PCCP

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Cite this: *Phys. Chem. Chem. Phys.*, 2025, **27**, 24777

Photoinduced decarboxylation in fluorescent proteins: charge-transfer states and structure–function relationship

Janko Čivić, *[a] Hideaki Mizuno [b] and Jeremy N. Harvey *[a]

Received 16th October 2025,
Accepted 3rd November 2025

DOI: 10.1039/d5cp03991f

rsc.li/pccp

Photoinduced decarboxylation of acidic amino acids near the chromophore is observed in various fluorescent protein variants, including avGFP, PA-GFP, DsRed, LSSmOrange, and PSLSSmKate. While a mechanism has been proposed and theoretically supported for GFPs, its generalization to other variants remains unclear. This study investigates the decarboxylation mechanism across different fluorescent proteins, with a particular focus on the red fluorescent proteins DsRed and PSLSSmKate, to establish a structure–function relationship. A benchmark study was conducted to assess various excited-state methods, basis sets, and embedding approaches. Excited-state QM/MM calculations using TD-DFT, ADC(2), and CC2 methods were employed to explore the excited states of multiple fluorescent protein variants. Our findings provide new insights into the decarboxylation mechanism in fluorescent proteins, contributing to the broader field of protein photophysics and potentially informing the design of novel fluorescent probes.

1. Introduction

Since the discovery of the green fluorescent protein (avGFP) from jellyfish *Aequorea victoria* in 1962,[1] fluorescent proteins (FPs) have become indispensable tools in cellular imaging. They are encoded by a single gene, and their chromophore forms through an autocatalytic post-translational process that requires no external cofactors, making them well suited for use in living systems.[2]

Nearly all FPs share a characteristic β-barrel fold that shields the chromophore from the environment. In avGFP, the chromophore forms from three residues (Ser, Tyr, Gly) and consists of a π-conjugated hydroxybenzylidene-imidazolinone system.[3] Despite this conserved architecture, mutations in or near the chromophore can produce variants with diverse photophysical properties (Fig. 1).

Apart from fluorescence, many FPs undergo light-induced chemical transformations of the chromophore or surrounding residues.[4] These transformations correspond to changes in physical properties, which may be undesirable when they cause loss of fluorescence, but can also be advantageous in certain contexts, for example in advanced imaging techniques such as super-resolution microscopy.[5] Because these processes occur on ultrafast timescales and involve electronically excited states, they are challenging to study experimentally, and the mechanisms of many of these reactions remain incompletely understood.

One specific transformation is the photoinduced decarboxylation of acidic residues near the chromophore. This reaction has been reported in a range of FPs, including avGFP itself.[6–12] For GFPs, irradiation at high energies leads to decarboxylation of Glu222, and both one-photon and multiphoton excitation routes have been demonstrated.[6,13,14] The generally accepted mechanism (Fig. 2) involves as a key step excited-state electron transfer (ET) from deprotonated Glu222 to the neutral chromophore, generating a glutamate radical that subsequently loses

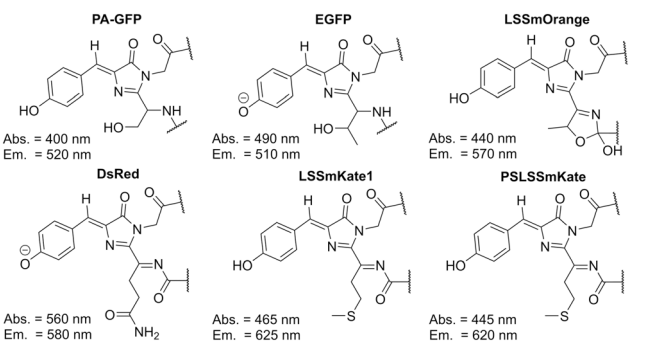

Fig. 1 Chromophores of fluorescent protein variants investigated in this work.

[a] *Department of Chemistry, KU Leuven – Celestijnenlaan 200F, Box 2404, 3001 Leuven, Belgium. E-mail: janko.civic@kuleuven.be, jeremy.harvey@kuleuven.be*
[b] *Department of Chemistry, KU Leuven – Celestijnenlaan 200G, Box 2403, 3001 Leuven, Belgium*

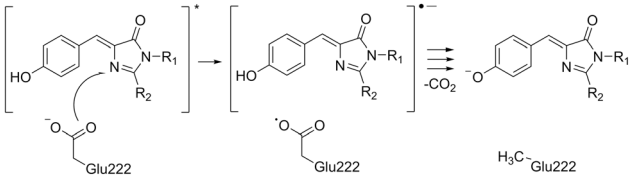

Fig. 2 Overview of the proposed mechanism for photoinduced decarboxylation in GFPs.[6]

CO_2. Computational studies support this picture: Ding *et al.*[15] showed that C–C bond cleavage can proceed spontaneously once Glu222 is oxidized, while Grigorenko *et al.*[16] identified glutamate-to-chromophore charge transfer (CT) excited states at energies consistent with experiments.

Whether this mechanism generalizes to other FPs remains an open question. In red FPs such as DsRed, photoinduced decarboxylation of Glu215 has been observed following pulsed laser excitation.[8] However, because the DsRed chromophore is anionic, ET from a negatively charged glutamate would yield a dianionic intermediate, which should in principle be unfavorable. Alternative mechanisms have been proposed, for example involving an initial ET step from the chromophore to a nearby positively charged residue.[17]

Another informative variant for understanding photoinduced decarboxylation is LSSmKate1, which does not undergo this process but can be converted into a variant that does through just two point mutations (PSLSSmKate).[9] Reported crystal structures suggest that the differences between these variants arise from rearrangements in the hydrogen-bond network around the chromophore. To our knowledge, no computational studies have yet been performed on these variants, but such calculations could provide valuable insights into the structure–function relationship underlying photoinduced decarboxylation.

In this work, we address these open questions by systematically comparing six FP variants: four that undergo photoinduced decarboxylation and two that do not (Fig. 1). Using QM/MM, we optimized their structures and computed vertical excitation energies with different electronic structure methods. We then analyzed the excited states to identify glutamate-to-chromophore CT states and assess whether their energetic accessibility can explain why some variants undergo photoinduced decarboxylation while others do not. A benchmark study was also performed to evaluate the sensitivity of these CT states to QM/MM methodological choices, including basis set size, embedding schemes, and QM region size. Together, our analysis provides new insights into the mechanism of photoinduced decarboxylation and suggests possible strategies for the rational design of photoactivatable FPs.

2. Computational methods

2.1. System preparation

For this study, the structures of six fluorescent protein variants were prepared for QM/MM calculations using their respective crystal structures: PA-GFP (PDB: 3GJ1,[7] chain A), EGFP (PDB: 2Y0G[18]) DsRed (PDB: 1ZGO,[19] chain A), LSSmOrange (PDB: 4Q7R,[20] chain A), LSSmKate1 (PDB: 3NT9,[21] chain A), and PSLSSmKate (PDB: 4NWS,[9] chain A). All structures were prepared following the same protocol, outlined here.

For residues with double occupancy, only the dominant conformation was selected. Crystallized water molecules within 8 Å of the chromophore were retained, while all other co-crystallized molecules were removed. Protonation of the structure was performed using the reduce program from AmberTools23.[22] Protonation states of residues near the chromophore were manually checked and compared with literature reports. The protonated protein was solvated in an octahedral box, with a minimum distance of 12 Å between the protein and the box edge, in 0.15 M NaCl. Additional sodium or chloride ions were added to neutralize the system's charge. The TIP3P[23] water model was used, along with the ff19SB[24] force field for the protein residues. TIP3P was chosen for practical reasons, as the recommended OPC[25] water model contains dummy atoms that complicate QM/MM calculations. Partial charges of the chromophore were developed using the restrained electrostatic potential (RESP) method.[26] The chromophore structure was extracted from the PDB file and capped with methylamine (NME) and acetyl (ACE) groups. Its geometry was optimized using B3LYP/6-31G*, while the positions of the terminal carbon atoms in the capping groups were kept fixed. The electrostatic potential was calculated with HF/6-31G* in the gas phase, and the RESP fitting of partial charges was performed using the resp program from AmberTools23.[22] These QM calculations were performed with Gaussian16.[27] Bonded parameters were based on the amber general force field (GAFF2).[28]

The structure of the solvated protein was relaxed through a multi-step procedure (Table S1) using sander or pmemd.cuda from Amber22.[29–32] The system was first minimized, then heated to 300 K, and equilibrated to a pressure of 1 bar. It was then cooled under constant volume to 50 K before final minimization. During all steps, the chromophore and residues within 3 Å of the chromophore were restrained with a force constant of 500 kcal mol^{-1} $Å^{-2}$. Temperature control was achieved with a Langevin thermostat, while pressure was maintained with an isotropic Berendsen barostat. Periodic boundary conditions (PBCs) were applied, using the particle mesh Ewald method[33] for electrostatic interactions with a 10 Å cutoff for short-range interactions. The SHAKE algorithm[34] was employed to constrain bonds involving hydrogens during the heating, cooling, and pressure equilibration steps. Next, constraints on the chromophore region were removed, and the non-periodic system was minimized using QM/MM, with atoms beyond 30 Å from the chromophore constrained. Non-bonded interactions were calculated for all atom pairs, and all MM atoms were included for electrostatic embedding. The QM calculation employed the composite PBEh-3c method[35] through the Amber22 QM/MM interface[36] with Orca.[37] The QM region included the chromophore, the side chain of the nearby glutamic acid residue prone to decarboxylation, and other interacting residues (Table S2). Geometry optimization

proceeded until the root mean square of the Cartesian gradient elements was less than 0.01 kcal mol^{-1} $Å^{-1}$.

2.2. Excited-state calculations

After obtaining the QM/MM optimized structures for all systems, the structure of PA-GFP was used to perform a benchmark study of several excited-state methods, basis sets, embedding potentials, and QM region sizes. Four QM region sizes were evaluated: (1) containing the chromophore and the Glu222 side chain; (2) extending to include the whole Glu222 residue and parts of residues forming hydrogen bonds ($\leq$3.2 Å) with its side chain; (3) further including parts of residues forming hydrogen bonds ($\leq$3.2 Å) with the chromophore's conjugated region; and (4) expanding to include residues with any heavy atom contacts ($\leq$3.65 Å) with either the side chain of Glu222 or the conjugated region. Capping hydrogen atoms, along with carbonyl and amine capping groups, were added as needed. Table S3 provides a complete list of residues for each region and variant.

Two different embedding potentials were tested: an electrostatic potential using Amber point charges from the system preparation, and a polarizable potential where the environment is represented by multipoles and polarizabilities developed using the PyFraME[38] Python package. In polarizable embedding calculations, to reduce computational cost, only solvent water molecules within 3 Å of any protein atom were retained, while the remaining waters were removed. Grabarek and Andruniów demonstrated that including more distant water molecules has minimal impact on calculated excitation energies of GFP chromophores.[39] The protein was divided into smaller fragments using the molecular fractionation with conjugate caps (MFCC) scheme.[40,41] For all residues not included in the QM region, atom-centered multipoles up to the second order and anisotropic dipole–dipole polarizabilities were determined using the localized properties (LoProp) method,[42] with the LoProp script[43] for the Dalton program[44,45] and B3LYP, employing an atomic natural orbital (ANO) recontraction of the 6-31+G* basis set. Charges of MM sites closer than 1.4 Å to the QM region were shifted to the nearest MM sites, while other embedding parameters on those sites were removed.

Excited-state QM/MM calculations were conducted using time-dependent density functional theory (TD-DFT) with the CAM-B3LYP[46] and ωB97X-D[47] functionals, as well as the CC2[48] and ADC(2)[49] methods. The resolution of identity (RI) approximation was employed for all methods. Several basis sets of the def2-type[50–52] were tested. Grimme D3(BJ) dispersion[53,54] was included in TD-DFT calculations. Additionally, the frozen-core and restricted virtual space (RVS) approximation was tested for ADC(2) and CC2 by freezing virtual orbitals above 60 eV, a threshold previously suggested as suitable for FPs.[55] These calculations were executed with Turbomole 7.8.[56–59] The character of excited states was analyzed with the TheoDORE analysis package.[60]

To investigate the impact of thermal fluctuations on the vertical excitation energies in PA-GFP, ground-state QM/MM MD simulations were performed for 10 ps with a 1 fs timestep, starting from the final structure of the constant-pressure equilibration (step 4 in Table S1). The QM region included only the chromophore with carbonyl and amine capping groups, described using the PBEh-3c method.[35] The SHAKE algorithm[34] was used to constrain bonds involving hydrogen atoms in the MM region. PBCs were used with particle-mesh Ewald for computing long-range electrostatic interactions of the MM region with a 10 Å cutoff for short-range nonbonded interactions. The particle mesh Ewald method for treating long-range electrostatic QM–MM and QM–QM interactions in periodic systems is not supported for the selected combination of methods, so a 15 Å real-space cutoff was used for all nonbonded interactions involving the QM region. Excited-state CAM-B3LYP/def2-TZVP calculations were carried out using QM region 2 (Table S3) and electrostatic embedding, on 100 evenly spaced snapshots from the last 5 ps of simulation.

Selected approaches were also applied on QM/MM optimized fluorescent protein structures of LSSmOrange, EGFP, DsRed, LSSmKate1, and PSLSSmKate. The quantum regions included the chromophore, the Glu residue at the decarboxylation-prone position, and portions of other residues that form hydrogen bonds with the conjugated part of the chromophore or the Glu side chain (Table S3).

3. Results and discussion

3.1. PA-GFP benchmark study

3.1.1. QM/MM geometry optimization. The structure of PA-GFP after QM/MM geometry optimization remained close to the starting crystal structure (Fig. 3 and Fig. S1). Hydrogen bonds connect Glu222 to the phenol ring of the chromophore *via* water molecules and Ser205. Additionally, there are hydrogen bonds between Glu222 and a water molecule, as well as

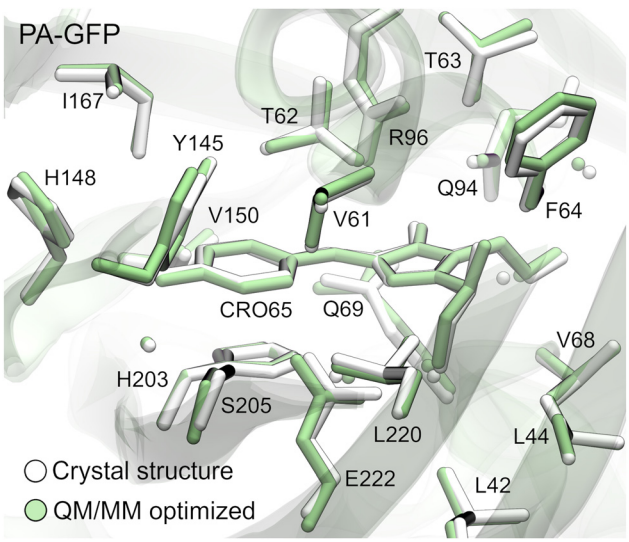

Fig. 3 Comparison of QM/MM optimized (green) and crystal (white, PDB: 3GJ1[7]) structure of PA-GFP. The chromophore and nearby amino acid side chains are shown as sticks and nearby water molecules as spheres. Hydrogen atoms are omitted for clarity.

between the hydroxyl group on the chromophore's N-terminus and Glu222. Arg96 and Gln94 form hydrogen bonds with the imidazolinone ring of the chromophore. The most notable difference compared to the crystal structure is the increase in the N–O distance between His148 and the chromophore by 0.18 Å, reaching 3.22 Å, with an unfavorable N–H–O angle of 115° for hydrogen bonding.

3.1.2. Excited state calculations. An overview of all calculations is provided in the SI (Table S4). In general, across all approaches, Glu222–chromophore CT states are identified at energies 4 to 6 eV above the ground state, consistent with previous computational studies[15,16,61] and reinforcing the idea of their importance in photoinduced decarboxylation. However, we note that the predicted excitation energies can vary by more than 1 eV depending on computational choices. This variability highlights the importance of the detailed benchmarking study we carried out, as discussed in the following paragraphs.

Given the large number of calculations performed, and to minimize subjectivity, it was necessary to define a quantitative and automated criterion for identifying Glu222–chromophore CT states. To this end, we used the TheoDORE analysis package[60] to analyze the one-electron transition density matrix, focusing on predefined fragments of the system. This approach enables quantification of the CT character of each excited state. Fig. 4 shows an example hole/electron population plot generated with TheoDORE, which provides a quantitative representation of the information one could obtain from plotting the natural transition orbitals (NTOs) of the excited states. We classified a state as a Glu222–chromophore CT state if the Glu222 side chain contributes more than 40% to the hole population and the conjugated part of the chromophore contributes more than 80% to the electron population. These thresholds were determined empirically. In most cases we find three distinct Glu222–chromophore CT states, corresponding to ET from each of the three highest occupied molecular orbitals of the carboxylate group. However, we sometimes observed that the CT character was distributed over multiple states. This observation motivated the choice of a somewhat lower threshold (40%) for the hole population of Glu222. Nevertheless, in most cases, this contribution was significantly higher, and the conclusions of the subsequent analysis are not sensitive to the exact threshold chosen.

Calculating higher excited states using correlated methods such as CC2 and ADC(2) with a triple-zeta basis set and a QM

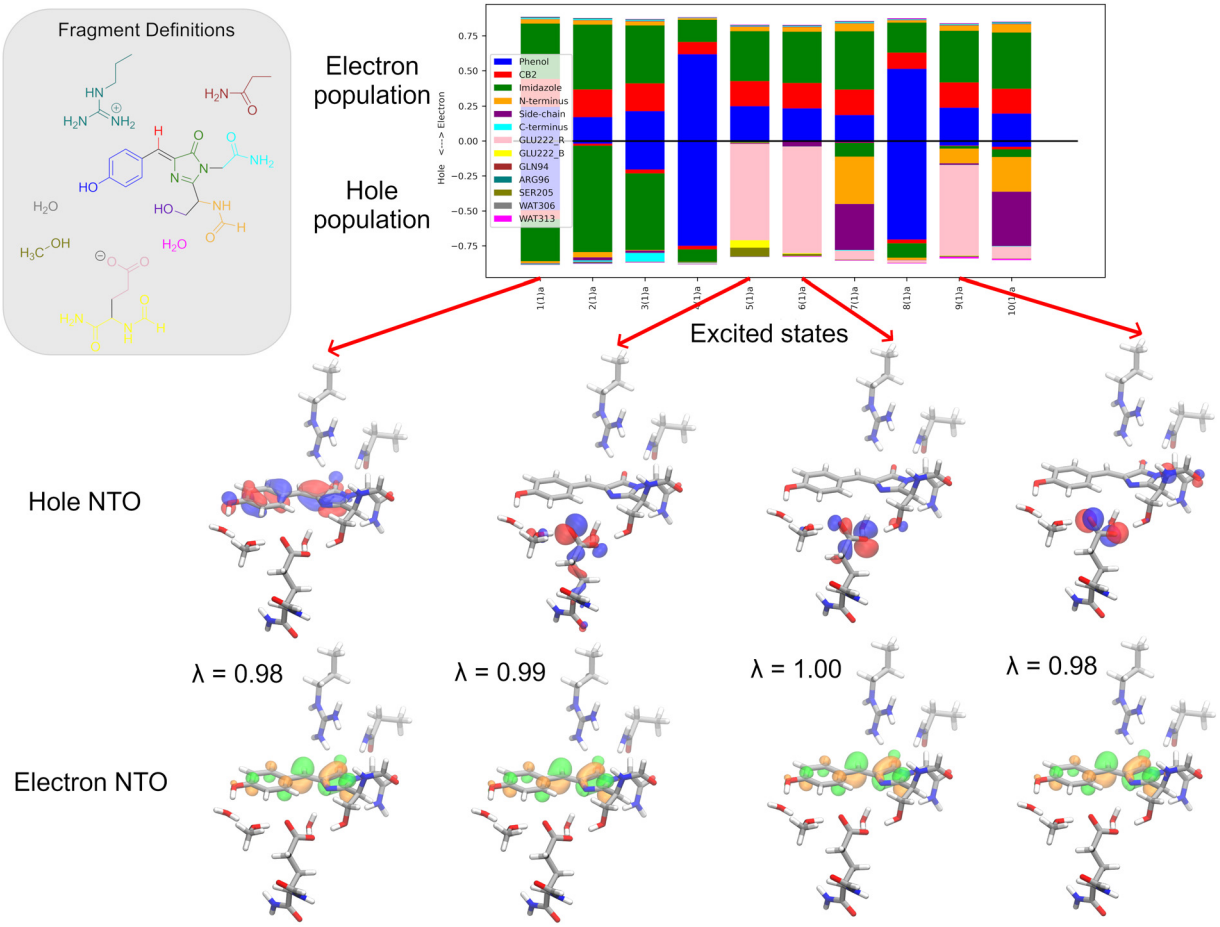

Fig. 4 Fragment hole/electron population plot generated with TheoDORE[60] for a QM/MM excited-state calculation of PA-GFP (QM region 3 from Table S3 and polarizable embedding for the MM region) at the CC2/def2-TZVP level of theory. The most significant NTO pair is shown for the first excited state and the Glu222–chromophore CT states.

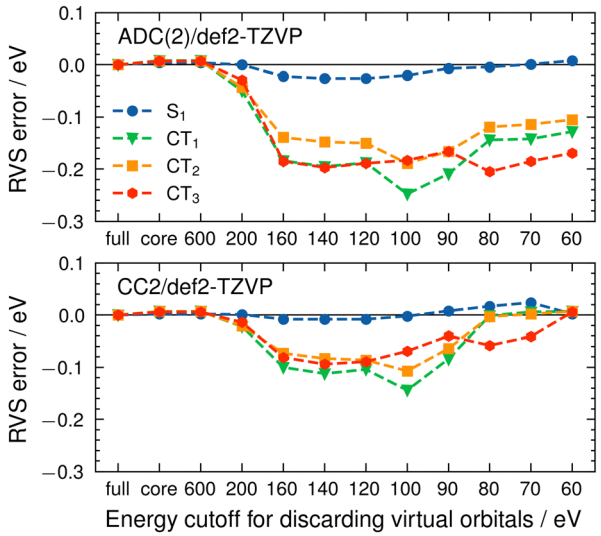

Fig. 5 Errors introduced by the RVS approximation in energies of the first excited state and the Glu222-chromophore CT states in the PA-GFP model with the smallest quantum region and electrostatic embedding, with varying energy cutoffs for discarding virtual orbitals. Based on data in Tables S6 and S7.

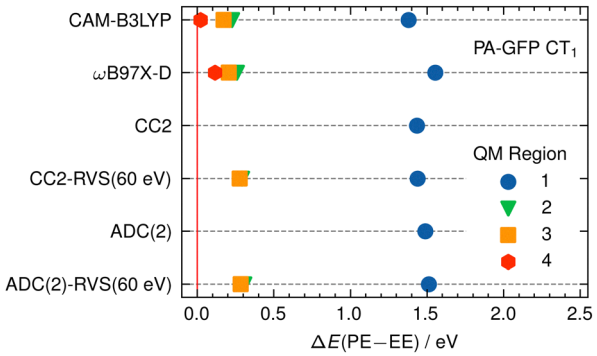

Fig. 6 Impact of polarizable (PE) compared to electrostatic embedding (EE) on the lowest Glu222-chromophore CT state in PA-GFP, calculated with def2-TZVP using different methods and QM region sizes. QM regions are defined in Table S3 and Fig. S2.

system of approximately 100 atoms is not possible without additional approximations. A common approach is the restricted virtual space (RVS) approximation, where a selected number of high-lying molecular orbitals are excluded from the excited-state energy calculations. We assessed its impact on PA-GFP using the smallest QM region and electrostatic embedding by systematically varying the energy cutoff for orbital exclusion (Fig. 5 and Tables S6, S7). For the first excited state, the effect is negligible for both methods, even when discarding more than half the orbitals, consistent with previous findings.[55] In contrast, the Glu222–chromophore CT states are more sensitive. ADC(2) shows larger errors (up to −0.25 eV) than CC2 (up to −0.14 eV). CC2 also benefits from error cancellation, giving very small errors at a 60 eV cutoff. Importantly, these errors remain similar when the QM region is enlarged or when polarizable embedding is applied (Table S5). Overall, while the RVS approximation introduces non-negligible errors for CT states, it remains suitable for semi-quantitative purposes where deviations of a few tenths of an eV are acceptable, as in our case, where precise experimental data on the positioning of the CT states is lacking.

Next, we examined the effect of increasing the basis set from def2-TZVP to def2-TZVPD, which includes diffuse functions (Fig. S3). Due to the increased computational cost, this analysis was performed only at the TD-DFT level. The results show that the inclusion of diffuse functions leads to only a minor change of approximately ±0.05 eV in both the first excited and Glu222-chromophore CT state energies. This effect is consistent across different embedding types and QM region sizes. Given the substantial increase in computational cost, the use of diffuse functions is not beneficial in this context.

Another aspect we examined is the impact of modeling the protein environment using polarizable embedding compared to electrostatic embedding (Fig. 6 and Fig. S4). For S_1, the impact of polarizable embedding relative to electrostatic embedding is somewhat unpredictable in terms of the sign of the shift (ranging between −0.1 and 0.2 eV), depending on the QM region (Fig. S4). For the CT states, a more uniform trend is observed, with polarizable embedding raising the energy irregardless of the computational method. For the smallest QM region the shift is very large, around 1.5 eV, but it reduces to approximately 0.1 eV for the largest QM region. For QM region 3 (104 atoms), the largest system we can use with correlated methods, the effect is between 0.2 and 0.3 eV. This is not negligible and suggests that the use of polarizable embedding is worthwhile, which we assume to be the more accurate approach. However, the shift is systematic, so even electrostatic embedding calculations are useful for qualitative insights or relative comparisons.

Additionally, we tested a larger solvent shell of 10 Å for polarizable embedding to assess the effect of the 3 Å cutoff. This increased the excitation energy of the lowest CT state by 0.12 eV (CAM-B3LYP/def2-TZVP level of theory with QM region 3), further increasing the difference relative to electrostatic embedding. This indicates that the 3 Å cutoff introduces a modest but non-negligible effect. However, as a larger solvent shell results in a considerable increase in computational cost, we restricted all subsequent calculations with polarizable embedding to the 3 Å cutoff. The sensitivity to the cutoff may deserve more attention in future studies.

Finally, we discuss the impact of the QM region choice. Definitions of the QM regions can be found in Table S3 and Fig. S2. The results with polarizable embedding indicate that the shifts in S_1 induced by increasing the QM region are generally small, between −0.05 and 0.05 eV. In contrast, a very different trend is observed for the CT states. Expanding the QM region from 1 to 2 induces a significant shift of up to −0.7 eV (Fig. 7 and Fig. S5). Interestingly, further increases in the QM region have negligible effects (less than ±0.05 eV), despite the fact that the increase from QM region 2 to QM region 3 adds 34 atoms, and further to QM region 4 additional 75 atoms. This suggests that accurately modeling the CT states requires at least

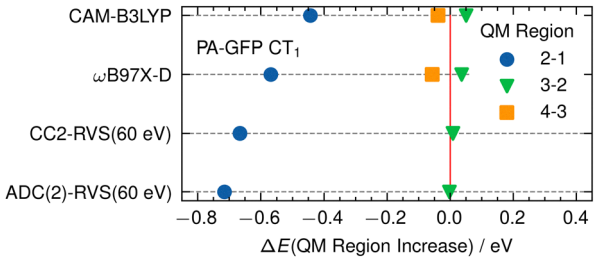

Fig. 7 Impact of increasing the QM region on the lowest Glu222-chromophore CT state in PA-GFP, calculated with def2-TZVP and polarizable embedding. QM regions are defined in Table S3 and Fig. S2.

Table 1 Energies of the first excited state and the Glu222–chromophore CT states of PA-GFP computed with various methods. All calculations used the def2-TZVP basis set with QM region 3 (Fig. S2), and polarizable embedding for the MM region

Method	S_1/ eV	CT / eV
CC2	3.31	4.69, 4.69, 5.39
ADC(2)	3.21	4.42, 4.51, 5.13
CAM-B3LYP	3.43	5.08, 5.13, 5.23, 5.65
ωB97X-D	3.44	5.30, 5.37, 5.83
Experiment	3.10[7]	∼4.65[14]

an accurate description of the environment around Glu222, while the chromophore environment is of lesser importance.

Now that the impact of all the different choices has been discussed, we present the results of some selected calculations in Table 1 in order to compare the results with experimental data. TD-DFT overestimates the first excited state energy by around 0.3 eV, consistent with previous reports.[62] CC2 and ADC(2) perform better.

Due to the limited data, it is hard to exactly position the experimental energy of the Glu222-chromophore CT state. However, some inference is possible. The most comprehensive experimental study of this process is by Langhojer *et al.*[14] They studied the T203V mutant of GFP where, similarly to PA-GFP, the neutral chromophore form is dominant. They report efficient decarboxylation with sequential 400 nm (3.10 eV) followed by 800 nm (1.55 eV) excitation populating a higher excited state. The CT state could be populated either directly from S_1 or from a higher locally excited chromophore state, but in either case, its energy should be around 4.65 eV. Both ADC(2) (Fig. S9) and CC2 (Fig. S10) predict an energy close to this, encouraging the use of these methods to study other variants. The NTO orbital pairs for CT states calculated with CC2 are shown in Fig. 4. TD-DFT overestimates the CT state energy, with the CAM-B3LYP functional (Fig. S7) yielding results closer to those of correlated methods than ωB97X-D (Fig. S8). Nevertheless, due to its low computational cost, TD-DFT can be useful for obtaining qualitative insights.

3.1.3. QM/MM MD simulations. To investigate how sensitive the excitation energies of the CT states are to thermal fluctuations, we performed 10 ps of ground-state QM/MM MD simulations of PA-GFP. The results (Fig. 8) show that the lowest CT state exhibits a larger standard deviation in vertical excitation energy (0.25 eV)

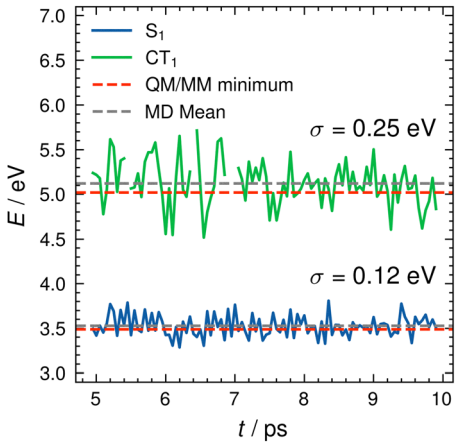

Fig. 8 Vertical excitation energies of the S_1 and lowest Glu222-chromophore CT state during ground-state QM/MM MD simulations of PA-GFP.

compared to the S_1 state (0.12 eV). The mean vertical excitation energy is close to the value obtained from the QM/MM optimized structure, suggesting that calculations on a QM/MM optimized structure provide a reasonable estimate of the mean.

Having established a reliable computational protocol, we applied it to other systems where the decarboxylation mechanism remains less well understood. These calculations were performed on the CC2/def2-TZVP level of theory with polarizable embedding, an equivalent of the third QM region, and the RVS approximation with freezing all virtual orbitals above 60 eV.

3.2. Application to other variants

3.2.1. LSSmOrange. First, we focus on LSSmOrange.[63] Compared to the chromophore in PA-GFP, the neutral chromophore in LSSmOrange contains an additional 2-hydroxy-dihydrooxazole ring, which extends the conjugated system of double bonds (Fig. 1). LSSmOrange has also been observed to undergo photoinduced decarboxylation of Glu215 upon exposure to 400 nm (3.10 eV) pulses from a femtosecond laser.[10] IR spectra of the unconverted and converted forms suggest that Glu215 which gets decarboxylated is anionic in the native protein.

After the QM/MM optimization, the chromophore environment remained close to the crystal structure (Fig. 9 and Fig. S13). The only notable difference was the displacement of a structural water molecule (residue 433 in the crystal structure) by 3.6 Å, allowing it to form a hydrogen bond with Glu215. This suggests that this water molecule may occupy multiple positions, or that additional water molecules near the chromophore were not resolved in the crystal structure.

Fragment decomposition of electron and hole populations (Fig. 10 and Fig. S14) from excited states calculated using CC2/def2-TZVP identified Glu215–chromophore CT states at 4.06, 4.49, and 4.91 eV. Experimental data to determine whether excitation occurs *via* one- or two-photon absorption is not available. However, assuming a two-photon process, as reported for PA-GFP with a 400 nm femtosecond laser,[14] an energy of 6.2 eV is delivered to the system. Since the calculated CT state

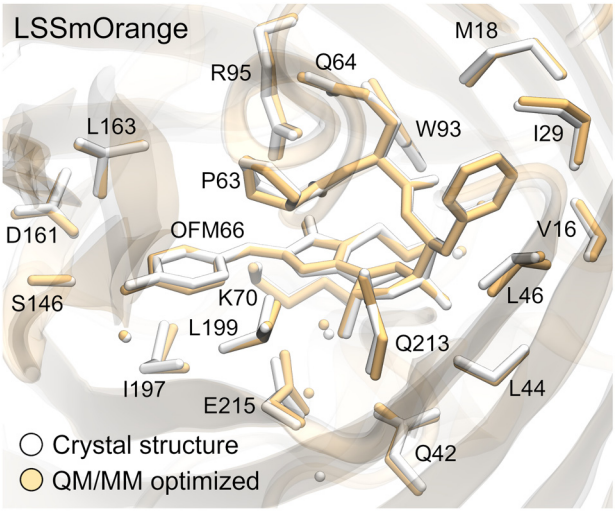

Fig. 9 Comparison of QM/MM optimized (orange) and crystal (white, PDB: 4Q7R[20]) structure of LSSmOrange. The chromophore and nearby amino acid side chains are shown as sticks and nearby water molecules as spheres. Hydrogen atoms are omitted for clarity.

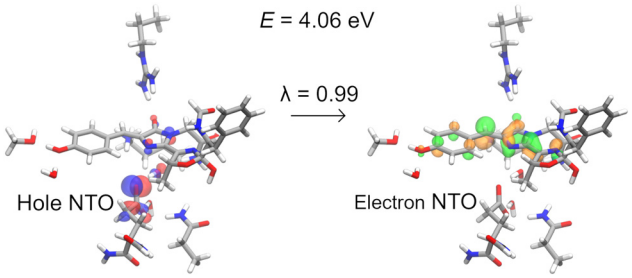

Fig. 10 NTOs of the lowest Glu215-chromophore CT state in LSSmOrange calculated with CC2/def2-TZVP. The rest of the system, not shown here, was described with polarizable embedding.

energies lie well below this value, a mechanism involving ET from Glu215 to the chromophore, analogous to that proposed for PA-GFP appears plausible. Interestingly, compared to PA-GFP, the energy of the lowest CT state is lowered by more than 0.5 eV, likely due to differences in the chromophore and/or the environment surrounding the Glu residue prone to decarboxylation.

3.2.2. EGFP. EGFP is not known to undergo photoinduced decarboxylation. Unlike PA-GFP, it has an anionic chromophore and neutral Glu222. From chemical intuition, a neutral Glu is a worse electron donor, and an anionic chromophore a worse electron acceptor, making ET from Glu222 to the chromophore unlikely. We tested whether our methodology would capture this effect. While most previous studies considered His148 to be neutral,[64,65] a relatively recent neutron diffraction experiment suggests it may be protonated.[66] Since this could influence Glu222–chromophore CT state energies, we examined both options. After QM/MM optimization the chromophore environment remained close to the crystal structure (Fig. S16). However, a protonated His148 led to a large displacement of Arg168, suggesting that the neutral form may be more likely. Fragment decomposition of electron and hole populations (Fig. S17 and S20) of excited states calculated with CC2/def2-TZVP shows no Glu222–chromophore CT states up to 5.49 eV (neutral His148) and 5.58 eV (protonated His148). Using CAM-B3LYP/def2-TZVP (Fig. S19 and S22), such CT states appear only at 5.92 eV (neutral His148) and 6.02 eV (protonated His148), nearly 1 eV higher than in PA-GFP calculated at the same level of theory. These results confirm that the absence of photoinduced decarboxylation in EGFP can be attributed to the high energy required for ET from Glu222 to the chromophore.

3.2.3. DsRed. Overall, the optimized structure of DsRed closely resembles the crystal structure (Fig. 11, and Fig. S23, S24). The most notable difference is the conformation of the *N*-acylimine group, which adopts a *gauche* geometry with a C═N−C═O dihedral angle of 87°. In contrast, the crystal structure features two conformations with similar occupancies: one s-*cis* and one s-*trans*. A similar conformational change was previously observed in another computational study of DsRed.[67] This discrepancy has been attributed to the refinement of the original crystal structure, while a more accurate QM-based refinement results in the *gauche* conformation.[68] Additional structural differences include the shortening of the H-bond between the chromophore and Lys163 from 3.1 to 2.6 Å, and the lengthening of the H-bond between the phenolate oxygen of the chromophore and a water molecule from 2.7 to 3.1 Å.

Habuchi *et al.*[8] reported photoinduced decarboxylation of Glu215 in DsRed upon illumination with a nanosecond-pulsed 532 nm laser. This step was proposed to precede chromophore *cis*/*trans* isomerization by creating free volume around the chromophore. The decarboxylation was supported by mass spectrometry, which revealed a 44 Da mass loss consistent with CO_2 release, and by FT-IR spectroscopy, which suggested that

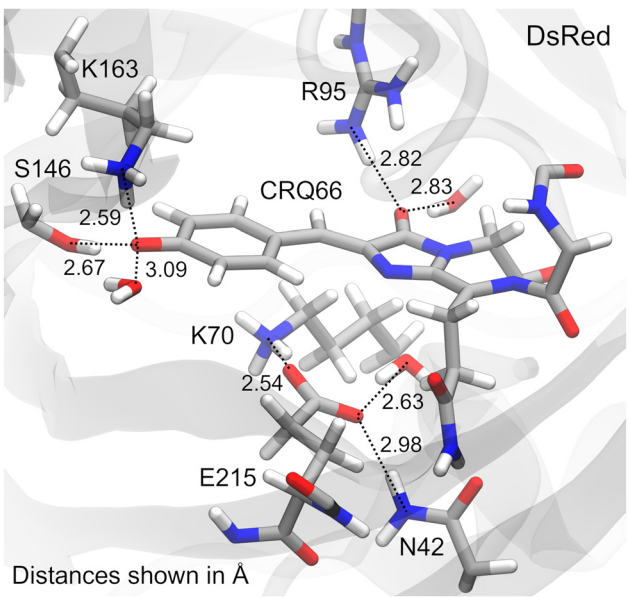

Fig. 11 QM/MM optimized structure of DsRed. The QM region used in the excited-state calculations is shown as sticks. The remainder of the system was described with polarizable embedding.

the decarboxylated residue is an anionic glutamate. Kinetic measurements further indicated that the process involves two consecutive one photon absorptions. Drobizhev *et al.*[17] later showed that decarboxylation in DsRed2, a variant with similar optical properties to DsRed, requires absorption of three 790 nm photons (totaling 4.7 eV). Given the expected lower electron affinity of the anionic chromophore, they argued that the initial triggering step is not ET from Glu215 to the chromophore (as proposed for GFPs), but rather ET from the chromophore to a nearby positively charged residue, such as Lys70.

To evaluate this hypothesis, we calculated vertical excitation energies on the optimized DsRed structure (Fig. 11) using CC2/def2-TZVP. Fragment decomposition of electron and hole populations of all ten computed excited states is shown in Fig. S25. Notably, state 9 (4.91 eV) exhibits dominant CT character from Glu215 to the anionic chromophore (Fig. 12). The calculated excitation energy of 4.91 eV is close to the experimentally reported threshold for decarboxylation (~4.7 eV) and is not much higher than the 4.69 eV we calculated for PA-GFP. This suggests that ET from Glu215 to the anionic chromophore is possible in DsRed, supporting a mechanism similar to that proposed for GFPs.

Although this may seem unexpected given the chromophore's negative charge, the presence of several nearby positively charged residues (Fig. 11) likely provides significant stabilization of the dianionic state. Such dianionic states in red fluorescent proteins have previously been proposed as intermediates potentially involved in bleaching and phototoxicity,[69] although it remained unclear whether other amino acid residues could serve as electron donors.

Among the ten calculated states, only state 10 appeared to exhibit ET from the chromophore to another residue (Fig. S25). However, closer inspection of the NTOs (Fig. S27) indicates that this state is likely an artifact caused by electron spill-out into the MM environment, a known issue in QM/MM calculations and previously reported for polarizable embedding calculations of FPs.[70] We note that the previously discussed state 9 does not exhibit such artifacts. Due to the computational cost of CC2, we did not compute more than ten excited states. However, calculation of 20 excited states using CAM-B3LYP/def2-TZVP with the same QM region and polarizable embedding confirmed the presence of Glu215-chromophore CT states, but still did not reveal any CT states from the chromophore to the environment besides those attributable to electron spill-out (Fig. S28). To assess whether nearby aromatic residues (Trp93, Trp141, Phe177) could act as electron acceptors, we individually added their side chains to the QM region and repeated the excited-state calculations. No states involving ET to these residues were observed (Fig. S29–S31).

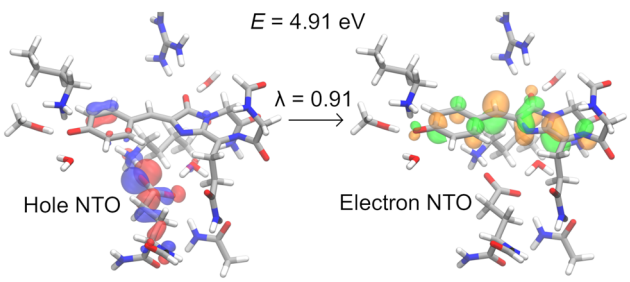

Fig. 12 NTOs of the Glu215-chromophore CT state in DsRed calculated with CC2/def2-TZVP. The rest of the system, not shown here, was described with polarizable embedding.

In summary, our results suggest that photoinduced decarboxylation in DsRed likely follows a similar mechanism as in GFPs, involving ET from Glu215 to the anionic chromophore. ET from the chromophore to surrounding residues appears unlikely. Our focus was specifically on the initial ET step, which is likely the trigger for decarboxylation. However, it would also be interesting to investigate the subsequent steps, as their mechanisms remain poorly understood, even in the case of GFPs.

3.2.4. LSSmKate1 and PSLSSmKate. As our final investigation, we focus on LSSmKate1 and PSLSSmKate. LSSmKate1 is a red FP, featuring a chromophore similar to DsRed but in a neutral protonation state.[21] LSSmKate1 does not undergo photoinduced decarboxylation when illuminated with 405 nm light from a high intensity LED setup, but introducing mutations Gly143Ser and Tyr67Lys results in photoinduced decarboxylation of Glu215.[9] This mutant is named PSLSSmKate. We applied our methodology to both variants with the goal of explaining the differences in their properties.

As in the previous systems, the choice of the protonation state of the chromophore and of the Glu residue prone to decarboxylation is crucial. Based on the absorption around 450 nm, it is reasonable to assume that the chromophore exists predominantly in a neutral state in both variants. In LSSmKate1, Glu215 is likely deprotonated since, as based on the crystal structure, it forms an H-bond with the positively charged Arg195. After QM/MM geometry optimization, the structure surrounding the chromophore remained similar to the crystal structure (Fig. 13 and Fig. S33), with the most significant change being the displacement of a structural water (residue 285 in the crystal structure) by 4.39 Å to form a hydrogen bond with Glu215. A similar movement was also observed for LSSmOrange.

In PSLSSmKate, the authors suggest that Glu215 is neutral because, in the crystal structure, one oxygen of its carboxylate group lies within H-bonding distance (3.12 Å) of the N2 nitrogen of the chromophore's imidazolinone ring.[9] The relatively low resolution of the crystal structure (2.70 Å) may limit the accuracy of the reported interatomic distances, so we have prepared PSLSSmKate systems with Glu215 modeled in both neutral and deprotonated states. In both cases, after QM/MM geometry optimization, the structure around the chromophore remained generally similar to the crystal structure (Fig. 13 and Fig. S38), making it difficult to determine which Glu215 protonation state is more likely based on structure alone. In the neutral Glu215 system, the H-bond between Glu215 and the N2 nitrogen of the chromophore's imidazolinone ring increased to 3.38 Å with an angle of 146°, indicating a weaker H-bond than

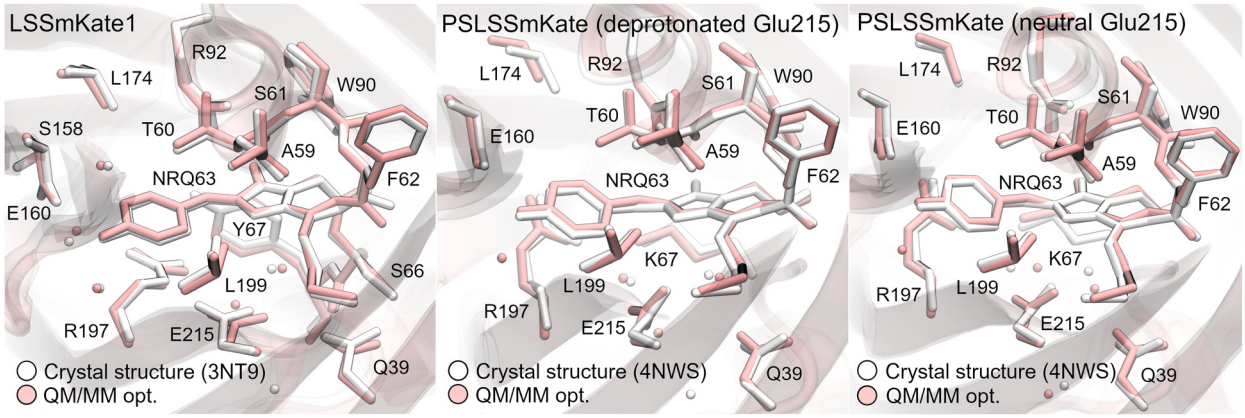

Fig. 13 Comparison of QM/MM optimized (pink) and crystal (white) structures of LSSmKate1[21] (left) and PSLSSmKate[9] with neutral (middle) and deprotonated (right) Glu215. The chromophore and nearby amino acid residues are shown as sticks and nearby water molecules as spheres. Hydrogen atoms are omitted for clarity.

initially inferred from the crystal structure. Additionally, Glu215 was no longer connected to Arg197 *via* a structural water molecule. In the deprotonated Glu215 system, Glu215 remained connected to both Arg197 and Lys70 through a structural water molecule, with the most notable change being the displacement of a structural water (residue 327 in the crystal structure) by 4.23 Å to form an H-bond with Glu215, similar to the movements observed in LSSmKate1 and LSSmOrange.

The CC2/def2-TZVP excitation spectra of LSSmKate1 and the two PSLSSmKate systems were compared, focusing on the energies of the Glu215–chromophore CT states (Table 2 and Tables S34, S39, S42). In the case of PSLSSmKate with deprotonated Glu215, the Glu215–chromophore CT states are more than 1.1 eV lower in energy than in LSSmKate1, despite both systems having the same structure and protonation states of the chromophore and Glu215. The energy of the lowest CT state is 3.66 eV and its non-zero oscillator strength of 0.02 may suggest it is possible to populate it directly with prolonged 405 nm (3.06 eV) illumination. This indicates that the loss of direct contact between Arg195 and Glu215 in PSLSSmKate (Fig. 13) significantly lowers the CT state energy, suggesting that only differences in the immediate environment of Glu215 are sufficient to explain the distinct photoactivation properties of the two variants. For neutral Glu215, however, no CT states are found up to 5.17 eV, and in that case our results cannot account for the differences between the two variants. Assuming deprotonated Glu215 in PSLSSmKate reproduces the experimental 0.1 eV blue shift of the first excited state relative to LSSmKate1, whereas assuming neutral Glu215 instead predicts a 0.2 eV red shift. Taken together, these results point toward deprotonated Glu215 as the more likely protonation state, although a more complex mechanism involving proton-coupled ET with neutral Glu215 cannot be ruled out.

The CT state energies in LSSmKate1 are very similar to those calculated for DsRed, which has been reported to undergo photoinduced decarboxylation. This possible discrepancy may be explained by differences in the experimental laser setups: an LED diode array was used for LSSmKate1, whereas a nanosecond pulsed laser was employed for DsRed. If the process requires multiphoton absorption, it is possible that the LED source cannot efficiently induce two-photon absorption, while the pulsed laser can. Thus, reexamining LSSmKate1 with a high-intensity pulsed laser could potentially reveal some degree of photoactivation.

Table 2 Energies of the first excited state and the Glu215–chromophore CT states of LSSmKate1 and PSLSSmKate, computed with CC2/def2-TZVP and polarizable embedding for the MM region

System	S_1 / eV	CT / eV
LSSmKate1	3.02	4.83, 4.86
PSLSSmKate[a]	3.12	3.66, 4.09, 4.55
PSLSSmKate[b]	2.85	>5.17

[a] Deprotonated Glu215. [b] Neutral Glu215.

4. Conclusions

Our first goal was to assess how different QM/MM methodological choices affect the energies of CT states. For PA-GFP, we find that CT states are considerably more sensitive than locally excited S_1 states. In particular, inclusion of the environment around the glutamic acid residue prone to decarboxylation in the QM region is crucial. We also find that the RVS approximation for correlated methods introduces larger errors for CT states than for S_1, with CC2 being less sensitive than ADC(2). More broadly, these results suggest that the character of the excited state must be taken into account when benchmarking computational approaches.

Comparison of six FPs allowed us to propose a structure–function relationship for photoinduced decarboxylation. Our results reinforce the previously suggested mechanism in which ET from deprotonated glutamic acid to the neutral chromophore is the key step in PA-GFP, and show that the same mechanism is also possible in DsRed despite its anionic chromophore. Differences between LSSmKate1 and PSLSSmKate further illustrate

how the local environment of the glutamate residue significantly influences the energies of the CT states, providing a possible explanation for their distinct photoactivation properties. These findings suggest that in the rational design of photoactivatable FPs, the microenvironment of the glutamic acid residue is a promising tuning target.

Key open questions remain, particularly how the CT state is initially populated. Addressing this requires going beyond our model of calculating vertical excitation energies at the ground-state minimum, ideally by performing non-adiabatic excited-state QM/MM MD simulations. The steps following CO_2 loss also remain poorly understood and warrant further study.

Altogether, this work advances both the computational modeling of FPs and the mechanistic understanding of photo-induced decarboxylation. It provides a possible strategy for a more rational and targeted design of photoactivatable proteins. Additionally, the six prepared FP systems could be very useful for future benchmarking studies. Finally, we believe that the use of TheoDORE[60] for the rapid analysis of large numbers of excited states could be useful for exploring the mechanisms of other photoinduced reactions.

Author contributions

Janko Čivić: conceptualisation, data curation, formal analysis, investigation, methodology, writing – original draft, writing – review & editing. Hideaki Mizuno: supervision, writing – review & editing. Jeremy N. Harvey: conceptualisation, supervision, writing – review & editing.

Conflicts of interest

There are no conflicts to declare.

Data availability

The data supporting this article have been included as part of the supplementary information (SI). supplementary information: detailed overview of system preparation steps, including explicit definitions of the QM regions for each variant. Additional tables and figures supporting the main results are also included. For all variants, we provide TheoDORE fragment decomposition plots of electron and hole populations, together with natural transition orbitals for each excited state. See DOI: **https://doi.org/10.1039/d5cp03991f**.

Additional files supporting the paper and enabling replication of the results are available through the KU Leuven Research Data Repository (**https://doi.org/10.48804/OTLZEI**); including classical molecular dynamics parameter files; initial and optimized system structures; trajectory of 100 snapshots from QM/MM MD simulations of PA-GFP; and QM region coordinates with their corresponding embedding potentials used in excited-state QM/MM calculations.

Acknowledgements

Janko Čivić acknowledges the support of the Research Foundation – Flanders (FWO) through grant 1120025N. Hideaki Mizuno acknowledges the support of KU Leuven through a Category 1 research grant (C14/22/087) and of the FWO through project grant G0A8523N. During manuscript writing ChatGPT was used in order to improve the readability and language of the text. After using this tool, the output was reviewed and edited as needed. The authors take full responsibility for the content of the publication.

References

1 O. Shimomura, F. H. Johnson and Y. Saiga, *J. Cell. Comp. Physiol.*, 1962, **59**, 223–239.
2 M. Chalfie, Y. Tu, G. Euskirchen, W. W. Ward and D. C. Prasher, *Science*, 1994, **263**, 802–805.
3 O. Shimomura, *FEBS Lett.*, 1979, **104**, 220–222.
4 A. Acharya, A. M. Bogdanov, B. L. Grigorenko, K. B. Bravaya, A. V. Nemukhin, K. A. Lukyanov and A. I. Krylov, *Chem. Rev.*, 2017, **117**, 758–795.
5 E. Betzig, G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz and H. F. Hess, *Science*, 2006, **313**, 1642–1645.
6 J. van Thor, T. Gensch, K. Hellingwerf and L. Johnson, *Nat. Struct. Biol.*, 2002, **9**, 37–41.
7 J. N. Henderson, R. Gepshtein, J. R. Heenan, K. Kallio, D. Huppert and S. J. Remington, *J. Am. Chem. Soc.*, 2009, **131**, 4176–4177.
8 S. Habuchi, M. Cotlet, T. Gensch, T. Bednarz, S. Haber-Pohlmeier, J. Rozenski, G. Dirix, J. Michiels, J. Vanderleyden, J. Heberle, F. C. De Schryver and J. Hofkens, *J. Am. Chem. Soc.*, 2005, **127**, 8977–8984.
9 K. D. Piatkevich, B. P. English, V. N. Malashkevich, H. Xiao, S. C. Almo, R. H. Singer and V. V. Verkhusha, *Chem. Biol.*, 2014, **21**, 1402–1414.
10 E. Fron, H. De Keersmaecker, S. Rocha, Y. Baeten, G. Lu, H. Uji-i, M. Van der Auweraer, J. Hofkens and H. Mizuno, *J. Phys. Chem. B*, 2015, **119**, 14880–14891.
11 F. V. Subach, V. N. Malashkevich, W. D. Zencheck, H. Xiao, G. S. Filonov, S. C. Almo and V. V. Verkhusha, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 21097–21102.
12 J. D. Slocum and L. J. Webb, *J. Phys. Chem. Lett.*, 2017, **8**, 2862–2868.
13 A. F. Bell, D. Stoner-Ma, R. M. Wachter and P. J. Tonge, *J. Am. Chem. Soc.*, 2003, **125**, 6919–6926.
14 F. Langhojer, F. Dimler, G. Jung and T. Brixner, *Biophys. J.*, 2009, **96**, 2763–2770.
15 L. Ding, L. W. Chung and K. Morokuma, *J. Phys. Chem. B*, 2013, **117**, 1075–1084.
16 B. L. Grigorenko, A. V. Nemukhin, D. I. Morozov, I. V. Polyakov, K. B. Bravaya and A. I. Krylov, *J. Chem. Theory Comput.*, 2012, **8**, 1912–1920.
17 M. Drobizhev, C. Stoltzfus, I. Topol, J. Collins, G. Wicks, A. Mikhaylov, L. Barnett, T. E. Hughes and A. Rebane, *J. Phys. Chem. B*, 2014, **118**, 9167–9179.

18 A. Royant and M. Noirclerc-Savoye, *J. Struct. Biol.*, 2011, **174**, 385–390.
19 J. L. Tubbs, J. A. Tainer and E. D. Getzoff, *Biochemistry*, 2005, **44**, 9833–9840.
20 S. Pletnev, D. M. Shcherbakova, O. M. Subach, N. V. Pletneva, V. N. Malashkevich, S. C. Almo, Z. Dauter and V. V. Verkhusha, *PLoS One*, 2014, **9**, e99136.
21 K. D. Piatkevich, V. N. Malashkevich, S. C. Almo and V. V. Verkhusha, *J. Am. Chem. Soc.*, 2010, **132**, 10762–10770.
22 D. A. Case, H. M. Aktulga, K. Belfon, D. S. Cerutti, G. A. Cisneros, V. W. D. Cruzeiro, N. Forouzesh, T. J. Giese, A. W. Götz, H. Gohlke, S. Izadi, K. Kasavajhala, M. C. Kaymak, E. King, T. Kurtzman, T.-S. Lee, P. Li, J. Liu, T. Luchko, R. Luo, M. Manathunga, M. R. Machado, H. M. Nguyen, K. A. O'Hearn, A. V. Onufriev, F. Pan, S. Pantano, R. Qi, A. Rahnamoun, A. Risheh, S. Schott-Verdugo, A. Shajan, J. Swails, J. Wang, H. Wei, X. Wu, Y. Wu, S. Zhang, S. Zhao, Q. Zhu, T. E. I. Cheatham, D. R. Roe, A. Roitberg, C. Simmerling, D. M. York, M. C. Nagan and K. M. J. Merz, *J. Chem. Inf. Model.*, 2023, **63**, 6183–6191.
23 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.
24 C. Tian, K. Kasavajhala, K. A. A. Belfon, L. Raguette, H. Huang, A. N. Migues, J. Bickel, Y. Wang, J. Pincay, Q. Wu and C. Simmerling, *J. Chem. Theory Comput.*, 2020, **16**, 528–552.
25 A. Sengupta, Z. Li, L. F. Song, P. Li and K. M. J. Merz, *J. Chem. Inf. Model.*, 2021, **61**, 869–880.
26 C. I. Bayly, P. Cieplak, W. Cornell and P. A. Kollman, *J. Phys. Chem.*, 1993, **97**, 10269–10280.
27 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery Jr, J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, *Gaussian 16 Revision C.01*, 2016, Gaussian Inc., Wallingford CT.
28 J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 1157–1174.
29 D. Case, H. Aktulga, K. Belfon, I. Ben-Shalom, J. Berryman, S. Brozell, D. Cerutti, T. Cheatham III, G. Cisneros, V. Cruzeiro, T. Darden, R. Duke, G. Giambasu, M. Gilson, H. Gohlke, A. Goetz, R. Harris, S. Izadi, S. Izmailov, K. Kasavajhala, M. Kaymak, E. King, A. Kovalenko, T. Kurtzman, T. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, M. Machado, V. Man, M. Manathunga, K. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, K. O'Hearn, A. Onufriev, F. Pan, S. Pantano, R. Qi, A. Rahnamoun, D. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, A. Shajan, J. Shen, C. Simmerling, N. Skrynnikov, J. Smith, J. Swails, R. Walker, J. Wang, J. Wang, H. Wei, R. Wolf, X. Wu, Y. Xiong, Y. Xue, D. York, S. Zhao and P. Kollman, *Amber 2022*, University of California, San Francisco, 2022.
30 R. Salomon-Ferrer, A. W. Götz, D. Poole, S. Le Grand and R. C. Walker, *J. Chem. Theory Comput.*, 2013, **9**, 3878–3888.
31 A. W. Götz, M. J. Williamson, D. Xu, D. Poole, S. Le Grand and R. C. Walker, *J. Chem. Theory Comput.*, 2012, **8**, 1542–1555.
32 S. Le Grand, A. W. Götz and R. C. Walker, *Comput. Phys. Commun.*, 2013, **184**, 374–380.
33 H. G. Petersen, *J. Chem. Phys.*, 1995, **103**, 3668–3679.
34 D. J. Tobias and C. L. Brooks III, *J. Chem. Phys.*, 1988, **89**, 5115–5127.
35 S. Grimme, J. G. Brandenburg, C. Bannwarth and A. Hansen, *J. Chem. Phys.*, 2015, **143**, 054107.
36 A. W. Götz, M. A. Clark and R. C. Walker, *J. Comput. Chem.*, 2014, **35**, 95–108.
37 F. Neese, F. Wennmohs, U. Becker and C. Riplinger, *J. Chem. Phys.*, 2020, **152**, 224108.
38 J. M. H. Olsen and P. Reinholdt, *PyFraME: Python framework for fragment-based multiscale embedding*, 2021, DOI: **10.5281/zenodo.4899311**.
39 D. Grabarek and T. Andruniów, *J. Chem. Theory Comput.*, 2020, **16**, 6439–6455.
40 D. W. Zhang and J. Z. H. Zhang, *J. Chem. Phys.*, 2003, **119**, 3599–3605.
41 P. Söderhjelm and U. Ryde, *J. Phys. Chem. A*, 2009, **113**, 617–627.
42 L. Gagliardi, R. Lindh and G. Karlström, *J. Chem. Phys.*, 2004, **121**, 4494–4500.
43 O. Vahtras, LoProp for Dalton, 2014, https://zenodo.org/records/13276.
44 K. Aidas, C. Angeli, K. L. Bak, V. Bakken, R. Bast, L. Boman, O. Christiansen, R. Cimiraglia, S. Coriani, P. Dahle, E. K. Dalskov, U. Ekström, T. Enevoldsen, J. J. Eriksen, P. Ettenhuber, B. Fernández, L. Ferrighi, H. Fliegl, L. Frediani, K. Hald, A. Halkier, C. Hättig, H. Heiberg, T. Helgaker, A. C. Hennum, H. Hettema, E. Hjertenæs, S. Høst, I.-M. Høyvik, M. F. Iozzi, B. Jansík, H. J. A. Jensen, D. Jonsson, P. Jørgensen, J. Kauczor, S. Kirpekar, T. Kjærgaard, W. Klopper, S. Knecht, R. Kobayashi, H. Koch, J. Kongsted, A. Krapp, K. Kristensen, A. Ligabue, O. B. Lutnæs, J. I. Melo, K. V. Mikkelsen, R. H. Myhre, C. Neiss, C. B. Nielsen, P. Norman, J. Olsen, J. M. H. Olsen, A. Osted, M. J. Packer, F. Pawlowski, T. B. Pedersen, P. F. Provasi, S. Reine, Z. Rinkevicius, T. A. Ruden, K. Ruud, V. V. Rybkin, P. Salek, C. C. M. Samson, A. S. de Merás, T. Saue,

S. P. A. Sauer, B. Schimmelpfennig, K. Sneskov, A. H. Steindal, K. O. Sylvester-Hvid, P. R. Taylor, A. M. Teale, E. I. Tellgren, D. P. Tew, A. J. Thorvaldsen, L. Thøgersen, O. Vahtras, M. A. Watson, D. J. D. Wilson, M. Ziolkowski and H. Ågren, *WIREs Comput. Mol. Biosci.*, 2014, **4**, 269–284.

45 Dalton, a molecular electronic structure program, Release 2020.1, 2022, see **https://daltonprogram.org**.

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

47 J.-D. Chai and M. Head-Gordon, *Phys. Chem. Chem. Phys.*, 2008, **10**, 6615–6620.

48 O. Christiansen, H. Koch and P. Jørgensen, *Chem. Phys. Lett.*, 1995, **243**, 409–418.

49 J. Schirmer, *Phys. Rev. A*, 1982, **26**, 2395–2416.

50 F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297–3305.

51 F. Weigend, *Phys. Chem. Chem. Phys.*, 2006, **8**, 1057–1065.

52 D. Rappoport and F. Furche, *J. Chem. Phys.*, 2010, **133**, 134105.

53 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.

54 S. Grimme, S. Ehrlich and L. Goerigk, *J. Comput. Chem.*, 2011, **32**, 1456–1465.

55 R. Send, C.-M. Suomivuori, V. R. I. Kaila and D. Sundholm, *J. Phys. Chem. B*, 2015, **119**, 2933–2945.

56 TURBOMOLE V7.8 2023, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; available from **https://www.turbomole.org**.

57 S. G. Balasubramani, G. P. Chen, S. Coriani, M. Diedenhofen, M. S. Frank, Y. J. Franzke, F. Furche, R. Grotjahn, M. E. Harding, C. Hättig, A. Hellweg, B. Helmich-Paris, C. Holzer, U. Huniar, M. Kaupp, A. Marefat Khah, S. Karbalaei Khani, T. Müller, F. Mack, B. D. Nguyen, S. M. Parker, E. Perlt, D. Rappoport, K. Reiter, S. Roy, M. Rückert, G. Schmitz, M. Sierka, E. Tapavicza, D. P. Tew, C. van Wüllen, V. K. Voora, F. Weigend, A. Wodynski and J. M. Yu, *J. Chem. Phys.*, 2020, **152**, 184107.

58 T. Schwabe, K. Sneskov, J. M. Haugaard Olsen, J. Kongsted, O. Christiansen and C. Hättig, *J. Chem. Theory Comput.*, 2012, **8**, 3274–3283.

59 A. Marefat Khah, S. Karbalaei Khani and C. Hättig, *J. Chem. Theory Comput.*, 2018, **14**, 4640–4650.

60 F. Plasser, *J. Chem. Phys.*, 2020, **152**, 084108.

61 J.-Y. Hasegawa, K. Fujimoto, B. Swerts, T. Miyahara and H. Nakatsuji, *J. Comput. Chem.*, 2007, **28**, 2443–2452.

62 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.*, 2013, **15**, 4735–4743.

63 D. M. Shcherbakova, M. A. Hink, L. Joosen, T. W. J. Gadella and V. V. Verkhusha, *J. Am. Chem. Soc.*, 2012, **134**, 7913–7923.

64 T. Schwabe, M. T. P. Beerepoot, J. M. H. Olsen and J. Kongsted, *Phys. Chem. Chem. Phys.*, 2014, **17**, 2582–2588.

65 C. Filippi, F. Buda, L. Guidoni and A. Sinicropi, *J. Chem. Theory Comput.*, 2012, **8**, 112–124.

66 C. Shibazaki, R. Shimizu, Y. Kagotani, A. Ostermann, T. E. Schrader and M. Adachi, *J. Phys. Chem. Lett.*, 2020, **11**, 492–496.

67 N. H. List, J. M. H. Olsen, H. J. A. Jensen, A. H. Steindal and J. Kongsted, *J. Phys. Chem. Lett.*, 2012, **3**, 3513–3521.

68 Y.-W. Hsiao, E. Sanchez-Garcia, M. Doerr and W. Thiel, *J. Phys. Chem. B*, 2010, **114**, 15413–15423.

69 R. B. Vegh, K. B. Bravaya, D. A. Bloch, A. S. Bommarius, L. M. Tolbert, M. Verkhovsky, A. I. Krylov and K. M. Solntsev, *J. Phys. Chem. B*, 2014, **118**, 4527–4534.

70 D. Grabarek and T. Andruniów, *J. Chem. Phys.*, 2020, **153**, 215102.

