Supporting Information: Exciton Structure and Energy Transfer in the Fenna-Matthews-Olson Complex

Erling Thyrhaug[1], Karel Žídek[1], Jakub Dostál[1,2], David Bína[3], Donatas Zigmantas[1].*

[1] - Department of Chemical Physics, Lund University, P.O. Box 124, 22100 Lund, Sweden

[2] - Institut für Physikalische und Theoretische Chemie, Universität Würzburg, Am Hubland, D-97074 Würzburg, Germany

[3] - Biology Centre CAS, Branišovská 31, and Faculty of Science, University of South Bohemia, Branišovská 1760, 370 05 České Budějovice, Czech Republic

Corresponding Author

* donatas.zigmantas@chemphys.lu.se

C. tepidum culture and FMO preparation: Cultivation of *C. tepidum* (strain TLS, DSM 12025) was performed in 800 mL batches of modified Pfennig medium (Wahlund et al.[1]) kept at 45°C, continuously illuminated with 60 W incandescent light bulbs. FMO preparation was based on Wen et al.[2] Briefly, cells were harvested after 3 days of cultivation by centrifugation at 6000 g, resuspended in 20 mM Tris-HCl, pH 8 and broken using EmulsiFlex-C5 (Avestin Inc., Canada) at 20000 psi. Following removal of unbroken cells by low speed centrifugation the membranes were collected by ultracentrifugation at 200000 g for 2 hours. FMO was released from membranes using 0.4 M Na_2CO_3 added in two steps over the course of 2 days. Following removal of debris by ultracentrifugation, the soluble protein fraction was dialyzed against 20 mM Tris, pH 8 for 72 hours, concentrated and purified using size exclusion and anion exchange chromatography until the OD_{271} / OD_{371} ratio dropped under ~0.6.

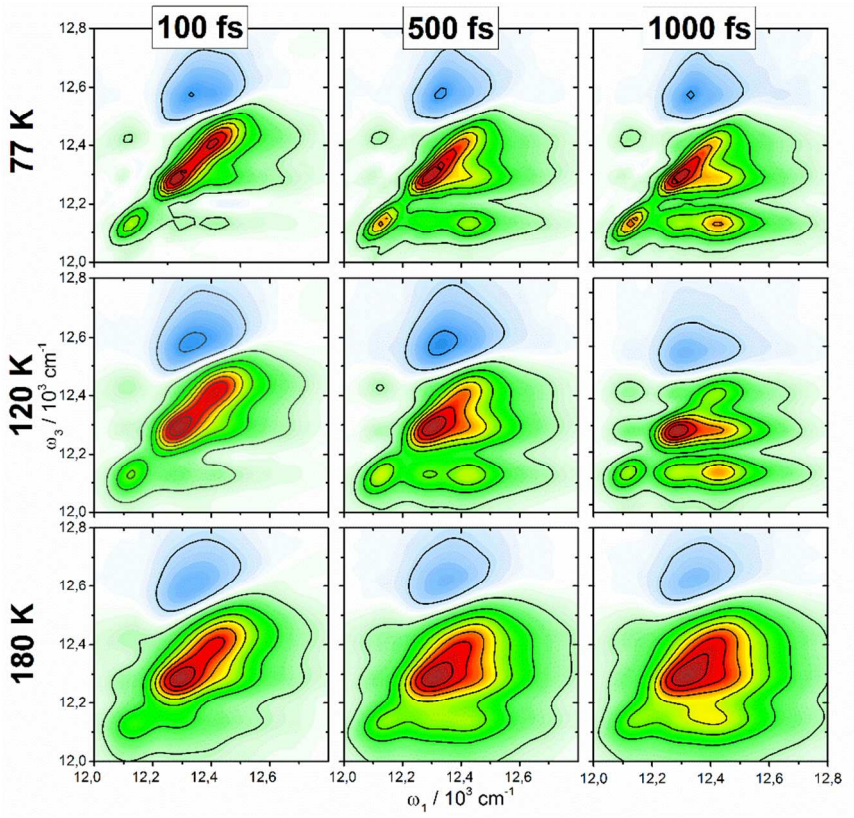

Figure S 1: Comparison of 2DES spectra, measured with magic angle polarization settings at different population times and temperatures.

Comparison of different approaches to PAS generation: Several different approaches can be found to separate fluorescence, pump-probe, or 2D spectrum into polarized components, thus allowing for the removal of unwanted signals (or at the very least – maximization of the desired signal). We are aware of at least three (for the present case) functionally similar, but conceptually different approaches to spectral decomposition into polarization associated spectra (PAS). These are: the anisotropy/PAS approach of Albrecht[3], the polarization tensor equation for four-wave mixing of Wright[4] and Hochstrasser[5-6], and the symmetry approach of Abramavicious & Mukamel[7]. These lead in many cases to (within a sign difference) identical results, as fundamentally the same operations are applied to the experimental data. In analysis of incoherent

dynamics and static excitonic structure (as is the case in this study), we favor the Albrecht approach due to its straightforward geometric interpretation. It is however worth noting some conceptual differences in the different approaches.

The Albrecht approach separates the isotropic spectrum (by use of the fundamental fluorescence-, pump-probe-, or 2DES- anisotropy) into components belonging to orthogonal molecular axes. In practice the absorptive contributions parallel and perpendicular to the probed transition (i.e. emission or pump-probe detected signal) are extracted. In the original paper an isotropic ensemble and orthogonal transitions were assumed.[3] Neither are fundamental limitations, and generalization to oriented ensembles and non-orthogonal transitions are straightforward, though somewhat cumbersome. In ultrafast- and multi-dimensional-spectroscopy two major limitations appear: 1) working via the fundamental anisotropy, only pulse sequences with pairwise identical pulses can be treated. This only arise as an issue in specific 2DES (and possibly in other multidimensional techniques) pulse-polarization sequences, as this requirement is automatically fulfilled for e.g. fluorescence and transient absorption. 2) Coherent signals do not appear in the theory. If coherences are induced by the experiment, these appear in the PAS spectra, however no direct interpretation of these quantum beats is provided. In the PAS spectra used in the main text pure correlated vibrational coherences (in the language of Jonas and coworkers[8]) will be suppressed in the S_Y spectrum, while electronic or vibronically mixed coherences between states of different polarization will appear as (off-phase) beatings between the two PAS spectra.

The approach by Abramavicious and Mukamel exploits certain symmetries (at time = 0) between pulse sequences in 2D experiments to cancel or enhance specific signals. This is clearly convenient for identifying spectral combinations that provides the desired type of signal, e.g. to

maximize coherent or chiral signals. No geometric information is provided however, and the approach may thus be inconvenient for exciton structure characterization.

In Hochstrassers' and Wright's approach all relevant information is collected in a tensor equation. This allows one to directly calculate the signal in a given "pixel" in the 2DES map, given the contributing transition moments and the polarization of the optical pulses. While this is a completely general approach, it does not in general provide any straight-forward method for identifying polarization conditions or combination spectra that maximizes (or minimizes) a given signal. To do so one must solve the tensor equation directly, which may not lead to the most straight-forward or elegant solution, depending on the problem at hand. Further it is a "point-by-point" calculation approach, and thus reveals little about the spatial relations in complex exciton structures.

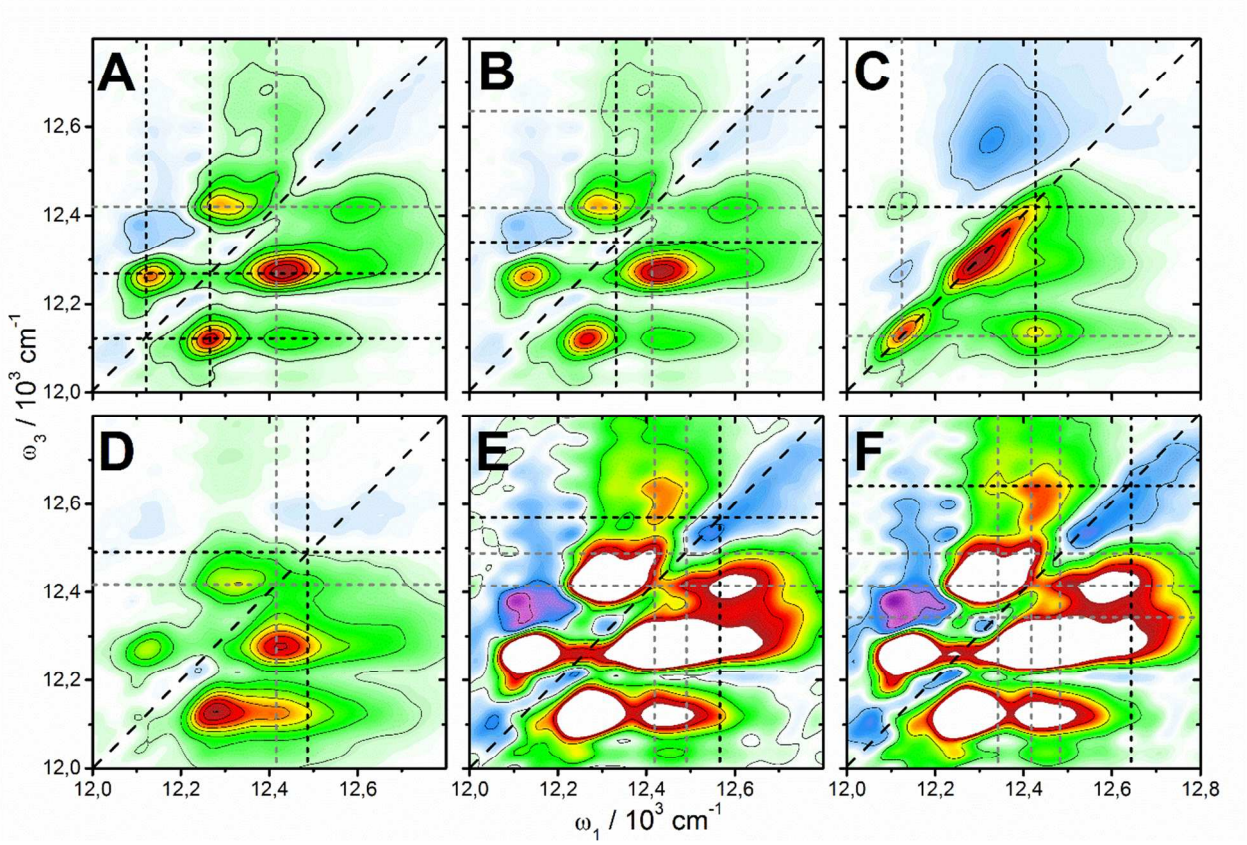

Figure S 2: Examples of energy-level triangulations in PAS maps. Panels A, B, E and F: S_Y spectrum at 40 fs population time. Panel C: S_Z spectrum at 1 ps population time. Panel D: S_Y spectrum at 1 ps population time. Amplitude is rescaled in Panels E and F to better reveal the cross-peaks. A: excitons 1 and 2 (black lines) and exciton 4 (gray). B: Exciton 3 (black) and excitons 4 and 6 (gray). C: Exciton 4 (black) and exciton 1 (gray). D: Exciton 5 (black) and exciton 4 (gray). E: Exciton 6 (black) and excitons 4 and 5 (gray). F: Exciton 7 (black) and excitons 3-5 (gray).

Table S 1: Comparison of exciton energies with literature data

Exciton	Current Work - PAS	Current Work - Fit	Vulto *et al.*
1	12.12	12.121	12.113
2	12.27	12.275	12.262
3	12.34	12.348	12.355
4	12.41	12.415	12.414
5	12.48	12.487	12.448
6	12.57	12.581	12.611
7	12.64	12.685	12.649

Exciton energies as read from PAS spectra, after kinetic fit, and literature values from Vulto and coworkers[9]. Energies in 10^3 cm^{-1}. The second data column, with energies resulting from the kinetic fit, represent a refinement of the exciton energies estimated by visual inspection in the first data column.

DAS Analysis: A four-component sum-of-exponential DAS decomposition of the 2DES data is shown in Figure S 3. The two longest components are well separated, and DAS analysis provides reliable decay rates and spectral shapes. These components have been assigned to inter-unit energy transfer and the ground-state recovery respectively. The 1.2 ps component is similarly relatively well captured by DAS, mainly showing the decay of exciton 3 (and an associated above-diagonal excited state absorption band) and the corresponding rise of the population of exciton 1. The remaining dynamics that are necessary to capture are the numerous intra-unit transfer processes, which have relatively closely spaced rates. This is demonstrated in the short-lifetime DAS in Figure S 3, where a large number of components appear with an identical lifetime of 200 fs. Adding more decay-components (without more advanced analysis approaches) in order to capture the individual transfer processes does not result in a meaningful

improvement in the fit statistics. In order to extract the individual level-to-level transfer rates rather than the sum-total observed in DAS we thus fit the data directly to a kinetic model.

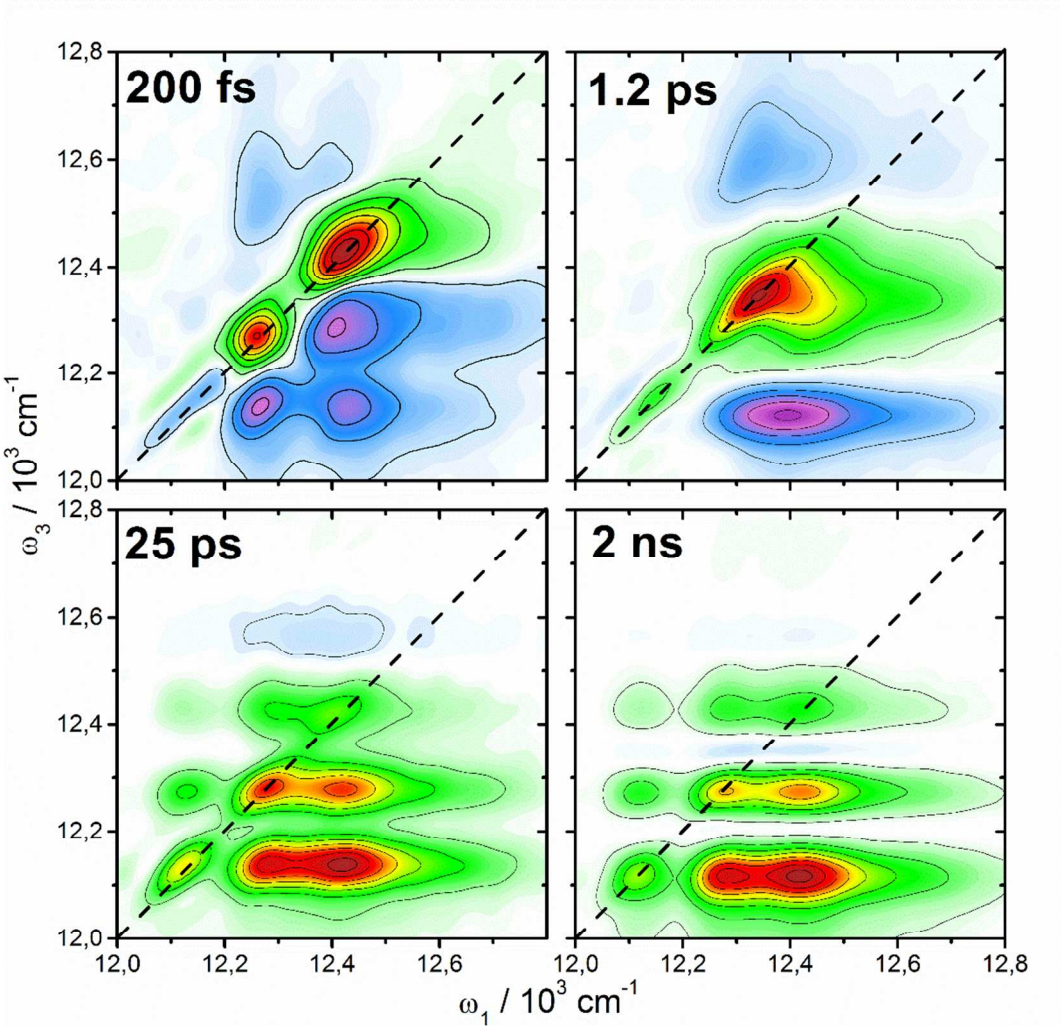

Figure S 3: Four-component DAS fit of magic angle data at 77 K. The components corresponding to the relaxation of exciton 3 (1.2 ps here), inter-unit energy transfer (25 ps) and ground-state recovery (~2 ns) are well-defined due to large separation in apparent lifetimes. The remaining intra-unit transfer processes appear as a single 200 fs component due to presence of multiple relaxation processes with similar lifetimes.

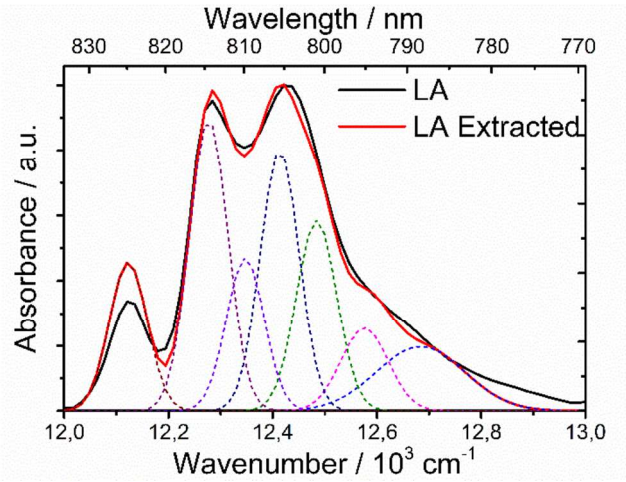

Figure S 4: Linear absorption spectrum (black) and LA spectrum calculated from the kinetic fit (red). Excitonic transitions contributing to the kinetic fit shown as dashed lines.

Fitting Procedure: The dynamics resolved by transient absorption (TA) spectroscopy is most often interpreted in terms of a rate equation model. The assumptions are made that 1: the decay of the population of one particular excited state (transient species) i is directly proportional to the population itself and 2: that the population of this state is increased by the population transfer from the other states $j \neq i$, which is seen as a decay of these states. The connectivity between N species and one ground state is described by a set of "intrinsic rate constants" k_{ij}. The mathematical formulation of this model is then a set of homogeneous linear differential equations of the first order

$$\frac{\mathrm{d}\vec{p}(t)}{\mathrm{d}t} = K\vec{p}(t), \qquad \text{(S1)}$$

where $\vec{p}(t)$ is (N+1)-component time-dependent vector of populations of individual states. K is a matrix with intrinsic rate constants as its off-diagonal elements. The diagonal elements of K are the "apparent rates" $k_{ii} = -\sum_{j \neq i} k_{ji}$ describing the lifetime of the individual states. As a consequence, the time-evolution of any state can (in almost all cases) be written as a linear combination of exponential functions decaying with different apparent rate constants.

Global analysis – one of the most common method of processing TA spectra – heavily relies on such a formulation. By fitting a set of decaying exponentials at each point of the acquired TA map one can retrieve most prominent apparent rates. By looking at their amplitude distribution in the spectrum one can interpret the spectra of individual states and infer intrinsic rate constants of the transfer processes connecting them. As it has been repeatedly shown before,[10] TA spectra unfortunately do not contain enough information to allow for unambiguous interpretation, since an infinite number of model schemes (combinations of intrinsic rates with spectra of individual states) result in an identical overall evolution of the TA spectrum.

The situation is however different for the 2DES data, because the extra dimension in the absorptive 2D spectra provide necessary additional information, and as a result not every model scheme leading to identical TA spectra generates identical 2D spectra. Therefore, careful analysis of the absorptive 2D spectra can be used to exclude multiple model schemes that would be otherwise considered as a valid interpretation of the TA spectra.

The spectral decomposition used in this paper proceeds as follows. The initial energy of the levels (relative to the ground state - GS), the transition dipole moments, homogeneous linewidths and all downwards rate constants are guessed. Spectral species are constructed as described in detail below. The upwards rate constants are calculated as $k_{up} = k_{down} \cdot \exp(-\Delta E/k_b T)$ to fulfil the detailed balance condition (ΔE is the energy difference between the two states connected by the rate constants, k_b is the Boltzmann constant, and T the absolute temperature). The rate equations are solved numerically using a combination of the classical Runge-Kutta and implicit Euler methods. Multiple initial conditions are used, depending on the excitation wavelength, each time a single state is populated (one spectral species). The subsequent time evolution describes dynamics observed in a single column of the 2D spectrum at the excitation frequency ω_1. The

intensity of each column is further multiplied by the corresponding square of the transition dipole moment to correct for the relative intensities of individual transitions.

The 2D spectral shape of every transient spectral species is assumed to consist of a set of two-dimensional Gaussian bands, arranged in a column and centered at the ω_3 frequencies of the individual states. The amplitude of individual bands is yet to be determined and can be both positive (to reflect the GSB and SE contributions) and negative (ESA contributions). In the fit multiple spectral species can contribute to each column, when multiple states are populated with increasing population time, depending on the energy relaxation pathways.

Altogether the fitted model function S can be written as

$$S = L_3 \,\mathrm{diag}(\mu_i^2)\, F \exp(K' t)\, \mathrm{diag}(\mu_i^2)\, L_1^{\mathrm{T}}, \qquad \text{(S2)}$$

where K' is a (N x N) matrix constructed from the rate constant matrix K by omitting the row and column of rates to or from the ground state. The symbol exp() denotes a matrix exponential. $\mathrm{diag}(\mu_i^2)$ is a diagonal matrix containing squares of transition dipole moments of individual excited states as its non-zero elements. The composition of spectra associated to the individual transient species as a set of Gaussian bands (as described above) is characterized by the matrix product $\mathrm{diag}(\mu_i^2)\, F$. Each element $\mu_{ii}^2 f_{ij}$ (i, j = 1...N) of the matrix product determines the amplitude of the Gaussian band centered at the energy of species i appearing in the spectrum associated to the species j. Matrices $L_1 \in \mathbf{R}^{n_1 \mathrm{x} N}$ and $L_3 \in \mathbf{R}^{n_3 \mathrm{x} N}$ are convolution matrices defining the appearance of individual transitions as 2D Gaussian bands in the actual 2D spectrum containing n_1 and n_3 samples along excitation an detection axis, respectively.

Note that this model cannot properly describe the ESA signal appearing outside the grid of considered states or substantially Stokes-shifted SE. The linewidth of the individual bands is given by the initially guessed homogeneous width, convolved with lifetime broadening,

determined from the rate constants. The unknown amplitudes of individual Gaussian bands of every spectral species are chosen in such a way that the resulting calculated 2D spectrum approximates the experimental data the best. Since the band intensities are linear parameters of the model their best values are uniquely determined from the experimental data by the linear least-square fitting method. Note that a perfect match between the experimental and calculated spectra would still require correct guess of the non-linear parameters (rates, energies, dipole moments and homogeneous linewidths).

The linear absorption spectrum is calculated from the energy levels, dipole moments and linewidths. Fit to the experimental absorption are used as an additional criterion determining the goodness of the fit.

In the following steps all the non-linear parameters are systematically scanned in order to fit the experimental data as well as possible. This is done by a combination of a variant of the particle swarm optimization[11] with Levenberg-Marquardt algorithm. Note that the separation of parameters into linear and non-linear, where the former are determined with respect to the latter is the principle of the variable projections method.[12]

The fitting algorithm was implemented in C++ using Armadillo linear algebra library. The detailed description of the model together with its mathematical properties will be summarized in a separate forthcoming publication.

References

(1) Wahlund, T. M.; Woese, C. R.; Castenholz, R. W.; Madigan, M. T. A Thermophilic Green Sulfur Bacterium from New-Zealand Hot-Springs, Chlorobium-Tepidum Sp-Nov. *Arch. Microbiol.* **1991,** *156*, 81-90.

(2) Wen, J. Z.; Zhang, H.; Gross, M. L.; Blankenship, R. E. Membrane Orientation of the Fmo Antenna Protein from Chlorobaculum Tepidum as Determined by Mass Spectrometry-Based Footprinting. *Proc. Natl. Acad. Sci. U.S.A.* **2009,** *106*, 6134-6139.

(3) Albrecht, A. C. Polarizations and Assignments of Transitions - Method of Photoselection. *J. Mol. Spectrosc.* **1961,** *6*, 84-108.

(4) Zilian, A.; Wright, J. C. Polarization Effects in Four-Wave Mixing of Isotropic Samples. *Mol. Phys.* **1996,** *87*, 1261-1271.

(5) Hochstrasser, R. M. Two-Dimensional Ir-Spectroscopy: Polarization Anisotropy Effects. *Chem. Phys.* **2001,** *266*, 273-284.

(6) Zanni, M. T.; Ge, N. H.; Kim, Y. S.; Hochstrasser, R. M. Two-Dimensional Ir Spectroscopy Can Be Designed to Eliminate the Diagonal Peaks and Expose Only the Crosspeaks Needed for Structure Determination. *Proc. Natl. Acad. Sci. U.S.A.* **2001,** *98*, 11265-11270.

(7) Abramavicius, D.; Voronine, D. V.; Mukamel, S. Unravelling Coherent Dynamics and Energy Dissipation in Photosynthetic Complexes by 2d Spectroscopy. *Biophys. J.* **2008,** *94*, 3613-3619.

(8) Tiwari, V.; Peters, W. K.; Jonas, D. M. Electronic Resonance with Anticorrelated Pigment Vibrations Drives Photosynthetic Energy Transfer Outside the Adiabatic Framework. *Proc. Natl. Acad. Sci. U.S.A.* **2013,** *110*, 1203-1208.

(9) Vulto, S. I. E.; de Baat, M. A.; Louwe, R. J. W.; Permentier, H. P.; Neef, T.; Miller, M.; van Amerongen, H.; Aartsma, T. J. Exciton Simulations of Optical Spectra of the Fmo Complex from the Green Sulfur Bacterium Chlorobium Tepidum at 6 K. *J. Phys. Chem. B* **1998,** *102*, 9577-9582.

(10) Dioumaev, A. K. Evaluation of Intrinsic Chemical Kinetics and Transient Product Spectra from Time-Resolved Spectroscopic Data. *Biophys. Chem.* **1997,** *67*, 1-25.

(11) Kennedy, J.; Eberhart, R. Particle Swarm Optimization. *Proc. IEEE Int. Conf. Neur. Net.* **1995**, 1942-1948.

(12) Golub, G.; Pereyra, V. Separable Nonlinear Least Squares: The Variable Projection Method and Its Applications. *Inverse Probl.* **2003,** *19*, R1-R26.