PLoS one

Modelling Multi-Pulse Population Dynamics from Ultrafast Spectroscopy

Luuk J. G. W. van Wilderen, Craig N. Lincoln, Jasper J. van Thor*

Division of Molecular Biosciences, Faculty of Natural Sciences, South Kensington Campus, Imperial College London, London, United Kingdom

Abstract

Current advanced laser, optics and electronics technology allows sensitive recording of molecular dynamics, from single resonance to multi-colour and multi-pulse experiments. Extracting the occurring (bio-) physical relevant pathways via global analysis of experimental data requires a systematic investigation of connectivity schemes. Here we present a Matlab-based toolbox for this purpose. The toolbox has a graphical user interface which facilitates the application of different reaction models to the data to generate the coupled differential equations. Any time-dependent dataset can be analysed to extract time-independent correlations of the observables by using gradient or direct search methods. Specific capabilities (i.e. chirp and instrument response function) for the analysis of ultrafast pump-probe spectroscopic data are included. The inclusion of an extra pulse that interacts with a transient phase can help to disentangle complex interdependent pathways. The modelling of pathways is therefore extended by new theory (which is included in the toolbox) that describes the finite bleach (orientation) effect of single and multiple intense polarised femtosecond pulses on an ensemble of randomly oriented particles in the presence of population decay. For instance, the generally assumed flat-top multimode beam profile is adapted to a more realistic Gaussian shape, exposing the need for several corrections for accurate anisotropy measurements. In addition, the (selective) excitation (*photoselection*) and anisotropy of populations that interact with single or multiple intense polarised laser pulses is demonstrated as function of power density and beam profile. Using example values of real world experiments it is calculated to what extent this effectively orients the ensemble of particles. Finally, the implementation includes the interaction with multiple pulses in addition to depth averaging in optically dense samples. In summary, we show that mathematical modelling is essential to model and resolve the details of physical behaviour of populations in ultrafast spectroscopy such as pump-probe, pump-dump-probe and pump-repump-probe experiments.

Introduction

The on-going development of high power, tuneable, multi-beam, and increasingly stable ultrafast laser sources as well as detector technology continues to generate exciting experimental applications. High power (>4 W) and 1–10 KHz repetition rate Ti:Sapphire regenerative amplifiers now allow the configuration of instruments that include several synchronously pumped optical parametric amplifiers (OPA's). It is not uncommon to combine four OPA's that provide independently tuneable output from the UV to mid-infrared, providing femtosecond and also synchronised picosecond sources. Here, we focus on the use of femtosecond spectroscopies of biomolecular dynamics and particularly the analytical power of detailed modelling of populations that interact with multiple pulses that have well known characteristics. The use of multiple ultrafast pulses in protein studies has proven to be a useful tool in revealing mechanistic information about transient species and their connectivities in proteins such as bacteriorhodopsin [1], proteorhodopsin [2], photoactive yellow protein [3], green fluorescent protein [4,5], calmodulin [6], phytochrome [7], and small molecules like cyanin [8], acetylene [9] and carotenoids [10–12]. Real world experimental data from VIS-VIS-infrared

pump-dump-probe experiments on phytochrome is used [7], and new corrections for the beam shape are proposed to ultimately extract physically relevant values such molecular directional information (via anisotropy). Two-pulse polarised infrared spectroscopy (such as VIS-infrared) is a powerful tool in the determination of (time-dependent) structural information. A good example is the accurate measurement of the angle between the diatom ligand and heme group in myoglobin [13–15]. Here we develop theory unexplored so far that describes the interaction of a second (polarised) visible pulse with a transient population. Multipulse spectroscopy has been used on phytochrome to directly test the validity of different reported connectivity schemes (i.e. a multi-exponential decay from the excited state [16], or a combination of sequential and parallel decays [17]). Accurate control of the time of arrival of the second laser pulse relative to the first pulse, allows for interactions of the second pulse with transient states or populations that may ordinarily not be observed in a conventional pump-probe experiment (if the transient decays faster than it is formed). In addition, by careful experimental design (of the timings and power densities), it is possible to distinguish between a species that directly decays to the ground state, and the formation of an intermediate prior to ground state

formation. This example demonstrates that multi-pulse spectroscopy is a powerful method to map connectivity schemes, and the averaged photolysed fraction plays an important role in the distinction between two (contrasting) models. The interaction of intense laser light with matter is complex, and theory describing these phenomena is similarly intricate. In this work we focus on the latter, and we present theory that describes the finite bleach effect (not to be confused with a similar term that is associated with loss of emission due to a chemical reaction) of, single and multiple, intense polarised femtosecond pulses on an ensemble of randomly oriented particles. In the present context the finite bleach effect therefore refers to the effective orientation of a population by polarised laser pulse(s).

The advancement of laser technology goes hand-in-hand with detector development. Current high speed optical multichannel analyser devices make it possible to use dispersive detection to monitor photochemical reactions on a single shot basis and over hundreds of channels. The extraction of physical, chemical, or biochemical information from such large datasets requires extensive modelling which can be challenging and time consuming. To facilitate this process, we present a software tool that allows the modelling of correlations between reactions and their (transient) products by global analysis (described in section *I*). This type of analysis requires a model to be applied to the data that describes the connectivity between the transient species in addition to explicit chemical kinetics equations that describe each reaction. It is important to emphasise that knowledge of the system under study is not a prerequisite for a global analysis, and that careful interpretation of the results can lead to physical relevant information. A complete understanding of a complex (photo-) reaction often necessitates the use of multiple techniques and approaches, as some methods are better suited to (populate and) resolve one or more reaction products than others. For instance, a radical species can be resolved by electron paramagnetic resonance spectroscopy (EPR), while its spectroscopic signature in time-resolved visible or infrared spectroscopy measurements may be more difficult or even impossible to assign. In turn, the spectroscopic data provides detailed time resolved structural and dynamical information that is not resolved by EPR spectroscopy and is complementary as such. Currently, time-resolved (spectroscopic) data can be analysed by different software packages such as Glotaran [18], Origin (OriginLab, Northhampton, MA, USA), Pro-K (Applied Photophysics Ltd, Leatherhead, UK), Bio-Kine32 (Bio-Logic, Claix, France), using different approaches. The program that we have developed has several key differences and additions with respect to the existing packages: First, it includes advanced global analysis of time resolved spectroscopic data, in addition to the modelling of the finite bleach effect for experiments that include single or multiple visible pulses in the presence of population decay. The first requires a connectivity scheme that can be *graphically* designed. Moreover, direct search- *and* gradient based fitting methods can be easily applied to the data. A graphical user interface is included that eases the learning curve for global fitting. In addition, the left singular values from an SVD analysis can be fitted without user-required data manipulation, and it includes new methods that allow molecular modelling to simulate anisotropy and Gaussian beam profiles. The software is Matlab (Mathworks, Inc.) based and therefore cross-platform and transparent (the source code is available). Furthermore, documentation is supplied and all buttons give instant tips. The global analysis toolbox is extensively discussed in section *I*, and is freely downloadable at http://www3.imperial.ac.uk/people/j.vanthor/research.

In section *II* we build on theoretical work on photoselection done by several groups [13–15,19–21] and extend it to include a

Gaussian beam profile and modelling of excited photolysis levels and anisotropy for multipulse experiments that only relatively recently has become possible due to technological progress. Importantly, we give graphical representations of ensemble averaged photolysed fractions and its corresponding anisotropy over a more realistic Gaussian beam profile of a real-world experiment. The introduction of a Gaussian beam shape as opposed to the generally assumed flat-top multimode configuration reveals the need for the inclusion of several corrections necessary for accurate information about the orientation of the vibrational transition dipole moment in relation to the visible transition dipole moment. In addition, we will show that the presented theory can be used to design an optimal experiment. We show that the extended theory for calculation of the finite bleach for multiple pulses in the presence of population decay is essential in order to extract the species associated spectral and kinetic information for multi-pulse experiments.

In essence, the described toolbox makes it possible to globally fit all data, or a noise-reduced subset of it (by applying SVD), and has extensions to include a direct search method in addition to a gradient based one, and makes it possible to graphically design a reaction model (via Matlab's Simbiology toolbox). Furthermore experimental parameters such as instrument response function and chirp can be extracted and the interaction of populations with multiple (Gaussian shaped) pulses can be modelled. The discussion commences in section *I* with a description of the theoretical foundations of the discussed global analysis program, and an overview of its capabilities is given by the analysis of a simulated dataset. The dual resonance interaction such as that of a visible laser pulse with an infrared beam is specifically treated in section *II*, as careful design of this experiment can give directional molecular information. This is an advantage over visible pump-probe spectroscopy that is essentially blind for this information (for a linear absorber), unless multiple optical transitions having a specific anisotropy are detected.

Results and Discussion

I GLOBAL ANALYSIS

In the following section we specifically focus on the analysis of spectroscopic data, but it is generally applicable to the analysis of kinetic datasets (such as mass or pH as function of time). It is not uncommon that experiments produce large amounts of time-resolved data that need to be properly analysed and interpreted. Consequently, the analysis of such datasets is very time consuming. Global analysis is a powerful method that enables the extraction of time-independent correlations (such as a number of spectra that represent a –meta- stable transition states in a spectroscopic experiment) present in the data. This approach greatly enhances and facilitates the interpretation of observed dynamics by reducing the dimensionality of the problem, i.e. by describing data in time-independent spectra with their associated time constants. This procedure is widely used in chemometrics [22,23] and spectroscopy [18,24,25]. It is important to stress that the results of a global analysis can potentially lead to an over-simplification and over-interpretation of the data, as the stochastic nature of some reactions (like radiative or vibrational transitions) may give rise to a range of time constants [26,27]. A time constant distribution can be visualised by for instance lifetime density mapping, but this approach gives at best semi-quantitative information [28]. In most cases a single rate constant that represents the complete manifold is assumed [18,29], and the presented work will therefore use the latter approach. Note that the use of singular value decomposition and global fitting assumes that the time resolved data consists of a

linear combination of time-independent components. This requirement is not met for instance if vibrational cooling leads to a time-dependent change of a vibrational frequency, in which case this would need to be modelled explicitly [30–32]. However, most femtosecond infrared spectrometers do not have sufficient spectral resolution to resolve such dynamics. In addition, vibrational cooling processes appear to be significantly faster in large biomolecules as compared to small molecules in solution [1,31,33]. Therefore, in many cases femtosecond dynamics of biomolecules may also be approximated by global fitting and SVD provided that the possibility of vibrational cooling in the data is carefully assessed.

In this work we present a software toolbox to perform global analysis in the Matlab environment. It has a user-friendly graphical user interface making it straightforward to use, and functions in Windows, Linux and Mac. The code is available, making it transparent. The philosophy behind the design of the program is to make as much as possible use of the environment's built-in features. For instance, a reaction model describing the connectivity scheme can be *graphically* designed and the corresponding coupled differential equations are automatically generated. Secondly, it is not only possible to analyse data by a gradient based algorithm, but also with a direct search method that is much less sensitive to converge to a local minimum. In addition, time traces resulting from (weighted) singular decomposition (SVD), that evaluates the number of components and number and values of time constants present in the data, can be fitted as well. First we briefly describe the used methodology with emphasis on (ultrafast) spectroscopy. Next, an example dataset is analysed to demonstrate the toolbox's capabilities.

Although a number of software packages are available to employ global analysis on time-resolved data, we focus on including a presentation of a step-by-step guide to ensure good-fitting practice. Discrete steps that represent the minimum of analysis necessary to unravel the essential kinetics are therefore given in the supplementary information.

Fitting procedure. The fitting procedure uses two different methods, i.e. a gradient based and a direct search method, to fit (time-dependent) data $X(t)$ that may exhibit some degree of uncertainty. These fitting procedures are based on the calculation and minimisation of the sum of residuals using in least-squares fitting (the employed algorithm is trust-region-reflective):

$$\min \sum \left(X(t) - \bar{X}(t) \right)^2 \qquad (1)$$

where $\bar{X}(t)$ is the idealised dataset. A widely used method to minimise an objective function is the non-linear least squares analysis (see [34] for an overview on different methods). Here, a gradient based method is employed, where, the differentiability of the objective function is essential since the first derivative of the error function needs to be (iteratively) calculated. If the problem is multidimensional (when for instance multiple time-dependent experiments are available, such as at multiple wavelengths), a surface needs to be scanned for a global minimum. High local gradients can however lead to a local instead of a global minimum, which is a known problem for gradient based methods [34]. The use of different (i.e. random) starting values may be necessary for the determination of the global solution, although infinite starting values need to be scanned in theory. A direct search method can be more powerful in finding a global minimum, as it has a lower tendency of remaining in a particular (local) minimum. A grid (called a *mesh*) is built up around the current point on the objective value surface (which is *not* required

to be differentiable or continuous), and consequently all directions are iteratively scanned via a pattern size that is contracting and expanding [35,36]. If a lower objective value (i.e. the residual sum of squares as in equation 1) is encountered, the new value becomes the new current point, after which the algorithm starts again, but with an increased mesh size. The mesh size is decreased if all scanned points have a higher objective value than the current one. The employed algorithm is *patternsearch* which has proven to be very effective in global optimisation problems ranging from neural networks to peptide structure prediction [37–39].

Problem definition. In ultrafast spectroscopy exponential functions are typically used to describe first order processes such as fluorescence and transient absorption signals (note that any function, such as a power law or stretched exponential can in principle be used with the presented package):

$$\bar{X}(t) = y_0 + \sum_i^n A_i \exp^{-k_i t} \qquad (2)$$

With y_0 representing a baseline offset, A_i the amplitude, and k_i the rate constant for every n exponentials (or components) that determines the concentration (or population). In a transient absorption experiment the parameters in the formula above are therefore referred to as spectra (i.e. A_i), and their respective lifetimes (i.e. k_i). Multi-exponential decays are often observed, and it may not be obvious if these components are dependent (for instance in consecutive order) or independent (for instance depopulating in parallel) of each other. Connectivity schemes describing the physical relationship between different components or species can be very complicated, rendering the underlying coupled differential equations that describe the evolution of each species individually (given by time-dependent concentration profiles) equally complex. The approach used here is to make use of numerical solvers (the standard solver used is *ode15s*) that solve the set of given ordinary differential equations for the time-dependent concentration (or population) of each component A_i in the connectivity scheme numerically (the precision that is used is 10^{-6}, but can be increased to 10^{-12}). The solution is produced by making use of the SimBiology toolbox. For example, the rate equation for the first component A_1 in the simple reaction $A_1 \rightarrow A_2$ is given by the first-order rate law:

$$\frac{d[A_1]}{dt} = k \left[A_1^0 \right] \qquad (3)$$

where A_1^0 is the concentration at time zero. The concentration of all components at any given point in time equals $\sum_{i=1}^n [A_i] = 1$, with n the number of components. The solution to the above differential equation for the first reactant essentially returns equation 2.

If multiple orders of magnitude in time are recorded, it is not uncommon to observe multi-exponential behaviour, making it necessary to decide on the *model* to be applied to the data. A model consists of (coupled) differential equations that are numerically solved, eliminating the otherwise necessary step of manually writing out (and implementing) the set of equations. A set of models is supplied with the program, rendering it straightforward to change example models for data analysis. Different types of models can be applied: parallel, sequential, or target (user-specified, which can include any type of connectivity

scheme including equilibrium reactions). A parallel model consists of independently decaying exponential functions (representing *species*) for which the amplitudes (or *concentration)* will then be determined by global fitting. A sequential model is built from one exponential (with unit amplitude) that determines the starting amplitude of the next and so on. A target model consists of any combination of the two (including reversible reactions). Several standard models (in *.xml* format [40]) are supplied with the program, but it is possible to build and export any model by making use of the *sbiodesktop* function (part of the SimBiology toolbox) that supplies a graphical user interface which allows for the *graphical* design of a connectivity scheme. Because the toolbox makes use of the .xml format that is compatible with the Systems Biology Markup Language, it is also possible to import models from other programs that export in this standard SBML format, see [41] for more information and a list of compatible programs.

Extraction of time-independent correlations. The goal of the global analysis procedure is to extract the time-independent spectra and their associated rate constants, defined by the model, from the data. Importantly, the choice of model has only consequences for the resulting spectra and not for the fitted rate constants (the order of rate constants is irrelevant for the fit quality). The rate constants are shared between all measured wavelengths independently, while their associated amplitudes are allowed to vary for each wavelength. The working method of the toolbox is to generate the final spectra by first generating concentration profiles (depending on the model), and consequently fitting those to the data. The spectra (the exponential pre-factors in equation 2) and an offset parameter are calculated by QR-decomposition (via the *mldivide* function) which uses linear fitting in a least-squares sense. QR-decomposition consists of the breakdown of a matrix into an orthogonal and a triangular matrix:

$$X = Q \cdot R \tag{4}$$

with $Q^T \cdot Q = I$ and R an upper triangular matrix. The solution is *unique* if all diagonal elements of R are positive. Because the exponential functions that are fit to the data contain linear and non-linear parameters, they are treated separately in the toolbox. The rate constants and offset in time zero are shared by all traces, and are non-linearly fitted in least squares sense with the *lsqnonlin* function.

Singular value decomposition (SVD). Consider a dataset $X(t)$ with the rows containing observables (for instance a spectrum) at each (increasing) time point. If the data for example is assumed to exhibit (multi-) exponential kinetics, a global analysis can be performed by fitting all traces simultaneously with a set of (shared) rate constants. Alternatively, a selection of time traces that are determined by SVD can be fitted (i.e. those that have the largest singular values). SVD is generally used to estimate the number of signal components (the *rank*), of which one determines a subset that can reconstruct the data via a linear combination [34,42,43].

The fitting of SVD traces (the left singular vectors, see below) is also used to provide the values and number of time constants in a dataset with greatest certainty, whereas the resulting basis spectra do not necessarily have physical meaning. This procedure assumes that data can be decomposed into a linear combination of unique orthogonal components by deconstructing data $X(1\ldots m,1\ldots n)$ containing m rows (corresponding to the time points) and n columns (corresponding to the spectra):

$$X = U \cdot S \cdot V' \tag{5}$$

Where S is a diagonal matrix (i.e. a matrix where all elements outside the diagonal are zero) of size $m \times n$ containing the singular values (in decreasing order with increasing row number), $U(1\ldots m,k)$ contains the traces in the time domain (the left singular vectors), and $V(k,1\ldots n)$ the spectral domain (the right singular vectors) that correspond to singular value $S(k,k)$. The singular values can also be used to weigh U [44,45], taking the amplitude of the singular value as measure for the dominance of the (wavelength independent) time traces:

$$U_w(1\ldots m,k) = U(1\ldots m,k) \cdot S(k,k) \tag{6}$$

In general, the largest singular values contain the relevant signal, while the smallest represent noise. Noise can be recognised by the absence of structure in the left and right singular vectors (i.e. random in the spectral and time domain). The number of singular values to include is not always obvious, because dominant noise components might overshadow data that has physical significance. Typically, a selection is made by inspecting the fall-off of singular values of the most significant components, rejecting higher components having singular values that are in excess of a linear trend for the most significant components. The scaling of the left singular value (see equation 6) significantly increases the speed to convergence during fitting, and may further assist to identify signal over noise. The presented program allows the fitting of all data, a selection of the left singular values resulting from SVD (i.e. U), or the weighted SVD time traces (i.e. U_w).

Impulse response function and chirp. Ultrafast spectroscopy data may exhibit a wavelength-dependence of time zero, and have a time resolution that is generally limited by the duration of the used laser pulses. The application of global analysis to these data requires therefore the modelling of chirp and impulse (or instrument) response function (IRF), respectively. The IRF, defined as the convolution of the ultrafast reaction-inducing laser pulse and the probe pulse, determines the lower limit of the shortest time scale on which a signal can be detected. If a Gaussian laser pulse (having full width half maximum $\varDelta$ and centre location μ, see figure 1 and equation 14) induces a response in a sample that exhibits exponential kinetics of relaxation rate k, the convoluted IRF_{conv} reads [18,46]:

$$IRF_{conv}(\tau) = \frac{1}{2}\exp^{-k\tau} \exp^{k\left(\mu + \frac{k\tilde{\Delta}^2}{2}\right)} \left(1 + erf\frac{\tau - \left(\mu + k\tilde{\Delta}^2\right)}{\sqrt{2}\tilde{\Delta}}\right) \tag{7}$$

With $\tilde{\Delta} = \Delta / \left(2\sqrt{2\ln 2}\right)$, and k the exponential relaxation rate. The IRF is typically also distorted in time (called chirp) due to self-phase modulation and group velocity dispersion in the material in which the super-continuum (i.e. probe) is generated [47]. A long temporal chirp may prevent the detection of the fastest occurring physically relevant process in a spectral region. This is generally caused by the choice of time points (which tends to cover a logarithmic base) that effectively changes the available time resolution with wavelength. In other words, the reaction inducing pump pulse arrives at the sample at time τ_0, but the time of arrival of the probe pulse is delayed for every wavelength. The
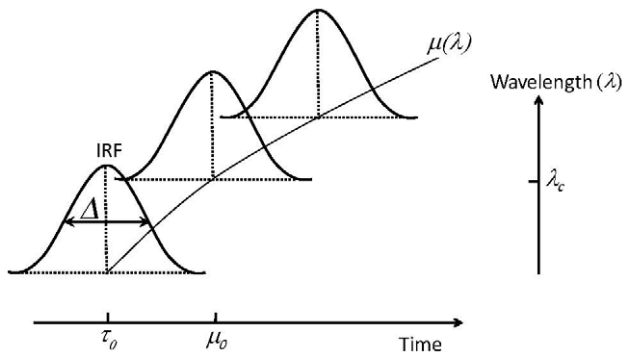
Figure 1. Schematic showing the IRF and chirp parameters. The IRF is represented by a Gaussian pulse arriving at τ_0 having full width half maximum Δ, and with its intensity out of the plane of the page. Each wavelength has its own IRF, and the location (in time) $\mu(\lambda)$ of each IRF changes as a function of wavelength (chirp). The chirp is symbolised by the curved line through the centres of each IRF, meaning that in this case the order parameter $i > 1$ in equation 8. The central wavelength λ_c is defined as the middle of the probed spectral window, and μ_0 as the middle of the time domain of the IRF dominated region. Note that both λ_c and μ_0 do not need to be exactly in the middle, as this depends on the order of the used IRF. The order $i = 0$ if the chirp is small (i.e. not resolved) or can be ignored.
doi:10.1371/journal.pone.0017373.g001

wavelength-dependent chirp can be modelled by a polynomial function [46]:

$$\mu = \mu_0 + \sum_{i}^{n} \mu_i \left(\frac{\lambda - \lambda_c}{100} \right)^i \qquad (8)$$

With n symbolising the order of the dispersion function, and μ_0 is the location of the IRF at wavelength λ_c. A typical experiment may exhibit a chirp in the order of tens of fs/nm, if uncompressed white light is used. The presented toolbox estimates the chirp by calculating the time shift between the signal (at each wavelength) and an instantaneous, infinitely short step function. This time delay is given by the time at which the maximum gradient occurs in the cross-correlation between each signal and the step function. The dispersion function (eq. 8) is then fitted to the resulting time delays and defines the chirp. The toolbox ultimately applies the chirp to the ideal dataset $\bar{X}(t)$ that is used to solve equation 1.

Example: Global and SVD analysis of model data. To show the capabilities of the program, a synthetic 3D data set is generated from three spectra (and two time constants) that sequentially evolve from one state to the next (see figure 2A). A second dataset is generated by artificially adding noise to the same dataset (resulting in the data shown in figure 2D). The noisy dataset simulates an actual experiment, and demonstrates some characteristic features of an ultrafast spectroscopy experiment which may suffer from detector noise and laser pointing instabilities that can cause spectral 'jumps'. The data is generated by adding 10% white noise to the starting spectra (i.e. the noise is 10% of the maximum value of each 'basis' spectrum) that are modelled by Gaussian distribution functions, and by adding 10% white noise to the concentration profile of each component.

We start the global analysis by determining of the number of components needed to satisfactorily fit the data. This is achieved by SVD analysis of the data (see above, and figure 3) that comprises the fitting of a number of SVD traces (the left singular

vectors resulting from SVD). An indication of the number of relevant time constants is given by the minimum number of singular values that are needed to extract the signal out of the data, effectively filtering the noise. The noise-reducing power of SVD is illustrated by the small effect the introduced noise has on the singular values, where the noise is small. Note also that the sign of the left and right singular vectors is irrelevant for the interpretation of the data as it has no physical meaning. For both data sets 3 singular values are significantly higher than the rest (see figures 3A and C), i.e. 3 components are likely required to reconstruct the signal. Note that the fourth singular value (smaller than 10^{-8}), reflects the precision of the numerical solver of the coupled differential equations that generates the linear combination of components. The white noise added to data in figure 2D is shown to be very close to being random, as the higher singular values are near those for the noise-free data. A fit of the first three left singular vectors results in time constants of about 1.8 s, 31 s and a long-lived component. A change in model from parallel to sequential has negligible influence on the quality of the fit, so only the results from the parallel model are shown in figure 3.

The time constants resulting from the fitting of the left singular vectors can be used as starting values, or fixed values if chosen, for a global analysis on the full data set, of which the results are shown in figure 2. Note that the time constants of the SVD and global analysis are very similar. Before exploring the results of the global fit in detail, some important general remarks about global fitting are discussed. Firstly, it is recommended that for global fitting the procedure should be initiated with as few parameters as possible (preferably with at least one component less than anticipated). This is particularly relevant for gradient based fitting methods (the default option in the toolbox), as convergence to a global minimum is not always guaranteed, and an unlucky choice of starting values may end up in a local minimum. The advantage of using such a conservative or restrained approach makes it less likely to 'over-fit' the data (i.e. by using too many components that are not present). The latter would lead to over-interpretation of the results. Critical interpretation of the results may also uncover time constants or spectra that do not necessarily have physical meaning. In practice, however, the interpretation of the results may be ambiguous, and extra care should therefore be taken to not include too many parameters. An objective measure for the goodness-of-fit is the sum of squares, which is a single number that decreases with fit quality and generally with the addition of more components. If only a small gain (i.e. a decrease) is achieved by adding an extra component, a careful evaluation of the previous fit is necessary to decide if the data is of sufficient quality to permit the inclusion of the extra component, *and* if this has any physical meaning (by investigation of their rates and spectra).

The SVD analysis shown in figure 3 has clearly indicated that not only 3 singular values are sufficient to reconstruct the majority of the signal present in the data, but also that three time constants satisfactorily fit the singular value's corresponding left singular vectors. Similarly, if three components are used to globally fit the full datasets of figure 2, it is observed that the parallel model (panels C and F) results in compensating amplitudes for the spectra that have 1.8 and 31 s time constants (look for instance at pixel 50). In general, the compensation of amplitudes of two components points to the application of an inadequate model. It could however also mean that too many time constants are used. Inspection of the sum of squares value indicates that this is not likely: it decreases from 0.093 to 0.0434 to 0.043259, for 2, 3 and 4 components (by applying the parallel model to the noisy data). Thus, a more than two-fold gain is achieved by adding one component with respect to two, while only a minor improvement is obtained by adding the
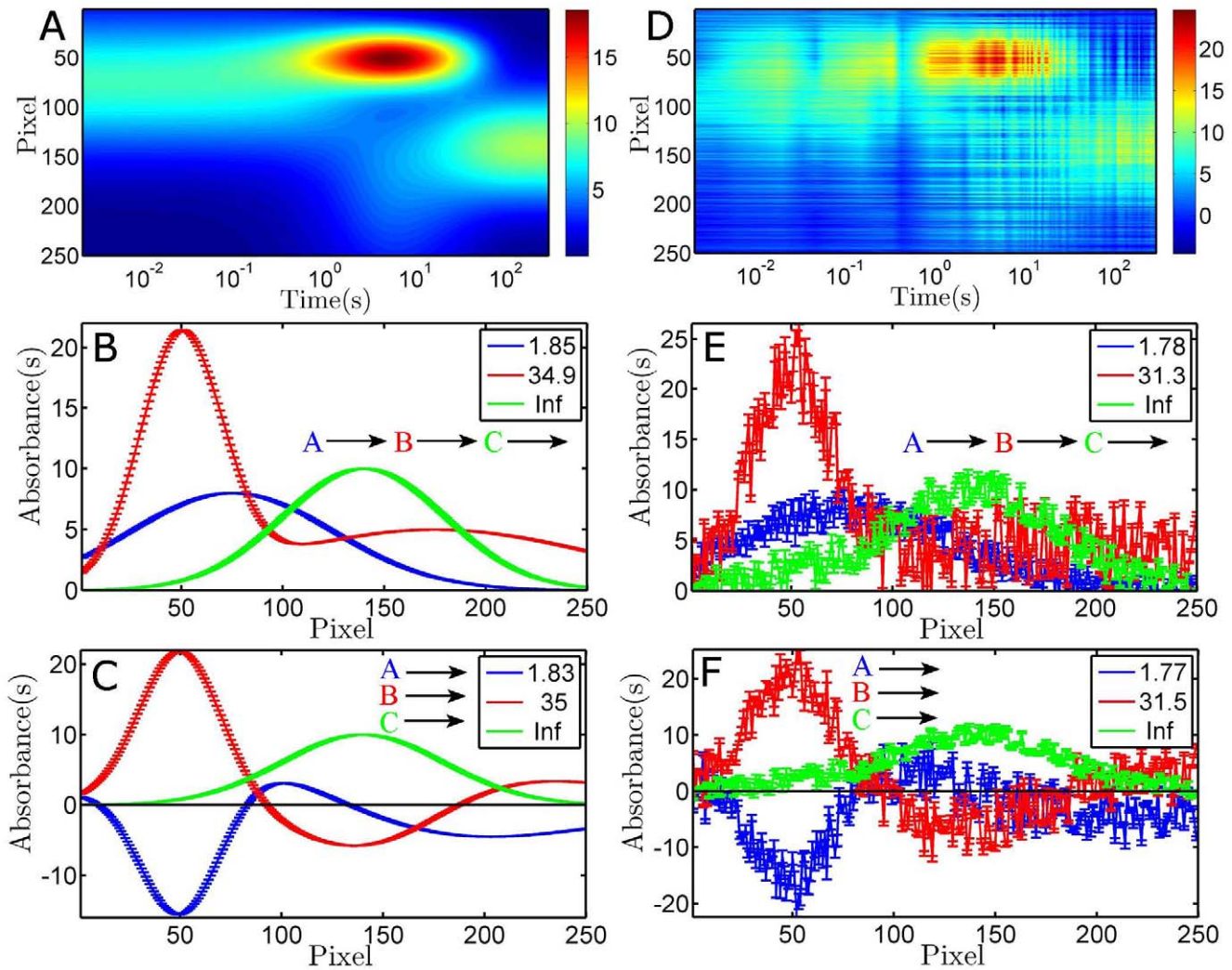
Figure 2. Global analysis of two simulated datasets representing transient absorption experiments. Two simulated datasets (A and D) with arbitrary wavelength (in pixels) and time delays, generated by allowing 1 species to sequentially form a second species and consecutively a final one. The spectra have (sums of) Gaussian line shapes, and the used time constants are 1.87 s and 34.8 s for the first two components to form a third (infinite or long-lived) component on a 300 second time base (91 time points at logarithmic intervals). The colour bar denotes the height of the signal. Panel D consists of the same data as panel A with ±10% added noise. Global analysis of the noiseless and noisy data is presented in panels B/C and E/F, where the pairs indicate the application of sequential and parallel models (indicated by the connectivity scheme in each panel), respectively. Typically, observation of compensating amplitudes such as those at pixel 50 in panels C and F for the 1.8 s and 35 s time constants indicates that a parallel model is unlikely to be correct. All spectra contain error bars (note that they are superimposed in panels B and C) and the legend shows the resolved time constants, including an Infinite one that is too long to be resolved on the used time base.
doi:10.1371/journal.pone.0017373.g002

fourth component (note that the number of digits chosen only illustrates the relative improvement). This strongly favours the presence of only 3 components in the analysed data. A second goodness-of-fit marker is the absence of systematic structure in the residuals in the time and spectral domain (i.e. exhibiting random noise only). Oscillations in the time domain of the residuals that do not appear random in nature indicate the need for one or more extra components. The corresponding structure in the spectral residuals gives an indication of the spectral identity of the component(s) that is (are) not optimally fit. The fits of a three-component model have random noise in the residuals (not shown), and the addition of a fourth component results in a negligible decrease in the noise of the residuals (and sum of squares). In conclusion, the global analysis gives objective criteria that favour a three-component sequential model, which is exactly how the

simulated data was created. In real-life experiments additional clues may be available to judge the physical relevance of the found spectra and time constants for an applied model, and also from complementary experiments. Global analysis is an extremely useful and powerful way of dissecting, visualising and interpreting data to ultimately resolve its physical significance.

II ENSEMBLE-AVERAGED PHOTOLYSED FRACTION: THEORETICAL BACKGROUND

Global analysis is a powerful tool that facilitates the analysis of time-dependent data obtained from any technique. However, the next section specifically elaborates on multi-pulse ultrafast (polarised) laser spectroscopy. A general assumption in such experiments on a sample in the condensed phase is that a sample consists of randomly oriented particles. Because a laser pulse is
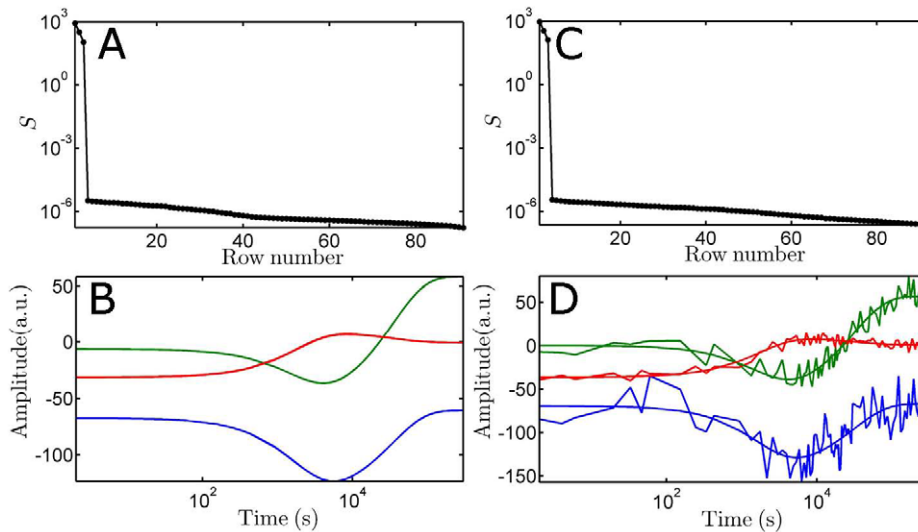
Figure 3. SVD analysis of the synthetic data shown in figure 2. Panels A and C show the singular values (the diagonal of S, note the logarithmic ordinate), which are used to get an indication of the number of occurring time constants, as function of the matrix row number (which is determined by the number of time points). Panels B and D show the first 3 left singular vectors and their corresponding fit (by using a parallel model). The fitted time constants are for panel B: 1.84 s, 34.90 s and a long-lived component, and for panel D: 1.78 s, 30.94 s and 19627 s (having errors of 0.04 s, 5.7 s and long-lived, respectively). In panel B the fits are superimposed on the data, while in panel D the fit is the smooth line through the noisy time trace.

doi:10.1371/journal.pone.0017373.g003

polarised by nature it will preferentially excite those particles that have their transition dipole moment aligned with the laser's polarisation. The selective excitation has therefore effectively oriented the sample. If a second pulse arrives it will thus interact with a non-random set of particles. Here we visualise the evolution of a partially oriented population as function of time, and its interaction with a second laser pulse. We adhere to the previously used notation [7,48] and ignore rotational diffusion between the two laser pulses. This condition is met for ultrafast laser experiments with a second laser pulse coming on a picosecond or even nanosecond timescale in some cases. Crucially, we explicitly include population decay occurring before the second pulse arrives.

The toolbox includes a graphical user interface that permits detailed insight into finite bleach effects of intense polarised femtosecond pulses, and is applicable to real world experiments. In the following discussion we focus on VIS-pump infrared-probe experiments because of the possibility to extract directional information about the vibrational transition dipole moment in relation to the visible transition dipole moment. We present a review of the photoselection theory from the literature [13–15,19–21] and extend this theory to model experiments with multiple pulses and explicit beam profiles. By using example values of real world experiments on phytochrome [7], the calculated results are plotted to demonstrate how the excited photolysis levels of an ensemble of randomly oriented particles are determined by the power density and beam profile characteristics. We also include the shape of the laser beam, a characteristic which has previously been ignored (i.e. a flat-top multimode beam is generally assumed), and extend current theory to include anisotropy calculation for multipulse experiments.

Photolysed fraction.   Before describing the time-dependent orientation distribution of a random ensemble of particles, we discuss the results for the ideal case (i.e. when the first pulse arrives). The ultimate goal is to extract structural information that gives the orientation of the vibrational transition dipole moment

$\hat{\mu}_2$ with respect to the polarisation of the visible laser $\hat{e}_1$. The first one has a fixed orientation relation with the optical transition along $\hat{\mu}_1$ (see figure 4).

In other words, a VIS laser pulse interacts with a fraction of the randomly oriented ensemble that is governed by a certain probability distribution function, and the probe pulse, having a polarisation that is different from the pump pulse, is needed to resolve the actual orientation of the direction of the transition dipole moment [48]:

$$\langle P_2(\hat{\mu}_2 \cdot \hat{e}_1) \rangle = \langle P_2(\hat{\mu}_2 \cdot \hat{\mu}_1) \rangle \langle P_2(\hat{\mu}_1 \cdot \hat{e}_1) \rangle \quad (9)$$
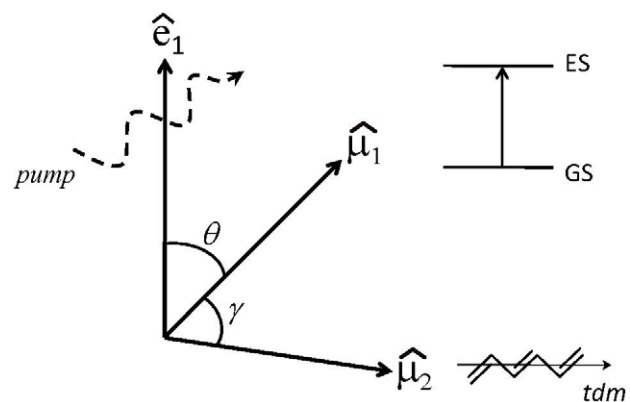

Figure 4. Schematic drawing defining the direction of the laser polarisation and two molecular parameters. The orientation between the pump beam polarisation $\hat{e}_1$ and the optical transition $\hat{\mu}_1$ (symbolised by an energy level diagram that connects the ground and excited states) is described by $x = \hat{\mu}_1 \cdot \hat{e}$, (i.e. the cosine of angle θ). The angle γ between $\hat{\mu}_1$ and the transition dipole moment $\hat{\mu}_2$ (symbolised by a linear molecule and its associated tdm) is fixed in space.

doi:10.1371/journal.pone.0017373.g004

with $P_2(x) = (3x^2 - 1)/2$ the second order Legendre polynomial, $x = \hat{\mu}_1 \cdot \hat{e}_1$ (symbolising the cosine of the angle θ between the reaction inducing laser pulse $\hat{e}_1$ and the optical molecular transition dipole moment $\hat{\mu}_1$), and the square brackets denoting the ensemble average over all molecular orientations. The left-hand side represents the measured anisotropy and the right-hand side the actual orientation and the optical anisotropy, respectively. The relation above explains why the measured anisotropy in pump-probe experiments where both beams have equal polarisation is always 0.4 (for infinitely low excitation power, see below), because the desired molecular information $\langle P_2(\hat{\mu}_2 \cdot \hat{\mu}_1) \rangle$ is not probed. In order to calculate the optical anisotropy, the last factor in the right-hand side of equation 9 needs to be derived first.

Consider therefore a sample with molecules that are randomly oriented. The fraction of molecules that is excited or photolysed by a laser pulse depends on the molecular properties of the particles [7,13,14,48]:

$$n_1(x) = 1 - \exp^{-N_1 x^2} \qquad (10)$$

where

$$N_1 = 3\bar{J}\sigma\varphi \qquad (11)$$

With $\bar{J} = \int J dt$ where J represents the total photon flux density, σ the absorption cross-section, and φ the quantum yield of the reaction. Equations 10 and 11 are only valid for a linear absorber. If the molecule is a circular absorber [13], equations 10–11 become:

$$n_1(x) = 1 - \exp\left(-3\bar{J}\sigma\varphi(1 - x^2)/2\right) \qquad (12)$$

The power dependency of the photolysed fraction is graphically illustrated by real world experimental values of phytochrome [7], with $\sigma = 1.08 \cdot 10^{-16}$ cm^2/molecule, $\varphi = 0.1$, and a pump beam diameter of 120 μm at 640 nm with a power of 1.65 μJ (see figure 5A). These results can be generated by inserting these experimental values in the graphical user interface that is included in the presented toolbox. The resulting orientational distribution shows that a certain fraction is photolysed for every orientation x, and that very high powers (in multiples of the used power) are needed to obtain full photolysis for a certain orientation. It is not possible to photolyse those molecules that have their transition dipole perpendicularly oriented (at $x = 0$) with respect to the polarisation of the laser (for a linear absorber).

The ensemble-averaged photolysed fraction $\langle n_1 \rangle$ of the randomly oriented molecules can then be analytically calculated by integration or approximated by Taylor expansion:

$$\langle n_1 \rangle = \frac{1}{2} \int_{-1}^{1} dx\, n_1(x)$$

$$= N_1 \sum_{k=0}^{\omega} \frac{(-N_1)^k}{(k+1)!(2k+3)} \qquad (13)$$

An additional illustration of the impossibility to photolyse the disadvantageously oriented molecules is shown in figure 5B ($\langle n_1 \rangle$ as function of $\bar{J}$) where the curve never reaches unity. If higher excited photolysis levels are needed for particular experiment where the photon flux is limiting, it is obvious that tighter focussing of the beam onto the sample increases the resulting excited photolysis levels (depicted in figure 5C).

Gaussian beam profile. The calculation of the photolysed fraction generally assumes that the beam has constant power density which abruptly drops to zero at the beam diameter (see figure 6A). Laser beams are in reality more accurately modelled by a Gaussian profile G as opposed to a top-hat beam profile:

$$G = \exp^{-\frac{D_x^2 + D_y^2}{2s^2}} \qquad (14)$$

In order to correct for a Gaussian beam shape, the used laser pulse power and thus $\bar{J}$ need to be multiplied by equation 14. Note that the standard deviation $s = \Delta/\left(2\sqrt{2\ln(2)}\right)$, where Δ represents the full width half maximum, is the same for both the x and y-direction
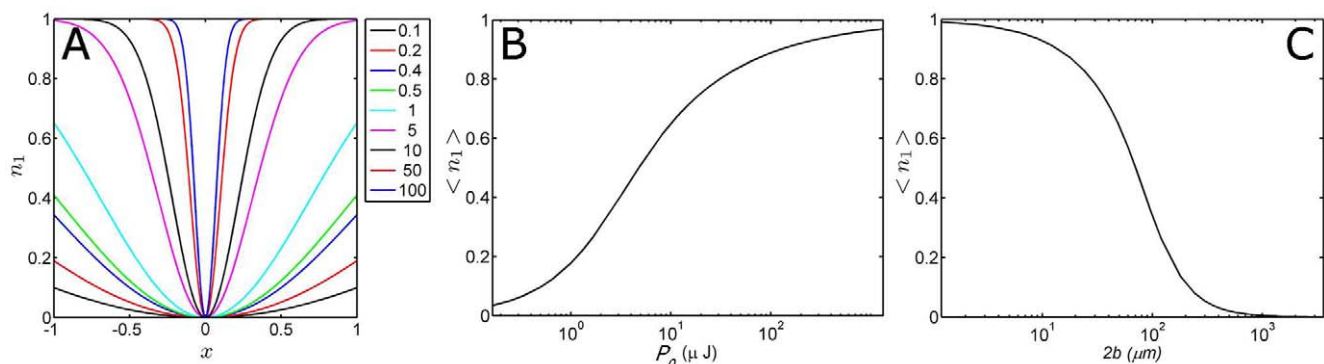

Figure 5. Excited photolysis fraction as function of molecular orientation, laser pump power and beam diameter. Panel A depicts the fraction photolysed molecules as function of the orientation between the optical excitation (with a flat-top, 'multi-mode' pulse profile) and the molecular orientation for different excitation powers (in multiples of the laser power). The ordinate represents the fraction that is photolysed for every molecular orientation. In other words, each orientation x for the molecules in the ensemble relative to the orientation of the pump pulse is evaluated, and the interaction with the optical beam leads to the photolysis of a finite fraction of those molecules. Panel B shows the integrated photolysed fraction (thus for all orientations) from panel A as function of the excitation power of the laser pulse P_0, and as function of beam diameter $2b$ in panel C. The results shown in these panels make use of equation 15.
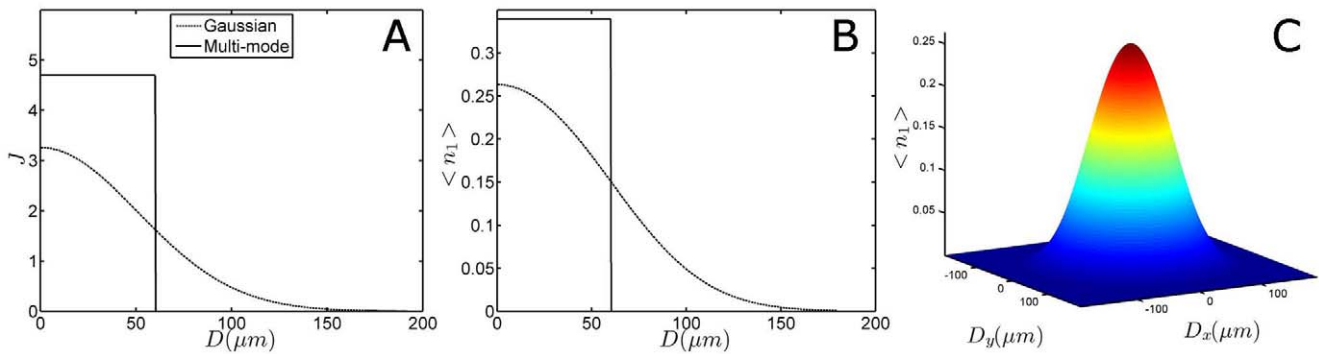doi:10.1371/journal.pone.0017373.g005

Figure 6. Beam profile dependence of excited photolysis fraction. Panel A shows the number of photons (x 10^{20}) per m² for a top-hat and a Gaussian shaped laser pulse at the used laser excitation power (1.65 μJ). Panel B depicts the resulting calculated photolysed fraction $\langle n_1 \rangle$, for both profiles. Both panels share the legend, and their abscissa is given by the distance from the centre of the laser beam D. Panel C shows how $\langle n_1 \rangle$ (also denoted by the colour scale) changes over a 2-dimensional Gaussian beam profile.
doi:10.1371/journal.pone.0017373.g006

in the above formula, assuming a symmetric beam. Previously, a correction factor of $\varepsilon_{2D} = \ln 2$ was applied to convert the maximum laser intensity of a top-hat (multi mode) profile to a Gaussian distribution [7]. This factor originates from the equality of the (two-dimensional) intensity integral of the two differently shaped pulses. The photon density over the two-dimensional beam profile (in any D-axis direction) then becomes:

$$\bar{J}(D) = \frac{\varepsilon_{2D} P_0}{E_0 2\pi s^2} \exp^{-\frac{D_x^2 + D_y^2}{2s^2}}$$ (15)

where P_0 represents the laser pulse power per (beam) area and the E_0 energy per photon. Equations 10, 11 and 13 for the top-hat profile can easily be modified to model Gaussian beams if N_1 is adapted accordingly, and the results are shown in figures 6A and B for $J(D)$ and $\langle n_1(D) \rangle$ as function of the distance from centre of the Gaussian and flat-top beam profiles in one dimension. A more complete graphical presentation of the ensemble-averaged photolysed fraction over a realistic beam profile is finally shown in figure 6C.

Anisotropy. The analytical expressions above show that the excited photolysis level changes over the beam profile. Here we will show that, consequently, the optical anisotropy changes over the beam profile as well. In practice, an optimum anisotropy experiment consists of the measurement of three different relative orientations between pump and probe beam, i.e. parallel, perpendicular and magic angle (54.7°). The measured (macroscopic) anisotropy r is determined by the difference in absorption in parallel $\Delta A_{//}$ and perpendicular $\Delta A_\perp$ directions of the pump beam with respect to the probe beam:

$$r = \frac{\Delta A_{//} - \Delta A_\perp}{\Delta A_{//} + 2\Delta A_\perp}$$ (16)

The magic angle, or anisotropy free, orientation can be reconstructed by $\Delta A_{MA} = (\Delta A_{//} + 2\Delta A_\perp)/3$. As outlined above, the anisotropy is a measure of the microscopic orientation of the polarisation orientation of the laser $\hat{e}_1$, the interacting optical transition $\hat{\mu}_1$, and the probed transition dipole moment $\hat{\mu}_2$ [48]. The above is valid for linear absorbers. In case degenerate transitions exist, such as for hemes [14,15], the anisotropy is

approximated by [49,50]:

$$r = \frac{1}{10} \left(3\cos^2 \psi - 2 \right)$$ (17)

with ψ representing the angle between the linear transition and the plane in which the degenerate transition occurs (and ignoring rotational relaxation).

The ensemble-averaged factor $\langle P_2(\hat{\mu}_1 \cdot \hat{e}_1) \rangle$, which determines the maximum obtainable anisotropy value, is approximated by

$$\langle P_2(\hat{\mu}_1 \cdot \hat{e}_1) \rangle = \frac{\frac{1}{2} \int_{-1}^{1} dx \, n_1(x) P_2(x)}{\frac{1}{2} \int_{-1}^{1} dx \, n_1(x)}$$ (18)

The analytical solution of this equation is a power series of which the coefficients are determined via the Cauchy multiplication rule, and for a low power density it can be approximated by the first 3 terms:

$$\langle P_2(\hat{\mu}_1 \cdot \hat{e}_1) \rangle = \frac{2}{5} \left(1 - \frac{9}{70} N_1 + \frac{19}{2100} N_1^2 + \frac{36}{67375} N_1^3 + \dots \right)$$ (19)

The optical anisotropy as function of N_1 (and thus power density, see figure 7A) is thus obtained, which shows that higher powers lead to less selective excitation (i.e. if multiple relative orientations x reach $n_1 = 1$, which occurs in our phytochrome experiment as described above with approximately 5 times the used laser power, see figure 5A). Ultimately these high powers lead to a reduced observed anisotropy which obviously needs to be avoided.

The substitution of equations 10, 11 and 13 and 15 in equation 19 give the relation between the optical anisotropy and the averaged photolysed fraction (see figure 7C). Note that figures 7A and C are independent of the beam profile characteristics, while figure 7B is only valid for a Gaussian profile. The flat-top profile assumes a fixed N_1 over the beam, which in turn fixes the optical anisotropy $\langle P_2(\hat{\mu}_1 \cdot \hat{e}_1) \rangle$. Only at very low laser pulse intensities it is possible to reach the maximum optical anisotropy of 0.4 (see equation 19). The desired molecular orientation information, defined as the angle γ between $\hat{\mu}_1$ and $\hat{\mu}_2$, can then be calculated
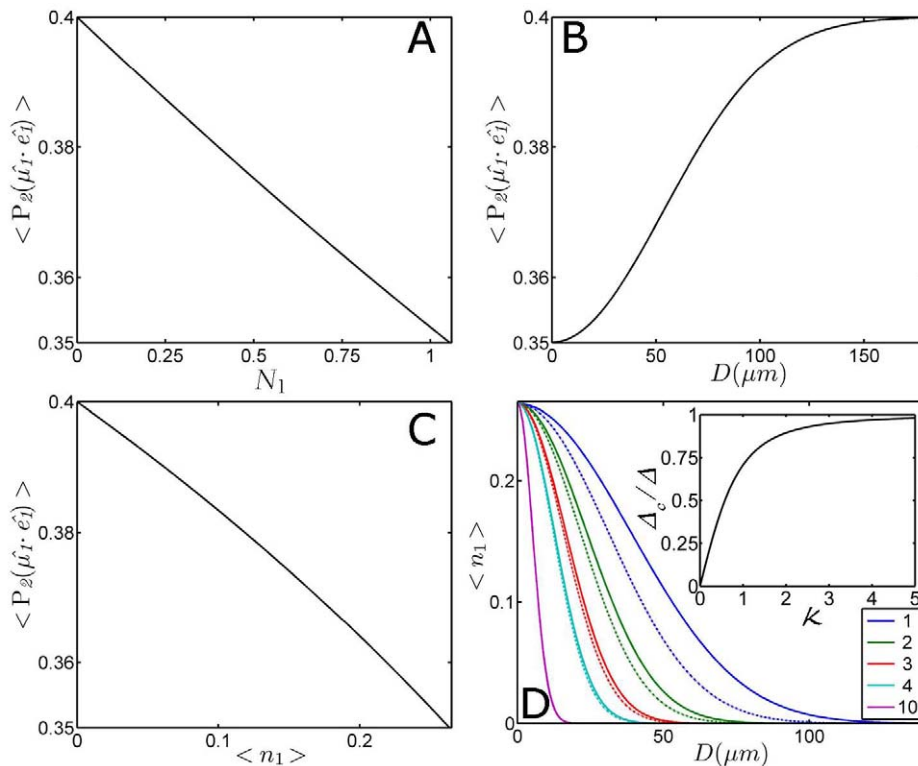
Figure 7. Optical anisotropy and pump/probe diameter ratio dependencies. Panels A–C show the optical anisotropy $\langle P_2(\hat{\mu}_1 \cdot \hat{e}_1)\rangle$ as function of power density N_1, (Gaussian) beam profile, and photolysed fraction $\langle n_1 \rangle$, respectively. Panels A and C are beam profile independent, while panel B is only valid for a Gaussian beam, with the abscissa representing the distance from the centre of the laser beam. Panel D shows the ensemble averaged photolysed fraction as function of the beam profile with varying pump/probe beam diameter ratios (in legend). The continuous lines represent $\langle n_1 \rangle$, and the dashed lines $\langle n_1^{cor}(x)\rangle$. The inset shows the correction factor Δ_c/Δ as function of the pump/probe beam ratio κ.
doi:10.1371/journal.pone.0017373.g007

from the measured anisotropy (the left-hand side of equation 9), and equation 19:

$$\langle P_2(\hat{\mu}_2 \cdot \hat{e}_1)\rangle = \frac{1}{2}(3\cos^2\gamma - 1)\langle P_2(\hat{\mu}_1 \cdot \hat{e}_1)\rangle \qquad (20)$$

Beam size dependence. Earlier we have shown how extreme focussing can lead to very high excited photolysis levels (figure 5C). A pump-probe experiment consists of two pulses, each generally having a different beam diameter. Ideally, such an experiment is done with a much larger diameter of the pump beam with respect to the probe beam to obtain the largest signal. In practice, pump beams are often not more than 1.5 or 2 times larger than the probe beam, which can lead to substantial deviations from the assumed photolysed fractions (see figure 7D) and hence also from the optical anisotropy (see figure 7B). The correction factor can be calculated by the product of the Gaussian beam profile (equation 15) of the pump-and probe-beams (provided both beams overlap in time and space), which ultimately results in a correction for the photolysed fraction (in the D_x-direction):

$$\langle n_1^{cor}(x)\rangle = \langle n_1(x)\rangle \exp^{\dfrac{-D_x^2}{2s_x^2}} \qquad (21)$$

The product of the two Gaussian pump and probe beams, having ratio $\kappa = s_{pump}/s_{probe}$, results in a new Gaussian profile (having Δ_c)

for the probe:

$$\Delta_c = \Delta\sqrt{\frac{\kappa^2}{\kappa^2 + 1}} \qquad (22)$$

The correction factors become 0.71, 0.89 and 0.95 for $\kappa = 1$, 2 and 3, respectively (see the inset in figure 7D). A reasonable value for the pump/probe beam ratio is therefore 3 or greater, in order to prevent substantial deviations of the excited photolysis level over the beam profile.

Depth averaging for absorbing samples. Because the excited photolysis level changes with photon flux, it may change over sample depth as well for highly absorbing samples. A high optical density (also called optically thick) attenuates the transmitted intensity (see figure 8A), consequently changing the photolysed fraction and optical anisotropy as function of sample depth [14]:

$$\bar{J}(z) = \bar{J}_0 10^{-\alpha z/t} \qquad (23)$$

With $\bar{J}_0$ the intensity at the front of the sample, α the sample absorbance, t the sample thickness, and z the sample depth. The attenuated intensity for a real world experiment on a sample with $\alpha = 0.4$ is shown in figure 8B.
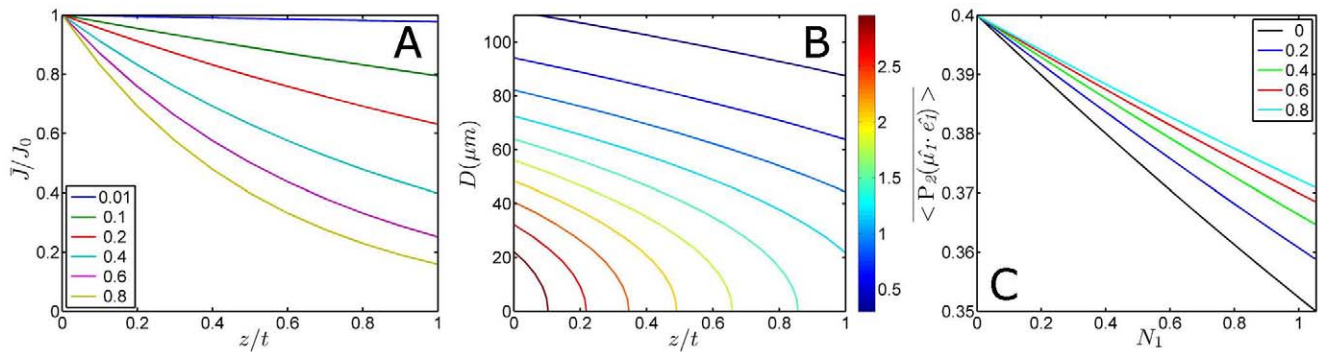
Figure 8. Influence of sample absorbance and thickness on photon density and depth-averaged optical anisotropy. Panel A illustrates the normalised photon density as function of the thickness ratio z/t (0 represents the surface, 1 the total thickness of the sample) for different absorbances (α in legend). Panel B is a contour plot of the transmitted photon density $\bar{J}(z)$ (photons/m^2) over the thickness of the sample (having α) for a Gaussian beam. The colour bar denotes the height of the contours x10^{20}, and the step size in height between the lines is equidistant. Panel C shows the depth-averaged optical anisotropy $\langle P_2(\hat{\mu}_1 \cdot \hat{e}_1) \rangle$ as function of N_1 for a selection of samples with different absorbance, and for a sample without attenuation correction ($\alpha = 0$).
doi:10.1371/journal.pone.0017373.g008

Note that for an optically 'thin' sample the contours are parallel to the abscissa. The parameters $\bar{J}(z)$ and $\bar{J}_0$ in equation 23 can be used for $N_1(z)$ and N_1 (see equation 11), allowing the calculation of the optical anisotropy via the power series in equation 19. The latter means that the excited photolysis fraction needs to be calculated for every layer, which ultimately leads to the depth-average of the optical anisotropy [14]:

$$\overline{\langle P_2(\hat{\mu}_1 \cdot \hat{e}_1) \rangle} = \frac{\int_0^t dz \langle n_1(z) \rangle \langle P_2(\hat{\mu}_1 \cdot \hat{e}_1) \rangle}{\int_0^t dz \langle n_1(z) \rangle} \quad (24)$$

Figure 8C shows the dependency of the depth average on the laser power for different sample absorbances. As for the pump/probe beam diameter ratio and the power density, significant deviations of the observed anisotropy occur, and low powers in combination with low sample absorbances are preferred to avoid the need for corrections.

Multipulse spectroscopy. The expressions (10, 13, 18, 20, and 24) are also applicable to a multipulse experiment, such as a pump-dump-probe experiment [1,4,5,7,51,52], of which a time course is drawn in figure 9.

If the dump pulse is also resonant with the ground state, an additional correction is needed for the photolysed fraction that has returned to the ground state at the moment the second pulse arrives. Therefore, calculation of the finite bleach requires the modelling of the population decay that occurs between pulse *I* and *II*. Here, we develop theory that specifically includes this decay. After the arrival of the first exciting pulse *I*, a certain fraction of the resonant ground state is photolysed. The second pulse *II* after delay τ consequently excites the remaining ground state fraction. The ensemble averaged photolysed fraction for the second pulse is thus dependent on the time-dependent ground state recovery fraction $1 - \beta(\tau)$ [7]:

$$n_2(x) = \left(1 - \beta(\tau)\left(1 - \exp^{-N_1 x^2}\right)\right)\left(1 - \exp^{-N_2 x^2}\right) \quad (25)$$

where the subscripts denote the first or second pulse. Eq. 25 is only valid if the rotational diffusion time is much longer than the time difference between the pump and dump pulses, and if the second pulse is not resonant with the higher lying excited states. From this

expression it can be derived that an instantaneous second pulse (i.e. $\beta(0) = 1$) with equal excitation density gives the same ensemble averaged population as a single pulse with twice the power density (if for instance $N_1 = N_2 = 1$ and $\beta(0) = 1$ it follows
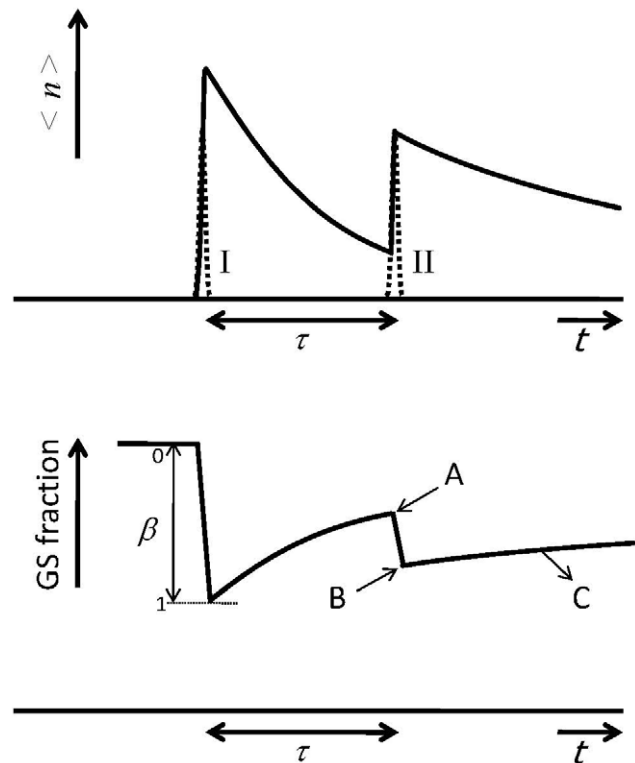

Figure 9. Illustration of a time course in a multipulse experiment. The top panel shows the ensemble averaged photolysed fraction as a function of time for an experiment that consists of two pulses *I* and *II* (drawn as dashed lines) that have the same power density and an inter-pulse time delay $\delta\tau$. The bottom panel sketches the evolution of the corresponding ground state fraction as function of time. The recovered ground state fraction is $1 - \beta$ (which is normalised to the initially photolysed fraction). Detailed information about the (averaged) photolysed fraction for three cases A, B and C in a real-life experiment is shown in figure 10 (in A, B and C, respectively).
doi:10.1371/journal.pone.0017373.g009

that $\langle n_{1+2}\rangle = \langle n_1\rangle + \langle n_2\rangle = 0.25 + 0.15 = 0.40$, identical to $\langle n_1\rangle = 0.40$ when $N_1 = 2$. A fully recovered population (i.e. $\beta = 0$) returns equation 10. The total (dump time τ and dump power N_2 dependent) excited state population then becomes (using Taylor expansion):

$$
\begin{aligned}
\langle n_2\rangle &= \frac{1}{2}\int_{-1}^{1} dx n_2(x) \\
&= N_2 \sum_{k=0}^{\omega} \frac{(-N_2)^k}{(k+1)!(2k+3)} \\
&\quad - \beta(\tau) N_1 N_2 \sum_{m=0}^{\omega}\sum_{k=0}^{\omega} \frac{(-N_1)^m(-N_2)^k}{(k+1)!(m+1)!(2k+5+2m)} \\
&= N_2 \sum_{m=0}^{\omega} (-N_1)^m \beta(\tau)^p \sum_{k=0}^{\omega} \frac{(-N_2)^k}{m!(k+1)!(2k+3+2m)}
\end{aligned}
\tag{26}
$$

where $p = \left(m+1-\sqrt{(m-1)^2}\right)\Big/2$. Equation 26 describes both the two-pulse as the one-pulse experiment simultaneously (most conveniently seen in the second line of the equation): with increasing dump times β goes to zero, and $\langle n_2\rangle$ consequently forms the solution to equation 13 (i.e. $\langle n_2\rangle$ moves over the curve from $N_1 + N_2$ to N_1).

To illustrate the use of the derived multipulse expressions a real world example is used, taken from [7]. This pump-dump-probe experiment uses the same intensity (and wavelength) for both pump and dump pulses, and three different pump-dump delay times have been used (1, 14 and 500 ps), which, depending on the *model* applied in the global analysis, correspond to different ground state recovery fractions (0.13, 0.64 and 0.9, respectively, for a parallel model). Assuming that no rotational diffusion occurs on the time scale of the delays that are used, the ground state is populated by the initially excited molecules (see figure 10A and case A in figure 9). The interaction of pulse *II* with the recovered ground stare population from case A is shown in figure 10B (corresponding to case B in figure 9). The orientation distribution of case B is very different from that in figures 5A and 10A, illustrating the necessity of the inclusion of the developed theory in similar experimental configurations. Figure 10C (black curve) shows how the averaged photolysed fraction (excited by the first pulse) evolves over time due to spontaneous decay (case C in figure 9). In our example a fraction of 0.9 eventually recovers to the ground state. Consequently, the photolysed fraction induced by the second pulse increases with time (blue curve). Note also that if both pulses arrive simultaneously, it is the excitation power of both pulses that needs to be added, and not the corresponding photolysed fractions.

The anisotropy in a multipulse experiment is analogous to equations 18 and 19, and can be calculated via:

$$
\langle P_2(\hat{\mu}_1 \cdot \hat{e}_1)\rangle = \frac{\frac{1}{2}\int_{-1}^{1} dx n_2(x) P_2(x)}{\frac{1}{2}\int_{-1}^{1} dx n_2(x)}
\tag{27}
$$

Note that the ensemble averaged excited photolysed fraction for the second pulse is dependent on the ground state recovery fraction $1-\beta$. Therefore, $\langle P_2(\hat{\mu}_1 \cdot \hat{e}_1)\rangle$ is plotted for different values of β in figure 10D. Obviously, the deviation from the single pulse experiment (i.e. the continuous blue line; $\beta = 0$) becomes

extremely severe for smaller ground state recovery fractions, hampering accurate anisotropy measurements. Again, it shows the importance of the use of low power densities in (accurate) anisotropy measurements. In addition, the effect of differences in photon density is also illustrated for the case where the first pulse has twice the intensity of the second pulse (represented by the dashed-dotted lines). Note also that the curve given by $\beta = 0$ and $N_1 = N_2$ corresponds to the curve drawn in figure 7A.

For a pump-dump-probe or pump-repump-probe experiment there is usually at least some measurable cross-section for both the ground state and the intermediate at the frequency of the pulse *II*, even if it is different from that of pulse *I*. Therefore, equations 13 and 26 must be applied in order to extract the time resolved signal that corresponds to the pure intermediate ΔA, which result from a pump-probe (*PP*), dump-probe (*DP*) and pump-dump-probe (*PDP*) experiment:

$$
\Delta A = PDP - PP - \alpha DP
\tag{28}
$$

Where $\alpha = \langle n_2\rangle/\langle n_1\rangle$. This approach has previously been applied to phytochrome [7]. The described population modelling is therefore required before the time resolved signal of a pure intermediate (just as for the pure ground state) may be analysed by global fitting and singular value decomposition.

Experimental implications. The theory described in this section allows for the design of an optimal laser experiment, and can extract the signal that corresponds to the dynamics of a pure intermediate that can either be an excited state (pump-dump-probe) or a ground state intermediate (pump-repump-probe). For instance, pump and probe beam diameters, optical density, and sample thickness may need to be adjusted for optimum detection of anisotropy in a polarised pump-probe experiment. Also, the photolysed fraction should be kept in the region of 0.1 if non-linear contributions are to be avoided.

Conclusions. Mathematical modelling and analysis is essential to resolve the details and model physical behaviour of populations in ultrafast spectroscopy experiments. A global analysis of such data can help to significantly reduce all time-dependent data to time-independent spectra and their associated time constants, and facilitating its interpretation. Here we present a software toolbox that permits the modelling of any time-dependent data by a kinetic scheme. It has been specifically developed for (dual resonance) ultrafast pump-probe spectroscopy, and includes finite bleach modelling. Based on the applied reaction model, the (bio-) physical pathways can be unravelled by globally analysing time-resolved experimental data. Throughout this work real world examples are used to illustrate molecular dynamics from pump-probe to multi-colour and multi-pulse experiments. In addition to the software, we present new theory that describes the finite bleach effect of (single and multiple) intense polarised femtosecond pulses on an ensemble of randomly oriented particles in the presence of population decay. This theory is essential in order to treat data from multi-pulse experiments, and to unravel transient populations and its characteristics in the spectral and time-domain. For single and multipulse experiments it is shown that the averaged photolysed fraction varies as function of power density and beam profile. In addition, it is graphically shown how the anisotropy changes over a Gaussian beam profile for single and multipulse experiments, and that this can lead to a significant underestimation of the actual orientation of the molecular transition dipole if compared with a top-hat 'multi-mode' profile. The interaction with multiple pulses in addition to depth averaging in optically dense samples is described. The
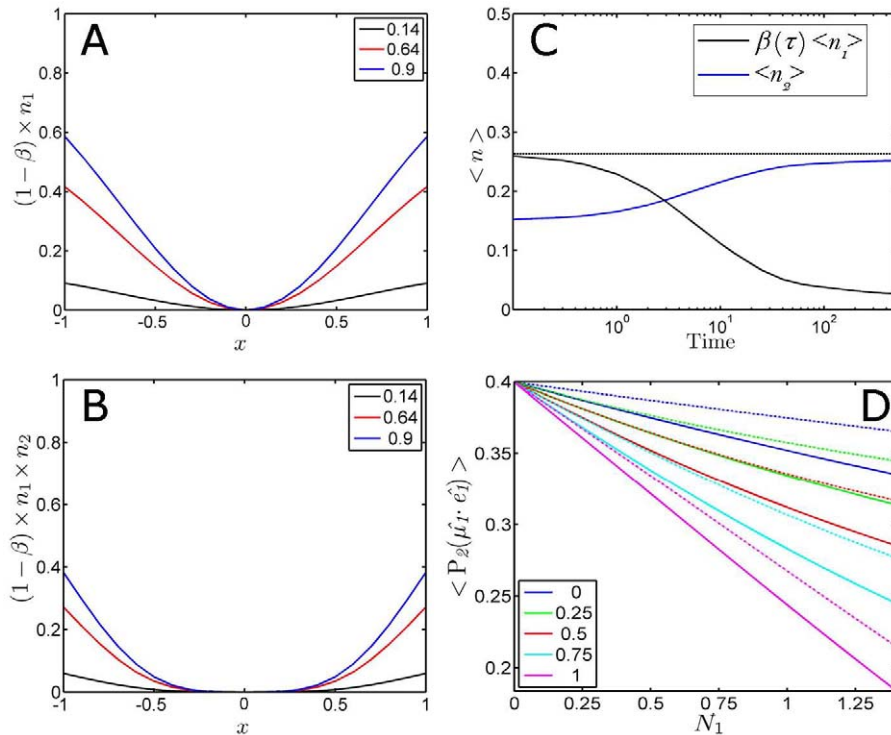
Figure 10. Photolysed fraction as function of orientation and time for multipulse experiments. Panel A shows $(1-\beta) \times n_1$ as function of x, representing the recovery to the ground state of the orientation distribution of different fractions of molecules that have been photoexcited by the first pulse (case A in figure 9). Three (increasing) ground state recovered fractions (in legend) are depicted, corresponding to (increasing) inter-pulse delays of 1, 14 and 500 ps. Panel B illustrates $(1-\beta) \times n_1 \times n_2$ as function of x, representing how the orientation distributions in panel A change after interaction with a second pulse (case B in figure 9). Panel C plots the time course of the (ensemble-averaged, i.e. $\langle n \rangle$) ground state recovered photolysed fraction for the first pulse, and that of the second pulse as function of the delay time $\delta\tau$ between the first and second pulse (case C in figure 9). The horizontal line is the (single-pulse) photolysed fraction at time zero. Panel D shows the optical anisotropy $\langle P_2(\hat{\mu}_1 \cdot \hat{e}_1) \rangle$ for a multipulse experiment as function of power density. Each line represents the power density dependence for a different β (in legend), and the continuous lines correspond to $N_1 = N_2$ and the dashed-dot lines to $N_1 = 2N_2$.
doi:10.1371/journal.pone.0017373.g010

modelling allows for the design of the optimum conditions for an ultrafast spectroscopy experiment. In summary, we have demonstrated that full understanding of macroscopic behaviour (aided by global analysis), requires a detailed understanding of its microscopic components (facilitated by modelling).

Supporting Information

File S1 A real world ultrafast spectroscopic experiment on phytochrome is further discussed in File S1 to illustrate the influence of several key parameters on a multipulse experiment. (DOC)

File S2 Documentation to the Population Dynamics Modelling Toolbox is supplied as supplementary information File S2, and includes a step-by-step guide to use the described graphical user interface, and a step-by-step guide to global fitting, describing

discrete steps we consider to represent the minimum of analysis necessary to unravel the kinetics. (DOC)

Author Contributions

Conceived and designed the experiments: LJGWvW CNL JvT. Performed the experiments: LJGWvW CNL JvT. Analyzed the data: LJGWvW CNL JvT. Contributed reagents/materials/analysis tools: LJGWvW CNL JvT. Wrote the paper: LJGWvW JvT. Wrote computer code: LJGWvW.

References

1. Gai F, McDonald JC, Anfinrud PA (1997) Pump-dump-probe spectroscopy of bacteriorhodosin: Evidence for a near-IR excited state absorbance. J Am Chem Soc 119: 6201–6202.
2. Rupenyan A, van Stokkum IHM, Arents JC, van Grondelle R, Hellingwerf KJ, et al. (2009) Reaction Pathways of Photoexcited Retinal in Proteorhodopsin Studied by Pump-Dump-Probe Spectroscopy. Journal of Physical Chemistry B 113: 16251–16256.
3. Larsen DS, Vengris M, van Stokkum IHM, van der Horst MA, de Weerd FL, et al. (2004) Photoisomerization and Photoionization of the Photoactive Yellow Protein Chromophore in Solution. Biophysical Journal 86: 2538–2550.
4. Kennis JTM, Larsen DS, van Stokkum IHM, Vengris M, van Thor JJ, et al. (2004) Uncovering the hidden ground state of green fluorescent protein. Proc Natl Acad Sci USA 101: 17988–17993.
5. van Thor JJ, Zanetti G, Ronayne KL, Towrie M (2005) Structural events in the photocycle of green fluorescent protein. Journal of Physical Chemistry B 109: 16099–16108.
6. Changenet-Barret P, Choma CT, Gooding EF, DeGrado WF, Hochstrasser RM (2000) Ultrafast dielectric response of proteins from dynamics stokes shifting of coumarin in calmodulin. Journal of Physical Chemistry B 104: 9322–9329.
7. van Wilderen LJGW, Clark IP, Towrie M, van Thor JJ (2009) Mid-Infrared Picosecond Pump-Dump-Probe and Pump-Repump-Probe Experiments to

Resolve a Ground-State Intermediate in Cyanobacterial Phytochrome Cph1. Journal of Physical Chemistry B 113: 16354–16364.

8. Dietzek B, Brueggemann B, Pascher T, Yartsev A (2007) Pump-shaped dump optimal control reveals the nuclear reaction pathway of isomerization of a photoexcited cyanine dye. Journal of the American Chemical Society 129: 13014–13021.

9. Pique JP, Engel YM, Levine RD, Chen Y, Field RW, et al. (1988) Broad Spectral Features in the Stimulated-Emission Pumping Spectrum of Acetylene. J Chem Phys 88: 5972–5974.

10. Wohlleben W, Buckup T, Hashimoto H, Cogdell RJ, Herek JL, et al. (2004) Pump-deplete-probe spectroscopy and the puzzle of carotenoid dark states. Journal of Physical Chemistry B 108: 3320–3325.

11. Larsen DS, Papagiannakis E, van Stokkum IHM, Vengris M, Kennis JTM, et al. (2003) Excited state dynamics of beta-carotene explored with dispersed multi-pulse transient absorption. Chemical Physics Letters 381: 733–742.

12. Papagiannakis E, Vengris M, Larsen DS, van Stokkum IHM, Hiller RG, et al. (2006) Use of ultrafast dispersed pump-dump-probe and pump-repump-probe spectroscopies to explore the light-induced dynamics of peridinin in solution. Journal of Physical Chemistry B 110: 512–521.

13. Ansari A, Jones CM, Henry ER, Hofrichter J, Eaton WA (1993) Photoselection in Polarized Photolysis Experiments on Heme-Proteins. Biophys J 64: 852–868.

14. Lim MH (2002) The orientation of CO in heme proteins determined by time-resolved mid-IR spectroscopy: Anisotropy correction for finite photolysis of an optically thick sample. B Kor Chem Soc 23: 865–871.

15. Lim M, Jackson TA, Anfinrud PA (2004) Orientational distribution of CO before and after photolysis of MbCO and HbCO: A determination using time-resolved polarized mid-IR spectroscopy. Journal of the American Chemical Society 126: 7946–7957.

16. van Thor JJ, Ronayne KL, Towrie M (2007) Formation of the early photoproduct lumi-R of cyanobacterial phytochrome cph1 observed by ultrafast mid-infrared spectroscopy. J Am Chem Soc 129: 126–132.

17. Dasgupta J, Frontiera RR, Taylor KC, Lagarias JC, Mathies RA (2009) Ultrafast excited-state isomerization in phytochrome revealed by femtosecond stimulated Raman spectroscopy. Proceedings of the National Academy of Sciences of the United States of America 106: 1784–1789.

18. van Stokkum IHM, Larsen DS, van Grondelle R (2004) Global and target analysis of time-resolved spectra. Biochimica Et Biophysica Acta-Bioenergetics 1657: 82–104.

19. Albrecht AC (1961) Polarizations and Assignements of Transitions: The Method of Photoselection. Journal of Molecular Spectroscopy 6: 84–108.

20. Steinber Iz, Ehrenber B (1974) Theoretical Evaluation of Effect of Photoselection on Measurement of Circular-Polarization of Luminescence. Journal of Chemical Physics 61: 3382–3386.

21. Wynne K, Hochstrasser RM (1993) Coherence Effects in the Anisotropy of Optical Experiments. Chemical Physics 171: 179–188.

22. Einax JW (2008) Chemometrics in analytical chemistry. Analytical and Bioanalytical Chemistry 390: 1225–1226.

23. Lavine B, Workman J (2008) Chemometrics. Analytical Chemistry 80: 4519–4531.

24. Holzwarth AR (1995) Time-Resolved Fluorescence Spectroscopy. Biochemical Spectroscopy 246: 334–362.

25. Satzger H, Zinth W (2003) Visualization of transient absorption dynamics - towards a qualitative view of complex reaction kinetics. Chemical Physics 295: 287–295.

26. Brochon JC (1994) Maximum-Entropy Method of Data-Analysis in Time-Resolved Spectroscopy. Numerical Computer Methods, Pt B 240: 262–311.

27. Heyne K, Herbst J, Stehlik D, Esteban B, Lamparter T, et al. (2002) Ultrafast dynamics of phytochrome from the cyanobacterium Synechocystis, reconstituted with phycocyanobilin and phycoerythrobilin. Biophysical Journal 82: 1004–1016.

28. Croce R, Muller MG, Bassi R, Holzwarth AR (2001) Carotenoid-to-chlorophyll energy transfer in recombinant major light-harvesting complex (LHCII) of higher plants. I. Femtosecond transient absorption measurements. Biophysical Journal 80: 901–915.

29. Roelofs TA, Lee CH, Holzwarth AR (1992) Global Target Analysis of Picosecond Chlorophyll Fluorescence Kinetics from Pea-Chloroplasts - a New Approach to the Characterization of the Primary Processes in Photosystem-Ii Alpha-Units and Beta-Units. Biophysical Journal 61: 1147–1163.

30. Nibbering ETJ, Elsaesser T (2000) Probing solvation dynamics with femtosecond vibrational spectroscopy. Applied Physics B-Lasers and Optics 71: 439–441.

31. Qian J, Schultz SL, Jean JM (1995) Observation of Intramolecular Vibrational Redistribution and Vibrational Cooling in S-2 Trans-Stilbene and 2-Phenyl-lindene in Solution. Chem Phys Lett 233: 9–15.

32. Hester RE, Matousek P, Moore JN, Parker AW, Toner WT, et al. (1993) Vibrational Mode-Selective Effects in the Picosecond Time-Resolved Resonance Raman-Spectrum of Singlet Excited Trans-Stilbene. Chem Phys Lett 208: 471–478.

33. Middleton CT, Cohen B, Kohler K (2007) Solvent and Solvent Isotope Effects on the Vibrational Cooling Dynamics of a DNA Base Derivative. J Phys Chem A 111: 10460–10467.

34. Istratov AA, Vyvenko OF (1999) Exponential analysis in physical phenomena. Review of Scientific Instruments 70: 1233–1257.

35. Torczon V (1997) On the convergence of pattern search algorithms. Siam Journal on Optimization 7: 1–25.

36. Audet C, Dennis JE (2003) Analysis of generalized pattern searches. Siam Journal on Optimization 13: 889–903.

37. Ihme M, Marsden AL, Pitsch H (2008) Generation of optimal artificial neural networks using a pattern search algorithm: Application to approximation of chemical systems. Neural Computation 20: 573–601.

38. Nicosia G, Stracquadanio G (2008) Generalized Pattern Search Algorithm for Peptide Structure Prediction. Biophysical Journal 95: 4988–4999.

39. Wienke D, Lucasius C, Kateman G (1992) Multicriteria Target Vector Optimization of Analytical Procedures Using a Genetic Algorithm .1. Theory, Numerical Simulations and Application to Atomic Emission-Spectroscopy. Analytica Chimica Acta 265: 211–225.

40. Bray T, Paoli J, Sperberg-McQueen CM, Maler E, Yergeau F (2008) Extensible Markup Language (XML) 1.0.

41. Hucka M, Bergmann F, Hoops S, Keating SM, Sahle S, et al. (2010) The Systems Biology Markup Language (SBML): Language Specification for Level 3 Version 1 Core (Release 1 Candidate).

42. Scharf LL (1991) The Svd and Reduced Rank Signal-Processing. Signal Processing 25: 113–133.

43. Henry ER, Hofrichter J (1992) Singular Value Decomposition - Application to Analysis of Experimental-Data. Methods in Enzymology 210: 129–192.

44. Joshi CP, Borucki B, Otto H, Meyer TE, Cusanovich MA, et al. (2006) Photocycle and photoreversal of photoactive yellow protein at alkaline pH: Kinetics, intermediates, and equilibria. Biochemistry 45: 7057–7068.

45. Schmidt M, Rajagopal S, Ren Z, Moffat K (2003) Application of singular value decomposition to the analysis of time-resolved macromolecular X-ray data. Biophysical Journal 84: 2112–2129.

46. Mullen KM, van Stokkum IHM (2007) TIMP: An R package for modeling multi-way spectroscopic measurements. Journal of Statistical Software 18.

47. Kovalenko SA, Dobryakov AL, Ruthmann J, Ernsting NP (1999) Femtosecond spectroscopy of condensed phases with chirped supercontinuum probing. Physical Review A 59: 2369–2384.

48. van Thor JJ, Ronayne KL, Towrie M, Sage JT (2008) Balance between ultrafast parallel reactions in the green fluorescent protein has a structural origin. Biophysical Journal 95: 1902–1912.

49. Moore JN, Hansen PA, Hochstrasser RM (1988) Iron-carbonyl bond geometries of carboxymyoglobin and carboxyhemoglobin in solution determined by picosecond time-resolved infrared spectroscopy. Proc Natl Acad Sci U S A 85: 5062–5066.

50. Borisov VB, Liebl U, Rappaport F, Martin JL, Zhang J, et al. (2002) Interactions between heme d and heme b595 in quinol oxidase bd from Escherichia coli: a photoselection study using femtosecond spectroscopy. Biochemistry 41: 1654–1662.

51. Changenet-Barret P, Choma C, Gooding E, DeGrado W, Hochstrasser RM (2000) Ultrafast Dielectric Response of Proteins from Dynamics Stokes Shifting of Coumarin in Calmodulin. J Physical Chemistry B 104: 9322–9329.

52. Logunov SL, Volkov VV, Braun M, El-Sayed MA (2001) The Relaxation dynamics of the excited electronic states of retinal in bacteriorhodopsin by two-pump probe femtosecond studies. Proc Natl Acad Sci U S A 98: 8475–8479.