Paper I

A polarizable embedding DFT study of one-photon absorption in fluorescent proteins

M. T. P. Beerepoot, A. H. Steindal, J. Kongsted, B. O. Brandsdal,
L. Frediani, K. Ruud and J. M. H. Olsen
Phys. Chem. Chem. Phys. **15** (2013), 4735–4743.

PCCP

RSC Publishing

PAPER

View Article Online
View Journal | View Issue

A polarizable embedding DFT study of one-photon absorption in fluorescent proteins†

Cite this: *Phys. Chem. Chem. Phys.*, 2013, **15**, 4735

Maarten T. P. Beerepoot,‡[a] Arnfinn Hykkerud Steindal,[a] Jacob Kongsted,[b] Bjørn Olav Brandsdal,[a] Luca Frediani,[a] Kenneth Ruud*[a] and Jógvan Magnus Haugaard Olsen[b]

Received 23rd December 2012,
Accepted 29th January 2013

DOI: 10.1039/c3cp44659j

www.rsc.org/pccp

A theoretical study of the one-photon absorption of five fluorescent proteins (FPs) is presented. The absorption properties are calculated using a polarizable embedding approach combined with density functional theory (PE-DFT) on the wild-type green fluorescent protein (wtGFP) and several of its mutants (BFP, eGFP, YFP and eCFP). The observed trends in excitation energies among the FPs are reproduced by our approach when performing calculations directly on the crystal structures or when using structures extracted from molecular dynamics simulations. However, in the former case, QM/MM geometry optimization of the chromophores within a frozen protein environment is needed in order to reproduce the experimental trends. An explicit account of polarization in the force field is not needed to yield the correct trend between the different FPs, but it is necessary for reproducing the experimentally observed red shift from vacuum to protein. This is the first computational study of a range of fluorescent proteins using a polarizable embedding potential.

1 Introduction

Intrinsically fluorescent proteins (FPs) are widely used tools to visualize cellular processes in living organisms. The well-known green fluorescent protein (GFP)[1,2] is frequently used as a model system for FPs. GFP was first isolated from the jellyfish *Aequorea victoria*, though many homologues have been found in other organisms. *Aequorea* contains blue-emitting luciferin which is a cofactor in the aequorin protein. The emitted blue light is subsequently absorbed by GFP which then emits green light through fluorescence.[1] The chromophore in GFP is the result of autocatalytic cyclization of three amino acid residues (Ser65-Tyr66-Gly67)[3] and is located within a protective β-barrel formed by the rest of the protein.[4] The absorption spectrum of the protein shows a major peak at 395 nm (3.14 eV) and a minor peak at 475 nm (2.61 eV).[5] These two peaks can be attributed to the neutral and anionic forms of the chromophore, respectively.[2] Excitation leads to an emission maximum at 508 nm (2.44 eV) (for the excitation at 395 nm) and 503 nm (2.46 eV) (for the excitation at 475 nm).[5]

The DNA sequence of the GFP gene has been successfully incorporated and brought to expression in other organisms with the use of recombinant DNA techniques.[6] Hence, the main applications of GFP in the biosciences are as a fusion tag (attached to another protein) to monitor protein localization and dynamics and as a reporter gene to study the transcription of a certain promoter.[1] The success of GFP can primarily be attributed to its ability to fold *in vivo* without the need for anything but oxygen[6] and to the observation that fusion with another protein does not affect the properties of the other protein significantly.[2]

Significant efforts have been devoted to the design of GFP mutants with advantageous spectroscopic properties, such as a shifted fluorescence emission wavelength, improved protein folding and enhanced linear and non-linear absorption cross sections.[3,5,7–13] Ideally, the biochemist's arsenal of FPs consists of several bright fluorophores that emit light at clearly discernible wavelengths, allowing for multi-colour imaging. Alteration of absorption properties can be achieved by mutations both inside and outside the chromophore. As an example of a modification within the chromophore, mutation of Tyr66 to other aromatic amino acid residues results in a shift of the absorption (and emission) maximum,[3,5] *e.g.* Y66H is the blue

[a] *Centre for Theoretical and Computational Chemistry, Department of Chemistry, University of Tromsø, N-9037 Tromsø, Norway. E-mail: kenneth.ruud@uit.no; Fax: +47 77644765; Tel: +47 77644082*
[b] *Department of Physics, Chemistry and Pharmacy, University of Southern Denmark, DK-5230 Odense M, Denmark*

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c3cp44659j

‡ Previous address: Leiden Institute of Chemistry, Gorlaeus Laboratories, Einsteinweg 55, 2333 CC Leiden, The Netherlands.

fluorescent protein (BFP) with an absorption maximum at 382 nm (3.25 eV)[5,9,11] and Y66W is the cyan fluorescent protein (CFP) with an absorption maximum at 435 nm (2.85 eV).[14] Moreover, mutating Ser65 into threonine suppresses the neutral peak in the absorption spectrum and leads to an enhanced fluorescence.[7,8] Hence, one of the mutants carrying the S65T mutation is called enhanced green fluorescent protein (eGFP).[15] The T203Y mutation in yellow fluorescent protein (YFP), which is an example of a modification outside the chromophore, is believed to cause a red shift of 39 nm (0.20 eV) in the low-energy band of the absorption spectrum through π–π stacking of the aromatic rings in Tyr66 and Tyr203.[4,12,16] Note that all of these mutants also carry additional mutations to improve the folding capacity and restore the fluorescent brightness.[1]

Linear absorption properties of wtGFP and its mutants have been calculated in numerous studies, both on the bare chromophore[17–19] and on the whole protein using combined quantum mechanics/molecular mechanics (QM/MM) approaches.[20–25] In an extensive study using several theoretical approaches on several chromophore models as well as on the whole protein, Filippi *et al.* found that especially the excitation energy of the neutral chromophore is challenging to compute.[18] Furthermore, they concluded in another study that the expected red shift from vacuum to protein is difficult to predict when applying QM/MM with a non-polarizable embedding potential.[24] Steindal *et al.* showed that polarizable embedding density functional theory (PE-DFT) is adequate to describe the wtGFP protein by comparing PE-DFT to full DFT calculations of a model of the chromophore together with close-lying amino acid residues.[23]

The aim of the present study is to explore to what extent theory can reproduce trends in linear absorption properties among different FPs, as a step towards the longer-term goal of predicting the effect of mutations on the linear and non-linear absorption properties of FPs. The absorption properties are calculated using the polarizable embedding (PE) approach combined with density functional theory (PE-DFT), as described in ref. 26 and previously applied to wtGFP in ref. 23. Structures are obtained from the protein data bank (PDB)[27] and from molecular dynamics (MD) simulations on wtGFP and a set of its mutants. The effect of QM/MM geometry optimization of the chromophores is investigated. Results are presented for one-photon absorption of the neutral chromophores of wtGFP, BFP, YFP and eCFP, and the anionic chromophores of wtGFP, eGFP and YFP. The effects of the protein surroundings on the absorption properties are detailed by comparing PE-DFT calculations on the whole protein to the corresponding DFT calculations on the bare chromophore using the vacuum-optimized geometry and the in-protein structure (NE: no embedding) as well as to non-polarizable embedding (NPE) calculations. Specifically, the vacuum results compared to NE give the indirect effect of the protein surroundings through the induced structural changes in the chromophore while NPE adds the direct electrostatic contributions due to the permanent charge distribution of the protein environment and finally PE adds many-body induction effects.

2 Methods

2.1 Protein preparation

The crystal structures presented in Table 1 were obtained from the protein data bank (www.pdb.org).[27] In contrast to previous work[23] we chose the 1GFL structure[28] rather than the 1EMB structure[4] as a model for wtGFP. Detailed comparison of the electron density map and the published structure of the latter showed abnormalities on two points in the chromophore: an unphysically long C–O bond length (1.51 Å) in the phenol group of Tyr66 (as also observed by Reuter *et al.*)[29] and a threonine residue instead of serine at position 65. The electron density map confirmed that the published PDB structure was not correct. No abnormalities were found in the 1GFL structure.

The crystal structures were prepared using the protein preparation wizard of the Schrödinger suite.[30] Hydrogen atoms were added and their positions optimized with the OPLS AA force field.[31,32] Hydrogen networks were optimized by changing the orientation and protonation state of all histidine, asparagine and glutamine residues as well as the protonation states of the other ionizable residues. In the case of wtGFP and YFP, two separate structures with neutral and anionic chromophores were prepared. Glu222 was deprotonated (anionic) in the structure with the neutral chromophore and protonated (neutral) in the structure with the anionic chromophore, in line with experimental evidence.[1] Crystal water molecules were used for the calculations based on the crystal structures. In the case of 1GFL, which shows a dimeric structure,[28] only chain A was used.

Isolated chromophore structures were prepared by extracting the three amino acid residues that make up the chromophore (65–67 for the FPs considered here) including a carbonyl group at the N-terminus and a nitrogen at the C-terminus that come from the previous and next amino acid, respectively. Both sides were capped with an extra hydrogen atom. The chromophores of the FPs under study are shown in Fig. 1.

2.2 MD simulations

The prepared crystal structures were solvated in an orthorhombic box of TIP3P[33] water molecules with a buffer distance of 10 Å between solute atoms and the edges of the simulation box. The MD simulations were carried out using Desmond[34] with the CHARMM27 force field.[35] For the chromophores, the parameters derived by Reuter *et al.*[29] were used. In the case of the BFP and

Table 1 The fluorescent proteins studied in this work with the amino acids that make up the chromophore, the PDB code, the possible charges (*Q*) of the chromophore and the experimental one-photon absorption wavelength in nm (eV)

Name	Chromophore	PDB ID	*Q*	OPA/nm (eV)
BFP	SHG	1BFP[b]	0	382 (3.25)[g]
YFP[a]	GYG	1F0B[c]	0, −1	392 (3.16), 514 (2.41)[c]
wtGFP	SYG	1GFL[d]	0, −1	395 (3.14), 475 (2.61)[g]
eCFP	TWG	2WSN[e]	0	435 (2.85)[h]
eGFP	TYG	2Y0G[f]	−1	488 (2.54)[i]

[a] YFP-H148Q in ref. 16. [b] Ref. 11. [c] Ref. 16. [d] Ref. 28. [e] Ref. 14. [f] Ref. 15. [g] Ref. 5. [h] Ref. 14. [i] Ref. 8.

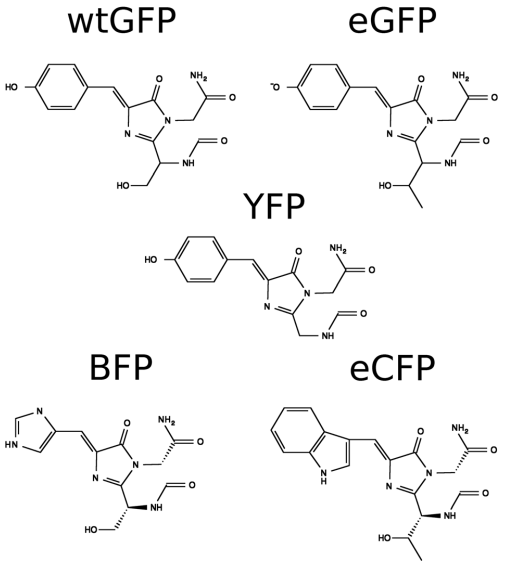


Fig. 1 The chromophores as used in the PE-DFT calculations for the five fluorescent proteins under study. Note that wtGFP and YFP can be neutral or anionic, BFP and eCFP only neutral and eGFP only anionic.[1]

eCFP chromophores, which do not have tyrosine at position 66, the parameters reported by Reuter *et al.* were complemented with CHARMM27 force field parameters for histidine and tryptophan. The simulations were performed within the NPT ensemble at 300 K and 1 atm. The temperature and pressure were controlled using the Nosé–Hoover thermostat method,[36,37] with a relaxation time of 1.0 ps and a frequency update every second step, and the Martyna–Tobias–Klein barostat method,[38] using isotropic coupling with a relaxation time of 2.0 ps and a compressibility of 4.5×10^{-5} bar^{-1}. Non-bonded interactions were treated with a cutoff radius of 9 Å. Electrostatic interactions beyond this threshold were treated with the smooth Particle Mesh Ewald method[39] with a tolerance of 10^{-9}. The equilibration time used was 10 ns followed by 15 ns production using the reference system propagator algorithm (RESPA) time-stepping scheme[40] with time steps set to 2 fs for bonded and near atoms and 6 fs for far atoms. Heavy atom–hydrogen covalent bonds were constrained using SHAKE[41] with a tolerance of 10^{-8} with a maximum of 8 iterations. Coordinates were sampled every 300 ps yielding 50 snapshots for each simulation. Solvent water molecules further away than 8 Å from a C_α were removed before further calculations on the MD structures.

2.3 Geometry optimization

The effect of QM/MM geometry optimization was investigated and compared with results of excitation energy calculations obtained with non-optimized structures. The QSite QM/MM program[42] was used to optimize the chromophore using DFT (B3LYP[43,44]/6-31+G*[45–47]) within a frozen protein environment described by the OPLS AA force field.[31,32] The QM region used in the QM/MM geometry optimization consisted of the aromatic system from residue 66 (phenol for wtGFP) and the imidazolidone ring of the chromophore plus the bridge between them. Hydrogen caps at the boundary were used to separate the QM and MM regions. The positions of nearby protein atoms (from residues 65 and 67) were allowed to move during the optimization to keep the link between the QM and MM regions stable. The geometries of the chromophore structures were also optimized in vacuum (with methyl-terminated side groups of the imidazolidone ring) at the B3LYP/6-31+G* level of theory using the Dalton2011 program.[48]

2.4 Embedding potentials

The classical embedding potentials used in the PE-DFT calculations consisted of distributed atom-centred multipole moments up to and including quadrupoles and anisotropic dipole–dipole polarizabilities. The former describe the permanent charge distribution of the environment while the latter describe many-body induction effects allowing for polarization of the environment. The potentials were derived using the molecular fractionation with conjugate caps (MFCC) procedure[49] as described by Söderhjelm and Ryde.[50] Accordingly, we decomposed the protein into amino acid-based fragments capped with $-COCH_3$ and $-NHCH_3$ groups built from neighbouring residues as well as the corresponding conjugate cap fragments (concaps) that consist of two neighbouring capping groups, *i.e.* $CH_3NHCOCH_3$ molecules. The concaps are used to model the immediate environment of the amino acid fragments and to correct errors due to the fragmentation procedure. Thus, a property P on atom A is calculated as

$$P^{\mathrm{A}} = \sum_{f=1}^{F} P_f^{\mathrm{A}} - \sum_{c=1}^{C} P_c^{\mathrm{A}} \qquad (1)$$

where P_f^{A} is the property on atom A in the fth amino acid fragment and P_c^{A} is the property on atom A in the cth concap. F and C are the number of capped amino acid fragments and concaps, respectively, that contain atom A. The distributed properties of each capped fragment and concap were calculated using the LoProp method[51] available in the Molcas program[52] at the B3LYP/6-31+G* level of theory using an atomic natural orbital (ANO) type recontraction of the standard 6-31+G* basis set. The choice of embedding potential was based on results from earlier work.[23]

2.5 Excitation energy calculations

Excitation energies were calculated using a development version of Dalton2011[48] using TDDFT[53] with the CAM-B3LYP exchange–correlation functional[54,55] and the 6-31+G* basis set.[45–47] Long-range corrected functionals such as CAM-B3LYP are known to describe anionic structures better than other exchange-correlation functionals.[56] Note that we considered only vertical excitation energies here, ignoring vibronic effects. PE-DFT calculations—with the protein included through an explicitly polarizable embedding potential—were compared

with calculations on the bare chromophore with no embedding (NE) and using the geometry optimized in vacuum. The importance of explicit polarization was examined by removing the polarizabilities from the embedding potential. Note that polarization of the environment is neither included explicitly nor implicitly in these non-polarizable embedding (NPE) calculations, since the parameters for each amino acid fragment are calculated in vacuum. The boundary between the quantum and classical regions was treated using a hydrogen link-atom approach. Moreover, multipoles and polarizabilities within a threshold distance of 1.4 Å from any QM atom were moved to the nearest atoms. The choice of basis set, link-atom approach and treatment of crystal water molecules in the PE-DFT calculations were based on extensive testing as discussed in an earlier work.[23] In all cases we only consider the first excitation energy.

3 Results and discussion

The calculated excitation energies of the FPs are presented both for the crystal structures (Section 3.2) and for snapshots extracted from the MD simulations (Section 3.3). In both cases, the results are compared to the experimentally obtained excitation energies to verify to what extent the trend in the excitation energies between the different fluorescent proteins is reproduced. However, we first examine the excitation energies of the chromophores in vacuum (Section 3.1).

3.1 Chromophores in vacuum

The excitation energies of the methyl-terminated aromatic system of the chromophores (geometries optimized in vacuum) are shown in Fig. 2. Note that the neutral chromophore models of wtGFP and YFP are the same, as are the anionic chromophore models of wtGFP, eGFP and YFP.

The calculations on the chromophores optimized in vacuum agree well with earlier B3LYP calculations on the same systems reported by Nifosì and Luo[17] for eCFP and the anionic chromophores. For BFP and neutral YFP/wtGFP, however, our calculations give an energy that is higher by around 0.2–0.3 eV. This can be rationalized by the choice of the CAM-B3LYP functional, which has been shown to yield higher excitation energies than the B3LYP functional.[57] In that study, List *et al.*[57] also found that CAM-B3LYP yields excitation energies comparable to RI-CC2 for neutral GFP-like chromophores. However, for charged chromophores it was shown to consistently overestimate excitation energies by about 0.3 eV. Moreover, Amat and Nifosì obtain red shifts between 0.5 and 0.8 eV for anionic chromophores using the CAM-B3LYP functional, significantly higher than their CASPT2 and CI calculations.[25]

The calculations on the vacuum chromophores account for the main differences between the FPs. In particular, BFP (with His66) gives a higher excitation energy than neutral wtGFP (with Tyr66), and eCFP (with Trp66) yields an excitation energy between neutral and anionic wtGFPs. Thus, a change in residue 66 is by far the most important cause of the differences in excitation energies between the FPs taken into account here.

3.2 The crystal structures

Nearly all excitation energies are overestimated compared to experiment, which can be rationalized by the use of the CAM-B3LYP functional as discussed above.[57] Overall, the results from calculations performed on the structures without geometry optimization do not reproduce the experimental trend among the different fluorescent proteins (see Fig. 2a). Therefore, closer inspection of the crystal structures is needed.

In the case of BFP, all three models (NE, NPE and PE) yield values that fail to reproduce the blue shift with respect to neutral wtGFP, instead giving a red shift of 0.2 to 0.3 eV. These findings indicate that the non-optimized crystal structures of the BFP chromophore (1BFP) are not suited for excitation energy calculations. A closer inspection of the chromophore (see Fig. 3) reveals a rather large difference in bond lengths of the crystal structure from those obtained by QM/MM geometry optimization, as well as from the non-optimized and QM/MM geometry-optimized MD structures (MD results are presented

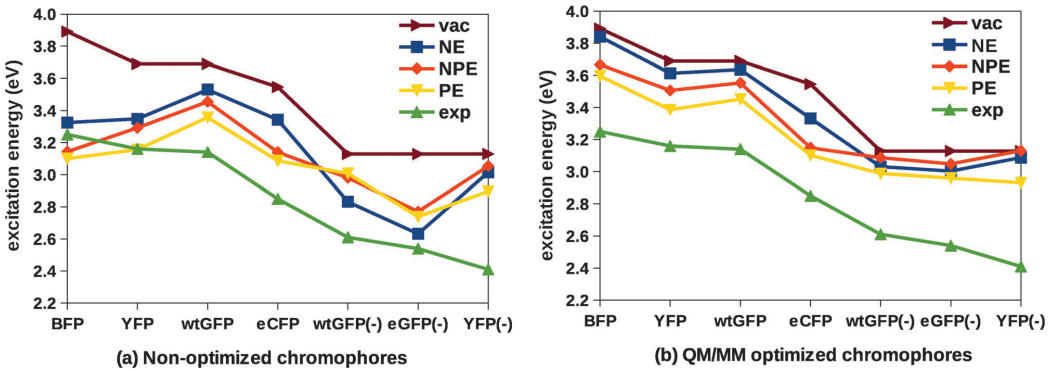


Fig. 2 Excitation energies of the fluorescent proteins based on crystal structures. The chromophores were taken directly from the crystal structure in (a) and QM/MM geometry-optimized in (b). Calculations were performed with no embedding (NE), non-polarizable embedding (NPE) and polarizable embedding (PE) at the CAM-B3LYP/6-31+G* level of theory. Included are also results from vacuum-optimized chromophores (vac) and experimental values (exp), which are the same in (a) and (b).

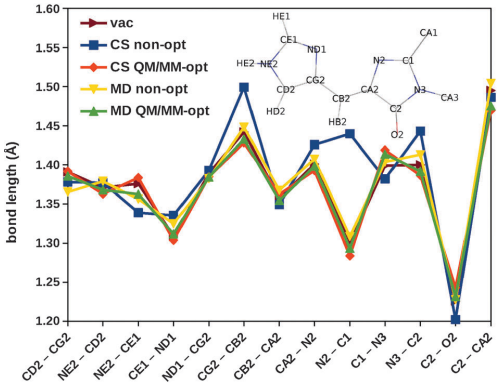


Fig. 3 Calculated bond lengths of the different models of the BFP chromophore: the vacuum structure (vac), the non-optimized crystal structure (CS non-opt), the QM/MM geometry-optimized crystal structure (CS QM/MM-opt), the averages from non-optimized MD snapshots (MD non-opt) and the averages from QM/MM geometry-optimized MD snapshots (MD QM/MM-opt). The atom labelling is shown in the inset.

in more detail in Section 3.3). All structures except those taken directly from the crystal structure of 1BFP possess very similar bond lengths. However, for the 1BFP structure both the CG2–CB2 and N2–C1 bonds are lengthened (for naming see the inset in Fig. 3), giving a longer bridge between the imidazole and imidazolidone rings as well as a different shape of the imidazolidone ring. The angles for the crystal structure chromophore of BFP are similar to the QM/MM geometry-optimized structure.† The CD2–CG2–CB2–CA2 dihedral angle is tilted 19° in the QM/MM geometry-optimized structure with respect to the non-optimized structure. In other words, the crystal structure has coplanar rings whereas the QM/MM geometry-optimized structure does not. These structural differences between non-optimized and optimized structures explain the large difference in the excitation energies between the BFP structure without geometry optimization (3.10 eV for PE) and with QM/MM geometry optimization (3.60 eV for PE). The too large N2–C1 bond length in the 1BFP structure was noted before and found to be the cause of the much lower excitation energy.[22] Structural data for all other fluorescent proteins are tabulated in the ESI.†

In the case of neutral YFP, a red shift of the order of 0.2 eV with respect to wtGFP is calculated from the non-optimized crystal structure (3.16 *vs.* 3.36 eV for PE), whereas the experimental excitation energies are nearly the same for the two proteins (3.16 *vs.* 3.14 eV). This discrepancy can be rationalized when taking into account that the YFP chromophore was in the anionic form during crystallization and X-ray diffraction,[16] which is also easily verified by looking at the bond length pattern of the crystal structure in comparison to the bond lengths of the optimized chromophores.† Indeed, the shorter OH–CZ and longer CZ–CE1 and CZ–CE2 bond lengths found in the non-optimized structure are characteristic of an anionic chromophore.

In a similar way, the figures in the ESI† show that wtGFP and eGFP chromophores are predominantly neutral in the crystal structures. This is surprising in the case of eGFP since the chromophore is said to be anionic.[8,15] Any faithful model for *anionic* wtGFP or eGFP should therefore include some sort of geometry optimization of the conjugated system.

The observations discussed above led to the conclusion that geometry optimization of the crystal structures is necessary prior to the excitation energy calculations, as also pointed out in ref. 25. Indeed, the trends in excitation energies are better reproduced for the QM/MM geometry-optimized structures (see Fig. 2b) and the following discussion will therefore be based on the results obtained using those structures.

The indirect effect of the protein surroundings—*via* the induced structural changes in the chromophore—red shifts the excitation energies (see Fig. 2). This effect is especially pronounced in the case of eCFP where the excitation energy is shifted by 0.21 eV whereas in the other FPs it is closer to 0.1 eV. Note that residue 66 in eCFP (Trp) is much larger than the corresponding residue in the other proteins (Tyr and His). A possible explanation is that the chromophore is actually too large for the cavity—made for tyrosine in wtGFP—which can lead to structural changes in the chromophore and therefore a relatively large difference between vacuum and no embedding (NE) calculations. Note that in addition to the structural changes in the aromatic part, the chromophores used in the NE calculations also contain parts of neighbouring amino acids whereas in the vacuum structures the aromatic part is terminated with methyl groups. Interestingly, adding the electrostatic interactions between the chromophore and a non-polarizable (frozen) environment (NPE) causes an additional red shift of the neutral species while it blue shifts the excitation energies of the anionic chromophores compared to NE. Stabilization of the ground state of the anionic chromophore by the surrounding protein might be the reason for the relatively higher excitation energy in the anionic case. Indeed, the hydrogen bonding network around the chromophore and the protonation state of Glu222 are different between the anionic and neutral proteins. Allowing mutual polarization between the chromophore and its environment (PE) lowers the excitation energies in all cases. Specifically, in the case of the anionic YFP chromophore it leads to a shift of 0.2 eV compared to around 0.1 eV in the other FPs.

Adding all the effects from the surrounding protein leads to red shifts ranging from 0.14 eV in the anionic wtGFP chromophore to 0.30 eV for neutral YFP and 0.44 eV in the case of eCFP, though the latter may be the result of structural abnormalities as discussed above. This gives an ordering of the excitation energies that is consistent with the experimental values except for the neutral YFP and wtGFP which are interchanged. Experimentally, the difference between the anionic and neutral peaks of YFP is 0.75 eV, while this difference is 0.53 eV for wtGFP. We are not able to reproduce this enlargement of the neutral–anionic shift for YFP. This is, however, not unexpected, since the red shift of the anionic peak for YFP is believed to originate from π–π stacking interactions with the neighbouring Tyr203

residue,[12] which requires a good description of dispersion effects, a well-known problem area for DFT.[58–61] One may also speculate that the wrong ordering of the excitation energies for the neutral YFP and wtGFP is partly related to the 1F0B crystal structure which is believed to contain the anionic chromophore (see the discussion above). Although the QM/MM geometry optimization accounts for the changes in the chromophore, the protein environment was frozen in the geometry optimization.

It is worth noting that polarization effects are important for the description of π–π stacking interactions, consistent with the observations made by Wachter *et al.*[12] Indeed, a point charge MM model cannot account for the spectral difference between anionic GFP and YFP.[25] Inclusion of Tyr203 in the QM region in both the QM/MM geometry optimization and the subsequent PE-DFT calculation lowered the excitation energy in both the neutral and anionic models but did not change the neutral–anionic shift significantly,† as could be expected considering the lack of a good description of dispersion in the excitation energy calculation. The distance between the two aromatic rings (around 3.4 Å) did not change significantly in the course of the QM/MM geometry optimization.

In summary, optimization of the crystal structure is needed to correctly calculate excitation energies of fluorescent proteins. After QM/MM geometry optimization of the chromophore in a frozen protein environment, the trend in excitation energies is generally well described, although we cannot reproduce the red shift resulting from π–π stacking in anionic YFP.

3.3 Molecular dynamics

The average excitation energies for the FPs after MD, with and without QM/MM geometry optimization of the chromophore, are presented in Fig. 4. We observe that all calculated excitation energies are overestimated compared to the experimental values, as also seen for calculations performed on the crystal structures (Section 3.2).

In contrast to the results for the crystal structures, differences between non-optimized and QM/MM optimized chromophores are much smaller (compare Fig. 4a and b). This indicates that the force field used in the MD simulations describes the chromophores adequately and that QM/MM geometry optimization is not strictly necessary. This is evident from a comparison of the NE values obtained from the non-optimized and QM/MM optimized chromophores which yields differences that are specifically due to the structural changes from the optimization. For the chromophores whose force field parameters have been designed for the GFP chromophore[29] we observe very little effect from the geometry optimization. Specifically, for the neutral chromophores in YFP and wtGFP the excitation energy is red shifted by 0.02 eV and for the anionic counterparts it is blue shifted by 0.04 eV while for the anionic eGFP chromophore it is blue shifted by 0.02 eV. Nevertheless, the phenol C–O bond changes from 1.40 to 1.36 Å after QM/MM geometry optimization of the neutral chromophores of wtGFP and YFP, which seems to point out an imperfection in the equilibrium bond length in the force field (1.411 Å). In fact, the bond length is taken from the CHARMM tyrosine parameter and not optimized specifically for the GFP chromophore.[29] In the case of BFP and eCFP there is a blue shift of 0.09 eV and a red shift of 0.08 eV, respectively, indicating that the quality of the force field parameters for these chromophores is somewhat poorer. Indeed, the force field developed by Reuter *et al.*[29] was optimized for chromophores with a phenol ring originating from a tyrosine amino acid. BFP, on the other hand, is made of an imidazole ring originating from histidine, and the chromophore in eCFP consists of an indole ring from tryptophan (see Table 1 and Fig. 1). Here we used standard CHARMM parameters for histidine and tryptophan, respectively, to describe the imidazole and indole rings (see Section 2.2).

It is not clear how well the force field performs in terms of the π–π stacking interactions in YFP. Inclusion of Tyr203 in the QM region of the excitation energy calculations based on the

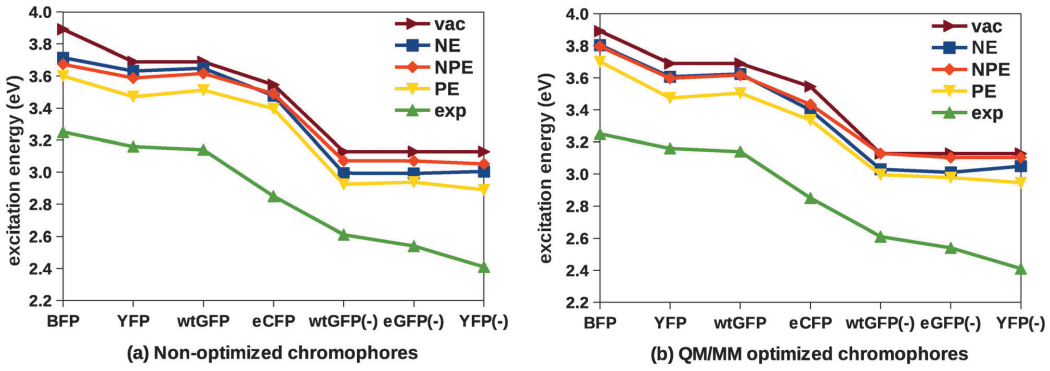


Fig. 4 Excitation energy averages of the fluorescent proteins based on 50 snapshots obtained from MD simulations. The chromophores were taken directly from the MD snapshots in (a) and QM/MM geometry-optimized in (b). Calculations were performed with no embedding (NE), non-polarizable embedding (NPE) and polarizable embedding (PE) at the CAM-B3LYP/6-31+G* level of theory. Included are also results from vacuum-optimized chromophores (vac) and experimental values (exp), which are the same in (a) and (b).

MD snapshots—both with and without QM/MM geometry optimization—leads only to slightly lower excitation energies in both the neutral and anionic chromophores.† Again this may be expected since extending the DFT region does not improve the description of dispersion in the excitation energy calculation, which is important to correctly describe π–π stacking effects.

The main features of the trend in excitation energies among the five FPs studied here are reasonably well reproduced compared to experiment. Including interactions with a frozen protein environment (NPE *vs.* NE) has a very small effect on the neutral chromophores. There seems to be a tendency to red shift the excitation energies, except in the case of eCFP where there is a slight blue shift, contrary to what was observed for the QM/MM-optimized chromophore based on the crystal structure (see Fig. 2). For the anionic chromophores we again see a blue shift, which is more pronounced than in the corresponding results based on the crystal structures. The effect from mutual polarization interactions between chromophore and protein surroundings red shifts all excitation energies between 0.07 and 0.16 eV compared to NPE with a tendency to yield slightly larger shifts in the anionic chromophores, especially for non-optimized chromophores. The geometry optimization slightly improves the trends among the anionic chromophores and gives the correct ordering, but the differences between the anionic FPs are not as pronounced as in experiment.

Experiments on GFP chromophores in vacuum indicate that the excitation energies are red-shifted going from vacuum to protein, both for the neutral and the anionic chromophores.[62–64] In contrast to a thorough theoretical investigation by Filippi *et al.*,[24] we do observe such a red shift going from vacuum to protein also for the anionic wtGFP. Indeed, for the PE-DFT calculations on anionic wtGFP, we calculate a red shift of 0.13 eV from vacuum (3.13 eV) to protein (3.00 eV) for the QM/MM geometry-optimized structure, and a red shift of 0.20 eV for the non-optimized structure (3.13 to 2.92 eV). Using a non-polarizable embedding potential (NPE), however, we observed a smaller red shift of 0.06 eV for the non-optimized structure and no red shift at all for the optimized structure (see ESI† for all tabulated data). To confirm that this is indeed due to differential polarization effects, which can only be modelled using a polarizable embedding potential, we performed an additional calculation on the QM/MM geometry-optimized crystal structure of anionic wtGFP where the environment was kept frozen in the response calculation. This corresponds to typical non-polarizable force fields where the ''average'' ground state polarization is taken into account implicitly through parametrization. Here, however, it is more accurate since the environment polarization is solved self-consistently and not just approximated *via* parametrization. The resulting excitation energy was 3.19 eV. This gives a blue shift of 0.06 eV compared to vacuum, whereas polarizable embedding gives a red shift of 0.13 eV. This confirms that differential polarization effects are essential for this system. Therefore, the lack of a polarizable embedding potential to describe the surrounding protein may very well be the reason why Filippi *et al.* did not observe the correct red shift for the anionic chromophore in wtGFP, as they also suggested in their conclusion.[24] We obtain a smaller red shift for neutral wtGFP going from vacuum to protein than that indicated by experiments based on solvent results extrapolated to vacuum,[62] in agreement with ref. 24.

It should be noted that conformational sampling is not a necessary step to reproduce trends in one-photon absorption between the fluorescent proteins, as the trends are already seen after QM/MM geometry optimization of the crystal structures (see Section 3.2). However, as recently detailed for the red fluorescent protein (DsRed), protein environmental effects are generally more significant in relation to two-photon absorption compared to one-photon absorption,[57] and it is therefore expected that sampling plays a more crucial role in two-photon absorption. Furthermore, two-photon absorption is more sensitive to the size of the conjugated system, so geometric distortions could be more relevant.

4 Conclusions and outlook

We have presented the first polarizable embedding study of a range of fluorescent proteins. Experimental trends in excitation energies among the fluorescent proteins could be reproduced. However, the CAM-B3LYP functional overestimates the excitation energies systematically by 0.3 to 0.5 eV, in agreement with other works using this functional on similar systems.[25,57] Raw crystal data are not suitable in all cases, but QM/MM geometry optimization of the chromophore in the protein is adequate to improve upon structural abnormalities. Conformational sampling can be successfully performed with a specially designed force field for the chromophore, although sampling is not necessary in the case of the one-photon absorption calculations presented here.

Explicit account of polarization in the force field is found to be needed to reproduce the experimentally obtained red shift from vacuum to protein in anionic wtGFP. However, our method fails to account for the red shift of YFP with respect to wtGFP in the anionic peak of the absorption spectrum, most likely because of intrinsic limitations of DFT.

Both sampling over different protein conformations and accounting for explicit polarization in the embedding potential are expected to be more relevant for the calculation of two-photon absorption. Apart from calculating two-photon absorption cross sections, subsequent steps include extending the number of FPs and attempting to reproduce more subtle differences in the excitation energies. The longer-term goal is to develop a faithful computational model that can predict the effects of mutations on the linear and non-linear absorption properties in fluorescent proteins.

Acknowledgements

This work has received support from the Research Council of Norway through a Centre of Excellence Grant (grant no. 179568/V30), from the European Research Council through a Starting Grant (grant no. 279619) to K.R. and from the Norwegian Supercomputer Program. J.K. thanks The Danish Council for

Independent Research (Sapere Aude), the Lundbeck Foundation, the Villum Foundation and the Danish Center for Scientific Computing for financial support.

References

1 R. Y. Tsien, *Annu. Rev. Biochem.*, 1998, **67**, 509–544.
2 M. Zimmer, *Chem. Rev.*, 2002, **102**, 759–781.
3 A. B. Cubitt, R. Heim, S. R. Adams, A. E. Boyd, L. A. Gross and R. Y. Tsien, *Trends Biochem. Sci.*, 1995, **20**, 448–455.
4 M. Ormö, A. B. Cubitt, K. Kallio, L. A. Gross, R. Y. Tsien and S. J. Remington, *Science*, 1996, **273**, 1392–1395.
5 R. Heim, D. C. Prasher and R. Y. Tsien, *Proc. Natl. Acad. Sci. U. S. A.*, 1994, **91**, 12501–12504.
6 M. Chalfie, Y. Tu, G. Euskirchen, W. W. Ward and D. C. Prasher, *Science*, 1994, **263**, 802–805.
7 R. Heim, A. B. Cubitt and R. Y. Tsien, *Nature*, 1995, **373**, 663–664.
8 B. P. Cormack, R. H. Valdivia and S. Falkow, *Gene*, 1996, **173**, 33–38.
9 R. Heim and R. Y. Tsien, *Curr. Biol.*, 1996, **6**, 178–182.
10 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.
11 R. M. Wachter, B. A. King, R. Heim, K. Kallio, R. Y. Tsien, S. G. Boxer and S. J. Remington, *Biochemistry*, 1997, **36**, 9759–9765.
12 R. M. Wachter, M.-A. Elsliger, K. Kallio, G. T. Hanson and S. J. Remington, *Structure*, 1998, **6**, 1267–1277.
13 S. J. Remington, *Methods Enzymol.*, 2000, **305**, 196–211.
14 M. Lelimousin, M. Noirclerc-Savoye, C. Lazareno-Saez, B. Paetzold, S. Le Vot, R. Chazal, P. Macheboeuf, M. J. Field, D. Bourgeois and A. Royant, *Biochemistry*, 2009, **48**, 10038–10046.
15 A. Royant and M. Noirclerc-Savoye, *J. Struct. Biol.*, 2011, **174**, 385–390.
16 R. M. Wachter, D. Yarbrough, K. Kallio and S. J. Remington, *J. Mol. Biol.*, 2000, **301**, 157–171.
17 R. Nifosì and Y. Luo, *J. Phys. Chem. B*, 2007, **111**, 14043–14050.
18 C. Filippi, M. Zaccheddu and F. Buda, *J. Chem. Theory Comput.*, 2009, **5**, 2074–2087.
19 M.-Y. Zhang, J.-Y. Wang, C.-S. Lin and W.-D. Cheng, *Int. J. Quantum Chem.*, 2012, **112**, 2607–2614.
20 M. A. L. Marques, X. López, D. Varsano, A. Castro and A. Rubio, *Phys. Rev. Lett.*, 2003, **90**, 258101.
21 I. Demachy, J. Ridard, H. Laguitton-Pasquier, E. Durnerin, G. Vallverdu, P. Archirel and B. Lévy, *J. Phys. Chem. B*, 2005, **109**, 24121–24133.
22 J.-Y. Hasegawa, K. Fujimoto, B. Swerts, T. Miyahara and H. Nakatsuji, *J. Comput. Chem.*, 2007, **28**, 2443–2452.
23 A. H. Steindal, J. M. H. Olsen, K. Ruud, L. Frediani and J. Kongsted, *Phys. Chem. Chem. Phys.*, 2012, **14**, 5440–5451.
24 C. Filippi, F. Buda, L. Guidoni and A. Sinicropi, *J. Chem. Theory Comput.*, 2012, **8**, 112–124.
25 P. Amat and R. Nifosì, *J. Chem. Theory Comput.*, 2013, **9**, 497–508.
26 J. M. Olsen, K. Aidas and J. Kongsted, *J. Chem. Theory Comput.*, 2010, **6**, 3721–3734.
27 H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, *Nucleic Acids Res.*, 2000, **28**, 235–242.
28 F. Yang, L. G. Moss and G. N. Philips Jr., *Nat. Biotechnol.*, 1996, **14**, 1246–1251.
29 N. Reuter, H. Lin and W. Thiel, *J. Phys. Chem. B*, 2002, **106**, 6310–6321.
30 *Schrödinger Suite 2011 Protein Preparation Wizard*, Schrödinger, LLC, New York, NY, 2011.
31 G. A. Kaminski, R. A. Friesner, J. Tirado-Rives and W. L. Jorgensen, *J. Phys. Chem. B*, 2001, **105**, 6474–6487.
32 W. L. Jorgensen, D. S. Maxwell and J. Tirado-Rives, *J. Am. Chem. Soc.*, 1996, **118**, 11225–11236.
33 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.
34 *Desmond Molecular Dynamics System, version 3.0*, D. E. Shaw Research, New York, NY, 2011; *Maestro-Desmond Interoperability Tools, version 3.0*, Schrödinger, New York, NY, 2011.
35 B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, *J. Comput. Chem.*, 1983, **4**, 187–217.
36 S. Nosé, *Mol. Phys.*, 1984, **52**, 255–268.
37 W. G. Hoover, *Phys. Rev. A*, 1985, **31**, 1695–1697.
38 G. J. Martyna, D. J. Tobias and M. L. Klein, *J. Chem. Phys.*, 1994, **101**, 4177–4189.
39 T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089–10092.
40 M. E. Tuckerman, B. J. Berne and A. Rossi, *J. Chem. Phys.*, 1991, **94**, 1465–1469.
41 J.-P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, *J. Comput. Phys.*, 1977, **23**, 327–341.
42 *QSite, version 5.7*, Schrödinger, LLC, New York, NY, 2011.
43 A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648–5652.
44 P. J. Stephens, F. J. Devlin, C. F. Chabalowski and M. J. Frisch, *J. Phys. Chem.*, 1994, **98**, 11623–11627.
45 W. J. Hehre, R. Ditchfield and J. A. Pople, *J. Chem. Phys.*, 1972, **56**, 2257–2261.
46 M. M. Francl, W. J. Pietro, W. J. Hehre, J. S. Binkley, D. J. DeFrees and J. A. Pople, *J. Chem. Phys.*, 1982, **77**, 3654–3665.
47 T. Clark, J. Chandrasekhar, G. W. Spitznagel and P. v. R. Schleyer, *J. Comput. Chem.*, 1983, **4**, 294–301.
48 DALTON, a molecular electronic structure program, Release Dalton2011 (2011), see http://daltonprogram.org/.
49 D. W. Zhang and J. Z. H. Zhang, *J. Chem. Phys.*, 2003, **119**, 3599–3605.
50 P. Söderhjelm and U. Ryde, *J. Phys. Chem. A*, 2009, **113**, 617–627.
51 L. Gagliardi, R. Lindh and G. Karlström, *J. Chem. Phys.*, 2004, **121**, 4494–4500.
52 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.
53 E. Runge and E. K. U. Gross, *Phys. Rev. Lett.*, 1984, **52**, 997–1000.
54 T. Yanai, D. P. Tew and N. C. Handy, *Chem. Phys. Lett.*, 2004, **393**, 51–57.
55 M. J. G. Peach, P. Benfield, T. Helgaker and D. J. Tozer, *J. Chem. Phys.*, 2008, **128**, 044118.
56 F. Jensen, *J. Chem. Theory Comput.*, 2010, **6**, 2726–2735.
57 N. H. List, J. M. Olsen, T. Rocha-Rinza, O. Christiansen and J. Kongsted, *Int. J. Quantum Chem.*, 2012, **112**, 789–800.
58 S. Kristyán and P. Pulay, *Chem. Phys. Lett.*, 1994, **229**, 175–180.
59 P. Hobza, J. Šponer and T. Reschel, *J. Comput. Chem.*, 1995, **16**, 1315–1325.
60 M. J. Allen and D. J. Tozer, *J. Chem. Phys.*, 2002, **117**, 11113–11120.
61 J. Černý and P. Hobza, *Phys. Chem. Chem. Phys.*, 2005, **7**, 1624–1626.
62 J. Dong, K. M. Solntsev and L. M. Tolbert, *J. Am. Chem. Soc.*, 2006, **128**, 12038–12039.
63 M. W. Forbes and R. A. Jockusch, *J. Am. Chem. Soc.*, 2009, **131**, 17038–17039.
64 K. Chingin, R. M. Balabin, V. Frankevich, K. Barylyuk, R. Nieckarz, P. Sagulenko and R. Zenobi, *Int. J. Mass Spectrom.*, 2011, **306**, 241–245.

Electronic Supplementary Information for "A polarizable embedding DFT study of one-photon absorption in fluorescent proteins"

M.T.P. Beerepoot, A.H. Steindal, J. Kongsted,
B. O. Brandsdal, L. Frediani, K. Ruud and J.M.H. Olsen

Table SI–I: Excitation energies (in eV) of the fluorescent protein crystal structures. Excitation energy calculations have been performed with no embedding (NE), non-polarizable embedding (NPE) and polarizable embedding (PE) at the CAM-B3LYP/6-31+G* level of theory. The vacuum calculations (vac) have been performed on chromophores optimized in vacuum, where the imidazolidone ring has been methyl-terminated. These results are plotted in Figure 2 in the main article.

			Non-opt			QM/MM-opt			
FPs	Q	vac	NE	NPE	PE	NE	NPE	PE	Exp
BFP	0	3.89	3.33	3.14	3.10	3.84	3.67	3.60	3.25
YFP	0	3.69	3.35	3.29	3.16	3.61	3.51	3.39	3.16
wtGFP	0	3.69	3.53	3.46	3.36	3.64	3.55	3.45	3.14
eCFP	0	3.55	3.34	3.14	3.09	3.33	3.15	3.10	2.85
wtGFP	-1	3.13	2.83	2.98	3.01	3.03	3.09	2.99	2.61
eGFP	-1	3.13	2.63	2.77	2.74	3.00	3.05	2.96	2.54
YFP	-1	3.13	3.02	3.05	2.90	3.09	3.13	2.93	2.41
DsRed[†]	-1	2.39				2.53	2.89	2.83	2.22

[†] Data from List *et al.* J. Phys. Chem. Lett. 2012, 3, 3513-3521

Table SI–II: Excitation energies (in eV) of the fluorescent proteins after MD. The results are averages of 50 snapshots and excitation energy calculations have been performed with no embedding (NE), non-polarizable embedding (NPE) and polarizable embedding (PE) at the CAM-B3LYP/6-31+G* level of theory. Standard deviations are shown between brackets. The vacuum calculations (vac) have been performed on chromophores optimized in vacuum, where the imidazolidone ring has been methyl-terminated. These results are plotted in Figure 4 in the main article.

			Non-opt			QM/MM-opt			
Name	Q	vac	NE	NPE	PE	NE	NPE	PE	Exp
BFP	0	3.89	3.72 (0.104)	3.67 (0.124)	3.60 (0.132)	3.80 (0.034)	3.80 (0.044)	3.70 (0.026)	3.25
YFP	0	3.69	3.63 (0.132)	3.59 (0.140)	3.47 (0.142)	3.61 (0.020)	3.60 (0.035)	3.48 (0.037)	3.16
wtGFP	0	3.69	3.65 (0.118)	3.62 (0.124)	3.51 (0.129)	3.63 (0.025)	3.62 (0.037)	3.51 (0.038)	3.14
eCFP	0	3.55	3.48 (0.122)	3.49 (0.134)	3.40 (0.145)	3.40 (0.024)	3.43 (0.048)	3.34 (0.056)	2.85
wtGFP	-1	3.13	2.99 (0.061)	3.07 (0.069)	2.92 (0.079)	3.03 (0.014)	3.13 (0.025)	3.00 (0.034)	2.61
eGFP	-1	3.13	2.99 (0.073)	3.07 (0.080)	2.94 (0.086)	3.01 (0.017)	3.10 (0.031)	2.98 (0.043)	2.54
YFP	-1	3.13	3.00 (0.064)	3.05 (0.068)	2.89 (0.075)	3.05 (0.027)	3.10 (0.028)	2.95 (0.030)	2.41

Table SI–III: Excitation energies (in eV) for neutral (E_n) and anionic (E_a) wtGFP and YFP and the magnitude of the neutral–anionic shift (ΔE). For YFP, data are shown for excitation energy calculations with only the chromophore in the QM region (YFP) and those with the chromophore and Tyr203 in the QM region (YFP extended). For the QM/MM geometry optimized structures, both the YFP chromophore and Tyr203 were present in the QM region of the geometry optimization. The MD results are averages of 50 snapshots. Excitation energy calculations have been performed with no embedding (NE), non-polarizable embedding (NPE) and polarizable embedding (PE) at the CAM-B3LYP/6-31+G* level of theory.

	wtGFP			YFP			YFP extended		
	E_n	E_a	ΔE	E_n	E_a	ΔE	E_n	E_a	ΔE
CS QM/MM-opt									
NE	3.64	3.03	0.61	3.61	3.09	0.52	3.51	2.98	0.53
NPE	3.55	3.09	0.46	3.51	3.13	0.38	3.43	3.02	0.41
PE	3.45	2.99	0.46	3.39	2.93	0.46	3.33	2.84	0.49
MD Non-opt									
NE	3.65	2.99	0.66	3.63	3.00	0.63	3.59	2.97	0.62
NPE	3.62	3.07	0.55	3.59	3.05	0.54	3.55	3.01	0.54
PE	3.51	2.92	0.59	3.47	2.89	0.58	3.45	2.88	0.57
MD QM/MM-opt									
NE	3.63	3.03	0.60	3.61	3.05	0.56	3.56	3.01	0.55
NPE	3.62	3.13	0.49	3.60	3.10	0.50	3.55	3.06	0.49
PE	3.51	3.00	0.51	3.48	2.95	0.53	3.45	2.93	0.52
Exp.	3.14	2.61	0.53	3.16	2.41	0.75	3.16	2.41	0.75

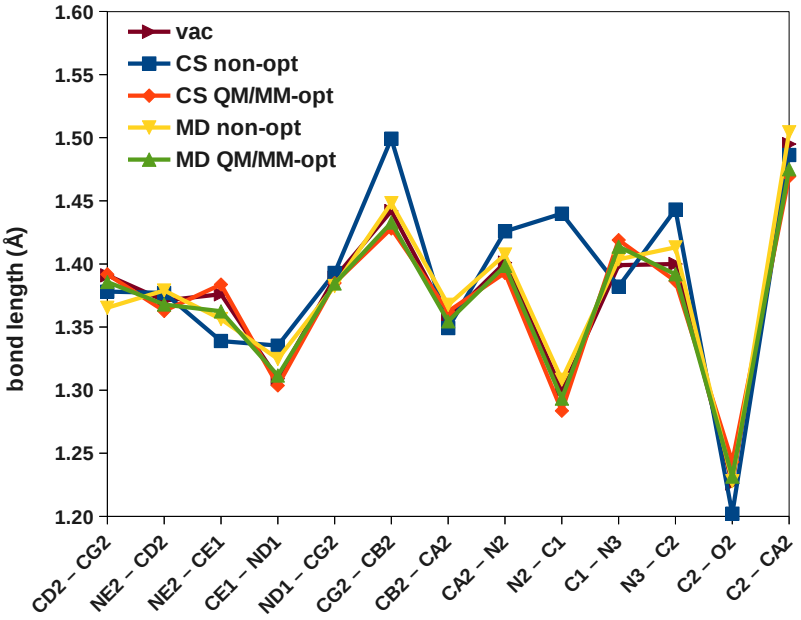


Figure SI–1: A selection of calculated bond lengths of the BFP chromophore obtained using different models: the optimized vacuum chromophore (vac), the non-optimized crystal structure (CS non-opt), the QM/MM geometry-optimized crystal structure (CS QM/MM-opt), the average over 50 non-optimized MD snapshots (MD non-opt) and the average over 50 QM/MM geometry-optimized MD snapshots (MD QM/MM-opt). The atom labeling of the chromophore is shown in the inset of Figure 3 in the main article.

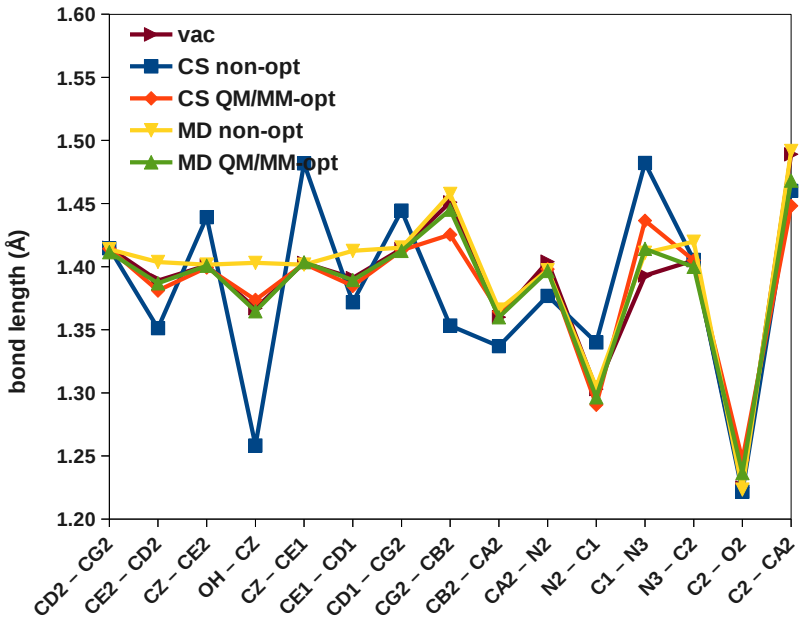


Figure SI–2: A selection of calculated bond lengths of the neutral YFP chromophore obtained using different models: the optimized vacuum chromophore (vac), the non-optimized crystal structure (CS non-opt), the QM/MM geometry-optimized crystal structure (CS QM/MM-opt), the average over 50 non-optimized MD snapshots (MD non-opt) and the average over 50 QM/MM geometry-optimized MD snapshots (MD QM/MM-opt). The atom labeling of the chromophore is shown in Figure SI–8a.

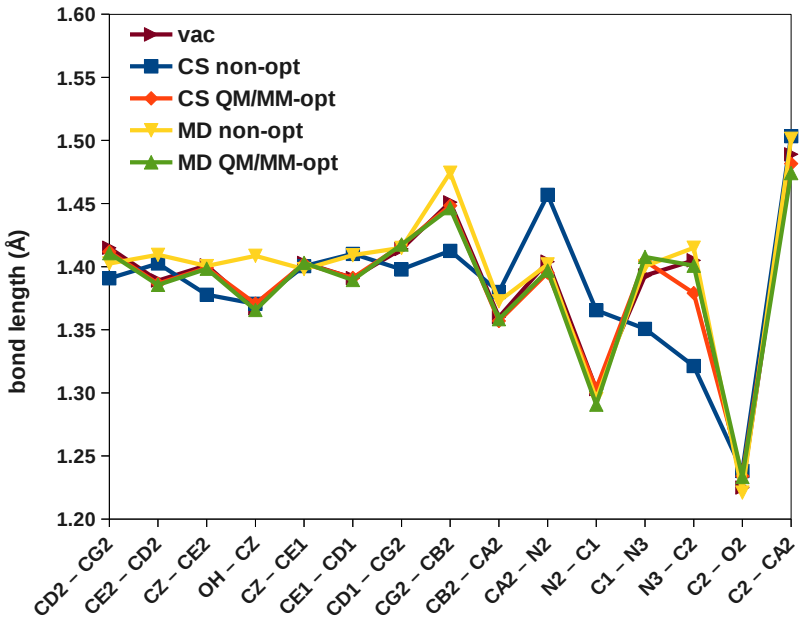


Figure SI–3: A selection of calculated bond lengths of the neutral wtGFP chromophore obtained using different models: the optimized vacuum chromophore (vac), the non-optimized crystal structure (CS non-opt), the QM/MM geometry-optimized crystal structure (CS QM/MM-opt), the average over 50 non-optimized MD snapshots (MD non-opt) and the average over 50 QM/MM geometry-optimized MD snapshots (MD QM/MM-opt). The atom labeling of the chromophore is shown in Figure SI–8a.

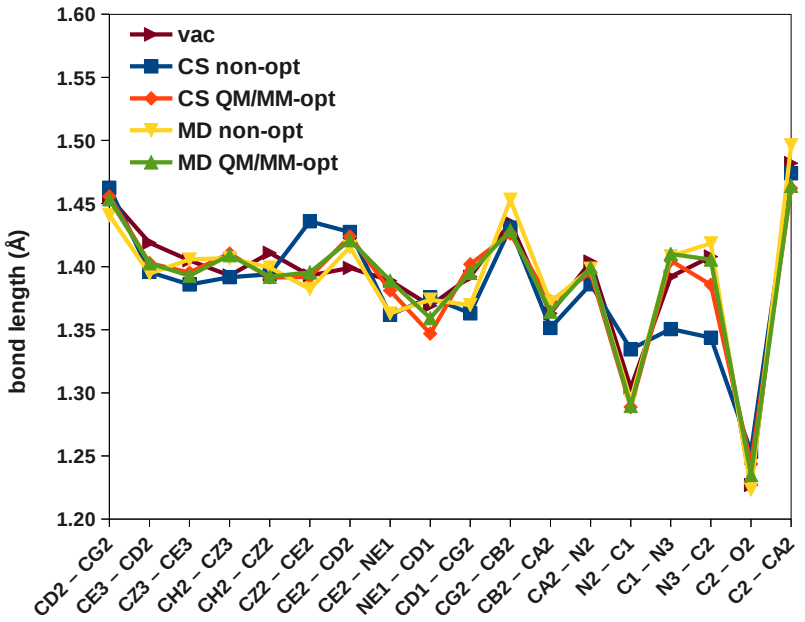


Figure SI–4: A selection of calculated bond lengths of the eCFP chromophore obtained using different models: the optimized vacuum chromophore (vac), the non-optimized crystal structure (CS non-opt), the QM/MM geometry-optimized crystal structure (CS QM/MM-opt), the average over 50 non-optimized MD snapshots (MD non-opt) and the average over 50 QM/MM geometry-optimized MD snapshots (MD QM/MM-opt). The atom labeling of the chromophore is shown in Figure SI–8b.

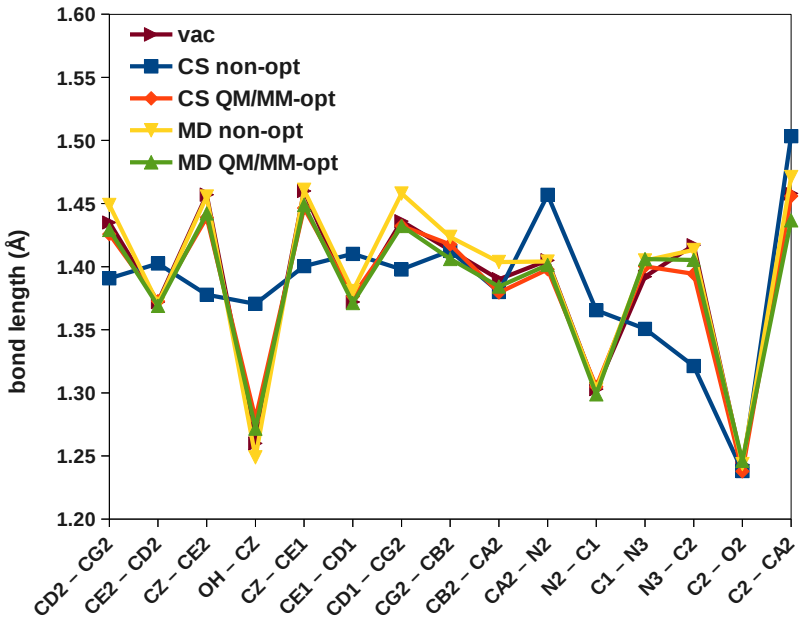


Figure SI–5: A selection of calculated bond lengths of the anionic wtGFP chromophore obtained using different models: the optimized vacuum chromophore (vac), the non-optimized crystal structure (CS non-opt), the QM/MM geometry-optimized crystal structure (CS QM/MM-opt), the average over 50 non-optimized MD snapshots (MD non-opt) and the average over 50 QM/MM geometry-optimized MD snapshots (MD QM/MM-opt). The atom labeling of the chromophore is shown in Figure SI–8a.

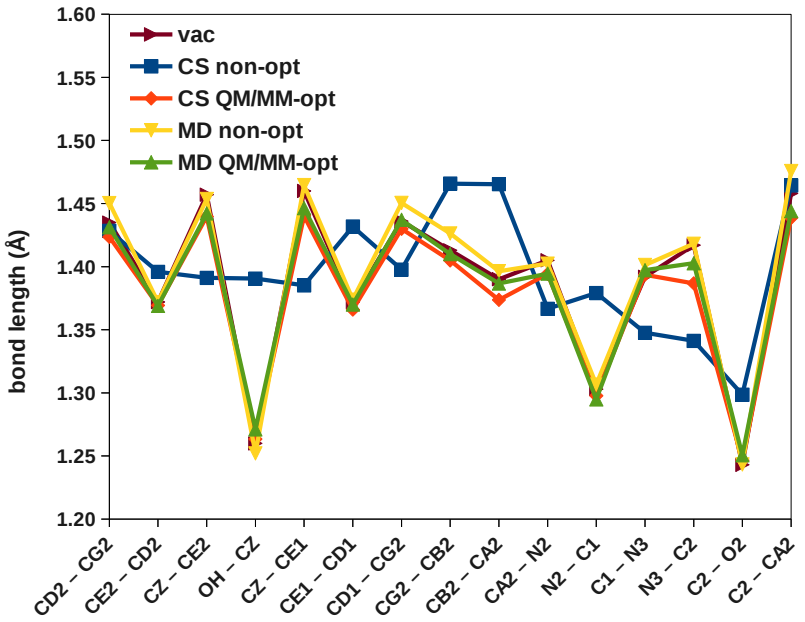


Figure SI–6: A selection of calculated bond lengths of the eGFP chromophore obtained using different models: the optimized vacuum chromophore (vac), the non-optimized crystal structure (CS non-opt), the QM/MM geometry-optimized crystal structure (CS QM/MM-opt), the average over 50 non-optimized MD snapshots (MD non-opt) and the average over 50 QM/MM geometry-optimized MD snapshots (MD QM/MM-opt). The atom labeling of the chromophore is shown in Figure SI–8a.

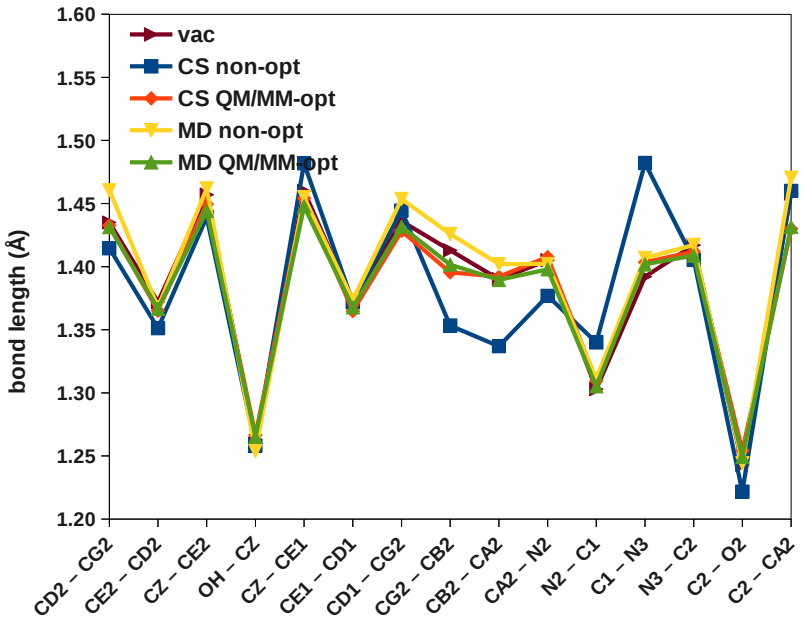
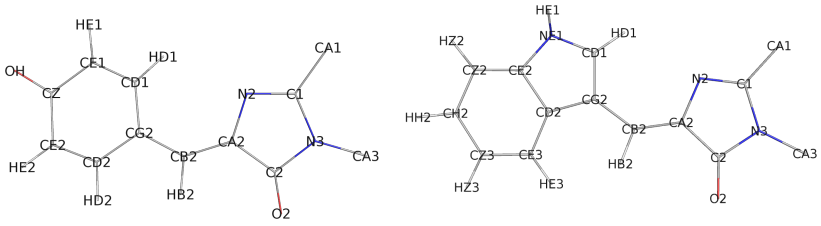


Figure SI–7: A selection of calculated bond lengths of the anionic YFP chromophore obtained using different models: the optimized vacuum chromophore (vac), the non-optimized crystal structure (CS non-opt), the QM/MM geometry-optimized crystal structure (CS QM/MM-opt), the average over 50 non-optimized MD snapshots (MD non-opt) and the average over 50 QM/MM geometry-optimized MD snapshots (MD QM/MM-opt). The atom labeling of the chromophore is shown in Figure SI–8a.



(a) (b)

Figure SI–8: Atom labeling for the chromophores of SI–8a wtGFP, YFP and eGFP and SI–8b eCFP. The atom labeling for the BFP chromophore is shown in the inset of Figure 3 in the main article.

Table SI–IV: A selection of bond lengths, angles and dihedral angles in the BFP chromophore. The vacuum calculation (vac) has been performed on the chromophore optimized in vacuum, where the imidazolidone ring has been methyl-terminated. The MD results are averages of 50 snapshots with standard deviations in parentheses.

	vac	CS Non-opt	CS QM/MM-opt	MD Non-opt	MD QM/MM-opt
Bond lengths					
CD2 - CG2	1.391	1.378	1.392	1.365 (0.025)	1.386 (0.003)
NE2 - CD2	1.371	1.377	1.363	1.379 (0.028)	1.368 (0.003)
NE2 - CE1	1.376	1.339	1.384	1.356 (0.025)	1.363 (0.003)
CE1 - ND1	1.309	1.335	1.304	1.325 (0.024)	1.312 (0.002)
ND1 - CG2	1.389	1.393	1.385	1.383 (0.029)	1.385 (0.003)
CG2 - CB2	1.442	1.499	1.428	1.448 (0.026)	1.433 (0.005)
CB2 - CA2	1.357	1.349	1.362	1.368 (0.021)	1.355 (0.002)
CA2 - N2	1.401	1.426	1.393	1.407 (0.027)	1.399 (0.002)
N2 - C1	1.300	1.440	1.284	1.308 (0.023)	1.293 (0.002)
C1 - N3	1.399	1.382	1.419	1.404 (0.024)	1.414 (0.004)
N3 - C2	1.400	1.443	1.386	1.413 (0.024)	1.392 (0.006)
C2 - O2	1.226	1.202	1.243	1.228 (0.020)	1.232 (0.004)
C2 - CA2	1.495	1.486	1.469	1.504 (0.024)	1.475 (0.007)
C1 - CA1	1.493	1.484	1.513	1.527 (0.026)	1.520 (0.004)
N3 - CA3	1.450	1.450	1.455	1.493 (0.030)	1.456 (0.003)
Angles					
ND1 - CG2 - CB2	126.2	122.9	123.9	120.6 (2.9)	124.9 (0.9)
CD2 - CG2 - CB2	124.3	129.3	126.2	129.1 (2.7)	126.0 (0.9)
CG2 - CB2 - CA2	130.4	122.9	125.4	131.4 (2.0)	128.7 (1.1)
CB2 - CA2 - C2	121.3	127.1	125.0	122.4 (2.7)	120.4 (1.2)
CB2 - CA2 - N2	129.4	123.6	125.6	130.1 (2.7)	130.0 (1.3)
Dihedral angles					
CD2 CG2 CB2 CA2	180.0	176.8	195.5	179.7 (9.9)	183.0 (8.4)
CG2 CB2 CA2 C2	180.0	183.7	178.8	180.4 (4.8)	181.8 (4.1)

Table SI–V: A selection of bond lengths, angles and dihedral angles in the YFP neutral chromophore. The vacuum calculation (vac) has been performed on the chromophore optimized in vacuum, where the imidazolidone ring has been methyl-terminated. The MD results are averages of 50 snapshots with standard deviations in parentheses.

	vac	CS Non-opt	CS QM/MM-opt	MD Non-opt	MD QM/MM-opt
Bond lengths					
CD2 - CG2	1.415	1.415	1.414	1.414 (0.026)	1.411 (0.002)
CE2 - CD2	1.389	1.351	1.381	1.404 (0.028)	1.387 (0.002)
CZ - CE2	1.401	1.439	1.399	1.402 (0.027)	1.401 (0.002)
OH - CZ	1.367	1.258	1.374	1.403 (0.032)	1.365 (0.005)
CZ - CE1	1.403	1.482	1.402	1.402 (0.028)	1.404 (0.002)
CE1 - CD1	1.391	1.372	1.385	1.413 (0.030)	1.389 (0.003)
CD1 - CG2	1.414	1.444	1.413	1.415 (0.027)	1.413 (0.002)
CG2 - CB2	1.451	1.353	1.425	1.458 (0.027)	1.445 (0.004)
CB2 - CA2	1.360	1.337	1.364	1.366 (0.021)	1.360 (0.002)
CA2 - N2	1.404	1.377	1.398	1.397 (0.023)	1.397 (0.002)
N2 - C1	1.303	1.400	1.291	1.305 (0.030)	1.297 (0.002)
C1 - N3	1.393	1.482	1.436	1.411 (0.024)	1.414 (0.004)
N3 - C2	1.405	1.405	1.405	1.420 (0.025)	1.400 (0.006)
C2 - O2	1.225	1.222	1.247	1.223 (0.022)	1.236 (0.003)
C2 - CA2	1.489	1.460	1.448	1.492 (0.033)	1.468 (0.005)
Angles					
CD1 - CG2 - CB2	123.8	119.9	123.3	123.7 (2.4)	123.7 (0.8)
CD2 - CG2 - CB2	118.5	121.5	116.9	118.3 (2.5)	118.2 (0.8)
CG2 - CB2 - CA2	130.1	134.2	133.0	132.0 (2.3)	130.3 (1.3)
CB2 - CA2 - C2	122.3	118.0	116.8	121.3 (3.3)	121.2 (1.6)
CB2 - CA2 - N2	128.5	131.1	132.6	130.2 (3.1)	129.3 (1.6)
Dihedral angles					
CD2 CG2 CB2 CA2	180.0	191.1	181.4	171.1 (10.8)	174.0 (7.7)
CG2 CB2 CA2 C2	180.0	173.2	182.4	183.5 (5.8)	184.0 (3.5)

Table SI–VI: A selection of bond lengths, angles and dihedral angles in the wtGFP neutral chromophore. The vacuum calculation (vac) has been performed on the chromophore optimized in vacuum, where the imidazolidone ring has been methyl-terminated. The MD results are averages of 50 snapshots with standard deviations in parentheses.

	vac	CS Non-opt	CS QM/MM-opt	MD Non-opt	MD QM/MM-opt
Bond lengths					
CD2 - CG2	1.415	1.391	1.412	1.403 (0.025)	1.411 (0.003)
CE2 - CD2	1.389	1.403	1.387	1.409 (0.026)	1.386 (0.003)
CZ - CE2	1.401	1.378	1.399	1.401 (0.030)	1.398 (0.003)
OH - CZ	1.367	1.371	1.371	1.409 (0.028)	1.366 (0.008)
CZ - CE1	1.403	1.400	1.402	1.398 (0.027)	1.403 (0.002)
CE1 - CD1	1.391	1.410	1.391	1.409 (0.027)	1.389 (0.002)
CD1 - CG2	1.414	1.398	1.417	1.415 (0.030)	1.418 (0.002)
CG2 - CB2	1.451	1.413	1.448	1.475 (0.025)	1.446 (0.005)
CB2 - CA2	1.360	1.380	1.357	1.373 (0.020)	1.358 (0.003)
CA2 - N2	1.404	1.457	1.395	1.402 (0.023)	1.397 (0.003)
N2 - C1	1.303	1.365	1.303	1.295 (0.022)	1.291 (0.002)
C1 - N3	1.393	1.351	1.405	1.400 (0.023)	1.408 (0.004)
N3 - C2	1.405	1.321	1.379	1.415 (0.031)	1.401 (0.006)
C2 - O2	1.225	1.238	1.234	1.221 (0.023)	1.233 (0.004)
C2 - CA2	1.489	1.503	1.482	1.501 (0.027)	1.474 (0.005)
Angles					
CD1 - CG2 - CB2	123.8	121.8	124.7	123.0 (2.5)	124.7 (0.7)
CD2 - CG2 - CB2	118.5	120.1	117.1	118.5 (3.0)	117.4 (0.6)
CG2 - CB2 - CA2	130.1	129.3	129.7	131.9 (2.3)	129.0 (0.8)
CB2 - CA2 - C2	122.3	121.2	123.9	123.1 (3.2)	123.5 (1.4)
CB2 - CA2 - N2	128.5	130.6	127.3	129.0 (3.0)	126.9 (1.2)
Dihedral angles					
CD2 CG2 CB2 CA2	180.0	179.8	175.5	179.4 (8.6)	178.6 (5.0)
CG2 CB2 CA2 C2	180.0	181.9	185.5	180.4 (5.6)	180.4 (3.9)

Table SI–VII: A selection of bond lengths, angles and dihedral angles in the eCFP chromophore. The vacuum calculation (vac) has been performed on the chromophore optimized in vacuum, where the imidazolidone ring has been methyl-terminated. The MD results are averages of 50 snapshots with standard deviations in parentheses.

	vac	CS Non-opt	CS QM/MM-opt	MD Non-opt	MD QM/MM-opt
CD2 - CG2	1.455	1.463	1.456	1.441 (0.024)	1.453 (0.004)
CE3 - CD2	1.419	1.396	1.403	1.395 (0.023)	1.403 (0.002)
CZ3 - CE3	1.405	1.386	1.395	1.406 (0.024)	1.393 (0.002)
CH2 - CZ3	1.393	1.392	1.410	1.407 (0.030)	1.409 (0.002)
CH2 - CZ2	1.411	1.394	1.391	1.399 (0.022)	1.392 (0.002)
CZ2 - CE2	1.393	1.436	1.392	1.382 (0.030)	1.396 (0.002)
CE2 - CD2	1.399	1.427	1.424	1.416 (0.030)	1.421 (0.003)
CE2 - NE1	1.389	1.362	1.381	1.363 (0.028)	1.389 (0.003)
NE1 - CD1	1.369	1.376	1.347	1.374 (0.029)	1.359 (0.004)
CD1 - CG2	1.392	1.363	1.402	1.370 (0.031)	1.395 (0.003)
CG2 - CB2	1.435	1.431	1.426	1.453 (0.028)	1.429 (0.006)
CB2 - CA2	1.363	1.351	1.372	1.372 (0.025)	1.364 (0.004)
CA2 - N2	1.404	1.386	1.395	1.398 (0.026)	1.399 (0.003)
N2 - C1	1.304	1.335	1.289	1.294 (0.031)	1.290 (0.002)
C1 - N3	1.392	1.351	1.405	1.408 (0.026)	1.410 (0.005)
N3 - C2	1.408	1.344	1.386	1.419 (0.027)	1.406 (0.008)
C2 - O2	1.227	1.253	1.244	1.223 (0.022)	1.235 (0.005)
C2 - CA2	1.482	1.474	1.462	1.496 (0.027)	1.464 (0.007)
C1 - CA1	1.493	1.472	1.517	1.518 (0.028)	1.522 (0.005)
N3 - CA3	1.451	1.465	1.454	1.488 (0.027)	1.456 (0.004)
Angles					
CD1 - CG2 - CB2	127.8	126.3	125.0	127.4 (2.9)	126.6 (1.0)
CD2 - CG2 - CB2	126.1	130.6	129.5	126.1 (3.1)	127.2 (1.2)
CG2 - CB2 - CA2	128.1	126.7	125.1	130.5 (2.5)	126.2 (1.0)
CB2 - CA2 - C2	123.4	126.4	127.0	125.4 (2.9)	124.6 (1.3)
CB2 - CA2 - N2	127.1	124.3	124.0	126.6 (3.0)	125.6 (1.1)
Dihedral angles					
CD2 CG2 CB2 CA2	180.0	173.2	175.2	183.5 (9.0)	181.2 (5.6)
CG2 CB2 CA2 C2	180.0	186.6	183.1	180.2 (5.4)	175.5 (4.8)

Table SI–VIII: A selection of bond lengths, angles and dihedral angles in the wtGFP anionic chromophore. The vacuum calculation (vac) has been performed on the chromophore optimized in vacuum, where the imidazolidone ring has been methyl-terminated. The MD results are averages of 50 snapshots with standard deviations in parentheses.

	vac	CS Non-opt	CS QM/MM-opt	MD Non-opt	MD QM/MM-opt
Bond lengths					
CD2 - CG2	1.435	1.391	1.426	1.449 (0.030)	1.429 (0.003)
CE2 - CD2	1.372	1.403	1.373	1.371 (0.030)	1.369 (0.003)
CZ - CE2	1.457	1.378	1.439	1.456 (0.029)	1.442 (0.003)
OH - CZ	1.260	1.371	1.280	1.249 (0.017)	1.272 (0.005)
CZ - CE1	1.460	1.400	1.446	1.461 (0.029)	1.449 (0.004)
CE1 - CD1	1.372	1.410	1.376	1.381 (0.029)	1.371 (0.003)
CD1 - CG2	1.436	1.398	1.432	1.458 (0.025)	1.432 (0.003)
CG2 - CB2	1.413	1.413	1.418	1.424 (0.022)	1.406 (0.005)
CB2 - CA2	1.390	1.380	1.379	1.404 (0.023)	1.384 (0.005)
CA2 - N2	1.405	1.457	1.398	1.404 (0.027)	1.401 (0.003)
N2 - C1	1.303	1.365	1.305	1.303 (0.028)	1.299 (0.002)
C1 - N3	1.392	1.351	1.400	1.405 (0.028)	1.406 (0.004)
N3 - C2	1.417	1.321	1.394	1.413 (0.025)	1.405 (0.006)
C2 - O2	1.243	1.238	1.238	1.244 (0.018)	1.246 (0.005)
C2 - CA2	1.458	1.503	1.456	1.471 (0.023)	1.437 (0.006)
Angles					
CD1 - CG2 - CB2	124.7	121.8	125.5	123.7 (3.0)	124.8 (0.9)
CD2 - CG2 - CB2	118.9	120.1	117.5	119.2 (3.1)	117.5 (0.7)
CG2 - CB2 - CA2	132.2	129.3	132.4	134.9 (2.3)	132.2 (1.4)
CB2 - CA2 - C2	122.7	121.2	122.7	122.4 (3.3)	120.3 (1.5)
CB2 - CA2 - N2	127.6	130.6	127.8	128.9 (3.2)	129.6 (1.5)
Dihedral angles					
CD2 CG2 CB2 CA2	180.0	179.8	176.3	177.2 (7.4)	177.0 (4.6)
CG2 CB2 CA2 C2	180.0	181.9	188.1	183.5 (5.4)	181.2 (3.8)

Table SI–IX: A selection of bond lengths, angles and dihedral angles in the eGFP chromophore. The vacuum calculation (vac) has been performed on the chromophore optimized in vacuum, where the imidazolidone ring has been methyl-terminated. The MD results are averages of 50 snapshots with standard deviations in parentheses.

	vac	CS Non-opt	CS QM/MM-opt	MD Non-opt	MD QM/MM-opt
Bond lengths					
CD2 - CG2	1.435	1.428	1.424	1.450 (0.032)	1.431 (0.003)
CE2 - CD2	1.372	1.396	1.369	1.372 (0.020)	1.369 (0.003)
CZ - CE2	1.457	1.391	1.440	1.454 (0.034)	1.442 (0.005)
OH - CZ	1.260	1.391	1.263	1.252 (0.017)	1.271 (0.004)
CZ - CE1	1.460	1.385	1.440	1.465 (0.031)	1.446 (0.003)
CE1 - CD1	1.372	1.432	1.366	1.374 (0.030)	1.370 (0.003)
CD1 - CG2	1.436	1.398	1.430	1.451 (0.024)	1.437 (0.003)
CG2 - CB2	1.413	1.466	1.405	1.426 (0.026)	1.410 (0.006)
CB2 - CA2	1.390	1.465	1.374	1.396 (0.028)	1.387 (0.005)
CA2 - N2	1.405	1.367	1.394	1.402 (0.025)	1.395 (0.005)
N2 - C1	1.303	1.379	1.298	1.307 (0.027)	1.295 (0.004)
C1 - N3	1.392	1.348	1.394	1.401 (0.024)	1.397 (0.006)
N3 - C2	1.417	1.341	1.387	1.418 (0.026)	1.403 (0.006)
C2 - O2	1.243	1.299	1.247	1.243 (0.022)	1.251 (0.005)
C2 - CA2	1.458	1.464	1.439	1.476 (0.029)	1.444 (0.006)
Angles					
CD1 - CG2 - CB2	124.7	124.9	123.2	123.1 (3.1)	125.1 (0.8)
CD2 - CG2 - CB2	118.9	113.2	118.7	119.9 (2.8)	117.8 (0.8)
CG2 - CB2 - CA2	132.2	128.1	130.0	134.7 (2.6)	130.5 (1.2)
CB2 - CA2 - C2	122.7	121.7	122.1	125.0 (2.9)	124.7 (1.7)
CB2 - CA2 - N2	127.6	127.9	128.4	126.5 (2.9)	125.8 (1.4)
Dihedral angles					
CD2 CG2 CB2 CA2	180.0	175.0	176.4	179.6 (6.1)	179.1 (4.8)
CG2 CB2 CA2 C2	180.0	185.5	184.3	185.7 (6.0)	182.2 (5.0)

Table SI–X: A selection of bond lengths, angles and dihedral angles in the YFP anionic chromophore. The vacuum calculation (vac) has been performed on the chromophore optimized in vacuum, where the imidazolidone ring has been methyl-terminated. The MD results are averages of 50 snapshots with standard deviations in parentheses.

	vac	CS Non-opt	CS QM/MM-opt	MD Non-opt	MD QM/MM-opt
Bond lengths					
CD2 - CG2	1.435	1.415	1.433	1.461 (0.026)	1.431 (0.003)
CE2 - CD2	1.372	1.351	1.365	1.368 (0.025)	1.367 (0.003)
CZ - CE2	1.457	1.439	1.450	1.462 (0.029)	1.444 (0.003)
OH - CZ	1.260	1.258	1.267	1.254 (0.018)	1.265 (0.004)
CZ - CE1	1.460	1.482	1.454	1.455 (0.027)	1.448 (0.005)
CE1 - CD1	1.372	1.372	1.365	1.374 (0.028)	1.368 (0.003)
CD1 - CG2	1.436	1.444	1.428	1.454 (0.031)	1.432 (0.003)
CG2 - CB2	1.413	1.353	1.395	1.426 (0.022)	1.402 (0.005)
CB2 - CA2	1.390	1.337	1.392	1.402 (0.026)	1.390 (0.004)
CA2 - N2	1.405	1.377	1.408	1.402 (0.028)	1.398 (0.004)
N2 - C1	1.303	1.340	1.310	1.311 (0.028)	1.305 (0.002)
C1 - N3	1.392	1.482	1.403	1.407 (0.029)	1.402 (0.004)
N3 - C2	1.417	1.405	1.412	1.417 (0.026)	1.409 (0.007)
C2 - O2	1.243	1.222	1.254	1.244 (0.019)	1.249 (0.004)
C2 - CA2	1.458	1.460	1.430	1.471 (0.026)	1.432 (0.006)
Angles					
CD1 - CG2 - CB2	124.7	119.9	124.8	123.7 (2.4)	124.5 (1.2)
CD2 - CG2 - CB2	118.9	121.5	117.9	119.4 (2.6)	117.9 (1.1)
CG2 - CB2 - CA2	132.2	134.2	135.6	135.5 (2.3)	132.5 (1.8)
CB2 - CA2 - C2	122.7	118.0	116.3	121.6 (2.9)	120.2 (1.8)
CB2 - CA2 - N2	127.6	131.1	133.2	130.0 (2.8)	130.1 (1.7)
Dihedral angles					
CD2 CG2 CB2 CA2	180.0	191.1	183.9	181.6 (5.5)	182.9 (3.6)
CG2 CB2 CA2 C2	180.0	173.2	184.8	188.0 (5.7)	188.2 (3.4)