A Combined Quantum Mechanics/Molecular Mechanics Study of the One- and Two-Photon Absorption in the Green Fluorescent Protein

Arnfinn Hykkerud Steindal,[a] Jógvan Magnus Haugaard Olsen,[b] Kenneth Ruud,[a] Luca Frediani,*[a] and Jacob Kongsted[b]

Received Xth XXXXXXXXXX 20XX, Accepted Xth XXXXXXXXX 20XX

First published on the web Xth XXXXXXXXXX 20XX

DOI: 10.1039/b000000x

We present for the first time a QM/MM study of the one- and two-photon absorption spectra of the GFP chromophore embedded in the full protein environment described by an advanced quantum mechanically derived polarizable force field. The calculations are performed on a crystal structure of the green fluorescent protein (GFP) using the polarizable embedding density functional theory (PE-DFT) scheme. The importance of treating the protein environment explicitly with a polarizable force field and higher-order multipoles is demonstrated, as well as the importance of including water molecules close to the chromophore in the protein barrel. For the most advanced description we achieve good agreement with experimental findings, with a peak at 405 nm for the neutral and a peak at 475 nm for the anionic form of the GFP chromophore. The presence of a dark OPA state, as suggested by other studies to explain the discrepancies between OPA and TPA spectra, is not supported by our calculations.

1 Introduction

The green fluorescent protein (GFP)[1], originally isolated from the jellyfish *Aequorea victoria*, is widely used as a biological marker. A fluorescent chromophore located inside the protective envelope of the barrel-shaped protein yields a characteristic green light at around 505 nm, depending on the excitation wavelength[2]. GFP owes its success to several factors: (i) the chromophore is generated within the protein without the need of co-factors, (ii) GFP does not seem to affect the function of the protein it is attached to and (iii) it can be expressed genetically, which opens the way for *in vivo* experiments. Despite the tremendous success of GFP as a biological marker, the underlying spectroscopic properties of the chromophore remain a challenge for experimentalists and theoreticians alike. One particular challenge relates to the understanding and modeling of the effect of the protein environment, as this seems to distinctively affect the properties of the chromophore: fluorescence is quenched in traditional solvents[3], and no significant frequency shift with respect to gas-phase measurement is observed[4,5].

Recently, some of these questions have – at least partially

[a] *Centre of Theoretical and Computational Chemistry, Department of Chemistry, University of Tromsø, N-9037 Tromsø, Norway. Fax: +47 77644765; Tel: +47 77644082; E-mail: luca.frediani@uit.no*

[b] *Department of Physics, Chemistry and Pharmacy, University of Southern Denmark, DK-5230 Odense M, Denmark.*

– been answered. It is for example generally accepted that the reason for the fluorescence of the chromophore in the protein environment is due to the hindrance of conformational rearrangements in the excited state due to the protein scaffolding, which in solvent is responsible for faster fluorescence timescales and quenching through non-radiative processes[6–9].

Among the least studied properties of the GFP chromophore are the multi-photon absorption properties, including two-photon absorption (TPA). There are several properties that make TPA an interesting tool for biological investigations: (i) the use of a less biologically invasive light source (IR or at the most visible instead of ultraviolet light) (ii) the quadratic dependence on the light intensity, which makes it possible to achieve three-dimensional imaging with high resolution; (iii) different selection rules with respect to one-photon absorption (OPA), leading to complementary spectra.

The electronic absorption spectrum of GFP is characterized by two main absorption peaks located at 398 nm and 475 nm. The former is attributed to the neutral form of the chromophore whereas the latter is attributed to its anionic form. It is commonly accepted that the barrel-shaped protein is responsible for providing a unique environment which ultimately leads to the peculiar fluorescence of GFP by e.g. limiting quenching of the excited-state chromophore. Recent investigations have also shown that the neutral form of the chromophore displays very limited solvatochromism in a wide range of solvents, whereas the anionic form is much more sensitive to environmental effects[10].

Several investigations of the TPA spectrum have been reported in the literature. To the best of our knowledge the first measurement of TPA for GFP has been reported by Xu and coworkers[11] in 1996 showing a spectrum similar to OPA and a maximum cross section of 8GM around 800nm. Later, Xia and coauthors[12] measured a TPA cross section of more than 100 GM units with a Two-Photon Induced Fluorescence (TPIF) technique. Rather than providing a full TPA characterization, this method was employed to assess the evolution of the GFP chromophore subsequent to absorption[13]. With the same technique Heikal et al.[14] obtained a maximum cross section of 60 GM at 970 nm. Similar results on other flourescent proteins were obtained by Blab et al.[15]. A full TPA characterization has recently been reported by Hosoi et al.[16] where much smaller cross sections were observed (2-3 GM) and for the bare chromophore (as well as for the eGFP mutant[1,17]) a shift in the TPA spectrum was attributed to the excitation to a different (hidden to OPA) excited state. Lately the OPA and TPA spectra of wild-type GFP and some common mutants have been revised by Drobizhev et al.[18]. Despite large discrepancies in the reported absolute cross section values all investigations seem to agree on the main spectral features with two absorption bands located at twice the wavelength of the $S_0 \rightarrow S_1$ transition for the neutral (800nm) and deprotonated (950nm) form.

Theoretical efforts are in this respect crucial in order to fully understand the TPA properties of the GFP chromophore as an isolated system, in solution but most importantly in its native environment. The TPA of a model chromophore has been investigated theoretically by Nifosì and Luo together with several other fluorescent protein (FP) chromophores[19], predicting the existence of strongly TP absorbing states in regions of the spectrum not yet experimentally investigated. Recent theoretical findings have proposed that the TPA shift with respect to the OPA spectrum can be explained with a breakdown of the Franck-Condon (FC) approximation rather than a "dark" OPA state[20]. A very recent theoretical study[21] which made use of both DFT and correlated wavefunction methods

for the chromophore, combined with non polarizable force fields for the protein has investigated the main spectral features of GFP.

In the present work, we present a combined quantum mechanics and molecular mechanics (QM/MM) study using the polarizable embedding density functional theory (PE-DFT) method[22], which is fully parallelized[23] and includes a self-consistent treatment of the environmental polarization at the ground- and excited-state level, in order to faithfully model the TPA properties of the GFP chromophore in its native environment. The results are compared with previous theoretical and experimental investigations and with gas-phase calculations. We present the theoretical framework of the method used in Sec. 2, and computational details are summarized in Sec. 3. Our findings are presented and discussed in Sec. 4. Concluding remarks will be drawn in the final section.

2 Theory

Multilevel methods, such as QM/MM, provide an effective way of dealing with large structured systems like GFP where only a small portion of the total system, the *chromophore*, needs to be described using *ab initio* methods. However, the remainder, i.e. the protein environment surrounding the chromophore, is important through its interactions with the chromophore. Here we employ the polarizable embedding (PE) scheme described by Olsen et al.[22,24] coupled with a DFT description of the quantum system, the PE-DFT method. The PE approach models the effects of an environment on a central core subsystem by including these effects directly in the density/wavefunction of the core system. This is achieved by combining classical electrostatics with quantum mechanics through effective operators. The PE method uses an atomistic description of the protein environment in order to accurately describe the embedding potential. Each atomic site is assigned a quantum-mechanically derived multipole moment expansion to model the electrostatic embedding potential and an anisotropic dipole-dipole polarizability tensor to account for many-body induction effects, thus allowing mutual polarization between the chromophore and the environment. The ground-state electronic density of the chromophore is optimized while taking into account the explicit electrostatic interactions and polarization effects from the protein environment in a fully self-consistent manner. Similarly, the excited states are solved self-consistently, taking into account the effects from the environment in the response formalism. Below we present a more detailed description of the PE-DFT method.

2.1 Ground-State Polarizable Embedding

Within the PE formalism the ground-state density is optimized using an effective Kohn-Sham (KS) operator given by

$$\hat{f}_{\text{eff}} = \hat{f}_{\text{KS}} + \hat{v}_{\text{PE}}, \tag{1}$$

where $\hat{f}_{\text{KS}}$ is the usual vacuum KS operator and $\hat{v}_{\text{PE}}$ is the polarizable embedding operator which contains the interactions with the embedding potential. The PE operator consists of two contributions

$$\hat{v}_{\text{PE}} = \hat{v}^{\text{es}}_{\text{PE}} + \hat{v}^{\text{ind}}_{\text{PE}}, \tag{2}$$

where $\hat{v}^{\text{es}}_{\text{PE}}$ provides the electrostatic potential from the permanent charge distribution of the environment and $\hat{v}^{\text{ind}}_{\text{PE}}$ describes many-body induction effects, i.e. it allows a mutual polarization between the quantum mechanically treated core and the environment. Within a second-quantization formalism[25] the electrostatic part of the PE operator is defined as

$$\hat{v}^{\text{es}}_{\text{PE}} = \sum_{s=1}^{S} \sum_{k=0}^{K} \frac{(-1)^{(k+1)}}{k!} \mathbf{Q}_s^{(k)} \sum_{pq} \mathbf{T}^{(k)}_{s,pq} \hat{E}_{pq}, \tag{3}$$

and it describes the interactions between the electrons in the chromophore and the permanent multipole moments in the environment. In the above equation S and K are the number of interaction sites in the environment and the level of the (local) multipole expansion at these points, respectively. Furthermore, $\hat{E}_{pq}$ is a one-electron excitation operator[25] where the sum pq is over the molecular orbitals. The $\mathbf{Q}_s^{(k)}$ factors are kth order multipole moments assigned to the sth site in the environment; e.g. $\mathbf{Q}_s^{(0)} = q_s$ (charge), $\mathbf{Q}_s^{(1)} = \mu_s$ (dipole), $\mathbf{Q}_s^{(2)} = \Theta_s$ (quadrupole) and so on. The $\mathbf{T}_{s,pq}^{(k)}$ factors are integrals over the kth order interaction tensors, defined as

$$\mathbf{T}_s^{(k)} = \nabla_s^{(k)} \frac{1}{|\mathbf{r}_s - \mathbf{r}|}, \tag{4}$$

where

$$\nabla^{(k)} = \left(\frac{\partial}{\partial r_x}\right)^{k_x} \left(\frac{\partial}{\partial r_y}\right)^{k_y} \left(\frac{\partial}{\partial r_z}\right)^{k_z}, \tag{5}$$

and where x, y, z indicates the cartesian directions. The polarization part of the PE operator is given by

$$\hat{v}_{\mathrm{PE}}^{\mathrm{ind}} = -\sum_{s=1}^{S} \mu_s^{\mathrm{ind}} \sum_{pq} \mathbf{T}_{s,pq}^{(1)} \hat{E}_{pq}, \tag{6}$$

where the induced dipoles μ_s^{ind} are obtained as the classical linear response due to the electric fields from all other sources

$$\mu_s^{\mathrm{ind}} = \alpha_s \mathbf{F}_{\mathrm{tot}}(\mathbf{r}_s) = \alpha_s \left(\mathbf{F}_{\mathrm{el}}(\mathbf{r}_s) + \mathbf{F}_{\mathrm{nuc}}(\mathbf{r}_s) + \mathbf{F}_{\mathrm{mul}}(\mathbf{r}_s) + \mathbf{F}_{\mathrm{ind}}(\mathbf{r}_s)\right), \tag{7}$$

where $\mathbf{F}_{\mathrm{tot}}(\mathbf{r}_s)$ is the total electric field at site s, i.e. the field from the nuclei ($\mathbf{F}_{\mathrm{nuc}}(\mathbf{r}_s)$) and electrons ($\mathbf{F}_{\mathrm{el}}(\mathbf{r}_s)$) of the chromophore and the permanent ($\mathbf{F}_{\mathrm{mul}}(\mathbf{r}_s)$) and induced ($\mathbf{F}_{\mathrm{ind}}(\mathbf{r}_s)$) multipole moments in the protein environment. An induced dipole moment depends on the field from all the other induced dipole moments and they are therefore obtained either iteratively in a self-consistent manner or directly through a matrix equation formulation. Furthermore, since the induced dipole moments depend on the electric field from the electrons in the chromophore, and therefore also the electron density, the PE operator is updated after each SCF iteration, thus leading to a fully self-consistent treatment of the polarization.

2.2 Response Functions in Polarizable Embedding

For a detailed description of linear and quadratic response theory in the DFT formalism we refer to the work by Sałek et al.[26,27] Furthermore, the PE-DFT response theory up to quadratic response fucntions has recently been presented by Olsen et al.[22]. Here, we will only present a brief summary.

The response functions are defined through a time-dependent perturbation expansion of the expectation value of a time-independent operator $\hat{A}$ given as

$$\langle t|\hat{A}|t\rangle = \langle t|\hat{A}|t\rangle^{(0)} + \langle t|\hat{A}|t\rangle^{(1)} + \langle t|\hat{A}|t\rangle^{(2)} + \dots, \tag{8}$$

where we use $|t\rangle$ to represent the time-dependent state. The first term on the right-hand side is the time-independent expectation value, i.e. $\langle t|\hat{A}|t\rangle^{(0)} = \langle 0|\hat{A}|0\rangle$, where $|0\rangle$ is the time-independent reference state, and the second and third terms describe the linear and quadratic response to the perturbation, respectively. The Fourier representation of the linear and quadratic response is given by

$$\langle t|\hat{A}|t\rangle^{(1)} = \int \langle\langle \hat{A}; \hat{V}^{\omega} \rangle\rangle_{\omega} \exp\big(-i\omega t\big) \mathrm{d}\omega, \tag{9}$$

$$\langle t|\hat{A}|t\rangle^{(2)} = \frac{1}{2} \iint \langle\langle \hat{A}; \hat{V}^{\omega_1}, \hat{V}^{\omega_2} \rangle\rangle_{\omega_1,\omega_2} \exp\big(-i(\omega_1+\omega_2)t\big) \mathrm{d}\omega_1 \mathrm{d}\omega_2, \tag{10}$$

where $\langle\langle \hat{A}; \hat{V}^{\omega} \rangle\rangle_{\omega}$ and $\langle\langle \hat{A}; \hat{V}^{\omega_1}, \hat{V}^{\omega_2} \rangle\rangle_{\omega_1,\omega_2}$ are the linear and quadratic response functions in frequency domain, respectively, and $\hat{V}$ is the perturbation operator.

In a matrix equation formulation, the linear response function can be defined as

$$\langle\langle \hat{A}; \hat{B} \rangle\rangle_{\omega} = \langle 0|[\hat{\kappa}^{\omega}, \hat{A}]|0\rangle = -\mathbf{A}^{\dagger} \boldsymbol{\kappa}^{\omega} = -\mathbf{A}^{\dagger} \left(\mathbf{E} - \omega \mathbf{S}\right)^{-1} \mathbf{B}. \tag{11}$$

Here we have expressed the time-dependent parameters according to $\hat{\kappa}^{\omega} = \sum_{pq} \kappa^{\omega}_{pq} \hat{E}_{pq} = \hat{\mathbf{q}}^{\dagger} \kappa^{\omega}$. Furthermore we have introduced $\mathbf{A} = \langle 0|[\hat{\mathbf{q}}, \hat{A}]|0\rangle$ and $\mathbf{B} = \langle 0|[\hat{\mathbf{q}}, \hat{B}]|0\rangle$ as well as the definition of the generalized overlap matrix $\mathbf{S} = \langle 0|[\hat{\mathbf{q}}, \hat{\mathbf{q}}^{\dagger}]|0\rangle$. The response vector κ^{ω} is determined by solving the linear response equation given by

$$(\mathbf{E} - \omega \mathbf{S})\, \kappa^{\omega} = \mathbf{B} \,. \tag{12}$$

Specific contributions to the environment only affects the electronic Hessian (the $\mathbf{E}$ matrix). The contributions due to the PE operator (Eq. (2)) appear in the $\mathbf{E}$ matrix as

$$\mathbf{E}_{\mathrm{PE}} \kappa^{\omega} = -\langle 0|[\hat{\mathbf{q}}, \hat{Q}^{\omega}_1 + \hat{Q}^{\omega}_2|0\rangle \,, \tag{13}$$

where $\hat{Q}^{\omega}_1$, which describes the response from a static environment, is defined as

$$\hat{Q}^{\omega}_1 = [\hat{\kappa}^{\omega}, \hat{v}^0_{\mathrm{PE}}] = \hat{v}^0_{\mathrm{PE}}(\kappa^{\omega}) \tag{14}$$

and $\hat{Q}^{\omega}_2$, which provides the dynamical response from the polarizable environment, is given by

$$\hat{Q}^{\omega}_2 = \hat{v}^{\omega}_{\mathrm{PE}} = \sum_{s=1}^{S} \mu^{\mathrm{ind}}_s(\tilde{\mathbf{F}}^{\omega}) \hat{\mathbf{T}}^{(1)}_s \,. \tag{15}$$

In $\hat{Q}^{\omega}_2$, the induced dipoles are recalculated according to the transformed electric field $\tilde{\mathbf{F}}^{\omega}$ defined by

$$\tilde{\mathbf{F}}^{\omega} = \langle 0|[\hat{\kappa}^{\omega}, \hat{\mathbf{T}}^{(1)}_s]|0\rangle = \langle 0|\hat{\mathbf{T}}^{(1)}_s(\kappa^{\omega})|0\rangle \,, \tag{16}$$

where $\hat{\mathbf{T}}^{(1)}_s = \sum_{pq} \mathbf{T}^{(1)}_{s,pq} \hat{E}_{pq}$.

The quadratic response function can be written as

$$\langle\langle \hat{A}; \hat{B}, \hat{C}\rangle\rangle_{\omega_1,\omega_2} = \kappa^{\mathbf{A}^{\dagger}} \mathbf{V}^{\omega_1,\omega_2} + \hat{P}_{12} \langle 0|[\hat{\kappa}^{\omega_1}, [\hat{\kappa}^{\omega_2}, \hat{A}]]|0\rangle \,, \tag{17}$$

where $\hat{P}_{12}$ is a permutation operator and the three response vectors are determined by solving three linear response equations

$$\kappa^{\mathbf{A}^{\dagger}} (\mathbf{E} - (\omega_1 + \omega_2)\mathbf{S}) = \mathbf{A}^{\dagger}, \tag{18}$$

$$(\mathbf{E} - \omega_1 \mathbf{S})\, \kappa^{\omega_1} = \mathbf{B}, \tag{19}$$

$$(\mathbf{E} - \omega_2 \mathbf{S})\, \kappa^{\omega_2} = \mathbf{C} \,. \tag{20}$$

The contributions to the quadratic response function from the polarizable environment that enter in the $\mathbf{E}$ matrices are analogous to those defined in Eq. (13)-(16). However, there are also contributions that enter the $\mathbf{V}^{\omega_1,\omega_2}$ vector (eq. 17)

$$\mathbf{V}^{\omega_1,\omega_2}_{\mathrm{PE}} = \hat{P}_{12} \left(\langle 0|[\hat{\mathbf{q}}, \hat{Q}^{\omega_1,\omega_2}_3 + \hat{Q}^{\omega_1,\omega_2}_4 + \hat{Q}^{\omega_1,\omega_2}_5]|0\rangle \right) \,. \tag{21}$$

These contributions are defined by

$$\hat{Q}^{\omega_1,\omega_2}_3 = [\hat{\kappa}^{\omega_1}, [\hat{\kappa}^{\omega_2}, \hat{v}^0_{\mathrm{PE}}]] = \hat{v}^0_{\mathrm{PE}}(\kappa^{\omega_2}, \kappa^{\omega_1}) \,, \tag{22}$$

$$\begin{aligned} \hat{Q}^{\omega_1,\omega_2}_4 &= 2[\hat{\kappa}^{\omega_1}, \hat{v}^{\omega_2}_{\mathrm{PE}}] \\ &= -2 \sum_{s=1}^{S} \mu^{\mathrm{ind}}_s(\tilde{\mathbf{F}}^{\omega_2})[\hat{\kappa}^{\omega_1}, \hat{\mathbf{T}}^{(1)}_s] \\ &= -2 \sum_{s=1}^{S} \mu^{\mathrm{ind}}_s(\tilde{\mathbf{F}}^{\omega_2}) \hat{\mathbf{T}}^{(1)}_s(\kappa^{\omega_1}) \,, \end{aligned} \tag{23}$$

$$\hat{Q}^{\omega_1,\omega_2}_5 = \hat{\tilde{v}}^{\omega_1,\omega_2}_{\mathrm{PE}} = -\sum_{s=1}^{S} \mu^{\mathrm{ind}}_s(\tilde{\mathbf{F}}^{\omega_1,\omega_2}) \hat{\mathbf{T}}^{(1)}_s \,, \tag{24}$$

where the transformed electric field in $\hat{Q}^{\omega_1,\omega_2}_4$ is defined as in Eq. (16) and the field in $\hat{Q}^{\omega_1,\omega_2}_5$ is defined as

$$\tilde{\mathbf{F}}^{\omega_1,\omega_2} = \langle 0|[\hat{\kappa}^{\omega_1}, [\hat{\kappa}^{\omega_2}, \hat{\mathbf{T}}^{(1)}_s]]|0\rangle = \hat{\mathbf{T}}^{(1)}_s(\kappa^{\omega_2}, \kappa^{\omega_1}) \,. \tag{25}$$

As in the case of the linear response function, there are contributions that give the response from a static environment, i.e. $\hat{Q}^{\omega_1,\omega_2}_1$ and $\hat{Q}^{\omega_1,\omega_2}_3$, while all other contributions are due to the dynamical response from the environment, i.e. $\hat{Q}^{\omega_1,\omega_2}_2$, $\hat{Q}^{\omega_1,\omega_2}_4$ and $\hat{Q}^{\omega_1,\omega_2}_5$.

The linear response function has poles where the absolute frequency $|\omega|$ is equal to an excitation energy ω_f. These can be determined by solving a generalized eigenvalue problem given by

$$(\mathbf{E} - \omega_k \mathbf{S})\, \mathbf{X}_k = 0 \,, \tag{26}$$

where the eigenvector $\mathbf{X}_k$ represents the k'th excited state. The OPA and TPA can now be determined as residues of the response functions. The residue of the linear response function yields

$$\lim_{\omega\to\omega_f}\left(\omega-\omega_f\right)\langle\langle\hat{A};\hat{B}\rangle\rangle_\omega = \mathbf{A}^\dagger\mathbf{X}_f\mathbf{X}_f^\dagger\mathbf{B} = \langle 0|\hat{A}|f\rangle\langle f|\hat{B}|0\rangle\ , \tag{27}$$

which for the choice of $\hat{A}$ and $\hat{B}$ equal to the position operators are related to the (length gauge) oscillator strength and hence to the OPA intensity. Similarly, the first residue of the quadratic response function is related to the TPA, where the central molecular quantity is the two-photon absorption transition amplitude tensor which, assuming that the frequency of the incoming light is equal to half of the excitation energy from the ground to the excited state, is given by

$$S_{\alpha\beta} = \sum_{k>0}\left(\frac{\langle 0|\hat{\mu}_\alpha|k\rangle\langle k|\hat{\mu}_\beta|f\rangle}{\omega_k-\omega_f/2}+\frac{\langle 0|\hat{\mu}_\beta|k\rangle\langle k|\hat{\mu}_\alpha|f\rangle}{\omega_k-\omega_f/2}\right)\ . \tag{28}$$

Here α, β indicates Cartesian directions, $\langle 0|$ is the ground state and $\langle f|$ the final excited state. The operators involved in the above equation are the electric dipole operators. The orientationally averaged microscopic transition probability δ^{TPA} (au) for the absorption of two identical photons is given by

$$\delta_{\mathrm{au}}^{\mathrm{TPA}} = \frac{1}{30}\sum_{\alpha,\beta}\left[S_{\alpha\alpha}S^*_{\beta\beta}F + S_{\alpha\beta}S^*_{\alpha\beta}H + S_{\alpha\beta}S^*_{\beta\alpha}G\right], \tag{29}$$

where $F,G,H=2$ for linearly polarized light and the conversion into the TPA cross section may finally be evaluated according to

$$\delta_{\mathrm{GM}} = \frac{(2\pi)^3\alpha a_0^5}{c}\frac{\omega^2}{\pi\Gamma}\delta_{\mathrm{au}}^{\mathrm{TPA}} \tag{30}$$

where α is the fine structure constant, a_0 is the Bohr radius (in cm), c is the speed of light (in cm/s), ω is the energy of the incoming photons (in au), and $\pi\Gamma$ is a normalization factor due to the Lorentzian-shape broadening of the excited state ($\Gamma = 0.1$ eV=0.0036749326 au).

3 Computational Details

The polarizable embedding density functional theory (PE-DFT) method[22] implemented in a development version of the DALTON quantum chemistry program[28] was used to calculate the one-photon (OPA) and two-photon absorption energies (TPA) in the green fluorescent protein (GFP). The CAM-B3LYP functional[29] (with $\alpha=0.19$, $\beta=0.46$, $\mu=0.33$) and the 6-31+G* basis set were used unless otherwise stated. The CAM-B3LYP functional was chosen for its good performance when calculating excitation energies[30,31]. Besides, it has been shown that long-range corrected functionals, such as CAM-B3LYP, are able to describe the electronic structure of anions more correctly than more commonly used DFT methods[32]. The GFP structure (1EMB) was obtained from the Protein Data Bank[2] without any modifications, except adding and optimizing the positions of hydrogen atoms. The protein structure was prepared using the protein preparation wizard in the Schrödinger suite of softwares[33]. Water molecules beyond 5 Å away from the chromophore were removed so that the closest seven water molecules were included.

A schematic representation of the GFP structure is shown in figure 1. The protein was divided into two regions treated at different levels of theory: the chromophore is treated using DFT while the rest of the protein is modeled using classical electrostatics through an embedding potential. The part of the protein treated at the quantum-mechanical level is shown in figure 2, with R_2 and R_3 being replaced by the large residues. To represent the protein environment we use an advanced polarizable force field derived from quantum-mechanical calculations. Each atom in the classically treated parts of the protein was assigned multipole moments and anisotropic polarizabilities.

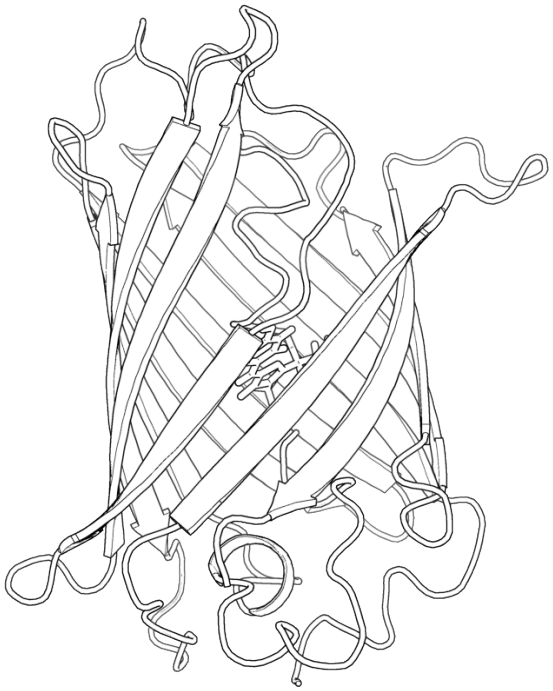

Fig. 1 A schematic representation of the Green Fluorescent Protein. The part of the protein treated at the DFT level of theory is represented by sticks, while the parts shown as cartoons are described classically.

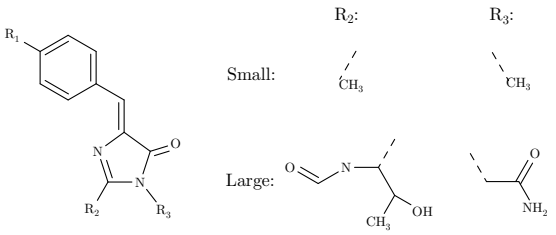

Fig. 2 A schematic representation of the chromophore of the green fluorescent protein. In the neutral chromophore R_1 is a OH group and in the anionic state it is deprotonated.

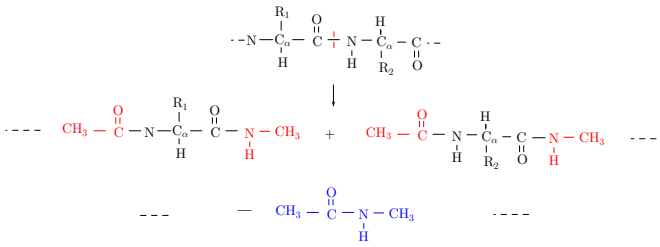

Fig. 3 Graphical representation of the protein fragmentation and capping

The multipoles and polarizabilities were obtained using the LoProp method[34] available in the Molcas program[35], using the conjugate caps (MFCC) procedure to fragment the protein[36]. A description of the fragmentation approach, also applied to the localized properties, has been presented by Söderhjelm and Ryde[37], and we will here only give a brief summary of the approach. The protein was fragmented into single amino acids and capped with -NHCH3 and -COCH3 on each side as shown in figure 3. The multipoles and polarizabilities were then calculated for these individual systems at the B3LYP/6-31+G* level. These parameters were also calculated for the corresponding coupled caps (blue structure in figure 3). In the end, the system was merged together by subtracting the parameters in the coupled conjugated caps (blue) from the capped amino acids.

We used the link-atom approach to treat the bonds between the QM treated chromophore and the remaining classically treated protein, by capping with hydrogen atoms. The C-Cα bond in PHE64 and the Cα-N in VAL68 were cut. In both cases, Cα is located in the classical region. To avoid having multipole moments or polarizabilities too close to the QM system, which could result in unphysical overpolarization, any parameters within a given threshold distance from the QM region are either moved to the closest classical atomic site outside the threshold distance or removed. The threshold was set to 1.4 Å unless stated otherwise. At the 1.4 Å threshold, the parameters on Cα and its two hydrogens were moved to nearby atomic sites. In the case of VAL68, the parameters were moved to the two closest carbon atoms; one in the protein backbone and one in the side chain. The parameters inside the

threshold in the PHE64 residue were moved to the nitrogen atom in the backbone. The smallest distance between the QM region and a classical site (the nitrogen in PHE64) was 1.56 Å after moving the parameters on the closest six atomic sites.

Calculations involving only the chromophore and a single amino acid have also been performed, and in these cases the positions of the hydrogen atoms were optimized for each system, while in the other cases the hydrogen atoms were optimized in the complete protein structure beforehand, using the OPLS-2005 force field[38]

4 Results and Discussion

4.1 Calculations on the isolated chromophore

Table 1 contains the lowest excitation energies and the corresponding TPA cross sections of the GFP chromophore in vacuum calculated using different basis sets. The chromophore structure is obtained from the crystal structure with optimized hydrogen atom positions. The excitation energies are lowered by roughly 0.1 eV in both the neutral and anionic chromophore when both diffuse and polarization functions are added to the parent Pople basis set. Using correlation-consistent basis sets does not change the picture much: with the aug-cc-pVTZ basis set, there is an additional shift relative to 6-31+G* of around 0.03 eV, but this basis set size is 3-4 times larger. Hence, we conclude that the 6-31+G* basis set is a good compromise between computational cost and accuracy.

These calculations were performed with structures as inside the protein. When optimizing the neutral chromophore in vacuum, we get a blue-shift of 0.4 eV compared to the non-optimized structure, leading to an excitation energy of 3.66 eV. In the optimized anionic structure we get an excitation energy of 3.06 eV, thus a blue-shift of $\sim$0.7 eV.

Table 1 The lowest electronic excitation energy (eV) in the GFP chromophore (large structure in Fig. 2) in vacuum calculated using CAM-B3LYP and different basis sets. The OPA oscillator strengths (f) / TPA cross sections (δ_{GM}) are given in parentheses, and nbf stands for number of contracted basis functions in the anionic chromophore.

Basis set	Neutral	Anionic	nbf
6-31G	3.37 (0.192/5.36)	2.58 (0.271/107.69)	259
6-311G	3.35 (0.195/5.36)	2.56 (0.269/126.00)	376
6-31G*	3.33 (0.192/5.36)	2.54 (0.262/127.63)	384
6-31+G	3.31 (0.195/6.20)	2.52 (0.281/184.63)	359
6-31+G*	3.26 (0.194/7.76)	2.49 (0.272/212.61)	484
6-31++G*	3.26 (0.194/7.82)	2.49 (0.271/214.42)	501
6-311+G*	3.25 (0.192/7.34)	2.49 (0.269/216.24)	601
cc-pVDZ	3.30 (0.189/3.86)	2.53 (0.259/133.95)	435
cc-pVTZ	3.25 (0.189/5.63)	2.50 (0.242/163.39)	988
aug-cc-pVDZ	3.21 (0.186/6.64)	2.47 (0.265/219.93)	728
aug-cc-pVTZ	3.22 (0.187/6.96)	2.46 (0.265/218.15)	1541

The OPA, calculated at the CAM-B3LYP/6-31+G* level of theory, for the geometry-optimized small structure shown in Fig. 2, allows us to perform direct comparisons with gas-phase calculations from the literature. Our calculations resulted in excitation energies of 3.75 eV and 3.17 eV for the neutral and anionic chromophore respectively. Filippi et al.[21] calculated OPA at the EOM-CCSD/cc-pVDZ level and reported excitation energies of 4.00 eV and 3.04 eV for the neutral and anionic chromophore, respectively. Thus, our CAM-B3LYP/6-31+G* results are underestimating the excitation energy for the neutral chromophore by 0.25 eV, but overestimating it for the anionic form, by 0.13 eV, compared to a higher level method. A similar trend was also observed for B3LYP/cc-pVDZ level calculations[21]. It is known that coupled-cluster calculations

using smaller basis sets of DZ quality may give too large excitations energies[39,40], bringing the results for the neutral species in better agreement with our CAM-B3LYP results, but with a possible deterioration of the agreement for the anionic species. Still, these differences are within the expected accuracy of the CAM-B3LYP functional[30,41] and the coupled-cluster approach.

4.2 Analysis of Different Link-Atom Approaches

The GFP chromophore is linked to the protein backbone by two peptide bonds: a bond between the carboxyl group on the chromophore and a valine (VAL68), and between the amine group and the phenylalanine amino acid (PHE64).

In order to perform a QM/MM study of such a system we employ a so-called link-atom approach for the two aforementioned bonds which cross through the boundary between the quantum and classical regions, as described in Sec. 3. The interactions between the classical sites and the hydrogen link-atom on the QM side are neglected whenever the site-H link-atom distance is lower than a given threshold. Alternatively they are assigned to the closest classical site situated beyond the predefined threshold. For those two scenarios, interaction removal (IR) and interaction displacement (ID), we have calculated excitation energies by making use of thresholds of 1.3 Å or 1.4 Å. The aforementioned thresholds are valid for the distances between any QM atom and classical site, but in practice, only classical sites close to the hydrogen link-atoms on the QM side are influenced.

The excitation energies in the neutral and anionic chromophore inside the protein barrel with different link-atom approaches are presented in Table 2. With the shortest threshold and, at the same time removing the interaction, the calculations on the anionic chromophore fail due to overpolarization of certain classical sites near the boundary. For the neutral chromophore with the same link-atom approach, there is one transition with HOMO→LUMO character at 2.79 eV and a transition with HOMO→LUMO+1 character at 3.01 eV. For the three other approaches (ID at 1.3 Å and 1.4 Å, and IR at 1.4 Å) we only observe a single transition in the same energy range. Most noticeable is the very high TPA cross section for the lowest-lying transition, which is more than one order of magnitude larger than in the other neutral chromophore results. A possible reason for the relatively large cross section, compared to the cross sections in the neutral chromophore with different link-atom approaches, might be an unphysical

Table 2 The two lowest electronic excitation energies (eV) in green fluorescent protein with different link-atom approaches. The multipoles and polarizabilities on classical sites closer than 1.3 or 1.4 Å to the QM region have either been removed (IR) or displaced (ID) to the closest neighboring site. The QM region was treated at the CAM-B3LYP/6-31+G∗ level while the classical region was modeled using anisotropic polarizabilities and multipoles up to octupoles. The OPA oscillator strengths (f) / TPA cross sections (δ_{GM}) are given in parentheses.

		Distance threshold (Å)	
		1.3	1.4
Neutral	IR	2.79 (0.040/292.02)	2.95 (0.181/9.62)
		3.01 (0.185/19.17)	3.45 (0.000/12.63)
	ID	2.94 (0.184/10.36)	2.93 (0.173/7.80)
		3.45 (0.001/14.76)	3.45 (0.001/15.59)
Anionic	IR	-	2.32 (0.275/465.36)
			2.62 (0.032/37.38)
	ID	2.31 (0.296/575.52)	2.35 (0.290/482.33)
		2.72 (0.019/53.03)	2.72 (0.017/33.83)

description of the system when the classically treated protein is too close to the chromophore. Overall, it is obvious that calculations done with the combination of the short threshold of 1.3 Å and removal of the interactions fail. It is, however, unclear, when inspecting the other link-atom approaches, if it is the short distance between the classical and quantum-mechanical regions or the removal of interactions, or a combination of these, that is the reason for the failure.

The differences between removal and displacement of sites at the 1.4 Å threshold are only minor when it comes to absolute excitation energies, but for the anionic form, when the parameters are removed from the MM sites close to the QM region, there are once again two transitions with contributions from HOMO→LUMO orbitals. At the 1.3 Å threshold, multipoles on Cα and one of its attached hydrogen atoms are removed or moved to the other hydrogen atom attached to Cα. Due to the problems with the combination of an 1.3 Å threshold and removal of interactions, the subsequent calculations have been performed with a threshold of 1.4 Å and displaced interactions.

4.3 Force Field Analysis

Results from one- and two-photon absorption calculations performed with different force fields are presented in Table 3. These calculations include permanent multipoles up to either (i) charges, (ii) dipoles, (iii) quadrupoles or (iv) octupoles in the force field. In addition, polarization has either been neglected (no pol.), modeled by isotropic atomic polarizabilities (iso. pol.) or through anisotropic atomic polarizabilities (aniso. pol.). Crystal structure water molecules close to the chromophore have also been included in the classical region and treated at the same level of approximation as the protein.

In the neutral chromophore without classically treated water molecules, a very small blue-shift of about 0.02 eV is observed when the multipole expansion is increased (moving in the vertical direction in Table 3). The absence of a noticeable shift from quadrupoles to octupoles indicate that the calculation is basically converged with respect to the multipole expansion. If water molecules are included, a blue-shift (0.05 eV) is observed when going from charges to dipoles. The trend is then reversed at higher multipoles order: the overall effect is then similar to the calculations without explicit water molecules. In both cases and regardless of the multipole order used, the effect of introducing polarization is a clear red shift of roughly 0.1 eV. Most of it is already present when an isotropic polarizability is employed.

As expected, the observed effects are larger for the anionic chromophore, which displays a 0.12 eV red-shift along the multipole series when polarization effects are included. The inclusion of crystal water also affects the excitation energy significantly, with a blue-shift between 0.12 eV (no pol.) and 0.22 eV (aniso. pol.). On the other hand, the presence of water makes the effect of increasing the multipole expansion less relevant: including dipoles red-shifts the excitation by about 0.1 eV but the red-shift is reduced once quadrupoles and octupoles are included. This reflects the observations made for the neutral form, and can be understood in light of the large dipole moment carried by the water molecules.

A red-shift of around 0.1 eV is observed in the excitation energies in neutral GFP when going from force fields without polarizabilities to force fields including these. Introducing anisotropic instead of isotropic polarizabilities further red-shifts the energies slightly, by $\sim$0.01 eV. The total red-shifts observed when going from no polarizabilities to anisotropic polarizabilities are also seen when crystal waters are included, but the shifts are smaller when going from no polarizabilities

Table 3 The lowest[a] excitation energy (in eV) in green fluorescent protein calculated with different force fields. The OPA oscillator strengths (f) / TPA cross sections (δ_{GM}) are given in parentheses.

		Neutral			Anionic		
	Multipole	no pol.	iso. pol.	aniso. pol.	no pol.	iso. pol.	aniso. pol.
No water	q	3.01 (0.126/4.43)	2.92 (0.160/13.69)	2.90 (0.152/15.26)	2.48 (0.271/165.12)	2.48 (0.320/501.12)	2.48 (0.319/501.12)
	$\rightarrow \mu]$	3.02 (0.138/3.45)	2.92 (0.168/10.20)	2.91 (0.165/10.57)	2.37 (0.223/147.01)	2.37 (0.285/451.06)	2.38 (0.289/464.84)
	$\rightarrow \Theta]$	3.02 (0.134/3.29)	2.94 (0.171/ 7.65)	2.93 (0.167/ 7.47)	2.20 (0.070/104.12)[b]	2.37 (0.281/483.99)	2.36 (0.274/478.28)
	$\rightarrow \Omega]$	3.02 (0.137/3.45)	2.94 (0.175/ 8.06)	2.93 (0.173/ 7.80)	2.25 (0.116/125.23)[c]	2.36 (0.290/483.18)	2.35 (0.290/482.33)
Incl. water	q	3.00 (0.112/11.92)	2.95 (0.148/21.88)	2.91 (0.140/25.07)	2.60 (0.269/239.73)	2.68 (0.311/566.25)	2.70 (0.308/553.37)
	$\rightarrow \mu]$	3.05 (0.126/ 8.62)	2.98 (0.162/15.41)	2.96 (0.157/17.72)	2.52 (0.251/212.18)	2.59 (0.312/629.13)	2.62 (0.310/643.79)
	$\rightarrow \Theta]$	3.01 (0.113/10.33)	2.97 (0.152/15.56)	2.94 (0.146/18.14)	2.50 (0.192/219.81)	2.65 (0.301/596.87)	2.68 (0.301/604.14)
	$\rightarrow \Omega]$	3.01 (0.114/10.12)	2.96 (0.154/14.77)	2.93 (0.148/16.83)	2.51 (0.208/223.42)	2.63 (0.303/593.98)	2.65 (0.304/605.10)

[a] Except for anionic chromophore without polarizabilites combined with dipoles and higher order multipoles, where the lowest transition had zero oscillator strength

[b] Another relevant transition at 2.48 eV (0.160/67.24)

[c] Another relevant transition at 2.54 eV (0.115/42.17)

to isotropic polarizabilities and even smaller when going from isotropic to anisotropic polarizabilities.

It is worth mentioning that the lowest transition for the anionic chromophore in the absence of polarization interactions and including multipoles (dipoles or higher) is most likely an artifact of the calculation as it is not dominated by the HOMO-LUMO transition. This is evident since the state is transparent to OPA (zero oscillator strength) and its TPA cross sections is negligible.

The environmental effect on the spectral intensities (OPA oscillator strengths and TPA cross sections) yields enhanced values, when a polarizable force field is employed: for the neutral chromophore it enhances the OPA by 30-40% and it makes the TPA cross sections 2-3 times larger than the values without polarization; for the anionic chromophore the effect on the OPA is still an enhancement albeit not as sizeable, whereas the TPA cross sections are again 2-3 times larger when polarization effects are included.

4.4 The Effect of Important Amino Acids

To investigate the effect of the protein environment on the OPA of the GFP chromophore further, selected amino acids were studied individually together with the chromophore. The results are shown in Figure 4 and Figure 5 for the neutral and the anionic form respectively.

Each amino acid was either treated at the same level as the chromophore or classically, at the level of anisotropic polarizabilities and multipoles up to and including octupoles. Results from calculations with crystal water molecules included in the structure are also given, and they were treated at the same level as the amino acid.

For the neutral GFP chromophore, no significant differences are observed between the two approaches. The largest deviations are observed for GLN69 and the positively charged

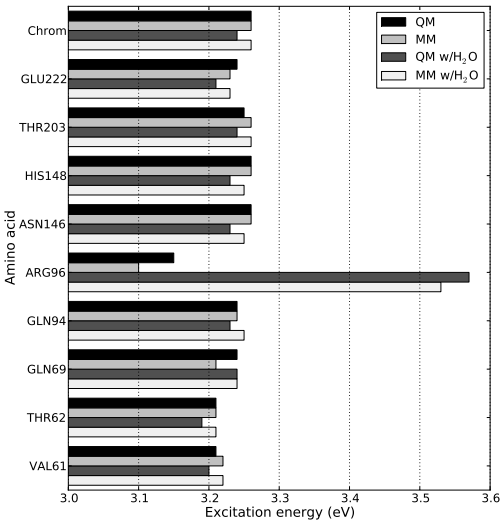

Fig. 4 The effect on the one-photon transition energies in the neutral GFP chromophore when including one amino acid to the system. *"Chrom"* is the chromophore alone, and *"QM"* and *"MM"* are therefore results from the same calculation, while *"QM w/H_2O"* and *"MM w/H_2O"* only differ in the level the water molecules were treated.

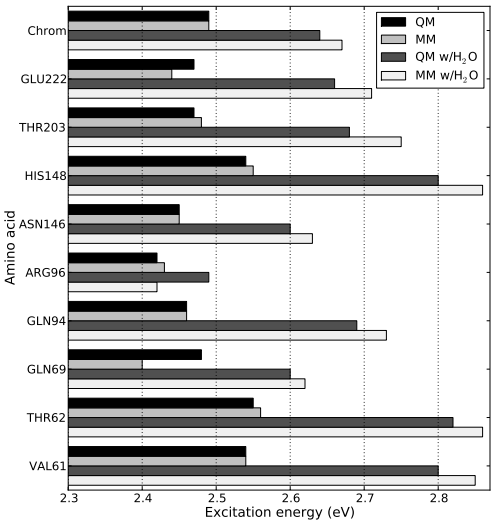

Fig. 5 The effect on the one-photon transition energies in the anionic GFP chromophore when including one amino acid to the system. *"Chrom"* is the chromophore alone, and *"QM"* and *"MM"* are therefore results from the same calculation, while *"QM w/H_2O"* and *"MM w/H_2O"* only differ in the level the water molecules were treated.

ARG96, where the red-shifts with PE are overestimated by 0.03-0.05 eV. Treating the water molecules surrounding the chromophore classically yields a small blue-shift of 0.02-0.04 eV compared to corresponding full QM system: the perfect match observed for GLN69 is then due to error cancellation.

For the anionic chromophore without water molecules (see Figure 5) there is generally very good agreement between the two cases (full QM system and a classically treated amino acid). The only noticeable exception is GLN69 where a red shift of 0.08 eV is obtained, in contrast to almost no shift observed for the full QM system. One reason for this may be that the amino acid is located very close to the chromophore backbone, the shortest distance being 1.2 Å between two hydrogen atoms, and it may therefore give rise to an unphysical behavior. Calculations where the force field parameters on the closest hydrogen atom were moved to the second closest atom (a carbon atom 1.8 Å away from the QM region) yielded, however, the same overestimation.

The most noticeable effect on the excitation energies of the anionic GFP chromophore due to the crystal water molecules is a typical blue-shift of 0.2 to 0.3 eV. The comparison of each full QM system with its classical counterpart shows that PE-DFT overestimates this effect by roughly 0.05 eV. This overestimation is also observed for the full protein when the water molecules are treated classically. Both observations can be understood by considering that both the electrostatic and the polarization interactions will be enhanced due to the presence of a charge on the chromophore. This implies that (a) the ef-

fect of the dipoles on the water molecules will be stronger and (b) differences between the full QM system and the combined system will also be enhanced.

In summary, the PE-DFT method describes the effects from specific amino acids on the excitation energies in the GFP chromophore quite well, compared with treating the whole system quantum-mechanically. The most noticeable exceptions are GLN69 for the anionic form, and ARG96 for the neutral form. GLN69 is probably not well treated due to its close proximity to the chromophores backbone, while for ARG96 a hydrogen bond between the oxygen at the imidazolin ring and the amino acid, combined with the positive charge of ARG96 (see Fig. 6) is poorly described by the PE potential. A small discrepancy between PE-DFT and pure DFT is also observed for the interaction of the anionic form with the embedded water molecules.

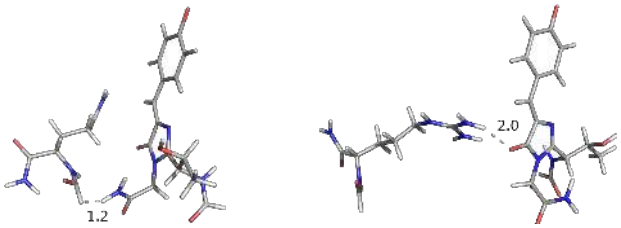

Fig. 6 The location of the two amino acids GLN69 (left) and ARG96 (right) around the GFP chromophore. The distances are given in Ångstrøm.

4.5 The Effect of Crystal Water Molecules

It is obvious, as seen in Fig. 4 and 5, that including the seven closest crystal water molecules in either the classical or QM region has a large effect on the excitation energies. We have therefore studied the effects of the specific water molecules more carefully (see table 4).

PE-DFT has no problems in describing the water molecules

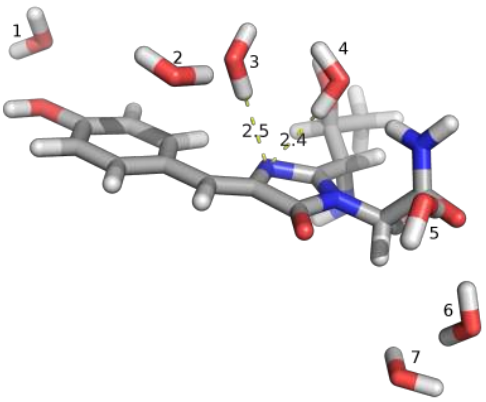

Fig. 7 The location of the seven water molecules around the neutral GFP chromophore. The orientations of the water molecules have been optimized in the complete protein structure. The numbering of the water molecules are as in table 4 and distances are given in Ångstrøm.

compared to a quantum-mechanical treatment when the chromophore is in its anionic form. PE-DFT overestimates the blue-shift due to the water molecule that is hydrogen bonded to the deprotonated oxygen (H_2O 1). The overestimated blue-shift is also seen when all seven water molecules are included in the PE potential, compared to a QM treatment of the water molecules.

The problems in describing water molecules included in the PE potential are more pronounced for the neutral chromophore. Water number 3 (see Fig. 7) gives rise to a small red-shift when treated classically but a blue-shift when treated quantum mechanically. A blue-shift is also observed when water number 4 is treated with QM, but no shift is observed when this molecule is treated classically. Both of these water molecules are hydrogen bonded to the imidazolin ring (see Fig. 7). As a consequence, the > 0.1 eV blue-shift found when going from GFP without water to GFP with QM treated water is not seen when the water molecules are treated classically.

Table 4 The lowest electronic excitation energy (in eV) in the GFP chromophore, with H_2O treated at classical or quantum mechanical level. The rest of the protein surrounding the chromophore are also included in the calculations. The numbering of the water molecules are given in Fig. 7. The OPA oscillator strengths (f) / TPA cross sections (δ_{GM}) are given in parentheses.

	Neutral		Anionic	
H_2O	PE	QM	PE	QM
None	2.93 (0.173/7.80)		2.35 (0.290/482.33)	
1	2.93 (0.161/8.45)	2.93 (0.182/10.49)	2.55 (0.291/562.20)	2.52 (0.296/614.19)
2	2.98 (0.172/6.04)	2.96 (0.174/6.19)	2.44 (0.313/547.90)	2.43 (0.305/536.49)
3	2.90 (0.157/9.54)	2.99 (0.192/12.81)	2.37 (0.305/507.03)	2.37 (0.298/497.16)
4	2.93 (0.163/5.94)	2.98 (0.188/9.37)	2.38 (0.287/516.30)	2.36 (0.276/496.23)
5	2.93 (0.165/6.92)	2.95 (0.180/8.21)	2.34 (0.280/476.63)	2.34 (0.277/479.84)
6	2.95 (0.175/7.98)	2.95 (0.182/8.57)	2.36 (0.286/489.71)	2.36 (0.287/494.60)
7	2.95 (0.175/7.19)	2.94 (0.175/6.71)	2.36 (0.288/492.97)	2.36 (0.288/496.23)
All	2.93 (0.148/16.83)	3.06 (0.237/52.69)	2.65 (0.304/605.10)	2.61 (0.300/634.89)

The one-photon oscillator strengths are not affected much by water for the anionic chromophore. For the neutral chromophore we get a reduction of the oscillator strength when waters are included in the PE potential but an increase when QM waters are included.

The calculated absorption spectra of GFP with and without water, treated at the classical and quantum mechanical levels, are presented in Fig. 8. The chromophore embedded in the classically treated protein gives a qualitatively correct spectrum compared to experiment, but red-shifted by around 25-30 nm. When seven classically treated water molecules close to the chromophore are included, the anionic peak is blue-shifted by 60 nm, whereas the neutral peak is still located at 423 nm. As a consequence, the resulting spectrum has one peak with a shoulder on the red-side with the present band width. By treating the water molecules quantum mechanically, the neutral peak is also blue-shifted resulting in a spectrum in perfect agreement with experimental spectra.

4.6 Two-Photon Absorption

In order to investigate the TPA of GFP, we have performed several calculations on the neutral and the anionic form of the chromophore. For each form, several conditions have been tested such as the ionization state of the chromophore (neutral or anionic), the presence of the protein environment, the presence of water molecules in the protein barrel described either at the classical or quantum mechanical level. Our results are reported in Tables 5 and 6, respectively, for the neutral and the anionic form. For each transition we have reported its main contributions (above 5% in square norm) in terms of orbital excitations, in order to characterize its nature. This is important when comparing excitations in different environments because different transitions are not equally affected by the varying conditions.

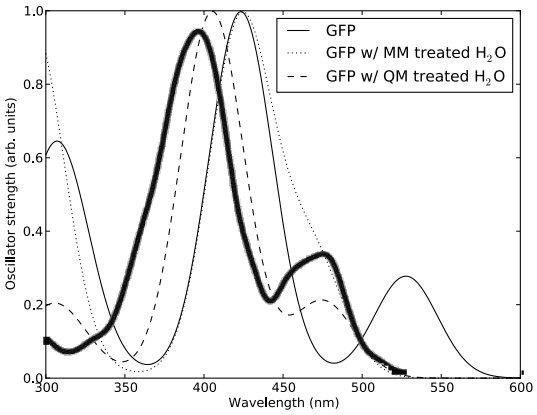

Fig. 8 The full calculated absorption spectrum of GFP without any water molecules (solid line), classically treated water molecules (dotted line) and QM treated water molecules (dashed line). The ratio between neutral and anionic GFP is set to 6:1 [1]. Gaussians have been made on each excitation with width of 20 nm. Thick solid line is experimental spectrum reprinted (adapted) with permission from ref. 5. Copyright 2001 by The American Physical Society.

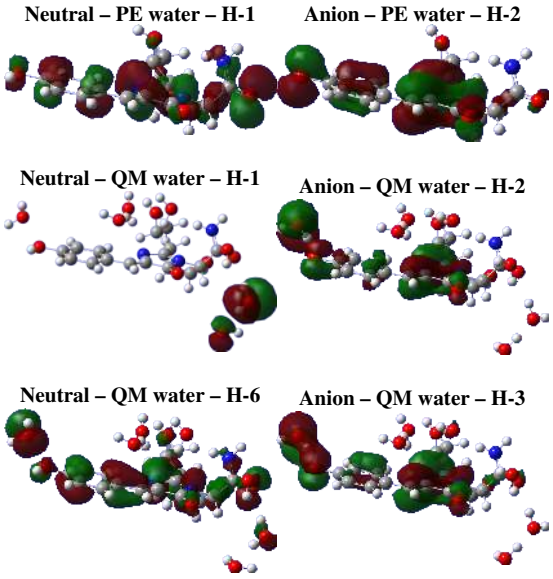

Fig. 9 A selection of molecular orbitals of the GFP chromophore

In all cases the first excitation is dominated by a HOMO→LUMO transition. Although this transition is according to our findings TP-active, it is not necessarily the one with the largest TPA cross section. For the neutral chromophore we find that the largest cross section among the first five transitions is obtained when the HOMO-1→LUMO is the dominating contribution. In case QM water molecules are present, the largest cross section is obtained for the HOMO-6→LUMO transition. A closer inspection reveals that this is qualitatively the same orbital which is now lower in energy due to the presence of the water molecules in the QM system (see Fig. 9). For comparison, we have also reported the one-photon oscillator strengths which are instead highest for the HOMO→LUMO dominated transition or at best the same as the HOMO-1→LUMO. This is a further confirmation that the two techniques are to some extent complementary also when rigorous symmetry arguments cannot be invoked.

For the anionic form of the chromophore, the HOMO→LUMO transition is again dominating the OPA spectrum. This transition is also quite important for the two-photon process. However, we have found that in the protein environment a significant cross section is displayed for the HOMO-2→LUMO transition (HOMO-3→LUMO when QM waters are present).

The comparison of our findings with the reported experimental data is in excellent agreement for what concerns the excitation energies: our "best" calculations for the neutral (Table 5, 3.06 eV) and anionic (Table 6, 2.61 eV) moieties (with GFP and QM water molecules) when compared to the latest experimental results [42] (3.04 and 2.63 eV). The OPA intensities are also very well reproduced by our oscillator strengths.

The TPA spectrum reported by Drobizhev et al. [18], alongside the spectral features shared with OPA, finds two additional features in terms of a shoulder at around 640 nm (3.87 eV) and a stronger peak (36 GM) at 550 nm (4.5 eV). Our findings confirm that the TPA spectrum in the low frequency re-

gion matches the OPA. We have also found a large TPA cross section at 610 nm (4.06 eV) for the neutral form and at 684 nm (3.62 eV) for the anionic form. These transition could explain the feature at 640 nm. The high energy peak (550 nm) is however due to a two-step process ($S_0 \rightarrow S_1$ followed by $S_1 \rightarrow S_n$) which cannot be reproduced by our TPA calculations. We observe however that the wavelength of 536-542 nm (around 4.59 eV) is quite well matched by the fifth excited state of the neutral chromophore (4.47 eV) in our calculations.

For a direct comparison the calculated two-photon absorption spectra are presented in Fig. 10, together with experimental spectrum from Drobizhev et al.[18]. It can be seen that the agreement on the TPA cross sections is however not very satisfactory. For the first peak of the neutral chromophore we obtain a cross section of 53 GM (peak at 810 nm) whereas for the anionic one (peak at 950 nm) we get 635 GM. Even by making use of the commonly used ratio of 6:1 the anionic peak si still larger than the neutral one. This is in contrast with OPA where our computed oscillator strengths match quite well the experimental data. Similar considerations can be drawn for the shoulder at 640 nm for which our calculated cross sections are much larger than the experimental measurements. It is worth mentioning that even on the experimental side the reported TPA cross sections vary from a few GM[16] to more than 100 GM[12] reflecting the challenge in such a measurement, however the relative intensity of the two low-energy peaks is by large consistent.

The orbital characterization of the TPA transitions reveals that the main component of the first excitation, which is both OPA and TPA active is the HOMO→LUMO transition as expected. This applies both for the neutral and for the anionic chromophore. At higher energy both the neutral and anionic form reveal a TPA active state which is in practice transparent to OPA. The signature of such state is characterized by the HOMO-1→LUMO transition for the neutral form (visual inspection shows that when water molecules are added to the QM system, the orbital is qualitatively the same although it is no longer the HOMO-1 orbital, being instead lower in energy). The corresponding transition for the anionic form is instead characterized by the HOMO→LUMO+2 transition (HOMO→LUMO+3 with QM water molecules).

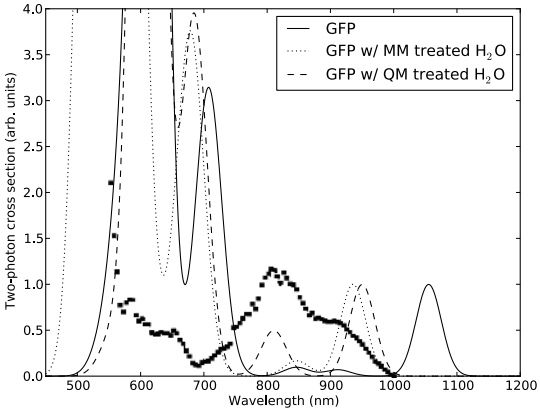

Fig. 10 The full calculated absorption spectrum of GFP without any water molecules (solid line), classically treated water molecules (dotted line) and QM treated water molecules (dashed line). The ratio between neutral and anionic GFP is set to 6:1[1]. Gaussians have been made on each excitation with width of 20 nm and the total spectra have been normalized to one in the visible region ($>$ 800 nm). Squares are the experimental spectrum, reprinted (adapted) with permission from ref. 18. Copyright 2007 American Chemical Society.

We observe also a large environmental enhancement of the cross section for both forms. For the neutral form the low energy peak is roughly 7 times larger going from gas phase (8 GM) to full embedding (53 GM). A twofold enhancement of the cross section is also observed for the other TPA active state. For the anionic form the first TPA cross section is en-

hanced three times, whereas the most active one gains two orders of magnitude through embedding.

It is worth mentioning also the environmental effect on the peak positions: for the neutral chromophore the embedding due to the PE (both GFP and water molecules) red-shifts the transition by 0.33 eV. When the solvation water molecules are treated at DFT level the transition is slightly less red-shifted (0.2 eV) displaying a value in excellent agreement with the most recent experimental observations[42]. For the anionic chromophore the transition is blue-shifted by 0.16 eV when PE is employed. However in this case promoting water molecules to the QM system only marginally affects the result reducing the blue-shift to 0.12 eV. A possible interpretation of the difference is that electrostatic effects, which are well reproduced by PE are more prominent for the anionic moiety, whereas the quantistic effect due to promoting water from PE to QM plays a more important role for the neutral one.

5 Conclusions

We have presented a study of the one-photon and two-photon absorption properties of the GFP chromophore in its native environment. The validity of our approach has been tested by comparing PE-DFT results with full DFT results when only one amino acid was present in addition to the chromophore. These results confirmed that polarizable embedding is a faithful method for describing the protein environment: differences between the full DFT and PE-DFT models are well below 0.1 eV, thus significantly smaller than the expected accuracy of DFT in predicting excitation energies. We have also shown that water inside the protein barrel leads to a significant blue-shift for the anionic form, but that it is less important for the neutral form, with the noticeable exception of the chromophore-ARG96 subsystem.

We have then turned our attention to the two-photon absorption (TPA) of the full system. The embedded chromophore shows a larger TPA cross section with respect to the bare chromophore, and the largest value is observed when water is included in the QM part by means of a supermolecular approach. The spectral features are very well reproduced by our results: our findings support well the most recent experimental observations by Drobizhev et al.[18] on GFP in the whole spectral range. The only observed discrepancy regards the intensity of the TPA peaks which we have not been able to reproduce. We believe that this point deserves further investigations by theoreticians and experimentalists alike.

A recent theoretical investigation with DFT and wavefunction methods for the chromophore and non polarizable MM embedding for GFP suggested that the absence of PE could be responsible for the mismatch between their observations and the experimental results[21]. Our finding confirm that picture but also highlight the importance of the supermolecular approach where nearby water molecules are included in the QM region. In this way the experimental peak positions are very well reproduced.

Finally, our study does not support the presence of a dark OPA state which should be instead TPA active, at least not in the form proposed by Hosoi et al.[16]. In their work they observe that the TPA spectrum of the bare anionic chromophore is blue-shifted by some 10nm (0.07eV). We do see TPA active states at higher energy than the lowest one but the energy difference between such states is larger than 1eV and is therefore not supporting the "dark state" interpretation proposed by Hosoi et al.[16]. This observation, together with a possible breakdown of the Frank-Condon approximation, suggested in another theoretical investigation[20], could be the source of the

observed wavelength shift between the OPA and TPA spectra.

Acknowledgments

This work has received support from the Research Council of Norway through a Centre of Excellence Grant (Grant No. 179568/V30) and from the Norwegian Supercomputer Program. J.K. thanks The Danish Councils for Independent Research (STENO and Sapere Aude programmes), the Lundbeck Foundation and the Villum foundation for financial support and the Danish Center for Scientific Computing (DCSC) for computational resources.

References

1 R. Y. Tsien, *Annu. Rev. Biochem.*, 1998, **67**, 509–544.

2 K. Brejc, T. K. Sixma, P. A. Kitts, S. R. Kain, R. Y. Tsien, M. Ormö and S. J. Remington, *Proc. Natl. Acad. Sci. U. S. A.*, 1997, **94**, 2306–2311.

3 H. Niwa, S. Inouye, T. Hirano, T. Matsuno, S. Kojima, M. Kubota, M. Ohashi and F. I. Tsuji, *Proc. Natl. Acad. Sci. U. S. A.*, 1996, **93**, 13617–13622.

4 L. Lammich, M. Å. Petersen, M. B. Nielsen and L. H. Andersen, *Biophys. J.*, 2007, **92**, 201–207.

5 S. B. Nielsen, A. Lapierre, J. U. Andersen, U. V. Pedersen, S. Tomita and L. H. Andersen, *Phys. Rev. Lett.*, 2001, **87**, 228102.

6 J. J. van Thor, *Chem. Soc. Rev.*, 2009, **38**, 2935–2950.

7 S. L. Maddalo and M. Zimmer, *Photochem. Photobiol.*, 2006, **82**, 367–372.

8 L. M. Tolbert, A. Baldridge, J. Kowalik and K. M. Solntsev, *Acc. Chem. Res.*, 2011.

9 S. R. Meech, *Chem. Soc. Rev.*, 2009, **38**, 2922–2934.

10 J. Dong, K. M. Solntsev and L. M. Tolbert, *J. Am. Chem. Soc.*, 2006, **128**, 12038–12039.

11 C. Xu, W. Zipfel, J. B. Shear, R. M. Williams and W. W. Webb, *Proc. Natl. Acad. Sci. U. S. A.*, 1996, **93**, 10763–10768.

12 A. D. Xia, S. Wada, H. Tashiro and W. H. Huang, *Arch. Biochem. Biophys.*, 1999, **372**, 280–284.

13 A. Volkmer, V. Subramaniam, D. J. S. Birch and T. M. Jovin, *Biophys. J.*, 2000, **78**, 1589–1598.

14 A. A. Heikal, S. T. Hess and W. W. Webb, *Chem. Phys.*, 2001, **274**, 37–55.

15 G. A. Blab, P. H. M. Lommerse, L. Cognet, G. S. Harms and T. Schmidt, *Chem. Phys. Lett.*, 2001, **350**, 71–77.

16 H. Hosoi, S. Yamaguchi, H. Mizuno, A. Miyawaki and T. Tahara, *J. Phys. Chem. B*, 2008, **112**, 2761–2763.

17 M. Zimmer, *Chem. Rev.*, 2002, **102**, 759–781.

18 M. Drobizhev, N. S. Makarov, T. Hughes and A. Rebane, *J. Phys. Chem. B*, 2007, **111**, 14051–14054.

19 R. Nifosì and Y. Luo, *J. Phys. Chem. B*, 2007, **111**, 14043–14050.

20 E. Kamarchik and A. I. Krylov, *J. Phys. Chem. Lett.*, 2011, **2**, 488–492.

21 C. Filippi, M. Zaccheddu and F. Buda, *J. Chem. Theory Comput.*, 2009, **5**, 2074–2087.

22 J. M. Olsen, K. Aidas and J. Kongsted, *J. Chem. Theory Comput.*, 2010, **6**, 3721–3734.

23 A. H. Steindal, J. M. H. Olsen, J. Kongsted, K. Ruud and L. Frediani, in preperation.

24 J. M. H. Olsen and J. Kongsted, *Adv. Quantum Chem.*, 2011, **61**, 107–143.

25 T. Helgaker, P. Jørgensen and J. Olsen, *Molecular Electronic-Structure Theory*, John Wiley & Sons, Chichester, 2000.

26 P. Sałek, O. Vahtras, T. Helgaker and H. Ågren, *J. Chem. Phys.*, 2002, **117**, 9630–9645.

27 P. Sałek, O. Vahtras, J. Guo, Y. Luo, T. Helgaker and H. Ågren, *Chem. Phys. Lett.*, 2003, **374**, 446–452.

28 DALTON, a molecular electronic structure program, Release Dalton2011 (2011), see http://daltonprogram.org/.

29 T. Yanai, D. P. Tew and N. C. Handy, *Chem. Phys. Lett.*, 2004, **393**, 51–57.

30 M. J. G. Peach, P. Benfield, T. Helgaker and D. J. Tozer, *J. Chem. Phys.*, 2008, **128**, 044118.

31 D. H. Friese, C. Hättig and K. Ruud, *Phys. Chem. Chem. Phys.*, 2012, **14**, 1175–1184.

32 F. Jensen, *J. Chem. Theory Comput.*, 2010, **6**, 2726–2735.

33 Schrödinger suite 2009 Protein Preparation Wizard, Schrödinger LCC, New York, NY 2009.

34 L. Gagliardi, R. Lindh and G. Karlström, *J. Chem. Phys.*, 2004, **121**, 4494–4500.

35 F. Aquilante, L. De Vico, N. Ferré, G. Ghigo, P.-Å. Malmqvist, P. Neogrády, T. B. Pedersen, M. Pitoňák, M. Reiher, B. O. Roos, L. Serrano-Andrés, M. Urban, V. Veryazov and R. Lindh, *J. Comput. Chem.*, 2010, **31**, 224–247.

36 D. W. Zhang and J. Z. H. Zhang, *J. Chem. Phys.*, 2003, **119**, 3599–3605.

37 P. Söderhjelm and U. Ryde, *J. Phys. Chem. A*, 2009, **113**, 617–627.

38 G. A. Kaminski, R. A. Friesner, J. Tirado-Rives and W. L. Jorgensen, *J. Phys. Chem. B*, 2001, **105**, 6474–6487.

39 M. R. Silva-Junior, S. P. A. Sauer, M. Schreiber and W. Thiel, *Mol. Phys.*, 2010, **108**, 453–465.

40 R. Send, V. R. I. Kaila and D. Sundholm, *J. Chem. Theory Comput.*, 2011, **7**, 2473–2484.

41 M. J. G. Peach, C. R. Le Sueur, K. Ruud, M. Guillaume and D. J. Tozer, *Phys. Chem. Chem. Phys.*, 2009, **11**, 4465–4470.

42 T. M. Creemers, A. J. Lock, V. Subramaniam, T. M. Jovin and S. Völker, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 2974–2978.

Table 5 Main orbital components of the excitation vectors for the first 5 excited states of the neutral GFP chromophore in different environments: gas-phase, PE protein, PE protein and water, PE protein and QM water. Excitation energies (eV), wavelengths (nm), OPA oscillator strengths and TPA cross sections for each transition are also shown. The employed PE consists of a multipole expansion up to quadrupoles (Ω) and anisotropic polarizabilities (α_{aniso})

Emb	H_2O	Ex	E (eV)	λ (nm)	osc. str.	δ_{GM}	Character
-	-	1	3.26	380	0.195	7.76	47%(H→L)
		2	3.63	342	0.000	2.21	28%(H-4→L)+15%(H-3→L)
		3	4.27	290	0.055	817.59	34%(H-1→L)+7%(H-3→L)
		4	4.52	274	0.001	61.67	33%(H-2→L)+9%(H→L+2)
		5	4.78	259	0.000	41.32	32%(H-9→L)
GFP	-	1	2.93	423	0.173	7.80	46%(H→L)
		2	3.45	359	0.001	15.98	20%(H-7→L)+10%(H-6→L)+7%(H-1→L)
		3	4.03	308	0.107	1370.86	24%(H-1→L)+6%(H-3→L)+5%(H-6→L)+5%(H-5→L)
		4	4.14	299	0.003	25.62	46%(H→L+1)
		5	4.31	288	0.004	179.66	38%(H-4→L)
GFP	PE	1	2.93	423	0.148	16.83	47%(H→L)
		2	3.49	355	0.001	6.93	13%(H-1→L)+13%(H-4→L)+8%(H-2→L)
		3	4.25	292	0.145	1079.93	21%(H-1→L)+11%(H-4→L)+7%(H-2→L)
		4	4.45	279	0.000	16.25	41%(H→L+1)
		5	4.57	271	0.000	201.38	12%(H→L+4)+11%(H-5→L)+5%(H-6→L)
GFP	QM	1	3.06	405	0.237	52.69	46%(H→L)
		2	3.64	341	0.001	7.57	20%(H-10→L)+16%(H-9→L)+7%(H-5→L)
		3	4.06	305	0.047	1748.84	19%(H-6→L)+18%(H-2→L)
		4	4.16	298	0.001	45.34	21%(H-4→L)+14%(H-3→L)+9%(H-2→L)
		5	4.47	277	0.000	43.16	22%(H-3→L)+15%(H-1→L)+5%(H-4→L)+5%(H-2→L)

Table 6 Main orbital components of the excitation vectors for the first 5 excited states of the anionic GFP chromophore in different environments: gas-phase, PE protein, PE protein and water, PE protein and QM water. Excitation energies (eV), wavelengths (nm), OPA oscillator strengths and TPA cross sections for each transition are also shown. The employed PE consists of a multipole expansion up to quadrupoles (Ω) and anisotropic polarizabilities (α_{aniso})

Emb	H_2O	Ex	E (eV)	λ (nm)	osc. str.	δ_{GM}	Character
-	-	1	2.49	498	0.272	212.61	48%(H→L)
		2	2.62	473	0.000	0.01	44%(H-1→L)
		3	3.53	351	0.000	28.05	44%(H→L+1)
		4	3.82	325	0.007	13.99	29%(H→L+2)+15%(H-2→L)
		5	3.83	324	0.014	10.06	27%(H-2→L)+16%(H→L+2)
GFP	-	1	2.35	528	0.290	482.33	47%(H→L)
		2	2.72	456	0.017	33.83	31%(H→L+1)+16%(H→L+2)
		3	2.84	437	0.000	0.01	42%(H-1→L)
		4	3.37	368	0.000	43.94	29%(H→L+2)+14%(H→L+1)
		5	3.50	354	0.002	940.65	39%(H-2→L)
GFP	PE	1	2.65	468	0.304	605.10	46%(H→L)
		2	3.62	342	0.000	335.29	14%(H→L+3)+10%(H→L+1)+10%(H→L+2)+5%(H-2→L)
		3	3.64	341	0.001	1432.91	18%(H-2→L)+13%(H-1→L)
		4	3.72	333	0.003	616.48	19%(H-4→L)+13%(H-2→L)+6%(H-1→L)+5%(H-5→L)
		5	3.90	318	0.000	76.67	22%(H-1→L)+16%(H-4→L)+5%(H-2→L)
GFP	QM	1	2.61	475	0.300	634.89	46%(H→L)
		2	3.44	360	0.000	6.07	24%(H→L+1)+5%(H→L+2)+13%(H→L+3)
		3	3.48	356	0.000	1.37	42%(H-1→L)
		4	3.62	342	0.002	2419.62	24%(H-3→L)+17%(H-2→L)
		5	3.76	330	0.002	76.65	10%(H→L+1)+8%(H→L+2)+14%(H→L+3)+6%(H→L+4)