Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics

Supplementary Information to *Excited state proton transfer in strongly enhanced GFP (sGFP2)*

Bart van Oort, Mirelle J. T. ter Veer, Marie Louise Groot, Ivo H. M. van Stokkum

This supplementary information consists of three sections:

1. Photoconversion in sGFP2
2. Comparison of kinetics of sGFP2 and GFPuv
3. Model based data analysis

1. Photoconversion in sGFP2

The X-ray structure of GFPuv (PDB 1B9C) [1] is shown in Figure S1. The amino acids labeled in grey are involved in hydrogen-bonding in GFPuv[2]. GFPuv has three mutations relative to wild type GFP (F99S, M153T, V163A) and sGFP2 seven (F64L, S65T, S72A, M153T, V163A, S175G and A206K) [3]. The six mutations of sGFP2 relative to GFPuv are labeled in red in Figure S1.

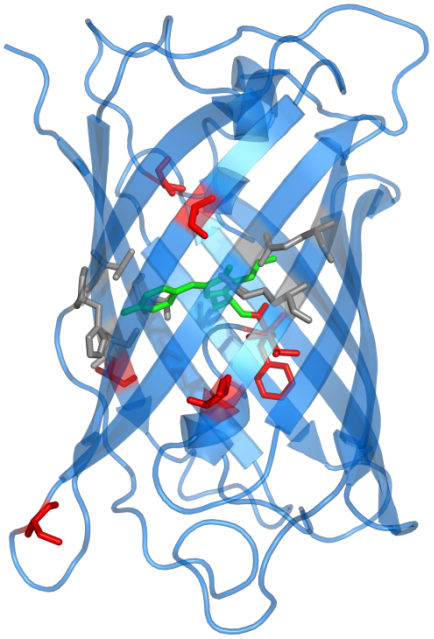

Figure S1. X-ray crystal structure of GFPuv (PDB 1B9C) [1]. The amino acids shown in grey are involved in hydrogen-bonding in GFPuv[2]. Those shown in red are mutated in sGFP2 relative to GFPuv. The chromophore is shown in green. This figure was created using the Pymol program[4].

In the ground state the chromophore can also be in the deprotonated B state. Several mutants with S65T retain the basic structure of the chromophore π electron system but render the A/B ratio and rate of proton transfer highly pH dependent [5]. With a small yield state B is also irreversibly formed from I*, which involves a conformational change around T203. A similar species is formed by photodecarboxylation of E222. Both modifications lead to permanent deprotonation of the chromophore (which is then in the anionic form) [6].

After some PDP experiments the steady state absorption spectra showed evidence of partial (irreversible) photoconversion: decreased absorption of the band at 398 nm, concomitant with an increase and a ~2-4 nm blue-shift of the absorption around 495 nm. The photoconversion is likely via decarboxylation of E222, similar to photoconversion in wtGFP [6], where it leads to a reduction of the ~400 nm absorption band and an increase of the band at 480 nm. This is similar to the changes in the sGFP2(E222Q) mutant relative to sGFP2[3]. It is therefore likely that the photoproduct (B) spectrally resembles sGFP2(E222Q), which absorbs only very weakly at 400 nm, and has a shorter excited state lifetime than sGFP2 [3]. B* will be dumped to B and not form GSI,

because both B and B* are deprotonated. In the fit the B* contribution will mix with I*, and lead to a smaller amplitude of the GSI spectrum and a faster decay of I*. Indeed the experiments with the strongest extent of photoconversion showed faster I* decays and slightly less GSI.

In strongly photoconverted samples it was possible to extract the B* SADS, by adding a component to the fit that is not coupled to the others. This fit (Figure S2) showed that B* contributes only weakly to the total ΔOD signal, due to its low absorption at 400 nm. Therefore B* is not taken explicitly into account in the fits in the main text, but it can explain the variation of the I* lifetime, and may have resulted in a slight underestimation of the intensity of the GSI spectrum.

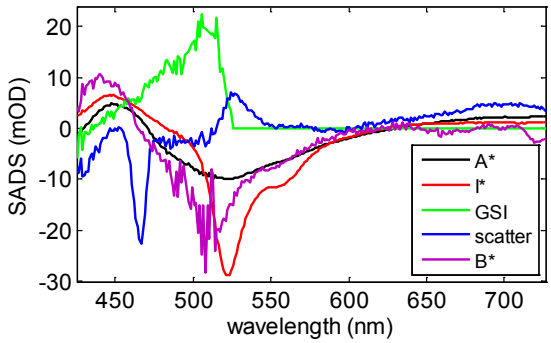

Figure S2. SADS estimated from a simultaneous target analysis of PP and PDP-DP data with 550 nm and 575 nm dump at 1 ps, including B* (purple). Black: A*, red: I* and green: GSI.

2. Comparison of kinetics of sGFP2 and GFPuv

GFPuv data (taken from ref [7]) showed clear biphasic ingrowth of the transient absorption at the I* minimum wavelength (510 nm), where the decay of A* and concomitant rise of I* was described by two exponentials of equal amplitude and lifetimes of 2.2 and 11 ps in H_2O (Figure S3). A monoexponential A*→I* ESPT with 5.9 ps lifetime (thin black line) does not fit these data. By contrast, in sGFP2 at the I* minimum wavelength (522 nm) a single lifetime of 5 ps was sufficient to describe the rise of I*.

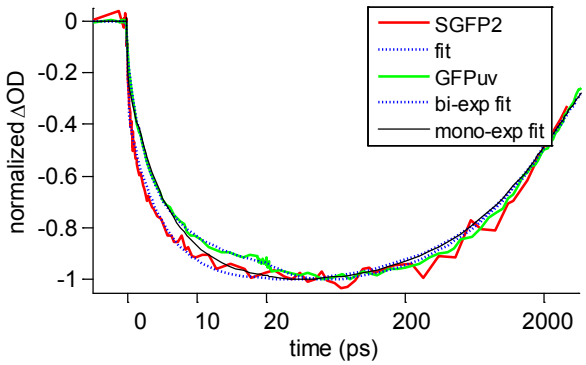

Figure S3. Absorption difference traces at the I* minimum wavelength (522 nm for sGFP2 (red) and 510 nm for GFPuv (green)). The time-axis is linear until 20 ps, and logarithmic thereafter.

3. Model based data analysis

The aim of data analysis is to obtain a model-based description of the full data set in terms of a model containing a small number of precisely estimated parameters, of which the rate constants and spectra are the most relevant. Description of the basic ingredient of kinetic models, the exponential decay, will be given first, followed by a description of how to use these ingredients for global and target analysis[8-10] of the full data. Our main assumption here is that the time and wavelength properties of the system of interest are separable, which means that spectra of species or states are constant. For details on parameter estimation techniques the reader is also referred to [8-11]. Software issues are discussed in[12].

A. Modeling an exponential decay

Here an expression is derived for describing an exponentially decaying component. The instrument response function (IRF) can usually adequately be modeled with a Gaussian with parameters μ and Δ for, respectively, location and full width at half maximum (FWHM):

$$IRF(t) = \frac{1}{\tilde{\Delta}\sqrt{2\pi}} \exp(-\log(2)(2(t-\mu)/\Delta)^2)$$

where $\tilde{\Delta} = \Delta/(2\sqrt{2\log(2)})$. The convolution (indicated by an *) of this IRF with an exponential decay (with rate k) yields an analytical expression which facilitates the estimation of the IRF parameters μ and Δ:

$$c^I(t,k,\mu,\Delta) = \exp(-kt) * IRF(t) = \frac{1}{2}\exp(-kt)\exp(k(\mu + \frac{k\tilde{\Delta}^2}{2}))\{1 + erf(\frac{t-(\mu + k\tilde{\Delta}^2))}{\sqrt{2}\tilde{\Delta}}\}$$

The wavelength dependence of the IRF location μ can be modeled with a polynomial.

$$\mu(\lambda) = \mu_{\lambda_c} + \sum_{j=1}^{j_{max}} a_j(\lambda - \lambda_c)^j$$

Typically, with visible difference absorption measurements over a large wavelength range, the order of this polynomial (j_{max}) is three. The reference wavelength λ_c is usually at the center of the spectrograph.

B. Global and target analysis

The basis of global analysis is the superposition principle, which states that the measured data $\psi(t,\lambda)$ result from a superposition of the spectral properties $\varepsilon_l(\lambda)$ of the components present in the system of interest weighted by their concentration $c_l(t)$.

$$\psi(t,\lambda) = \sum_{l=1}^{n_{comp}} c_l(t)\varepsilon_l(\lambda)$$

The $c_l(t)$ of all n_{comp} components are described by a compartmental model, that consists of first-order differential equations, with as solution sums of exponential decays. We will consider three types of compartmental models: (1) a model with components decaying monoexponentially in parallel, which yields Decay Associated Difference Spectra (DADS), (2) a sequential model with increasing lifetimes, also called an unbranched unidirectional model[13], giving Evolution Associated Difference Spectra (EADS), and (3) a full

compartmental scheme which may include possible branchings and equilibria, yielding Species Associated Difference Spectra (SADS). The latter is most often referred to as target analysis, where the target is the proposed kinetic scheme, including possible spectral assumptions. In this paper we present results from a target analysis. Here we present the rationale for the kinetic scheme that has been used.

With parallelly decaying components the model reads

$$\psi(t,\lambda) = \sum_{l=1}^{n_{comp}} c^I(k_l) DADS_l(\lambda)$$

The DADS thus represent the estimated amplitudes of the above defined exponential decays $c^I(k_l)$. When the system consists of parallelly decaying components the DADS are true species difference spectra. In all other cases, they are interpreted as a weighted sum (with both positive and negative contributions) of true species difference spectra. As an example, the DADS and exponential decays describing the PP data of Figure 2a are depicted in Figure S4. Here only the red DADS resembles a true species difference spectrum, *I**. The black DADS represents the transition $A^* \rightarrow I^*$. The green DADS, which decays in ~0.2ns, will later be shown to represent the inverse of the *GSI* difference spectrum.

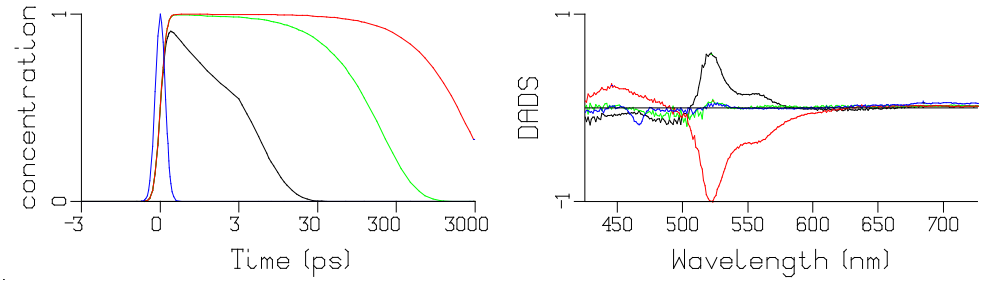

Figure S4. Exponential decays $c^I(k_l,t)$ and $DADS_l(\lambda)$ describing the PP data of Figure 2a. Key: black 5 ps, green 0.2 ns, red 2.8 ns. The time-axis is linear until 3 ps and logarithmic thereafter. The blue traces are attributed to scatter (with the shape of the IRF) and coherent artifact.

a sequential model reads

$$\psi(t,\lambda) = \sum_{l=1}^{n_{comp}} c_l^{II} EADS_l(\lambda)$$

where each concentration is a linear combination of the exponential decays, $c_l^{II} = \sum_{j=1}^{l} b_{jl} c^I(k_l)$, and the amplitudes[13] b_{jl} are given by $b_{11} = 1$ and for $j \le l$:

$$b_{jl} = \prod_{m=1}^{l-1} k_m \Big/ \prod_{n=1, n\neq j}^{l} (k_n - k_j)$$

When the system consists of sequentially decaying components $1 \rightarrow 2 \rightarrow \ldots \rightarrow n_{comp}$ the EADS are true species difference spectra. In all other cases, they are interpreted as a weighted sum (with only positive contributions) of true species difference spectra.

In general, the decay rates k_l are decreasing, and the lifetimes $1/k_l$ are increasing. As an example, the EADS and concentration profiles describing the PP data of Figure 2a are depicted in Figure S5. Here only the black and green EADS represent true species difference spectra, A^*, I^*. The red EADS is still a mixture.

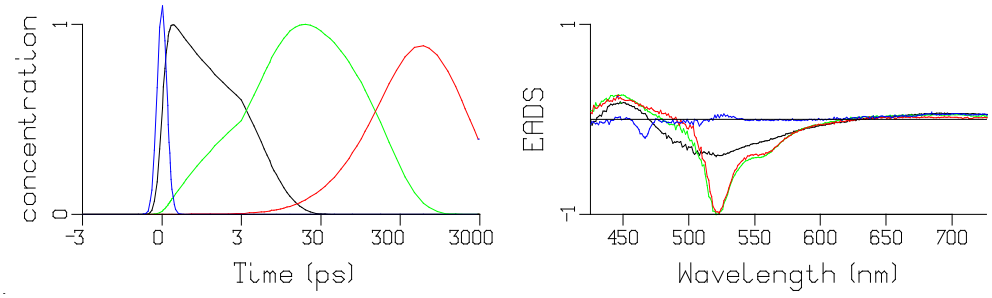

Figure S5. Concentration profiles $c_l^{II}(t)$ and $EADS_l(\lambda)$ describing the PP data of Figure 2a. Key: black 5 ps, green 0.2 ns, red 2.8 ns. The time-axis is linear until 3 ps and logarithmic thereafter. The blue traces are attributed to scatter (with the shape of the IRF) and coherent artifact.

The target model that we used in Figure 2 (depicted in panel f) consists of a sequential kinetic scheme, where we have interchanged the 0.2 and 2.8 ns lifetimes. Furthermore, we have assumed that the SADS of the final species (green) equals zero above 520 nm. This species is the ground-state intermediate *GSI*. Because of its inverted kinetics its concentration is very small, and its SADS is difficult to estimate.

In a more general target analysis a compartmental scheme is used to describe the concentrations of the compartments (states). Transitions to and from compartments are described by microscopic rate constants. The model that is fitted to the data now reads

$$\psi(\lambda,t) = \sum_{i=1}^{n_{comp}} c_i(t) SADS_i(\lambda)$$

where $c_i(t)$ now corresponds to the concentration of the *i*-th compartment. $SADS_i(\lambda)$ is the *i*-th Species Associated Difference Spectrum.

The concentrations of all compartments are collated in a vector $c(t) = \begin{bmatrix} c_1(t) & c_2(t) & \dots & c_{n_{comp}}(t) \end{bmatrix}^T = \begin{bmatrix} A^*(t) & I^*(t) & GSI(t) \end{bmatrix}^T$ which obeys the differential equation

$$\frac{d}{dt} c(t) = Kc(t) + j(t)$$

where the transfer matrix *K* contains off-diagonal elements k_{pq}, representing the microscopic rate constant from compartment *q* to compartment *p*. The diagonal elements contain the total decay rates of each compartment. The input to the compartments is $j(t) = IRF(t)\begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^T$.

The model can be extended to include a dump pulse. In that case the differential equation reads

$$\frac{d}{dt}c(t) = Kc(t) + j(t) + j_{dump}(t)$$

where $j_{dump}(t) = IRF(t - t_{dump})\begin{bmatrix} -x_A & -x_I & x_I \end{bmatrix}^T$. Here x_A represents the dump fraction from A^* and x_I represents the dump fraction from I^* to *GSI*. The K matrix reads:

$$K = \begin{bmatrix} -k_1 & 0 & 0 \\ k_1 & -k_2 & 0 \\ 0 & k_2 & -k_3 \end{bmatrix}$$

The target model in Figure 3 assumes that the *GSI* SADS is zero above 540 nm. Imposing this constraint hardly affects the rms error of the fit.

An important goal of target analysis is to achieve spectra that describe real physical states. However, many target models may provide fits of the same (mathematical) quality. Therefore alternative target models were tested, and compared on the basis of rates and SADS shapes and amplitudes. These must compare favourably with those found in related proteins (e.g. GFPuv[7]) and must agree with previous fluorescence experiments of sGFP2 [3, 14]. Furthermore the target models must yield consistent results for datasets with dumping at different delay times.

From the comparison of alternative models (see main text), it followed that the model in Figure 3 is the simplest to satisfactorily and consistently describe the data. The SADS most likely describe real physical states.

Notes and references

1. R. Battistutta, A. Negro and G. Zanotti, *Proteins*, 2000, **41**, 429-437.
2. M. Di Donato, L. J. G. W. van Wilderen, I. H. M. Van Stokkum, T. C. Stuart, J. T. M. Kennis, K. J. Hellingwerf, R. van Grondelle and M. L. Groot, *Phys. Chem. Chem. Phys.*, 2011, **13**, 16295-16305.
3. G.-J. Kremers, J. Goedhart, D. J. van den Heuvel, H. C. Gerritsen and T. W. J. Gadella, *Biochemistry*, 2007, **46**, 3775-3783.
4. W. L. DeLano, The PyMOL Molecular Graphics System, http://www.pymol.org.
5. P. J. Tonge and S. R. Meech, *J. Photochem. Photobiol. A-Chem.*, 2009, **205**, 1-11.
6. J. J. van Thor, T. Gensch, K. J. Hellingwerf and L. N. Johnson, *Nat. Struct. Biol.*, 2002, **9**, 37-41.
7. J. T. M. Kennis, D. S. Larsen, I. H. M. van Stokkum, M. Vengris, J. J. van Thor and R. van Grondelle, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 17988-17993.
8. A. R. Holzwarth, in *Biophysical Techniques in Photosynthesis*, eds. J. Amesz and A. J. Hoff, Kluwer, Dordrecht, The Netherlands, 1996, pp. 75-92.
9. I. H. M. van Stokkum, D. S. Larsen and R. van Grondelle, *Biochimica Et Biophysica Acta-Bioenergetics*, 2004, **1657**, 82-104.
10. I. H. M. van Stokkum, B. van Oort, F. van Mourik, B. Gobets and H. van Amerongen, in *Biophysical Techniques in Photosynthesis Vol. II*, eds. T. J. Aartsma and J. Matysik, Springer, Dordrecht, The Netherlands, 2008, pp. 223-240.
11. I. H. M. van Stokkum, in *Lecture notes for the Troisième Cycle de la Physique en Suisse Romande*, Department of Physics and Astronomy, Faculty of Sciences, Vrije Universiteit, Amsterdam, The Netherlands, 2005.
12. I. H. M. van Stokkum and H. E. Bal, *Concurrency and computation: practice and experience*, 2006, **18**, 263-269.
13. J. F. Nagle, L. A. Parodi and R. H. Lozier, *Biophys. J.*, 1982, **38**, 161-174.
14. G. J. Kremers, E. B. van Munster, J. Goedhart and T. W. J. Gadella, *Biophys. J.*, 2008, **95**, 378-389.