and measurement of ultrashort pulse shapes (in amplitude and phase) with femtosecond accuracy. *Applied Optics* **24,** 1270–1282 (1985).
11. Diels, J. C. & Rudolph, W. *Ultrashort Laser Pulse Phenomena* (2006).
12. Akturk, S., Gu, X., Kimmel, M. & Trebino, R. Extremely simple single-prism ultrashort-pulse compressor. *Optics express* **14,** 10101–10108 (2006).
13. Wang, Q. & Moerner, W. Single-molecule motions enable direct visualization of biomolecular interactions in solution. *Nature methods* **11,** 555 (2014).
14. Squires, A. H. & Moerner, W. Direct single-molecule measurements of phycocyanobilin photophysics in monomeric C-phycocyanin. *Proc. Natl. Acad. Sci.* **114,** 9779–9784 (2017).
15. Maus, M., Cotlet, M., Hofkens, J., Gensch, T., De Schryver, F. C., Schaffer, J. & Seidel, C. An experimental comparison of the maximum likelihood estimation and nonlinear least-squares fluorescence lifetime analysis of single molecckules. *Analytical chemistry* **73,** 2078–2086 (2001).
16. Loos, D., Cotlet, M., De Schryver, F., Habuchi, S. & Hofkens, J. Single-molecule spectroscopy selectively probes donor and acceptor chromophores in the phycobiliprotein allophycocyanin. *Biophys. J.* **87,** 2598–2608 (2004).
17. Goldsmith, R. & Moerner, W. Watching conformational- and photodynamics of single fluorescent proteins in solution. *Nat. Chem.* **2,** 179–185 (2010).

18. Palacino-González, E., Gelin, M. F. & Domcke, W. Theoretical aspects of femtosecond double-pump single-molecule spectroscopy. II. Strong-field regime. *Physical Chemistry Chemical Physics* **19,** 32307–32319 (2017).

19. Gelin, M. F., Egorova, D. & Domcke, W. Strong and long makes short: strong-pump strong-probe spectroscopy. *The journal of physical chemistry letters* **2,** 114–119 (2011).

20. Gelin, M. F., Egorova, D. & Domcke, W. Strong-pump strong-probe spectroscopy: effects of higher excited electronic states. *Physical Chemistry Chemical Physics* **15,** 8119–8131 (2013).

21. Good, P. I. *Permutation, parametric, and bootstrap tests of hypotheses* (Springer Science & Business Media, 2006).