Journal Name

ARTICLE TYPE

Cite this: DOI: 00.0000/xxxxxxxxxx

The key to the yellow-to-cyan tuning in the green fluorescent protein family is polarisation†

Riccardo Nifosì *[a], Benedetta Mennucci[b], and Claudia Filippi[c]

Received Date
Accepted Date

DOI: 00.0000/xxxxxxxxxx

Computational approaches have to date failed to fully capture the large (about 0.4 eV) excitation energy tuning displayed by the nearly identical anionic chromophore in different green fluorescent protein (GFP) variants. Here, we present a thorough comparative study of a set of proteins in this sub-family, including the most red- (phiYFP) and blue-shifted (mTFP0.7) ones. We employ a classical polarisable embedding through induced dipoles and combine it with time-dependent density functional theory and multireference perturbation theory in order to capture both state-specific induction contributions and the coupling of the polarisation of the protein to the chromophore transition density. The obtained results show that only upon inclusion of both these two effects generated by the mutual polarisation between the chromophore and the protein can the full spectral tuning be replicated. We finally discuss how this mutual polarisation affects the correlation between excitation energies, dipole moment variation, and molecular electrostatic field.

1 Introduction

The green fluorescent protein (GFP) family is an outstanding example of proteins which show extensive spectral tuning.[1–4] covering the whole visible spectrum and beyond. This unique behaviour, which has been paramount in enabling multicolor imaging,[5] is due both to different chromophore structures and to the interactions of such chromophore with different protein surroundings. Indeed, the same chromophore – a hydroxybenzilidene-imidazolinone (HBI) motif in the anionic protonation state, Fig. 1 – in different protein environments exhibits extensive perichromism of its π-π^* transition, with excitation peak varying from 450 (mTFP0.7[6]) to 525 nm (phiYFP[7]), respectively, at the short and long wavelength extremes of the anionic HBI-chromophore GFPs. Similar wide spectral tuning is at play in red FPs[8–10], containing a chromophore with an enlarged π-conjugated system.

Continuous scientific effort has been devoted over the last decade to understand how nature achieves this tuning by selecting and arranging the available amino acids (and water molecules) around the chromophore(s). A possible interpretation is through internal electrostatic fields. In particular, for GFPs, Drobizhev *et al.*[4] proposed a model explaining the spectral tuning on the basis of the chromophore excess dipole moment (i.e. the dipole moment variation upon excitation, $\Delta\mu$). Larger $\Delta\mu$ are associated with blue-shifted excitation energies and can also be correlated with wider absorption spectra (YFPs have narrower spectra than cyan/green), with the rationale that larger $\Delta\mu$ will be associated with larger variations of excitation energy due to the fluctuating internal electrostatic field. A suitable combination of one- and two-photon experimental measurements could in fact estimate both the $\Delta\mu$ values and the (average) molecular electrostatic field felt by the chromophore[11], exploiting the relation between $\Delta\mu$ and the excitation energy.

In addition to empirical correlations, many quantum chemical investigations have been presented in recent years to explain the molecular mechanism of the tuning in GFP and related proteins.[9,12–21] In these studies, a hybrid approach is usually employed, where the chromophore (possibly together with its immediate surroundings) is treated with quantum mechanical (QM) methods while the embedding system is treated by less computationally expensive methods, often a molecular-mechanics (MM) approach. Besides its computational convenience, this separation

[a] NEST, CNR - Istituto Nanoscienze and Scuola Normale Superiore, Piazza San Silvestro 12, 56127 Pisa, Italy
[b] Dipartimento di Chimica e Chimica Industriale, Universitaà di Pisa, Via Giuseppe Moruzzi 13, 56124 Pisa, Italy
[c] MESA+ Institute for Nanotechnology, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands

† Electronic Supplementary Information (ESI) available: Structural models (Figures S1 and S2); complete excitation-energy results (Figures S3 and S4) and corresponding values (Table S1-3); relation between excitation-energy shifts due to polarisation and dipole moment change, $\Delta\mu$ (Figure S5) and complete $\Delta\mu$ results (Table S4-S6, Figure S6); comment on per-residue contributions to polarisation energy terms (Section S3) and corresponding results (Figure S7 and Table S7) ; comment on the molecular electrostatic field in the chromophore cavity (Section S4) and corresponding results (Figures S8 and S9, Table S8); angle between $\Delta\mu$ and transition dipole moment (Table S9). See DOI: 00.0000/00000000.

has the advantage to provide intuitive explanations of the tuning mechanisms, particularly when the interactions are treated "classically" as in the MM methods. Some of us already investigated the tuning of "green" FPs using a QM/MM electrostatic embedding approach[17], showing however that the whole extent of the tuning is not fully reproduced even when accurate quantum chemistry methods (CASPT2 and SAC-CI) are employed. TD-DFT in a polarisable environment[19] has also been used for similar investigations; however, in the case of anionic chromophore proteins (wtGFP-, EGFP, and EYFP), the tuning is only partially reproduced with the shift between wtGFP- and EYFP being only 0.05-0.06 eV compared to the experimental 0.20 eV.

The lack of a satisfactory agreement with the experimental spectral tuning in these previous works has prompted us to address the problem combining the best quantum chemical methods applicable to this kind of chromophores, with the state of the art of classical MM approaches introducing a polarisable force field. When polarisation of the MM subsystem is included, two effects combine to modify the excitation energy of the system[22,23]. The first is a classical effect due to the (instantaneous) polarisation of the embedding protein to the different electronic charge densities of the ground and excited states (i.e. state-specific polarisation, or polSS in the following). The second term is a purely QM effect and has been ascribed to the instantaneous response of the MM polarisable sites to the *transition* density of the electronic excitation (i.e. linear-response polarisation, polLR in the following)[22,24–26]. Using the polarisable MM model[25,27], in which atomic polarisabilities give rise to induced dipole moments on the MM sites, we include these two polarisation effects performing TD-DFT and complete active space self-consistent field (CASSCF) and corresponding second-order perturbation (CASPT2) calculations. The proteins investigated here are listed in Table 1 (their chromophore pocket is shown in Fig. 1) and include the most red (phiYFP) and blue (mTFP0.7) shifted FP with an anionic HBI chromophore to date.

The consistency of the obtained results and the quantitative reproduction of the experimental shifts allow us for the first time to assess the connection between tuning, dipole-moment variation upon excitation, and electrostatic field in the chromophore cavity, thus giving a clear interpretation of the role of the protein in determining •he spectral properties of the embedded chromophore.

2 Theory and computational methods

2.1 Theoretical framework

Mutual polarisation effects between the QM and the MM subsystems are here introduced through induced dipoles sitting on the MM atoms, which are characterized by an isotropic atomic polarisability (in addition to a fixed point charge)[27,34]. We employ here three variants of this approach which we shall indicate as MMpol and which requires a self-consistent solution of the QM and MM polarisation problem. In the polGS scheme, the dipoles are polarised to the ground-state charge density of the chromophore and the excitation energy, $\Delta E_{01}^{\mathrm{polGS}}$, is then computed in these fixed MM dipoles. In the state-specific (polSS) scheme, the dipoles are separately optimised for each electronic state of the QM subsystem, in our case the ground and the excited state. The polSS excitation energy computed as the difference of the energy of the complete system in the two states, $\Delta E_{01}^{\mathrm{polSS}}$, fully accounts for the classical induction contribution to spectral tuning. The polSS scheme requires the QM density in the excited state and is thus amenable for wave function methods such as CASSCF/CASPT2. By contrast, in linear-response approaches for the determination of excitation energies (such as linear-response TD-DFT), the inclusion of a polarisable embedding leads to a different contribution to the spectral tuning, where the MM dipoles are polarised to the transition density of the QM subsystem. We shall name this last scheme polLR and the corresponding excitation energy will be denoted as $\Delta E_{01}^{\mathrm{polLR}}$.

Following Ref.[22], the polSS and polLR contributions can be approximately expressed as a function of the dipole moment change of the QM subsystem upon excitation ($\Delta\mu$) and its transition dipole moment (μ_{01}), respectively:

$$\Delta^{\mathrm{SS}} = \Delta E_{01}^{\mathrm{polSS}} - \Delta E_{01}^{\mathrm{polGS}} \simeq -\frac{1}{2} g_e \left|\Delta\mu\right|^2 \quad (1)$$

$$\Delta^{\mathrm{LR}} = \Delta E_{01}^{\mathrm{polLR}} - \Delta E_{01}^{\mathrm{polGS}} \simeq -g_e^d \left|\mu_{01}\right|^2 , \quad (2)$$

where g_e describes the static electronic response of the solvent (in our case, the surrounding protein matrix), while g_e^d is the dynamic response at the frequency of the excitation. A result equivalent to Eq. 1 was also recently re-derived in a more general context[26].

For the FPs of Table 1, we will perform QM/MMpol calculations of the excitation energy using the TD-DFT and the multireference correlated CASSCF/CASPT2 approaches to treat the QM chromophore. While the wave function calculations will allow us to directly access the polSS contributions, TD-DFT will instead give us the polLR term for the same systems. In order to account for the missing polLR term in the polSS CASSCF/CASPT2, we will use Eq. 2 and set

$$\left.\frac{\Delta^{\mathrm{LR}}}{\mu_{01}^2}\right|_{\mathrm{CASPT2}} = \left.\frac{\Delta^{\mathrm{LR}}}{\mu_{01}^2}\right|_{\mathrm{TD-DFT}} . \quad (3)$$

Likewise, one can estimate the Δ^{SS} missing from the polLR calculation. This can be achieved either using Eq. 1 or by a perturbation approach (i.e. without self-consistency), based on the change in charge density upon excitation. This is the so-called corrected linear response[35,36] or cLR scheme, here referred to as pol(LR+SS). We anticipate that this correction to the TD-DFT excitation energies is generally quite small due to the well known underestimation of the change in dipole moment upon excitation in TD-DFT (see for example Ref.[37]).

2.2 Protein models

We treat the different FPs following a consistent protocol. In particular, we keep the atoms close to their positions in the crystal structures (apart from the hydrogens, which are equilibrated by MM molecular dynamics simulations, and the atoms of the QM region). Although possible structural artefacts may arise from crystallisation, spectroscopic studies on FP crystals show a rather good agreement between absorption spectra in solution

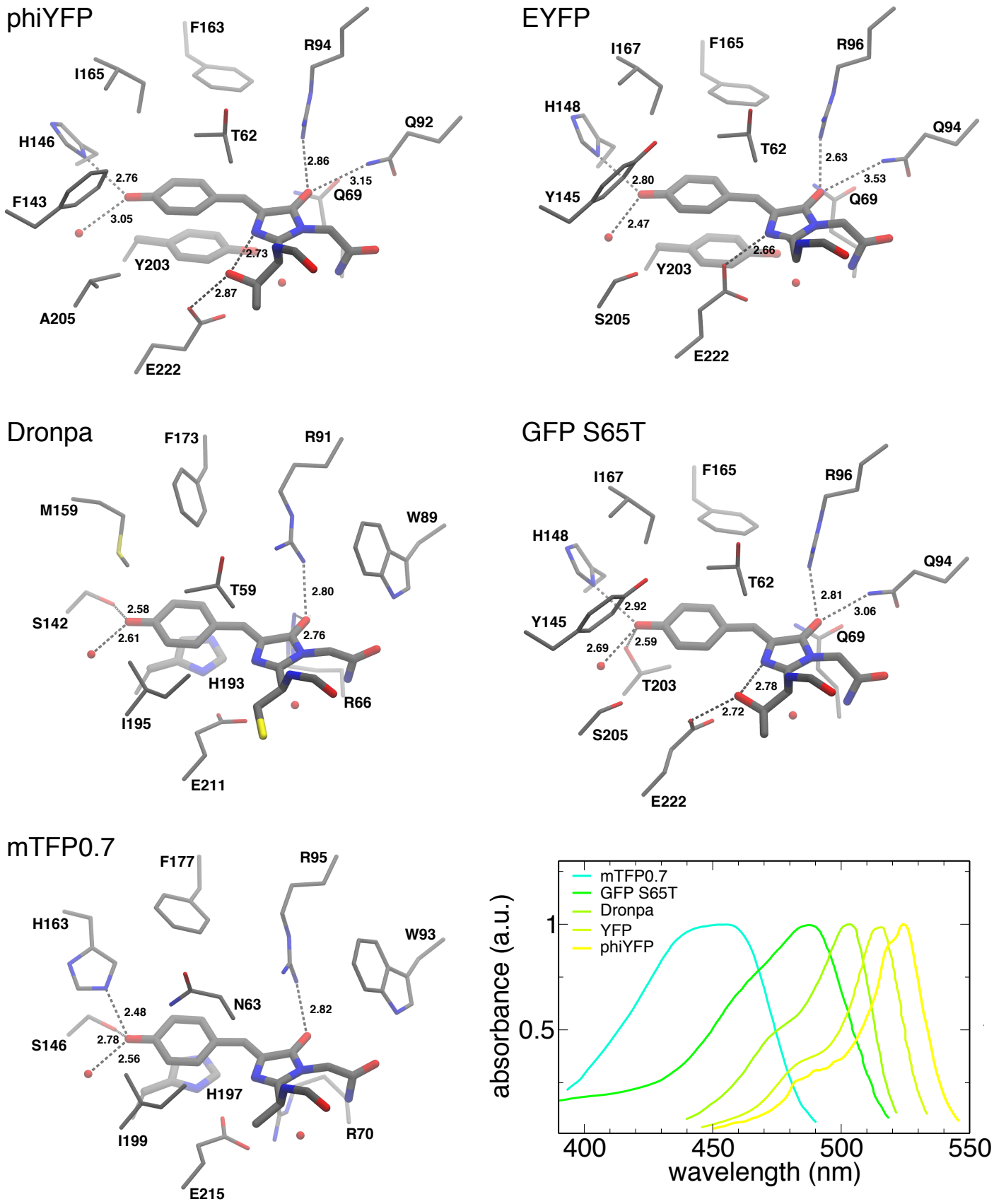

Fig. 1 Chromophore pockets from X-ray structures for the five FPs investigated in this work. The thicker parts indicate the QM subsystem. Figure produced using VMD[33]. The normalised absorption spectra are shown in the bottom right panel.

Table 1 Fluorescent proteins investigated in this work, their absorption properties, and X-ray structures.

Protein	Absorption peak[§]		$\varepsilon^{\dagger}$	X-ray structure	Temp of X-ray
	nm	eV	$M^{-1}cm^{-1}$		K
phiYFP*	525	2.36	124000	4HE4[7]	100
EYFP[‡]	515 (520)	2.41 (2.38)	83400	1YFP[28]	295
Dronpa	503	2.46	95000	2IOV[29]/2Z1O[30]	100
GFP-S65T	484 (490)	2.56 (2.53)	55000	1EMG[31]	295
mTFP0.7	453	2.74	60000	2OTB[6]	100

§ The values given in parenthesis for EYFP and GFP-S65T are from low-temperature measurements (1.6 K)[32].
†Extinction coefficient.
‡ wtGFP with mutations T203Y/S65G/V68L/S72A.
* Structure of phiYFPv, spectrally identical to phiYFP but with mutations for improved maturation in bacteria[7].

and *in cristallo*. For example, the solution absorption spectrum of (non-photoconverted and at room temperature) IrisFP[38] nicely matches the one in the crystal as far as the peak wavelength is concerned. In c3GFP, a folding mutant of *av*GFP, the absorption peak of the anionic state is slightly red shifted[39] from 480 to 500nm (corresponding to ∼0.1 eV) presumably due to dimerisation, and the same shift is recorded at high protein concentration in solution. Finally, the fluorescence spectrum of Citrine *in cristallo* is very similar to that in solution[40].

We start from the PDB structures of the five proteins listed in Table 1 and shown in Fig. 1. Among the several X-ray structures of Dronpa, we select the one with PDB code 2Z1O and, in order to assess the sensitivity of our results to the starting structure, we extend our analysis in the ESI to 2IOV, finding only very slight differences. In the cases where more than one polypeptide chain is present, chain A is considered. The Glu residue (Glu222 in GFP) close to the chromophore is protonated in phiYFP, EYFP, and GFP-S65T, and deprotonated in Dronpa and mTFP0.7. Following Ref.[17], we solvate the proteins in a water box with Na^+ and Cl^- ions added to a salt concentration of 0.1 M (and Na^+ in excess to neutralise the system). We then perform a force-field molecular equilibration to equilibrate the bulk solvent molecules and the hydrogen atoms of the protein with restraints to keep the protein structure and the crystallographic water molecules in virtually the same positions of the X-ray structure. To this aim, we start with a MD simulation of 300 ps at room temperature and impose restraints of 1000 kJ/mol/Å^2 on the positions of the non-hydrogen atoms of the protein. We then perform a simulated annealing run of 1 ns, gradually decreasing the temperature from 350 K and to 5 K. Finally, we perform 4000 steps of geometry optimisation, increasing the strength of the restraints to 106 kJ/mol/Å^2 so that the position of the non- hydrogen atoms are virtually the same as in the X-ray structure (maximum distance < 0.1 Å).

The orientation of the hydrogen atoms for the residues and water molecules close to the chromophore is particularly relevant in the spectral tuning. In almost all cases, these orientations can be established unambiguously from the X-ray structures and the MD simulations. However, for residue Tyr145 in GFP-S65T and EYFP, both possible hydrogen orientations of the OH group are plausible and also unrestrained MD simulations show that both configurations are occupied with roughly similar populations. We therefore analyse both geometries and refer to them in the Tables as $OH^{145}\downarrow$ and $OH^{145}\uparrow$ depending on whether the OH points, respectively, toward the chromophore phenolate or away from it. A very recent subatomic resolution X-ray structure of the S65T/F99S/M153T/V163A mutant indicates that the hydrogen points away from the chromophore ($OH^{145}\uparrow$)[41].

The resulting solvated proteins are then optimised in a QM/MM simulation with the ONIOM Amber99//PBE0 6-31G(d) approach, where the MM atoms are kept fixed and the QM subsystem is relaxed in DFT with the PBE0 functional and the 6-31G(d) basis set. The QM part comprises the chromophore (Fig. 1 and Fig. S1 in the ESI) and the stacked aromatic residue which is present in all investigated FPs except GFP-S65T where, in the absence of an aromatic side chain, the corresponding cavity is occupied by Thr203 and some water molecules. The stacked residue is Tyr203 in EYFP and phiYFP, His193 in Dronpa, and His197 in mTFP0.7 with the Histidine being protonated in both cases (charge +1). At the QM/MM boundaries, the C_β of these amino acids is replaced with a methyl group. For comparison, calculations with only the chromophore in the QM subsystem are reported in the ESI. The chromophore pockets in the final models are compared with the starting X-ray structures in Figure S2. The molecular dynamics simulations were carried out with Gromacs 5.1[42,43], while we employ Gaussian 09[44] for the ONIOM calculations.

2.3 Excitation energy calculations, $\Delta\mu$ and MMpol set up

For the calculations of the excitation energies, we use TD-DFT at the CAM-B3LYP[45]/6-31+G* level (results with LC-BLYP[46] are reported in the ESI) and state-average CASSCF/CASPT2 with the ANO basis set[47] and two roots with equal weights. The active space is a CAS(14,13) that includes all lowest π molecular orbitals on the chromophore except the one corresponding to the lone pair on the nitrogen atom connecting the system to the protein backbone (see ref.[23]). A two-step strategy is used for the MMpol CASSCF/CASPT2 calculations: the state-specific dipoles computed within CASSCF/MMpol are used as a static external potential in CASPT2. We use a modified version of Gaussian09, revision A.02[44] for TD-DFT and CASSCF/MMpol, and Molcas[48] for the CASPT2 calculations.

Dipole moments in CASPT2 are calculated by finite field perturbation theory (FFPT), i.e. by taking the derivative of the energy at vanishing uniform external electric field. In our calculation, we use a field of 10^{-4} a.u. The FFPT dipole and the one computed as an expectation value of the dipole operator are equal for variational wave functions but not for state-average CASPT2 wave

functions. In this case, the $\Delta\mu$'s from FFPT are more accurate because the energy is predicted at a higher order in perturbation theory than the wave function (see for example Ref.[49]).

For the non-polarisable QM/MM calculations (hereafter simply QM/MM), we use the Amber99 charges[50], setting to zero the charges of the atoms closest to the MM boundary and redistributing them to the next charge group, weighted by nuclear charge. In the QM/MMpol treatment, the polarisabilities and partial charges for the polarisable environment are taken from Amber pol12 parameters (AL model in Ref.[51]). The polarisabilities of the MM atoms located at distances greater than 15 Å (20 Å in some cases) to any QM atom are set to zero and their partial charges to the Amber99 values. This choice of polarisation cutoff has been shown to yield converged excitation energies[52].

3 Results

The different contributions to the tuning

Table 2 and the left panel of Fig. 2 report the excitation energies calculated with CASPT2 on the set of FPs in the various embedding schemes, namely, non-polarisable QM/MM and the two variants of QM/MMpol with the MM dipoles polarised to the ground-state charge density (polGS) and to the state-specific densities (polSS).

As already pointed out in Ref.[17], apart from being blue shifted, the QM/MM excitation energies fail to reproduce the full 0.38 eV experimental tuning, varying instead by only 0.29 eV. Although these discrepancies might not appear so severe at a first glance, we need to recall that we are using one of the most accurate QM methods for excited states (CASPT2) and we are looking at *trends*, so systematic errors should cancel out. Nonetheless, the agreement with the experiment is far from the conventional chemical accuracy threshold of $\sim$0.04 eV (1 kcal/mol). Therefore, the embedding by static charges appears to be missing important effects on the spectral tuning.

To further investigate this, we first switch on the mutual polarisation between the protein and the chromophore in its ground state using our MMpol approach. This improved polGS description of the chromophore-protein interactions in the ground state results however in a much broader tuning of the excitation energy (0.50 eV), worsening the agreement with experiments. The reason for this finding is that the ground state of the system is over-stabilised by the favourable interactions with the induced MM dipoles, at the expenses of the excited-state energy. When we allow the mutual polarisation to change in the excited state by relaxing the MM dipoles to the state-specific charge densities (polSS), the polGS over-tuning is corrected and a remarkable agreement between the predicted and the experimental tuning (0.38 eV) is obtained. Furthermore, not only the variation in excitation energy between the extremes (phiYFP and mTFP0.7) is well reproduced but the excitation energies of all the proteins are uniformly shifted with respect to the corresponding experimental values, so the relative tuning is consistently well reproduced as we move from yellow to cyan FPs.

We note however that these results are all blue shifted by a sizeable –albeit uniform– amount ($\sim$0.4 eV), a difference which cannot only be attributed to vibronic effects or to a broadening due to thermal movements but to effects that are not accounted for in our polSS description. What we are still missing is in fact the coupling between the MM polarisation and the electronic excitation, which is described by the linear-response term discussed above. Our strategy is therefore to evaluate this contribution in the framework where it arises intrinsically, i.e. in linear-response TD-DFT.

The right panel of Fig. 2 shows the results of the TD-DFT/CAM-B3LYP calculations (see also Table S2). If we polarise the MM dipoles to the ground-state DFT density and keep them fixed in the computation of the TD-DFT excitation energies, we obtain polGS results that are in line with the CASPT2 findings, i.e. we observe a larger extent of tuning with respect to QM/MM. The inclusion of the polLR contributions red shifts the polGS excitation energies by a rather uniform amount (0.22-0.27 eV) across the set of proteins. This uniformity can be easily explained through Eq. 2 and the moderate variation of the transition dipole moment, which spans the narrow range of 9.9-10.25 Debye as reported in Table 2.

Having computed Δ^{LR} and μ_{01} in TD-DFT, we are now in a position to use Eq. 3 and estimate the linear-response CASPT2 contributions missing from the corresponding polSS calculations. As reported in Table 2, the CASPT2 transition dipole moments are within 8% of the corresponding TD-DFT values, so the estimated polLR terms in CASPT2 are quite similar to the TD-DFT counterparts. Consequently, the linear-response contribution leads to a uniform red shift also in CASPT2 and preserves the correct tuning of the state-specific (polSS) calculations. The $\Delta E_{01}^{\mathrm{pol(SS+LR)}}$ CASPT2 energies display a considerably improved agreement with the experimental absorption maxima (see left panel of Fig. 2).

To verify the consistency of our approach, we plot in Fig. 3 the state-specific CASPT2 contribution to the excitation energies (Δ^{SS}) as a function of polGS $\Delta\mu$, and the linear-response TD-DFT term (Δ^{LR}) as a function of μ_{01}. We find that the data are well described by the relations of Eqs. 1 and 2 using $g_e^d = 0.0020$ and $g_e = 0.0025$ eV/Debye^2, pointing to the conclusion that the effective electronic response of the protein matrix is rather uniform across the protein set.

We further evaluate the robustness of our results by extending these calculations to the same protein structures but with only the chromophore in the QM subsystem. The results are reported in Figure S3 and Tables S1 and S2 in the ESI. With respect to the models including the π-stacked residue in the QM subsystem, the CASPT2 excitation energies are blue shifted from a minimum of 0.04 eV in mTFP0.7 to a maximum of 0.07 eV in EYFP, so the *tuning* changes by less than 0.03 eV. These shifts reveal the kind and the magnitude of effects beyond the coupling described within the MMpol approach (for recent studies of the π-stacking with the GFP chromophore see Ref.[53,54]).

For one of the residues closest to the chromophore, Tyr145 in GFP-S65T and EYFP, we also explore the effect of different orientations of the hydroxyl hydrogen atom. Indeed, in all but a couple of the available X-ray structures of GFP variants, the position of hydrogen atoms and the protonation state of the titratable

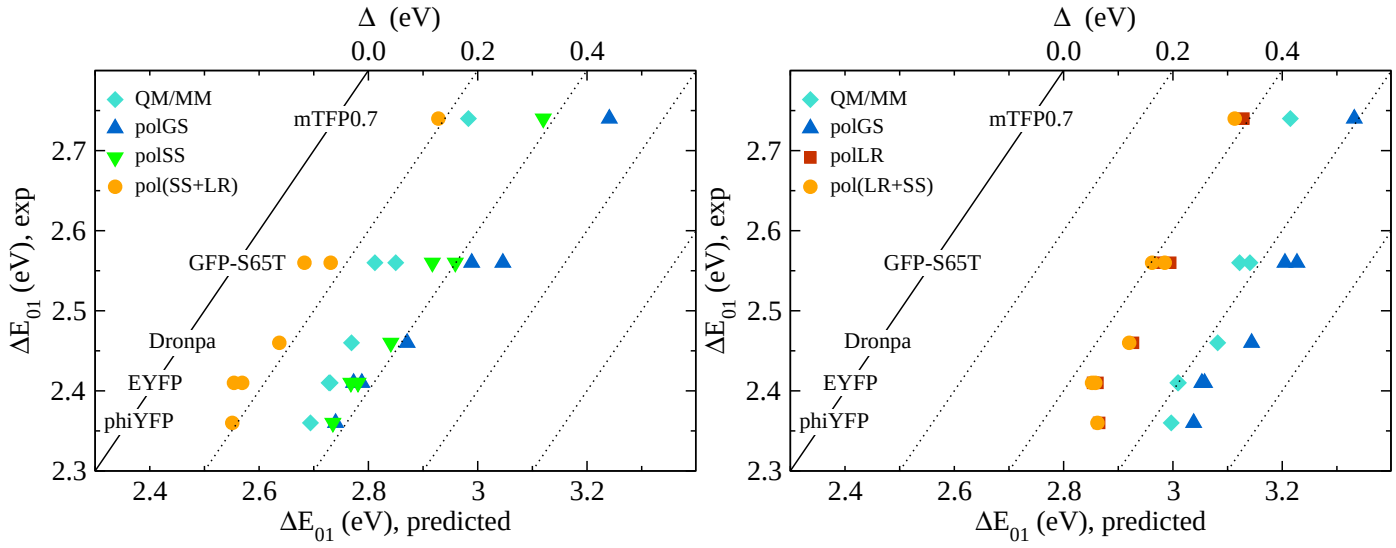

Fig. 2 CASPT2 (left) and TD-DFT (right) excitation energies computed within QM/MM and QM/MMpol compared with the experimental absorption maxima of the five proteins. For EYFP and GFP-S65T, we show both the $OH^{145}\uparrow$ and $OH^{145}\downarrow$ results, the latter being the blue-shifted ones. Dotted lines are placed at 0.2 eV intervals to help visualize the deviation (Δ top axis) from the experimental values in the continuous line.

Table 2 QM/MM and QM/MMpol excitation energies (eV) and transition dipole moments (Debye).

	CASPT2							TD-DFT	
Protein	$\Delta E_{01}^{QM/MM}$	ΔE_{01}^{polGS}	ΔE_{01}^{polSS}	Δ^{SS}	μ_{01}^{polGS}	Δ^{LR} ‡	$\Delta E_{01}^{pol(SS+LR)}$	Δ^{LR}	μ_{01}^{polLR}
phiYFP	2.694	2.740	2.735	-0.005	10.480	-0.184	2.551	-0.174	10.206
EYFP $OH^{145}\uparrow$	2.730	2.773	2.768	-0.006	10.541	-0.213	2.554	-0.199	10.179
EYFP $OH^{145}\downarrow$	2.728	2.788	2.781	-0.007	10.470	-0.212	2.569	-0.197	10.091
Dronpa (2z1o)	2.769	2.871	2.841	-0.030	10.154	-0.204	2.637	-0.217	10.467
GFP-S65T $OH^{145}\uparrow$	2.812	2.989	2.917	-0.072	10.450	-0.234	2.683	-0.236	10.503
GFP-S65T $OH^{145}\downarrow$	2.850	3.046	2.959	-0.088	10.297	-0.227	2.731	-0.232	10.405
mTFP0.7	2.983	3.241	3.120	-0.121	9.653	-0.193	2.928	-0.203	9.915

‡Estimate based on Eq. 3.

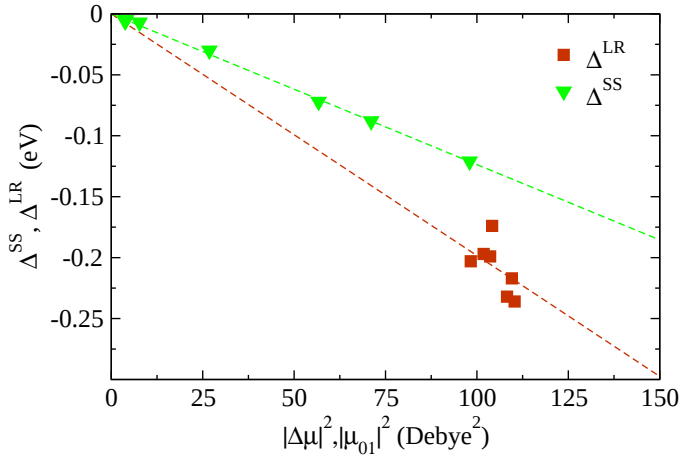

Fig. 3 Green triangles: Δ^{SS} (see Eq. 1) as a function of $\Delta\mu^{polGS}$ computed with CASSCF (values are listed in Tab. S5). Red squares: Δ^{LR} as a function of μ_{01}^{polLR} in TD-DFT (see Eq. 2). Dashed lines are linear homogeneous fits.

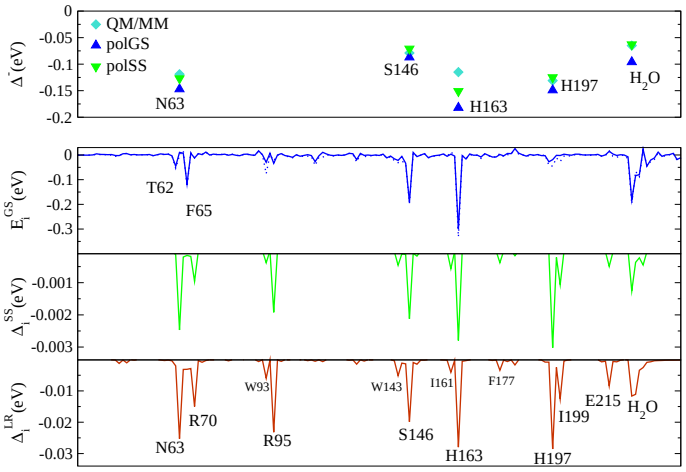

Fig. 4 Analysis of the contributions coming from various residues in mTFP0.7. Top panel: CASPT2 excitation-energy shifts in models lacking the indicated residue. For each scheme (non-polarisable QM/MM, polGS, and polSS), the shift with respect to the corresponding complete model is reported. Bottom panels: TD-DFT *per* residue contributions to the ground-state (GS) polarisation energy and to Δ^{SS} and Δ^{LR}. Calculations with only the chromophore as QM subsystem are reported.

residues must be inferred from the structure. We find that the OH flip has a minor influence on EYFP excitation energy, while the effect is more relevant for GFP-S65T, in line with the higher sensitivity to environment fluctuations due to the larger $\Delta\mu$ [4] (see below).

A sub-atomic X-ray structure of a GFP mutant with S65T was published during the writing of this paper, showing that Tyr145 OH points away from the chromophore [41], that a minor alternative configuration exists for Glu222 (*syn* and *anti* configurations of the hydroxyl group), and that His148 shares one of its protons with a close by residue. In addition, the chromophore bond lengths show some slight but relevant deviations from those of our computational models. Accounting for these structural features might improve the prediction for GFP-S65T, currently too red shifted with respect to the other proteins in the set.

Regarding the influence of the orientation of hydrogen atoms and water molecules outside the chromophore pocket, we observe that the results for the two models of Dronpa, constructed starting from two different X-ray structures, are very similar (Figs. S3 and S4 and Tables S1-S6) . This confirms that the important features lay in the vicinity of the chromophore and also proves the robustness of our approach in terms of small differences in the model structures.

As a final comment on the comparison with experiments, we note that our results on the vertical excitation energies are fully able to explain the broad variation of absorption spectra peaks (Fig. 1). Finer details regarding the shape of the spectra can be addressed by including vibronic effects as well as thermal motion. A limited number of absorption spectra of FPs have been recorded at low temperature, allowing the vibronic progression to be singled out [32,55]: with respect to the wavelength absorption peak at room temperature, only a small 5nm red shift is measured in EYFP and GFP-S65T (see Tab. 1).

Analysis of *per* residue contributions: the case of mTFP0.7

The most blue-shifted excitation energy of our FPs (i.e. that of mTFP0.7) stands out also for being well separated from the next most blue-shifted value in the set, which is for GFP-S65T. To identify the molecular factors giving rise to this behaviour, we perform CASPT2 calculations in mTFP0.7 models lacking specific residues interacting with the chromophore phenolate (the Arginine H-bonded to the imidazolidinone carbonyl is a residue present in all our FPs and as such is not investigated).

The CASPT2 excitation-energy shifts with respect to the complete model are reported in the top panel of Fig. 4 for the QM/MM and the polGS and polSS schemes. The bottom panel reports the *per*-residue decomposition of the ground state polarisation energy E^{polGS}, and of Δ^{LR} and Δ^{SS} obtained from TD-DFT MMpol calculations (for formal definitions, see Eqs. S1-S3 in the ESI). All deficient models have red-shifted excitation energies, as expected since we are removing residues H-bonded to the phenolate which would stabilise the charge on this part of the chromophore and, therefore, the ground state. In line with the previous analysis for all FPs, the polGS shifts are overestimated with respect to the QM/MM and polSS values which are usually in good agreement except for the His163 case, where the removed residue forms a tight H-bond with the chromophore phenolate oxygen (Ser146 and one water molecule are also H-bonded to the phenolate but with larger donor-acceptor distances, see Fig. 1). The importance of this residue is also apparent in the TD-DFT polarisation contributions (Fig. 4, bottom panels), suggesting that the improvement of the MMpol scheme mostly arises from a more accurate description of the effects of strongly interacting residues.

$\Delta\mu$: comparison between theory and experiment

Table 3 lists the CASPT2 values for the difference between the ground- and excited-state dipole moments ($\Delta\mu$) obtained in polGS and polSS by finite field perturbation theory, and compare them with the available experimental estimates from one- and two-photon absorption cross section measurements [4] and from Stark spectroscopy [55].

Table 3 Theoretical (CASPT2) versus experimental $\Delta\mu$ (Debye).

Protein	$\Delta\mu^{\text{polGS}}$	$\Delta\mu^{\text{polSS}}$	$\Delta\mu^{\text{MM}\,b}$	$\lvert\Delta\vec{\mu}^{\text{polSS}}_{\text{FFPT}}+\Delta\vec{\mu}^{\text{MM}}\rvert$	Exp.
phiYFP	2.16	2.49	0.43	2.06	
EYFP OH$^{145}\uparrow$	2.00	2.39	0.48	1.93	$(1.52/f_w)$=1.21^c
EYFP OH$^{145}\downarrow$	2.43	2.75	0.65	2.11	
Dronpa (2z1o)	4.38	4.70	1.27	3.49	
GFP-S65T OH$^{145}\uparrow$	4.85	5.30	1.70	3.60	$(4.5/f_w)$=3.6^d, $(7.0/f)$ = 3.9-5.0^e
GFP-S65T OH$^{145}\downarrow$	5.39	6.12	1.95	4.17	
mTFP0.7	6.15	7.82	2.38	5.54	$(6.87/f_w)$=5.45^c

^{b}Dipole contribution of the MM part calculated as the vector sum of the $\Delta\mu$ on each MM site.
^{c}From optical measurements[4]. f_w=1.26 is the Lorentz local field factor for water.
^{d}Value for EGFP[4].
^{e}From Stark spectroscopy[55]. According to Ref.[56], values of the local field factor f in the chromophore cavity appropriate for Stark spectroscopy are in the range 1.4-1.8.

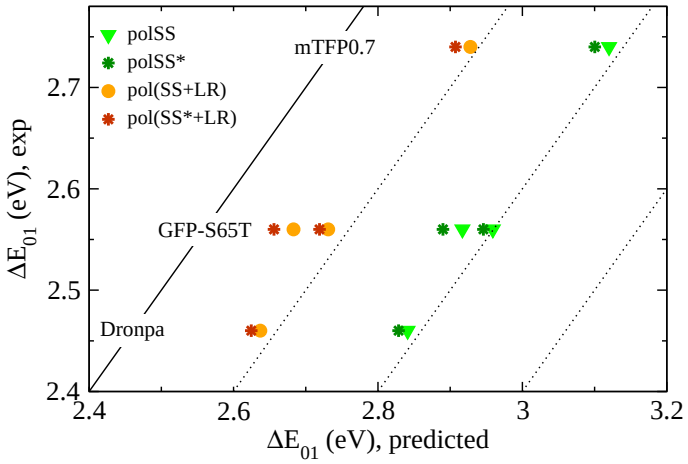

Fig. 5 CASPT2 polSS excitation energy with and without the correction for overpolarisation.

Starting from the state-specific FFPT dipole differences, we can further improve the estimate of $\Delta\mu$ by including the contribution of the MM induced dipole differences between the two states ($\Delta\mu^{\text{MM}}$, Table 3). This operation gives the *total* (i.e. chromophore plus protein matrix) dipole-moment variation. This $\Delta\mu$ is the appropriate quantity for comparison with the optical measurements since, in deriving the experimental values, the usual local-field correction (f_w in Table 3) is introduced to account for the polarisation of the water medium[4] and model the local field in the water cavity occupied by the whole protein. A detailed discussion of the local-field effects in the context of QM/MM polarisable models can be found in Ref.[57].

From the analysis of the dipole moment difference, we note that the CASSCF $\Delta\mu$ values (reported in Tables S5 and S6) are considerably larger than the more accurate CASPT2 ones, in particular for the proteins with a relatively large $\Delta\mu$ (i.e. Dronpa, GFP-S65T, and mTFP0.7). Thereby, the use of state-specific dipoles polarised to the CASSCF ground- and excited-state charge densities challenges the self-consistency of the polSS scheme since the perturbation correction is included at CASPT2 level. In other words, the MM induced dipoles in these models are overestimated due to the overpolarisation of the CASSCF charge densities.

To correct for such an overpolarization, we start from the (non-polarisable) QM/MM excitation energies and make the empirical observation that, at the CASSCF level, the effect of classical polarisation obeys the quadratic relation

$$\Delta E_{01}^{\text{polSS}} - \Delta E_{01}^{\text{QM/MM}} \approx c|\Delta\mu^{\text{polGS}}|^2\,, \tag{4}$$

with a coefficient c of about 0.0033 eV/Debye^2 (see Figure S5). Using this relation and value of c with the corresponding CASPT2 excitation energies and $\Delta\mu^{\text{polGS}}$ values, we obtain the estimated $\Delta E_{01}^{\text{polSS}}$ (red stars) of Fig.5, which shows that the overall correction due the lack of self-consistency in the MMpol scheme is very small and only slightly affects the final results.

Electrostatic field in the protein cavity

Recently, a renewed interest in the measurement of local electrostatic fields in complex molecular environments has led to the development and calibration of different techniques[11,56,58–60]. One such technique[60] is based on the first-order Stark effect and relates changes in excitation energy ($\Delta\Delta E_{01}$) in different GFP mutants to changes in the electrostatic field $\Delta\vec{F}$ as

$$\Delta\Delta E_{01} = -\Delta\vec{\mu}\cdot\Delta\vec{F}\,. \tag{5}$$

Given the large variation of $\Delta\mu$ in different proteins, the application of this formula is however restricted to limited energy shifts.

Another route to estimate the molecular field in the GFP cavity was worked out in Ref.[4] and is based on the value of the (measured) $\Delta\mu$ in different proteins. In a *static* field $\vec{F}$ and in the linear regime, one has

$$\Delta\vec{\mu}(\vec{F}) = \Delta\vec{\mu}^{vac} + \Delta\alpha^{vac}\cdot\vec{F}\,, \tag{6}$$

where $\Delta\alpha = \alpha_1 - \alpha_0$ is the variation of the polarisability tensor upon excitation and the *vac* superscript identifies quantities *in vacuum* (one can use a different reference system, in which case the field variation from the reference is measured). Hence, approximating the problem as unidimensional, an effective field can be estimated by inverting this equation and using for $\Delta\alpha^{vac}$ and $\Delta\mu^{vac}$ values from experiment or from theory.

We can connect these measurements with our calculations by evaluating the field generated by the MM partial charges and in-

duced dipoles (polGS scheme) in the protein cavity embedding the chromophore. As shown in the top panel of Fig. 6, the field (blue arrows) is far from homogeneous, even when considering only the component of the field parallel to $\Delta\mu$ ($F_{||}$), thereby challenging the nature of the effective field measured by the experiments. However, if we simply take the average of $F_{||}$ on the chromophore atoms (more precisely the two-ring chromophore atoms excluding the carbonyl groups and hydrogen atoms), we obtain a clear correlation between $F_{||}^{\mathrm{ave}}$ and both the excitation energies ΔE_{01} and the $\Delta\mu$'s for our protein set, as shown in the bottom panel of Fig. 6. While ΔE_{01} versus $F_{||}^{\mathrm{ave}}$ displays an apparent deviation from linearity, the linearity of $\Delta\mu$ versus $F_{||}^{\mathrm{ave}}$ is approximately maintained through the whole protein set. There are however some deviations from this linear relation (e.g. for EYFP), limiting the general accuracy by which the fields can be estimated via Eqs. 5 or 6 and which are ultimately due to inherent approximations such as the assumption of isotropic response of $\Delta\mu$, the neglect of chromophore structure changes, and the non-uniformity of the field in the cavity.[4]

The absolute magnitude of the calculated fields cannot be directly compared with the experiment because our QM subsystems contain the peptide bonds connecting the chromophore to the rest of the protein, which should instead contribute to the estimated electrostatic field (in other words, the definition of chromophore cavity is ambiguous because the chromophore is covalently linked to the protein). On the other hand, the field variations between the two proteins are suitable for comparison. In our models, the average $F_{||}^{\mathrm{ave}}$ drops by -30 MV/cm in going from mTFP0.7 to phiYFP, and by -20 MV/cm from mTFP0.7 to EYFP. This compares quite favourably with the measured variation (using Eq. 6) of -34 MV/cm from mTFP0.7 to Citrine[4], a GFP mutant with T203Y like EYFP and phiYFP. In the measurements by Slocum *et al.*[60] based on the linear Stark effect (Eq. 5), the field variation is -8 MV/cm when the T203Y mutation is introduced in superfolder GFP, a mutant similar to GFP-S65T. The excitation energy variation is less than in the mTFP0.7 to EYFP comparison, so that the linear regime can still be considered adequate. The variation in our models from GFP-S65T to EYFP is -13 MV/cm in reasonable agreement with this experimental value. Overall, the comparison between theory and experiments confirms the magnitude of the field variations needed to tune the chromophore excitation energy.

There is however a subtlety related to the fact that the environment polarisation is state specific so that, as a consequence of the sizeable change of chromophore dipole moment, also the reaction field in the cavity will change upon excitation. In our MMPol approach, we are able to evaluate the magnitude of this change and its variation in our protein set (green arrows in Fig. 6). As shown in the ESI (Eq. S11), we can incorporate this effect in Eq. 6 and find it to only marginally change the field estimation in GFP-like systems, also in view of the other involved approximations, while it needs to be considered in systems with large $\Delta\mu$ values.

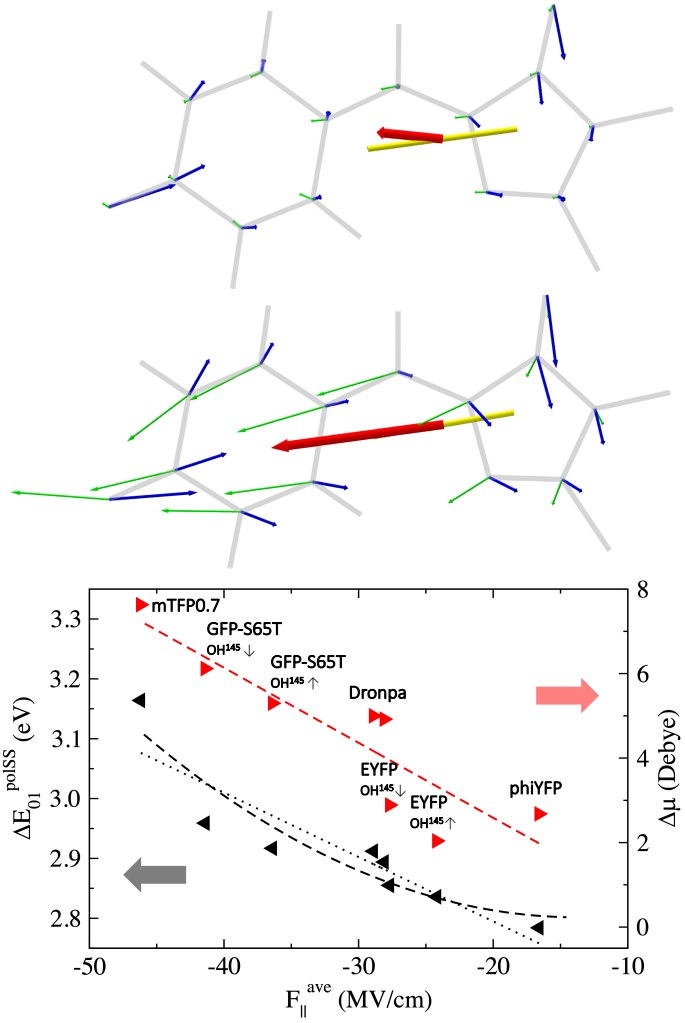

Fig. 6 Top panel: Electrostatic field on the chromophore atoms for phiYFP (top) and mTFP0.7 (bottom) computed within polGS ($\vec{F}^{\mathrm{polGS}}$, blue arrows). The variation of polarisation field upon excitation (i.e. the difference between the ground and excited states in polSS) is shown as green arrows and scaled up ten times for ease of visualisation. Red arrows and yellow cylinder are respectively $\Delta\vec{\mu}^{\mathrm{polGS}}$ (CASPT2 FFPT) and $\vec{\mu}_{01}$. For the complete protein set, see Figure S8. Bottom panel: CASPT2 polSS excitation energy (black triangles, left y axis) and $\Delta\mu^{\mathrm{polSS}}$ (FFPT) (red triangles, right y axis) as a function of $F_{||}^{\mathrm{ave}}$. Lines are linear and quadratic fits. Calculations with only the chromophore in the QM subsystem are reported.

4 Discussion and conclusions

Some of the determinants of spectral tuning in fluorescent proteins can be inferred from analysis of the structures[2,10]. For ex-

ample, the number and the strength of the hydrogen bonds to the chromophore phenolate clearly correlate with blue shifts of the excitation energy (see Fig. 1), as does the bond-length alternation (BLA) in the chromophore structure [17,23,61–63]. In a very simplified electrostatic picture, what will change more in the different proteins is the ground state of the chromophore, the excited state being less sensitive due to its lower polarisability [4]. The blue shift of the excitation energy going from yellow to cyan FPs is in fact mostly due to larger molecular fields in the chromophore cavity stabilising the ground state and, concurrently, increasing its dipole moment according to its ground-state polarisability. Since the dipole moment of the excited state is less responsive to the molecular field, the dipole moment change upon excitation, $\Delta\mu$, will be larger in module for the more blue-shifted proteins. Thereby, the molecular field will not only blue shift the excitation energy but also increase the dipole-moment change, $\Delta\mu$, as experimentally verified by Drobizhev *et al.* [4].

Despite this rather well understood structure-function relationship, predicting the protein tuning by computational approaches has proven less straightforward [17,19]. Here, we have shown that the accuracy in this prediction can indeed be achieved but only if the mutual polarisation between the chromophore and the surrounding protein matrix is properly included in the model. In fact, the polarisation accounts for two effects that have been so far either neglected or only partially included in multiscale calculations of the FP tuning, namely the electrostatic induction and a purely quantum effect due to the coupling of the classical but polarisable atomic sites with the transition density of the QM subsystem [22,24,26].

What contributes most in adjusting the tuning is the full account of the induction effect. It is certainly true that the electrostatic fields acting in the protein cavity affect the chromophore through changes in both its structure and electronic density, in such a way that the latter is significantly different among the various protein structures. However, it is only by adding the mutual polarisation effects that we can achieve the large excitation energy shifts measured in the FP family. This further tuning is missing in non-polarisable QM/MM models and this is the reason why, in these approaches, the energy variations are too limited.

Actually, if one considers only the polarisation in the ground state (polGS), this mutual polarisation is even too pronounced, resulting in the "overtuning" of the excitation energies, meaning that, with respect to the red-shifted proteins, the blue-shifted ones are too much so. The appropriate amount of tuning is achieved only if the state-specific polarisation (polSS) is properly accounted for. The electronic density change is well described by the variation of the permanent dipole moment upon excitation, $\Delta\mu$, which spans a broad range in our set of proteins, increasing markedly from the red-shifted to the blue-shifted ones. Upon excitation, the environment polarisation will instantaneously change (we are ignoring retardation effects that in the length scales of these systems are negligible), and such change will be larger when the $\Delta\mu$ is larger. As a consequence, the polSS correction is larger for the blue-shifted proteins and this corrects the overtuning.

The last polarisation contribution, polLR, is less intuitive, having no classical analogue. For our systems, the polLR correction corresponds to a sizeable $\sim$0.2 eV red shift. This term is proportional to the square of the transition dipole moment, μ_{01}, which displays a more moderate variation than $\Delta\mu$, both among the different proteins and the different intersystem coupling schemes (non-polarisable QM/MM, polGS and polSS). Consequently, the polLR correction systematically red shifts the predicted excitation energies of all FPs, which are are too high with respect to the experiment when only the induction contribution is included. The transition dipole moment is slightly smaller in the blue-shifted proteins (Tab. 2), a fact that can be correlated with the larger charge dislocation upon excitation. Therefore, though to a minor extent than the induction effect, also the polLR can contribute to the tuning.

We have also found that, for a polarisable model to be very quantitative, a proper QM method has to be selected. In fact, TD-DFT also with the inclusion of both polLR and polSS contributions reproduces the tuning only partially, at least for the exchange and correlation functionals tried in this work (CAM-B3LYP and LC-BLYP). The TD-DFT range of variations of $\Delta\mu$ among the FPs is in fact too narrow (Table S4), pointing to a response to the changing molecular field that is too limited. In contrast, CASSCF predicts a marked variation of the change of the dipole moment upon excitation both among the proteins in the set and among the QM/MM and the polGS/polSS schemes. Whereas the $\Delta\mu$'s predicted by CASSCF are actually larger than the experimental estimates, we achieve a very good agreement with experiments by including both the electron correlation in CASPT2 and the contribution of protein polarisation change. Finally, we analysed the fields in the chromophore cavity also in relation with experimental estimates, broadly confirming the field variations responsible for the tuning, but also pinpointing the limitations of such field estimations based on optical properties.

In conclusion, combining an improved description of the coupling between chromophore and environment with an accurate QM treatment of the ground and excited states has allowed us to clarify how small differences in protein structure and composition are sufficient to generate the observed large spectral changes thanks to an enhancement effect due to mutual polarisation. This is a different perspective on the chromophore-protein interaction that can open the way to new design tools for the control of the spectral properties of light-active proteins.

5 Acknowledgments

We thank Mikhail Drobizhev and Celestino Angeli for fruitful discussion. B.M. acknowledges funding by the European Research Council, under the grant ERC-AdG-786714 (LIFETimeS).

Notes and references

1 R. Y. Tsien, *Annual review of biochemistry*, 1998, **67**, 509–44.
2 A. A. Pakhomov and V. I. Martynov, *Chemistry & Biology*, 2008, **15**, 755–764.
3 R. Nifosì and V. Tozzini, *Fluorescent Proteins I*, 2012, 3–40.
4 M. Drobizhev, P. R. Callis, R. Nifosì, G. Wicks, C. Stoltzfus, L. Barnett, T. E. Hughes, P. Sullivan and A. Rebane, *Scientific*

Reports, 2015, **5**, 13223.
5 T. A. Weissman and Y. A. Pan, *Genetics*, 2015, **199**, 293–306.
6 J. N. Henderson, H. Ai, R. E. Campell and S. J. Remington, *PNAS*, 2007, **104**, 6672–6677.
7 N. V. Pletneva, V. Z. Pletnev, E. Souslova, D. M. Chudakov, S. Lukyanov, V. I. Martynov, S. Arhipova, I. Artemyev, A. Wlodawer, Z. Dauter and S. Pletnev, *Acta Crystallographica Section D*, 2013, **69**, 1005–1012.
8 M. Drobizhev, S. Tillo, N. S. Makarov, T. E. Hughes and A. Rebane, *J Phys Chem B*, 2009, **113**, 12860–12864.
9 N. H. List, J. M. H. Olsen, H. J. A. Jensen, A. H. Steindal and J. Kongsted, *Journal of Physical Chemistry Letters*, 2012, **3**, 3513–3521.
10 A. Hense, K. Nienhaus and G. U. Nienhaus, *Photochem. Photobiol. Sci.*, 2015, **14**, 200–212.
11 A. Rebane, G. Wicks, M. Drobizhev, T. Cooper, A. Trummal and M. Uudsemaa, *Angewandte Chemie International Edition*, 2015, **54**, 7582–7586.
12 A. Sinicropi, T. Andruniow, N. Ferré, R. Basosi and M. Olivucci, *Journal of the American Chemical Society*, 2005, **127**, 11534–11535.
13 S. S. Patnaik, S. Trohalaki, R. R. Naik, M. O. Stone and R. Pachter, *Biopolymers: Original Research on Biomolecules*, 2007, **85**, 253–263.
14 J.-Y. Hasegawa, K. Fujimoto, B. Swerts, T. Miyahara and H. Nakatsuji, *Journal of Computational Chemistry*, 2007, **28**, 2443–2452.
15 C. Filippi, F. Buda, L. Guidoni and A. Sinicropi, *Journal of Chemical Theory and Computation*, 2011, **8**, 112–124.
16 K. B. Bravaya, B. L. Grigorenko, A. V. Nemukhin and A. I. Krylov, *Accounts of chemical research*, 2012, **45**, 265–275.
17 P. Amat and R. Nifosì, *Journal of Chemical Theory and Computation*, 2013, **9**, 497–508.
18 V. R. Kaila, R. Send and D. Sundholm, *Physical Chemistry Chemical Physics*, 2013, **15**, 4491–4495.
19 M. T. P. Beerepoot, A. H. Steindal, J. Kongsted, B. O. Brandsdal, L. Frediani, K. Ruud and J. M. H. Olsen, *Physical Chemistry Chemical Physics*, 2013, **15**, 4735.
20 J. W. Park and Y. M. Rhee, *Physical Chemistry Chemical Physics*, 2016, **18**, 3944–3955.
21 L. J. Nåbo, J. M. H. Olsen, T. J. Martínez and J. Kongsted, *Journal of Chemical Theory and Computation*, 2017, **13**, 6230–6236.
22 S. Corni, R. Cammi, B. Mennucci and J. Tomasi, *Journal of Chemical Physics*, 2005, **123**, year.
23 C. Daday, C. Curutchet, A. Sinicropi, B. Mennucci and C. Filippi, *Journal of Chemical Theory and Computation*, 2015, 150916081745000.
24 L. Cupellini, C. Amovilli and B. Mennucci, *The Journal of Physical Chemistry B*, 2015, **119**, 8984–8991.
25 R. Guareschi, O. Valsson, C. Curutchet, B. Mennucci and C. Filippi, *Journal of Physical Chemistry Letters*, 2016, **7**, 4547–4553.
26 T. Schwabe, *Journal of Chemical Physics*, 2016, **145**, year.
27 C. Curutchet, A. MuñozLosa, S. Monti, J. Kongsted, G. D. Scholes and B. Mennucci, *Journal of Chemical Theory and Computation*, 2009, **5**, 1838–1848.
28 R. M. Wachter, M. A. Elsliger, K. Kallio, G. T. Hanson and S. J. Remington, *Structure*, 1998, **6**, 1267–1277.
29 A. C. Stiel, S. Trowitzsch, G. Weber, M. Andresen, C. Eggeling, S. W. Hell, S. Jakobs and M. C. Wahl, *Biochem J*, 2007, **402**, 35–42.
30 H. Mizuno, T. K. Mal, M. Wälchli, A. Kikuchi, T. Fukano, R. Ando, J. Jeyakanthan, J. Taka, Y. Shiro, M. Ikura and A. Miyawaki, *Proc Natl Acad Sci U S A*, 2008, **105**, 9227–32.
31 M. A. Elsliger, R. M. Wachter, G. T. Hanson, K. Kallio and S. J. Remington, *Biochemistry*, 1999, **38**, 5296–5301.
32 T. Creemers, A. Lock, V. S. T. Jovin and S. Völker, *Proc Nat Acad Sci USA*, 2000, **97**, 2974–2978.
33 W. Humphrey, A. Dalke and K. Schulten, *J Mol Graph*, 1996, **14**, 33–38.
34 M. A. Thompson, *The Journal of Physical Chemistry*, 1996, **100**, 14492–14507.
35 M. Caricato, B. Mennucci, J. Tomasi, F. Ingrosso, R. Cammi, S. Corni and G. Scalmani, *The Journal of chemical physics*, 2006, **124**, 124520.
36 H. Schr oder and T. Schwabe, *Journal of Chemical Theory and Computation*, 2018, **14**, 833–842.
37 D. Jacquemin, *Journal of Chemical Theory and Computation*, 2016, **12**, 3993–4003.
38 C. Duan, V. Adam, M. Byrdin, J. Ridard, S. Kieffer-Jaquinod, C. Morlot, D. Arcizet, I. Demachy and D. Bourgeois, *Journal of the American Chemical Society*, 2013, **135**, 15841–15850.
39 R. Battistutta, A. Negro and G. Zanotti, *Proteins: Structure, Function and Genetics*, 2000, **41**, 429–437.
40 B. Barstow, N. Ando, U. K. Chae and S. M. Gruner, *Biophysical Journal*, 2009, **97**, 1719–1727.
41 K. Takaba, Y. Tai, H. Eki, H.-A. Dao, Y. Hanazono, K. Hasegawa, K. Miki and K. Takeda, *IUCrJ*, 2019, **6**, 1–14.
42 D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. Berendsen, *Journal of Computational Chemistry*, 2005, **26**, 1701–1718.
43 M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl, *SoftwareX*, 2015, **1**, 19–25.
44 M. J. Frisch and et al., *Gaussian 09, Revision A.02*, Gaussian, Inc., Wallingford, CT, 2009.
45 T. Yanai, D. P. Tew and N. C. Handy, *Chemical Physics Letters*, 2004, **393**, 51–57.
46 H. Iikura, T. Tsuneda, T. Yanai and K. Hirao, *The Journal of Chemical Physics*, 2001, **115**, 3540–3544.
47 P.-O. Widmark, P.-Å. Malmqvist and B. O. Roos, *Theoretica chimica acta*, 1990, **77**, 291–306.
48 F. Aquilante, J. Autschbach, R. K. Carlson, L. F. Chibotaru, M. G. Delcey, L. De Vico, I. Fdez. Galván, N. Ferré, L. M. Frutos, L. Gagliardi *et al.*, *Journal of Computational Chemistry*, 2016, **37**, 506–541.
49 D. Tzeli and A. Mavridis, *The Journal of Chemical Physics*, 2003, **118**, 4984–4986.

50 J. Wang, P. Cieplak and P. A. Kollman, *Journal of Computational Chemistry*, 2000, **21**, 1049–1074.
51 J. Wang, P. Cieplak, J. Li, J. Wang, Q. Cai, M. Hsieh, H. Lei, R. Luo and Y. Duan, *The Journal of Physical Chemistry B*, 2011, **115**, 3100–3111.
52 C. Curutchet, V. I. Novoderezhkin, J. Kongsted, A. Muñoz-Losa, R. van Grondelle, G. D. Scholes and B. Mennucci, *The Journal of Physical Chemistry B*, 2013, **117**, 4263–4273.
53 E. Jacchetti, E. Gabellieri, P. Cioni, R. Bizzarri and R. Nifosì, *Physical Chemistry Chemical Physics*, 2016, **18**, 12828–12838.
54 M. G. Khrenova, A. V. Nemukhin and V. G. Tsirelson, *Chemical Physics*, 2019, **522**, 32–38.
55 G. Bublitz, B. A. King and S. G. Boxer, 1998, **8**, 9370–9371.
56 S. D. Fried, S. Bagchi and S. G. Boxer, *Journal of the American Chemical Society*, 2013, **135**, 11181–11192.
57 N. H. List, H. J. A. Jensen and J. Kongsted, *Phys. Chem. Chem. Phys.*, 2016, **18**, 10070–10080.
58 J. M. Kriegl, K. Nienhaus, P. Deng, J. Fuchs and G. U. Nienhaus, *Proceedings of the National Academy of Sciences*, 2003, **100**, 7069–7074.
59 T. Nakabayashi, K. Hino, Y. Ohta, S. Ito, H. Nakano and N. Ohta, *The Journal of Physical Chemistry B*, 2011, **115**, 8622–8626.
60 J. D. Slocum and L. J. Webb, *Journal of the American Chemical Society*, 2016, **138**, 6561–6570.
61 A. F. Bell, X. He, R. M. Wachter and P. J. Tonge, *Biochemistry*, 2000, **39**, 4423–4431.
62 T. Laino, R. Nifosì and V. Tozzini, *Chem Phys*, 2004, **298**, 17–28.
63 S. Olsen and R. H. McKenzie, *Journal of Chemical Physics*, 2011, **134**, year.