

# How far can a single hydrogen bond tune the spectral properties of the GFP chromophore?

## Supporting Information

Hjalte V. Kiefer,<sup>†</sup> Elie Lattouf,<sup>†</sup> Natascha W. Persen,<sup>†</sup> Anastasia V. Bochenkova,<sup>†,‡</sup> and Lars H. Andersen<sup>\*,†</sup>

*Department of Physics and Astronomy, Aarhus University, DK-8000 Aarhus C, Denmark, and Department of Chemistry, M.V. Lomonosov Moscow State University, 119991 Moscow, Russia*

E-mail: lha@phys.au.dk

## Experimental Techniques

The experimental setup consists of two major parts, the ions source and ELISA (Fig. 2A of the main text).<sup>1</sup> In the ion source, a solution of methanol and chromophores are sprayed using electrospray ionization. The charged droplets are evaporated in a heated capillary resulting in singly charged ions in the gas phase. From the capillary, the ions are guided in an octupole guide to a 22-pole radio-frequency trap in which they are stored for 50 ms. While stored, the ions thermalize with the helium buffer gas.

---

\*To whom correspondence should be addressed

<sup>†</sup>Department of Physics and Astronomy, Aarhus University, DK-8000 Aarhus C, Denmark

<sup>‡</sup>Department of Chemistry, M.V. Lomonosov Moscow State University, 119991 Moscow, Russia

The ions are extracted from the trap and accelerated from the high voltage platform to 22 kV, mass selected through the dipole magnet and stored in ELISA.

## Calculating the action absorption cross section

A laser pulse is fired after about 5 ms of storage. The measured absorption cross section is then obtained by subtracting the background counts,  $N_0$  not arising from the laser, from the number of counts in our time-gated signal window,  $N$  (colored area in Figs. 2 C and D in the main text). The signal is normalized to the number of photons and the number of ions in the ion bunch. The number of photons are proportional to the laser pulse energy times the wavelength  $N_{\text{photons}} \propto E_{\text{pulse}} \times \lambda$ , and the number of ions in the bunch is proportional to the number of background counts before firing the laser  $N_{\text{ion bunch}} \propto N_{\text{Bg}}$ . We arrive at the formula:

$$\sigma_{\text{Photoabsorption}} \propto \frac{N - N_0}{E_{\text{pulse}} \lambda N_{\text{Bg}}}. \quad (1)$$

Saturation effects might occur, in the case of detector saturation or depletion of the ion bunch. This is included as a correction to the neutral yield per photon (see also Fig. 2B in the main text).

$$\sigma_{\text{Photoabsorption}}^{\text{Saturation corrected}} \propto \frac{N - N_0}{E_{\text{saturation}} \cdot [1 - \exp(-\frac{E_{\text{pulse}}}{E_{\text{saturation}}})] \lambda N_{\text{Bg}}}. \quad (2)$$

In our measurements, the laser pulse saturation energy at 455 nm is found to be 60  $\mu\text{J}$ . In the experiment we stabilize the energy at 15  $\mu\text{J}$ , well below the saturation energy.

# Computational Methods

## Electronic structure calculations

The ground-state geometry optimization is carried out at the MP2/(aug)-cc-pVTZ level of theory, with the basis set augmented by a diffuse spdf shell on all oxygen atoms. Vibrational analysis is performed to confirm that the stationary points found are true minima. The vertical excitation energies are calculated at the extended multi-configuration quasi-degenerate perturbation theory (XMCQDPT2) level within the same basis set. The reference wave functions of the ground state and the target excited states are constructed within the complete active space self-consistent field (CASSCF) method, with the active spaces (16,14) comprising 16 electrons distributed over 14 orbitals. They include  $\pi$ -type valence orbitals of either a single conjugated system or both the neutral and the anionic parts to enable their simultaneous treatment within a single calculation. The mixed active space is chosen based on the analysis of the occupation numbers of the natural CASSCF orbitals, obtained through the calculations using the entire active spaces of each part treated separately. A state-averaging (SA) procedure is applied. The XMCQDPT2 effective Hamiltonians are constructed in the frame of the reference space spanned by 7 SA(7)-CASSCF zeroth-order wavefunctions. Energies of all semi-canonical orbitals used in the perturbation theory are defined as eigenvalues of corresponding blocks of a DFT-based effective Fock operator, diagonalized in the basis of the computed CASSCF molecular orbitals. The hybrid PBE0 functional is used to construct the DFT-based Fock matrix.

The vertical detachment and ionization energies of the anionic and neutral parts of the dimer are calculated at the canonical XMCQDPT2/SA(7)-CASSCF(14,14) and XMCQDPT2/SA(7)-CASSCF(12,13) levels of theory, using both the pure and mixed active spaces, respectively. For these types of calculations, the basis set is augmented by an additional diffuse p-type function with a particularly small exponent (-10) centered at the middle carbon atom of the anionic part. To mimic an electron detachment (ionization) process, either a single  $\pi^*$ -

orbital or the entire  $(p_x, p_y, p_z)$  diffuse shell is included in the pure and mixed active spaces, respectively. The ground and detached (ionized) states are included in the state-averaging procedure.

An excited-state gradient has been calculated at the XMCQDPT2/SS-CASSCF(14,13)/-(aug)-cc-pVTZ level of theory, using a pure active space of the anionic part of the dimer. The effective Hamiltonian is constructed in the frame of the reference space spanned by two CASCI zeroth-order wavefunctions of the ground and first excited states. The XMCQDPT2 excited-state gradient is calculated numerically as a two-sided derivative.

For all electronic structure calculations, the Firefly Version 8.0,<sup>2</sup> partially based on the GAMESS (US)<sup>3</sup> source code is used.

## Modeling of spectral shapes

The S0-S1 temperature-dependent absorption profiles of the anionic part in the dimer have been obtained by evaluating Franck-Condon overlap integrals within the double harmonic parallel mode approximation, using our own original program. The maximum vibrational excitation energy both in S1 (for transitions from the ground vibrational state in S0) and in S0 (for hot transitions) is set to 10 000  $\text{cm}^{-1}$ , so that the overall width of the simulated spectrum can be as large as 20 000  $\text{cm}^{-1}$ . In total, a maximum of 50 vibrational levels of each mode is considered in the electronic ground state, and the level of excitation for each mode does not exceed 11. These parameters ensure the inclusion of all relevant vibronic excitations, so that the sum of transition probabilities for each mode is always 1, except for the single lowest-frequency mode of , which yields the sum of 0.999 821 8 at 100 K (average excitation number 5) and 0.943 728 8 at 300 K (average excitation number 16). The final spectral shapes are obtained by convoluting the calculated stick spectra with Gaussian functions of particular widths.

The displacements between the ground-state and excited-state minima along each normal mode, i.e. the origin shifts, are estimated using the quadratic approximation of the excited

state potential energy surface at a Franck-Condon point. The displacements are then found based on the calculated excited-state gradient at the equilibrium geometry of the ground state. Since the ground-state optimization is carried out at the MP2 level of theory and the excited-state calculations employ the XMCQDPT2 theory, the final excited-state gradient is corrected by subtracting from it the ground-state gradient calculated at the same level of theory. The same procedure has also been used by us for calculating the S0-S1 spectral shape of the anionic chromophore inside the protein within the combined XMCQDPT2/EFP approach.<sup>4</sup>

To account for the broad experimental absorption profile of the anionic part in the dimer, which is also inhomogeneously broadened by the ground-state proton dynamics, we use two theoretical models. We first note that the most red-shifted absorption should correspond to the S0-S1 transition in the anionic part of the dimer at the equilibrium geometry configuration, where the proton is located at one or the other monomer in one of the two minima of the symmetric double-well potential. This has to be compared with the blue-shifted absorption of both formally deprotonated chromophores at the geometry configuration of the transition state, where the proton is located perfectly in the middle between the two monomers. Since we are interested in locating the most red-shifted 0-0 transition, we only consider the equilibrium geometry configuration of one of the two equivalent minima. Subsequently, we also aim at distinguishing between the Franck-Condon active modes that are predominantly localized at the anion in the dimer from those that are excited in the entire coupled system following electronic excitation localized in the anion. This partitioning allows us to compare the absorption profile that is calculated using the entire dimer system with the one originated from only the anionic part in the dimer. In other words, we are able to visualize the effect of a strong coupling between the two monomers. Such partitioning is also useful in aligning the absorption spectrum of the anion in the dimer with that of the GFP protein containing the anionic chromophore (see Fig. 3 in the main text).

The dimer system is partitioned in two parts, the anionic and the neutral, and the two

kinds of calculations are carried out using both the total and the partial Hessian matrix evaluation (the matrix of second derivatives of the total potential energy with respect to nuclear coordinates of the entire system and of only the anion in the dimer), followed by entire and partial vibrational analysis, respectively. The vibrational analysis is carried out in the ground electronic state, assuming no change in frequencies upon electronic excitation. The evaluated Hessian matrices (the entire 165x165 and the partial 81x81 matrices) are projected to eliminate rotational and translational contaminants and are diagonalized. Their all positive eigenvalues ensure that the located stationary point is a minimum. In case of the entire dimer, the two lowest-frequency modes of 1 and 11  $\text{cm}^{-1}$  (the wagging and scissoring bend modes of two monomers), highly delocalized over the entire dimer, are eliminated from the subsequent calculations of spectral shapes due to their highly anharmonic nature, where the harmonic estimation of origin shifts fails severely, greatly overestimating the corresponding shifts. A similar problem with another low-frequency mode of 80  $\text{cm}^{-1}$  that corresponds to the bending mode of the anion, which is also coupled with the neutral part through a hydrogen bond, has been solved by using the origin shift calculated through the partial analysis, where the corresponding bending mode is strictly localized on the anionic part. All other modes have been treated as usual. Since the total width of the spectral shape is defined by high-frequency modes, and low-frequency modes only provide a spectral blurring, the neglect of the most highly anharmonic low-frequency modes does not alter the width and the main features of the calculated spectrum. The additional broadening is effectively included by convoluting the calculated stick spectrum with Gaussian functions. Within the partial vibrational analysis, all localized modes on the anion are treated explicitly, where the harmonic approximation is valid for all Franck-Condon active modes.

In the case of the GFP protein, the details of the S0-S1 spectral shape analysis are described in Ref. 4, where the same XMCQDPT2 level of theory is used to calculate the excited-state gradient within the combined electronic embedding QM/MM scheme. The quantum mechanical part and the chromophore itself have been treated using extended

models of a total of 154 and 70 atoms, respectively.

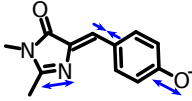
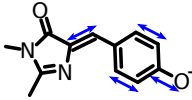
## **Supporting figures and tables**

Is included on the following pages

Table S1: Vertical excitation and detachment energies of the dimer calculated using mixed and pure active spaces within the XMCQDPT2/CASSCF method.

	Anionic Part	Neutral Part	Charge Transfer
Vertical Excitation Energies			
Mixed active space XMCQDPT2[7]/SA(7)- CASSCF(16,14)/(aug)-cc- pVTZ	456 nm (2.72 eV)	408 nm (3.04 eV)	351 nm (3.53 eV)
Pure active space XMCQDPT2[7]/SA(2)- CASSCF(16,14) /(aug)-cc-pVTZ	466 nm (2.66 eV)	420 nm (2.95 eV)	–
Vertical detachment energies and Ionization energies			
Mixed active space with a continuum p-type shell XMCQDPT2[7]/SA(7)- CASSCF(12,13)/(aug)-cc- pVTZ+ extremely diffuse p-type shell	367 nm (3.38 eV)	289 nm (4.29 eV)	–
Pure active space with a single $\pi^*$ -type continuum basis function XMCQDPT2[7]/SA(7)- CASSCF(14,14)/(aug)-cc- pVTZ+ extremely diffuse p-type shell	374 nm (3.32 eV)	283 nm (4.38 eV)	–

Table S2: Vibrational frequencies and Huang-Rhys factors (i.e., a half of a squared dimensionless origin shift, shown in the brackets) of the most active BLA modes of the anionic chromophore in the dimer and inside the protein. All Franck-Condon active modes are shown in Fig. S1

	Vibrational Mode	Dimer	Protein
C=N stretch		1579 cm <sup>-1</sup> (0.10)	1614 cm <sup>-1</sup> (0.28)
C=C bridge stretch		1692 cm <sup>-1</sup> (0.07)	1696 cm <sup>-1</sup> (0.02)



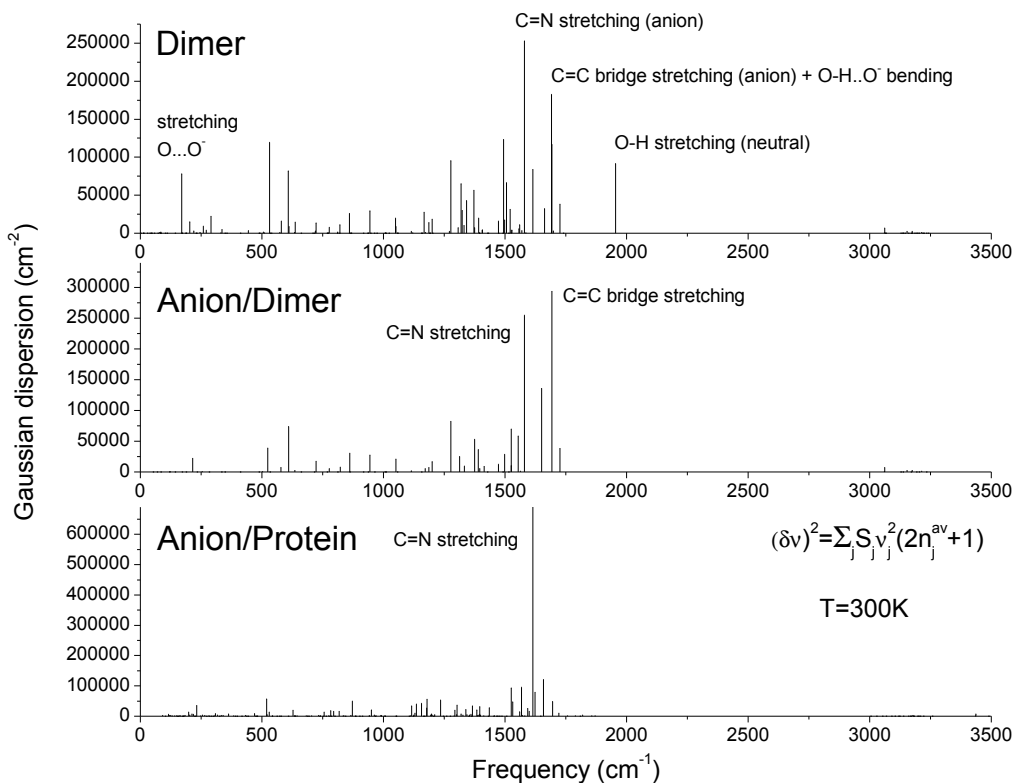


Figure S1: The contributions from each normal mode to the S0-S1 spectral broadening calculated within the two theoretical models of the dimer (upper and middle panels) as compared to those inside the GFP protein (lower panel) at 300 K. The total width of the spectrum can be approximated by a nearly Gaussian function and can be expressed as a sum of the contributions from each normal mode. The total Gaussian variance accounts for temperature, origin shifts, and frequencies. Each contribution is defined as a product of a so-called Huang-Rhys factor (which is a half of a squared dimensionless origin shift along the mode),<sup>5,6</sup> a square of a frequency, and a factor of  $(2n^{\text{av}} + 1)$ , where  $n^{\text{av}}$  is the thermal average occupation number of the mode. Note that high-frequency bond-length alternating (BLA) modes predominantly define the overall width in all cases. In the case of the dimer, a much larger number of the modes that are excited as a response to the electronic excitation localized on the anionic part of the dimer, as compared to the number of Franck-Condon active vibrational modes accounted for in the frame of the reduced model of the anion in the dimer, manifests a strong coupling between the monomers. Note that the O-H stretching mode localized on the neutral part is excited upon the S0-S1 excitation of the anionic part in the dimer. The most intense transition in the dimer refers to the 0-1 excitation with respect to the O-O stretching mode of  $170\text{ cm}^{-1}$  between the two monomers, whereas in all other cases the 0-0 transition has the largest strength. The electron density redistribution upon the S0-S1 excitation in the anionic part of the dimer results in lowering the H-bond interaction strength, and as a result the O-H and O-O stretching modes are excited as explicitly disclosed by the calculated origin shifts along these modes.

## References

- (1) Andersen, J. U.; Hvelplund, P.; Nielsen, S. B.; Tomita, S.; Wahlgreen, H.; Møller, S. P.; Pedersen, U. V.; Forster, J. S.; Jørgensen, T. J. D. *Rev. Sci. Instrum.* **2002**, *73*, 1284–1287.
- (2) Granovsky, A. A. Firefly version 8. <http://classic.chem.msu.su/gran/firefly/index.html>.
- (3) Schmidt, M. W.; Baldrige, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347–1363.
- (4) Bochenkova, A. V.; Andersen, L. H. *Faraday Discuss.* **2013**, *163*, 297–319.
- (5) Lax, M. *The Journal of Chemical Physics* **1952**, *20*, 1752–1760.
- (6) Yurenev, P. V.; Kretov, M. K.; Scherbinin, A. V.; Stepanov, N. F. *The Journal of Physical Chemistry A* **2010**, *114*, 12804–12812, PMID: 21080718.