

# Modeling Spectral Tuning in Red Fluorescent Proteins Using the Dipole Moment Variation upon Excitation

*Maria G. Khrenova,<sup>a,b</sup> Fedor D. Mulashkin,<sup>a</sup> Alexander V. Nemukhin<sup>a,c</sup>*

<sup>a</sup> Department of Chemistry, Lomonosov Moscow State University, Moscow, 119991, Russian Federation

<sup>b</sup> Bach Institute of Biochemistry, Federal Research Centre “Fundamentals of Biotechnology”, Russian Academy of Sciences, Moscow, 119071 Russian Federation

<sup>c</sup> Emanuel Institute of Biochemical Physics, Russian Academy of Sciences, Moscow, 119334, Russian Federation

\* Corresponding author:

Prof. Alexander V. Nemukhin, Lomonosov Moscow State University, Russian Federation

E-mail: [anem@lcc.chem.msu.ru](mailto:anem@lcc.chem.msu.ru), [anemukhin@yahoo.com](mailto:anemukhin@yahoo.com)

† *Electronic supplementary information (ESI) available.*

## Abstract

We describe a model for spectral tuning in red fluorescent proteins (RFPs) based on the relation between an electronic structure descriptor, the dipole moment variation upon excitation (DMV), and the excitation energy of a protein. This approach aims to overcome a problem of accurate prediction of excitation energies in RFPs, which span a very narrow window of band maxima. The latter roughly corresponds to the energy range of 0.1 eV, which is comparable with typical errors in calculations of the excitation energy by conventional quantum chemistry methods. In this work, we demonstrate a strong quantitative correlation between DMV values, obtained computationally with modest efforts, and excitation energies  $\Delta E_{\text{ex}}$  at the experimental excitation band maxima for a series of RFPs with the bands between 570 and 605 nm. Protein models are constructed by motifs of the relevant crystal structures, and atomic coordinates are optimized in quantum mechanics/molecular mechanics (QM/MM) calculations with QM-subsystems composed of large chromophore-containing regions. DMV values are evaluated with the electron density computed at the TDDFT level using several functionals and basis sets. We demonstrate that the results obtained with the CAM-B3LYP, BHLYP and M06-2X functionals demonstrate favorable correlations between DMV and  $\Delta E_{\text{ex}}$  with the mean absolute error less than 0.01 eV. Taking into account the solid theoretical grounds of the relation between the DMV and the excitation energy in fluorescent proteins, the described modeling strategy presents a rational tool for spectral tuning in these efficient markers for *in vivo* imaging.

## Introduction

It is hard to overestimate the significance of fluorescent proteins (FPs) for modern applications in bioimaging.<sup>1-4</sup> Of a special importance are the proteins with absorption and emission in the red to near infrared region required for tissue transparency.<sup>5</sup> Red fluorescent proteins (RFPs) are also requested as acceptors in the FRET pairs;<sup>6-10</sup> in this case, it is important to provide a large overlap between the fluorescence band of a donor and the excitation band of an acceptor. In this respect, fine spectral tuning, including rational design, is an important issue in the efforts to improve FP properties, expand their applications and propose the ways to create novel efficient biomarkers.<sup>11-16</sup>

From the theoretical side, rational design of FPs requires fast and reliable estimates of transition energies; however, applications of modern quantum chemistry molecular modeling methods can hardly result in excitation energies in fluorescent proteins with errors less than 0.1 eV.<sup>17-23</sup> The excitation band maximum of the green fluorescent protein (GFP) is 475 nm,<sup>24</sup> and the 0.1 eV uncertainty in energy corresponds to the uncertainty of about 20 nm on the wavelength scale. Thus, conventional calculations of the  $S_{0,\min} - S_1$  transition energy can principally discriminate between such species as the GFP and the yellow fluorescent protein (YFP),<sup>25</sup> a GFP variant with shifted  $\lambda_{\text{ex}}$  to 514 nm due to the introduced  $\pi$ -stacking interaction of the chromophore with the tyrosine residue. Accurate computational estimates of the electronic excitation energies in RFPs are considerably more difficult, since the 40 nm region in the red part of spectrum (575 – 605 nm) corresponds to just this range of 0.1 eV in the energetic scale. Therefore, the request for spectral tuning in RFPs encourages seeking alternative theoretical approaches to evaluate excitation energies, including the development of quantitative structure - property relationship (QSPR) models.<sup>26-28</sup>

In this respect, the use of the dipole moment variation upon excitation (DMV) electronic structure descriptor, which reflects redistribution of charge upon electron excitation,<sup>29</sup> is an attractive strategy. As shown by Drobizhev et al., the spectral tuning in fluorescent proteins, including RFPs, is associated with the internal electrostatic field of the protein that affects the excess dipole moment of the chromophore or its DMV.<sup>30-33</sup> Due to the strong electric field within the  $\beta$ -barrel of FPs, the second-order Stark effect is observed, employing that the relation between the electron excitation energy and DMV is quadratic. The Stark tuning of the absorption band maximum is further supported on a large library of the GFP family.<sup>34</sup> From theoretical perspective, this correlation has been successfully explored in modeling the yellow-to-cyan tuning in the GFP family,<sup>26</sup> as well as in understanding the origin of the shifts in the electronic excitation energies of different anionic FP chromophores in the  $\pi$ -stacked complexes with aromatic molecules.<sup>27,28</sup>

In this work, we examine a quantitative correlation between the computationally derived DMV values and the experimental electronic excitation energies at the excitation band maxima,  $\Delta E_{\text{ex}}$ , for a series of RFPs, which operate with the DsRed-type chromophore.<sup>35,36</sup> Their excitation band maxima range between 570 and 605 nm because of the extended conjugated  $\pi$ -system as compared with the GFP-type chromophore. Structures of full-atom protein molecules simulating 11 diverse RFPs are optimized at the quantum mechanics/molecular mechanics (QM(PBE0-D3/cc-pvdz)/MM(AMBER)) level. Calculations of electron densities in the ground and excited states are performed at the TDDFT level using 7 different DFT functionals and two basis sets followed by evaluation of DMV values. The results demonstrate a favorable correlation between the experimentally obtained excitation energies,  $\Delta E_{\text{ex}}$ , and the computed DMVs, assuming a quadratic dependence of  $\Delta E_{\text{ex}}$  on DVM.

## Material and methods

### *Protein models*

We select for our analysis the diverse RFPs (see Table 1) originated from different ancestors. The degree of similarity between the mRojo and RDSmCherry families is large (~90%), but between eqFP670, mKeima and the mRojo family it does not exceed 60%, according to the Muscle alignment analysis.<sup>37</sup> All these proteins share the DsRed-type chromophore and conservative residues in the chromophore-binding site.

Full-atom model systems were constructed using the available X-ray structures of proteins<sup>13,14,38-40</sup> or by introducing point mutations to the PDB entries following the description in the experimental papers.<sup>13,14,41,42</sup> Table 1 contains the names of the selected proteins, experimental data on their excitation bands and the relevant structural information. Hydrogen atoms were added using the Reduce program<sup>43</sup> and manually checked. Model systems were solvated with water molecules followed by preliminary structure optimization at the molecular mechanics level. Final optimization steps were performed using quantum mechanics/molecular mechanics (QM/MM) method. The QM subsystems are composed of 100 – 140 atoms depending on a specific protein. They include the entire chromophore, the nearby amino-acid residues and water molecules. The energies and forces in the QM subsystems were calculated at the PBE0-D3/cc-pvdz level.<sup>44-46</sup> The MM parts were modeled with the AMBER<sup>47</sup> force field and TIP3P<sup>48</sup> for water molecules. The electronic embedding scheme was utilized and all point charges from MM atoms were added to the one-electron part of the QM Hamiltonian. All calculations were performed with the NWChem program.<sup>49</sup>

**Table 1.** Fluorescent proteins considered in this work, their electronic excitation and structural properties. Mutations with respect to the PDB structures (third column) are specified in the last column.

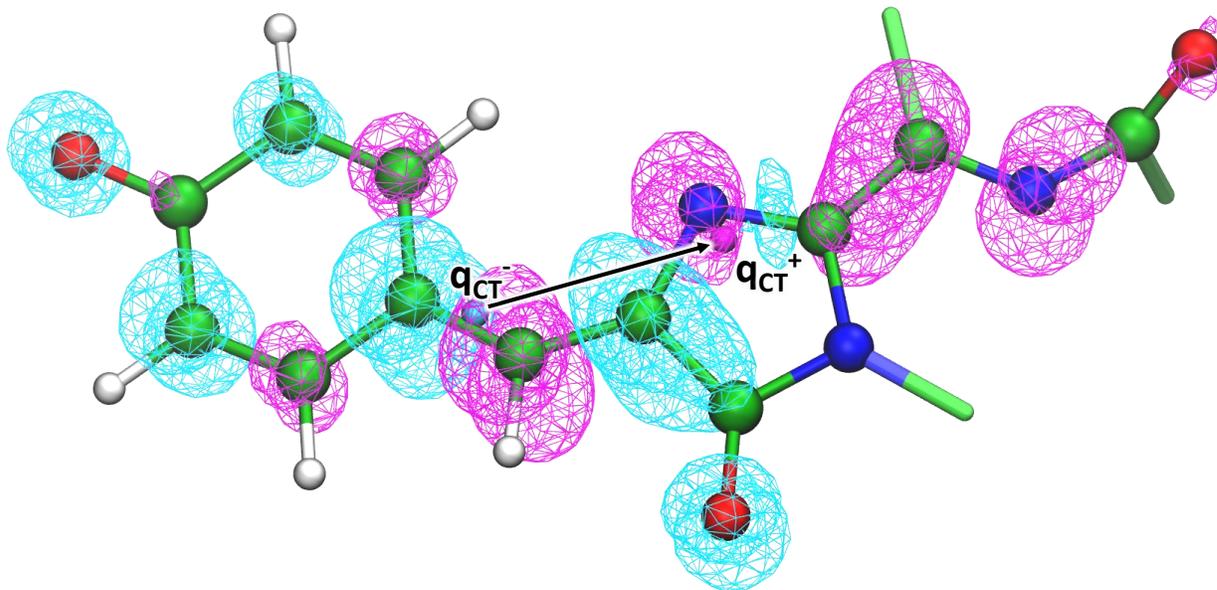
FP name	$\lambda_{\text{ex}}$ , nm / $\Delta E_{\text{ex}}$ , eV	PDB ID	Mutations
mRojoA	593 / 2.091 <sup>13</sup>	3NEZ <sup>14</sup>	-
mRojo-THSL	575 / 2.156 <sup>13</sup>	5H87 <sup>13</sup>	-
mRojo-VFAV	577 / 2.149 <sup>13</sup>	5H88 <sup>13</sup>	-
mRojo-TFAL	577 / 2.149 <sup>13</sup>	3NEZ <sup>14</sup>	V16T + V163L
mRojo-VYGV	571 / 2.171 <sup>13</sup>	5H89 <sup>13</sup>	-
mRojo-VYGL	570 / 2.175 <sup>13</sup>	5H89 <sup>13</sup>	V163L
mRojo-TYGV	575 / 2.156 <sup>13</sup>	5H89 <sup>13</sup>	V16T
RDSmCherry0.2	600 / 2.066 <sup>41</sup>	2H5Q <sup>38</sup>	K92N + K138C + K139R + S147T + I161G + Q163G + V195C + N196D + I197Y + T202L
RDSmCherry0.5	604 / 2.053 <sup>41</sup>	2H5Q <sup>38</sup>	V16S + A44C + K92N + K138C + K139R + A145P + S147T + I161G + Q163G + V195C + N196D + I197Y + T202L
mKeima	586 / 2.116 <sup>42</sup>	3IR8 <sup>39</sup>	-
eqFP670	605 / 2.050 <sup>40</sup>	4EDS <sup>40</sup>	-

### *DMV estimates*

To calculate the dipole moment variation upon excitation (DMV), the electron densities in the ground and excited states were calculated using the time-dependent density functional theory (TDDFT) approaches with the ORCA program.<sup>50</sup> We examined a set of 7 functionals including CAM-B3LYP,<sup>51</sup> LC-BLYP,<sup>52</sup> M06-2X,<sup>53</sup> BHLYP,<sup>54</sup> PBE0,<sup>44</sup> B1LYP<sup>55</sup> and B97M-D3BJ,<sup>56,57</sup> with the cc-pvdz<sup>46</sup> basis set. For the chromophores in the model systems occur in the anionic form, we also check the effect of adding diffuse functions to the basis set. Correspondingly, calculations with the CAM-B3LYP, M06-2X and BHLYP functionals were also carried out with the ma-def2-SVP<sup>58</sup> basis set.

The dipole moment variation upon excitation, DMV or  $\Delta\mu$ , is a product of the transferred charge between the ground and excited electronic states,  $q_{\text{CT}}$ , and the distance between barycenters of

the positive and negative parts of the electron density difference,  $q_{CT}^+$  and  $q_{CT}^-$  (Fig. 1). The  $q_{CT}$  value was obtained by the integration of the positive part of the  $S_{0,min}-S_1$  electron density difference. The DMV values were calculated using the Multiwfn software.<sup>59</sup>



**Figure 1.** Illustration of the electron density difference upon the  $S_{0,min} \rightarrow S_1$  transition for the FP chromophore. Green, red, blue and white colors refer to the carbon, oxygen, nitrogen and hydrogen atoms, respectively. The shown densities are calculated at the TDDFT CAM-B3LYP/cc-pvdz level. The magenta and cyan wireframe isosurfaces correspond to the increase and the decrease of electron density upon electronic excitation, respectively; the contour values are  $\pm 0.0015$  a.u. The black arrow identifies the negative charge transfer direction; it starts at the barycenter of the negative electron density difference ( $q_{CT}^-$ ) and ends at the barycenter of the positive electron density difference ( $q_{CT}^+$ ).

### ***Fitting procedure***

In this work, we rely on the correlation between the absolute value of DMV and the excitation energy,  $\Delta E_{ex}$ , obtained by the careful consideration of internal electric fields in fluorescent proteins. Assuming the second-order Stark effect, this relation reads<sup>33</sup>

$$\Delta E_{ex} = A \cdot DMV^2 + B \cdot DMV + C, \quad (1)$$

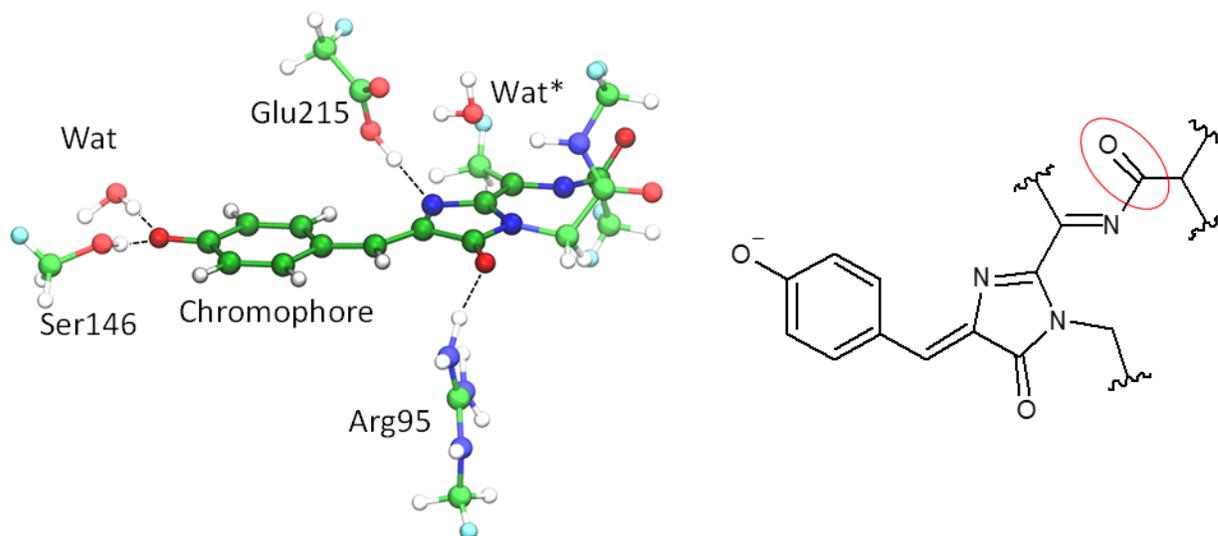
where  $A = -\frac{2}{hc\Delta\alpha_0}$ ,  $B = \frac{2}{hc\Delta\alpha_0}\Delta\mu_0$ ,  $C = \Delta E_0$ . The  $\Delta E_0$  is the hypothetical electronic transition energy of the chromophore in vacuum in the absence of electrostatic interactions with the protein. The  $\Delta\mu_0$  is the projection of DMV in vacuum on DMV in the protein cavity. The  $\Delta\alpha_0$  is the component of difference of polarizability tensors between two states along the DMV direction. Therefore, every coefficient in Eq. (1) has a well-defined physical meaning.

## Results

### Model systems

Atomic coordinates of the QM(PBE0-D3/cc-pvdz)/MM(AMBER) optimized structures in the ground electronic state of all model systems are available in Supporting Information. We use these coordinates for calculations of electron densities in the ground and excited states followed by evaluation of the DMV values.

The common motif of the chromophore-binding pocket is shown in the left part in Fig. 2 using residue numbering of the crystal structure PDB ID 3NEZ of mRojo. <sup>14</sup> In all systems, the nitrogen atom of the 5-membered imidazolinone ring is always hydrogen-bonded to the protonated side chain of the Glu215, and the oxygen atom of this ring forms a hydrogen bond with the positively charged Arg95 residue. The  $S_{0,\min} - S_1$  transition is associated with the electron density redistribution from the phenolate to the imidazolinone fragment, and these two hydrogen bonds are important for the  $S_1$  state stabilization. The phenolate oxygen forms two hydrogen bonds with a water molecule (Wat in Fig. 2) and a side chain of Ser145. The latter is highly conservative among FPs, however, other amino acid residues that can act as hydrogen bond donors can be found at this position, e.g., asparagine occupies this position instead of serine in eqFP670.

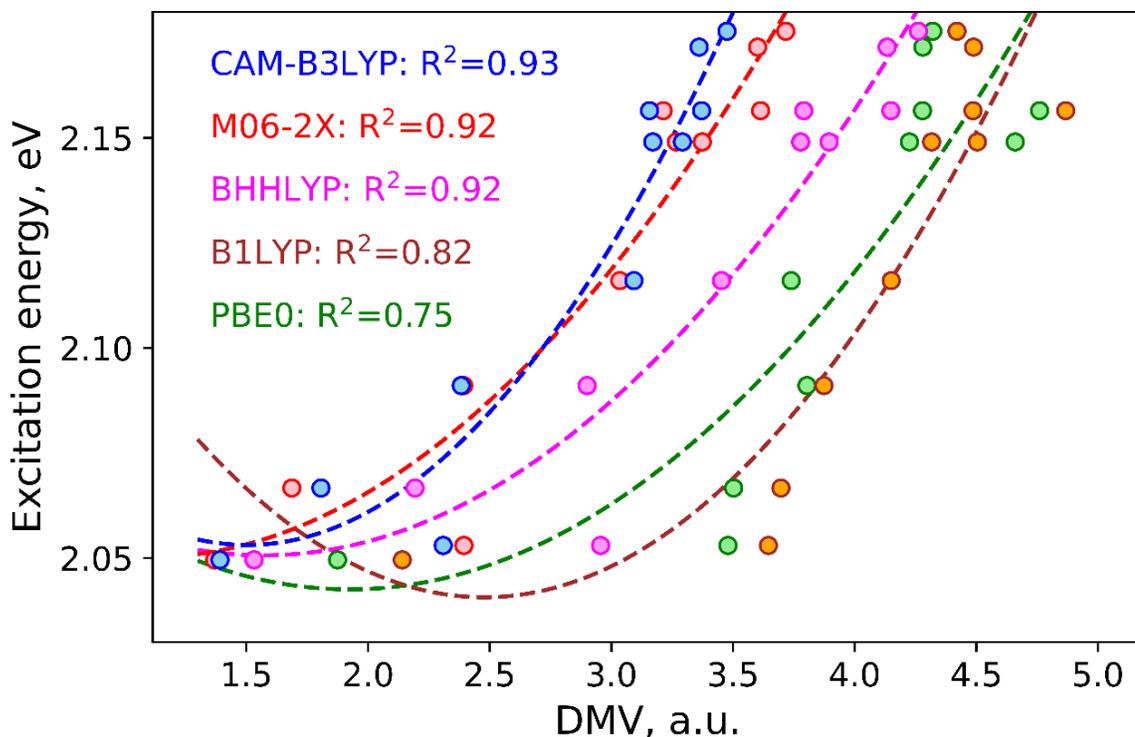


**Figure 2.** Left: The chromophore and the neighboring molecular fragments common for all considered systems. Residues are numbered according to the crystal structure of mRojoA PDB ID 3NEZ. <sup>14</sup> Color code: carbon – green, oxygen – red, nitrogen – blue, hydrogen – white, light blue – capping hydrogen atoms on the QM/MM border. Dark green, blue and red colors highlight the conjugated  $\pi$ -system of the chromophore. Right: Chemical structure of the chromophore. The C=O fragment that is out of the plane of the rest of the chromophore is marked with red ellipse.

According to the experimental data (see Table 1), the excitation band maxima in this set of RFPs covers the region between 570 nm and 605 nm. All proteins contain the chromophore with an extended conjugated  $\pi$ -system compared with the GFP chromophore. It is elongated by =N–C(=O)-fragment; however, due to protein cavities its C=O group is almost orthogonal to the plane of the rest part of the chromophore (see the right part in Fig. 2). This decreases the extent of  $\pi$ -conjugation as compared with the planar chromophore occurring in the gas phase.<sup>28</sup>

### ***Excitation energy - DMV correlation***

Fig. 3 and Table 2 present the main findings of the present work. We tested 7 DFT functionals in the TDDFT approach to compute electron densities and to evaluate DMV values. The favorable performance of the CAM-B3LYP functional for the GFP-like chromophores is well documented in the literature.<sup>26,28,60</sup> Additionally we utilized 6 more functionals. Two of them, LC-BLYP and B97M-D3BJ, demonstrate poor correlations between the calculated DMV and the observed excitation energy. As shown in Fig. 3, the CAM-B3LYP, BHHLYP, M06-2X, B1LYP and PBE0 functionals exhibit a fair correlation between computed DMVs and experimental excitation energies. Among them, CAM-B3LYP, BHHLYP and M06-2X demonstrate close results with the mean absolute error (MAE) not exceeding 0.01 eV (Table 2). For the B1LYP and PBE0 functionals, the correlation takes place; however, the  $R^2$  values are lower and the MAE values are greater. Table 2 shows that the fitting coefficients A, B, C are fairly close for the favorable CAM-B3LYP, M06-2X and BHHLYP functionals.



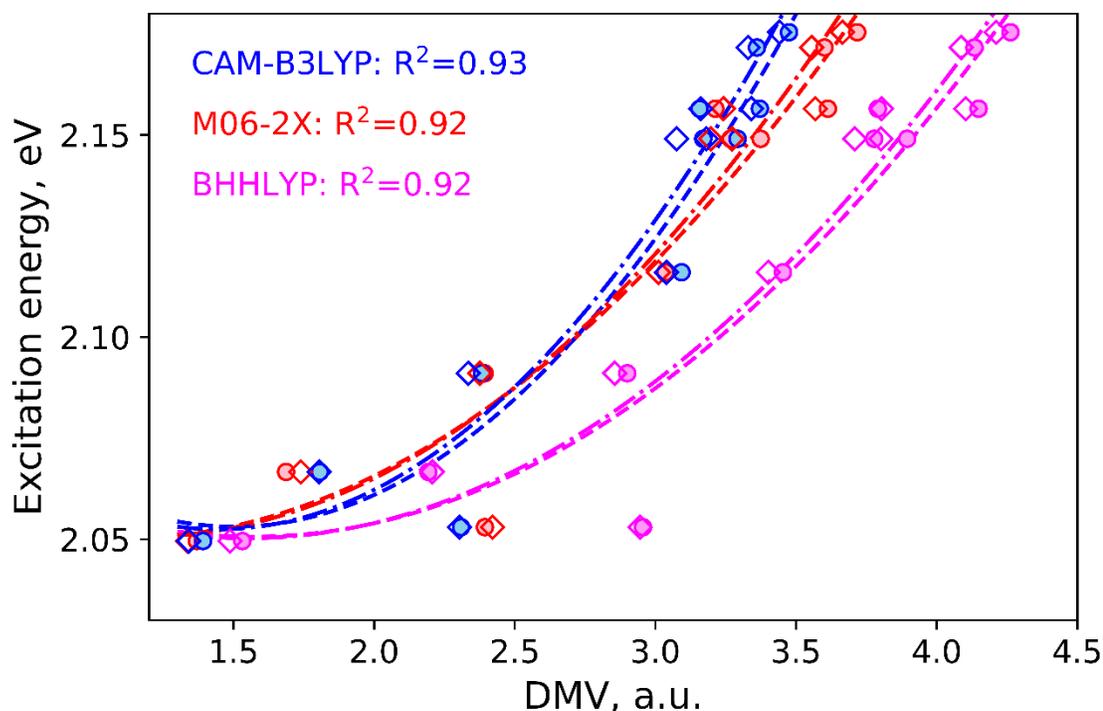
**Figure 3.** Experimental excitation energies,  $\Delta E$ , and calculated (the color-filled circles) DMV values computed at the TD-DFT/cc-pvdz level. The dashed lines show the fitting curves  $\Delta E = A \cdot \text{DMV}^2 + B \cdot \text{DMV} + C$ .

**Table 2.** Parameters of the nonlinear regression curves between the experimental excitation energies ( $\Delta E$ ) and calculated dipole moment variation upon excitation (DMV) and corresponding mean absolute errors (MAE).

TD-DFT functional	Parameters of $\Delta E = A \cdot \text{DMV}^2 + B \cdot \text{DMV} + C$ fitting			MAE, eV
	A, eV/a.u. <sup>2</sup>	B, eV/a.u.	C, eV	
cc-pvdz basis set				
CAM-B3LYP	$0.032 \pm 0.012$	$-0.096 \pm 0.061$	$2.125 \pm 0.071$	0.010
M06-2X	$0.0188 \pm 0.0096$	$-0.041 \pm 0.050$	$2.072 \pm 0.061$	0.0098
BHHLYP	$0.0179 \pm 0.0070$	$-0.056 \pm 0.042$	$2.095 \pm 0.059$	0.0095
B1LYP	$0.027 \pm 0.011$	$-0.134 \pm 0.079$	$2.21 \pm 0.13$	0.015
PBE0	$0.018 \pm 0.011$	$-0.067 \pm 0.073$	$2.11 \pm 0.12$	0.022
ma-def2-SVP basis set				
CAM-B3LYP	$0.032 \pm 0.012$	$-0.091 \pm 0.057$	$2.118 \pm 0.066$	0.010
M06-2X	$0.0208 \pm 0.0098$	$-0.048 \pm 0.050$	$2.079 \pm 0.060$	0.010
BHHLYP	$0.0184 \pm 0.0070$	$-0.057 \pm 0.042$	$2.094 \pm 0.058$	0.0097

### Impact of diffuse basis functions

It is important to verify the sensitivity of calculation results to such detail of the method as the presence of diffuse functions in the basis set, because the chromophore in the considered RFPs occurs in the anionic form. Therefore, we calculated DMV for the same systems with the extended basis set, ma-def2-SVP, that includes diffuse functions on all non-hydrogen atoms (Table 2, Fig. 4). We find that the values of DMV calculated with the same functional and different basis sets are similar; the mean differences do not exceed 0.045 a.e. This indicates that the double-zeta basis set with polarization functions should be an adequate tool for a proper DMV description, also showing that the results are stable with respect to basis set extension.

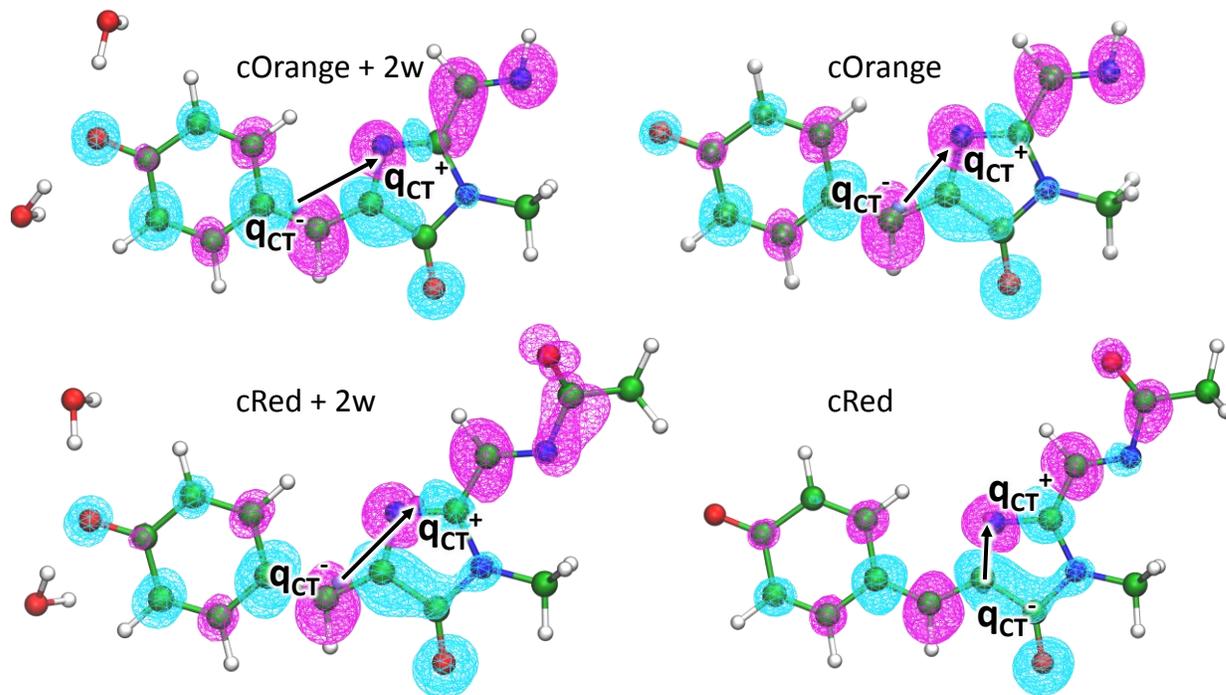


**Figure 4.** Experimental excitation energies,  $\Delta E$ , and calculated dipole moment variations upon excitation, DMV, calculated at the TDDFT/ma-def2-SVP (empty diamonds and dash-dotted fitting curves) and TDDFT/cc-pvdz (color-filled circles and dashed fitting curves) levels with different functionals. The fitting curves are  $\Delta E_{\text{ex}} = A \cdot \text{DMV}^2 + B \cdot \text{DMV} + C$ . The  $R^2$  values are the same for both basis sets and the same functional.

In the following subsection, we analyze the relevance of the obtained fitting parameters in Eq. (1) to properties of chromophores in the gas phase.

### Analysis of fitting parameters

We emphasize that Eq. (1) has a solid physical ground, and every coefficient is meaningful.<sup>33</sup> In particular, these coefficients have a strong relevance to the chromophore properties in vacuum. In this subsection, we show that the coefficients obtained in the present fitting procedure (see Table 2) provide meaningful parameters. To this goal, we performed additional calculations for the related systems shown in Fig. 5, i.e. the cRed and cOrange chromophores,<sup>28</sup> in the gas phase. The cRed chromophore has the same conjugated  $\pi$ -system as the DsRed-type chromophore considered in the present work. However, according to the quantum chemistry optimization of its structure in the gas phase, the C=O group in the extended fragment is in the plane of the chromophore, thus providing additional conjugation as compared with the chromophore in the protein cavity. The cOrange chromophore (Fig. 5) lacks this C=O group, and therefore has a smaller conjugated  $\pi$ -system compared with the DsRed chromophore. For both chromophores, we also consider their complexes with two water molecules (cRed+2w and cOrange+2w), that stabilize the negative charge on the phenolate oxygen.



**Figure 5.** Electron density difference upon the  $S_{0,\min} \rightarrow S_1$  transition calculated at the TDDFT CAM-B3LYP/cc-pvdz level for the isolated cRed and cOrange chromophores (right) and for these chromophores hydrogen-bonded to two water molecules (left). The magenta and cyan wireframe isosurfaces correspond to the increase and the decrease of electron density upon electronic excitation, respectively; the contour values are  $\pm 0.0015$  a.u. The black arrow identifies the negative charge transfer direction; it starts at the barycenter of the negative electron density difference ( $q_{CT^-}$ ) and ends at the barycenter of the positive electron density difference ( $q_{CT^+}$ ).

We calculated the average value of the C parameter of Eq. (1) using two basis sets (cc-pvdz and ma-def2-SVP) and three functionals CAM-B3LYP, BHHLYP and M06-2X functionals. The obtained parameter  $2.10 \pm 0.02$  eV refers to the electronic excitation energy for the hypothetical state of the gas-phase chromophore,  $\Delta E_0$  (Eq. (1)). This value is between the calculated best estimates of the vertical transition energies of the cRed and cOrange chromophores, both with additional water molecules, cRed+2w and cOrange+2w, respectively. The cRed+2w is red-shifted with the calculated transition energy value of 1.86 eV.<sup>28</sup> This is reasonable because of the planar structure of the chromophore in the gas phase. The decrease of the conjugated  $\pi$ -system in the case of cOrange+2w results in the blue shift of the electronic excitation energy to the value of 2.21 eV.<sup>28</sup> Thus, the electronic transition energy calculated for the hypothetical gas-phase chromophore is reasonable.

Another quantity is related to the excess of dipole moment or DMV,  $\Delta\mu_0$ , for the isolated chromophore. We compare the values obtained directly from the fitting curves with those calculated for the cOrange and cRed chromophores and in complexes with two water molecules. Similarly to the electronic excitation energy, we obtain  $\Delta\mu_0$  from the fitting curves as  $2.78 \pm 0.41$  a.u. The calculated DMVs calculated for the isolated systems composed of the chromophore and two water molecules are comparable with this value being  $2.80 \pm 0.08$  a.u. for cRed and  $2.51 \pm 0.09$  a.u. for cOrange. In the experimental paper,<sup>33</sup> Drobizhev et al. reported  $\Delta\mu_0 = 4.49 \pm 0.27$  D ( $1.77 \pm 0.11$  a.u.). This value is in a perfect agreement with the calculated DMVs of the isolated cRed and cOrange chromophores, being between  $1.65 \pm 0.02$  a.u. and  $1.96 \pm 0.12$  a.u., respectively (Table 3).

To conclude this section, we note that development of experimental techniques and computational chemistry approaches resulted in appearance of large libraries of novel fluorescent proteins. To assist future efforts, a computational problem of modeling fine spectral tuning in RFPs should be solved. Results of the present work, as well as of the previous studies<sup>26,28,31–33</sup> are encouraging toward this goal.

**Table 3.** Excess of the dipole moment derived from the fitting curves for the hypothetical isolated chromophore ( $\Delta\mu_0$ ) and calculated DMVs for the isolated chromophores cRed and cOrange and their complexes with two water molecules that stabilize the negative charge on the phenolate oxygen, cRed+2w and cOrange+2w.

TDDFT functional	$\Delta\mu_0$ , a.u.	DMV(chromophore + 2 w.), a.u		DMV(chromophore), a.u	
		cRed	cOrange	cRed	cOrange
cc-pvdz basis set					
CAM-B3LYP	3.01	2.73	2.46	1.654	1.917
M06-2X	2.18	2.78	2.40	1.647	1.822
BHHLYP	3.14	2.88	2.60	1.640	2.074
ma-def2-SVP basis set					
CAM-B3LYP	2.89	2.72	2.48	1.628	1.918
M06-2X	2.33	2.79	2.45	1.635	1.881
BHHLYP	3.10	2.91	2.64	1.669	2.140
Average	$2.78 \pm 0.41$	$2.80 \pm 0.08$	$2.51 \pm 0.09$	$1.65 \pm 0.02$	$1.96 \pm 0.12$

## Conclusion

We demonstrate that a favorable correlation between the computationally attainable electronic structure descriptor DMV and the experimentally obtained transition energies at the excitation band maxima,  $\Delta E_{\text{ex}}$ , takes place for such complex systems as red fluorescent proteins. For this correlation is grounded on a solid theoretical framework, justifying the quadratic dependence of  $\Delta E_{\text{ex}}$  on DMV, a proper computational level to estimate DMV is only required to suggest a useful tool for rational spectral tuning in these important fluorescent biomarkers. Accuracy of calculations depends on the quality of model structures simulating proteins and on the quality of computed electron density. We show that the use of QM(PBE0-D3/cc-pvdz)/MM(AMBER) optimized structures constructed by motifs of crystal structures of proteins and of the TDDFT approaches with the CAM-B3LYP, BHHLYP and M06-2X functionals perform properly and lead to excitation energies for the model systems with the mean absolute error less than 0.01 eV.

**Supporting Information** including atomic coordinates of the QM/MM optimized structures is available in the Zenodo and can be accessed via <https://doi.org/10.5281/zenodo.5182720>.

## Acknowledgements

This work was partly supported by the Russian Science Foundation (project #17-13-01051). We acknowledge the use of supercomputer resources of the Joint Supercomputer Center of the Russian Academy of Sciences and the equipment of the shared research facilities of HPC computing resources at Lomonosov Moscow State University.<sup>61</sup>

## References

- (1) Chernov, K. G.; Redchuk, T. A.; Omelina, E. S.; Verkhusha, V. V. Near-Infrared Fluorescent Proteins, Biosensors, and Optogenetic Tools Engineered from Phytochromes. *Chem. Rev.* **2017**, *117* (9), 6423–6446. <https://doi.org/10.1021/acs.chemrev.6b00700>.
- (2) Greenwald, E. C.; Mehta, S.; Zhang, J. Genetically Encoded Fluorescent Biosensors Illuminate the Spatiotemporal Regulation of Signaling Networks. *Chem. Rev.* **2018**, *118* (24), 11707–11794. <https://doi.org/10.1021/acs.chemrev.8b00333>.
- (3) Acharya, A.; Bogdanov, A. M.; Grigorenko, B. L.; Bravaya, K. B.; Nemukhin, A. V.; Lukyanov, K. A.; Krylov, A. I. Photoinduced Chemistry in Fluorescent Proteins: Curse or Blessing? *Chem. Rev.* **2017**, *117* (2), 758–795. <https://doi.org/10.1021/acs.chemrev.6b00238>.
- (4) Newman, R. H.; Fosbrink, M. D.; Zhang, J. Genetically Encodable Fluorescent Biosensors for Tracking Signaling Dynamics in Living Cells. *Chem. Rev.* **2011**, *111* (5), 3614–3666. <https://doi.org/10.1021/cr100002u>.
- (5) Weissleder, R.; Ntziachristos, V. Shedding Light onto Live Molecular Targets. *Nat. Med.* **2003**, *9* (1), 123–128. <https://doi.org/10.1038/nm0103-123>.
- (6) Savitsky, A. P.; Rusanov, A. L.; Zherdeva, V. V.; Gorodnicheva, T. V.; Khrenova, M. G.; Nemukhin, A. V. FLIM-FRET Imaging of Caspase-3 Activity in Live Cells Using Pair of Red Fluorescent Proteins. *Theranostics* **2012**, *2* (2), 215–226. <https://doi.org/10.7150/thno.3885>.
- (7) Erickson, M. G.; Moon, D. L.; Yue, D. T. DsRed as a Potential FRET Partner with CFP and GFP. *Biophys. J.* **2003**, *85* (1), 599–611. [https://doi.org/10.1016/S0006-3495\(03\)74504-4](https://doi.org/10.1016/S0006-3495(03)74504-4).
- (8) Goryashchenko, A. S.; Khrenova, M. G.; Savitsky, A. P. Detection of Protease Activity by Fluorescent Protein FRET Sensors: From Computer Simulation to Live Cells. *Methods Appl. Fluoresc.* **2018**, *6* (2). <https://doi.org/10.1088/2050-6120/aa9e47>.
- (9) Lam, A. J.; St-Pierre, F.; Gong, Y.; Marshall, J. D.; Cranfill, P. J.; Baird, M. A.; McKeown,

- M. R.; Wiedenmann, J.; Davidson, M. W.; Schnitzer, M. J.; Tsien, R. Y.; Lin, M. Z. Improving FRET Dynamic Range with Bright Green and Red Fluorescent Proteins. *Nat. Methods* **2012**, *9* (10), 1005–1012. <https://doi.org/10.1038/nmeth.2171>.
- (10) Mo, G. C. H.; Posner, C.; Rodriguez, E. A.; Sun, T.; Zhang, J. A Rationally Enhanced Red Fluorescent Protein Expands the Utility of FRET Biosensors. *Nat. Commun.* **2020**, *11* (1), 1848. <https://doi.org/10.1038/s41467-020-15687-x>.
- (11) Matlashov, M. E.; Shcherbakova, D. M.; Alvelid, J.; Baloban, M.; Pennacchietti, F.; Shemetov, A. A.; Testa, I.; Verkhusha, V. V. A Set of Monomeric Near-Infrared Fluorescent Proteins for Multicolor Imaging across Scales. *Nat. Commun.* **2020**, *11* (1), 239. <https://doi.org/10.1038/s41467-019-13897-6>.
- (12) Eason, M. G.; Damry, A. M.; Chica, R. A. Structure-Guided Rational Design of Red Fluorescent Proteins: Towards Designer Genetically-Encoded Fluorophores. *Curr. Opin. Struct. Biol.* **2017**, *45*, 91–99. <https://doi.org/10.1016/j.sbi.2016.12.001>.
- (13) Pandelieva, A. T.; Baran, M. J.; Calderini, G. F.; McCann, J. L.; Tremblay, V.; Sarvan, S.; Davey, J. A.; Couture, J.-F.; Chica, R. A. Brighter Red Fluorescent Proteins by Rational Design of Triple-Decker Motif. *ACS Chem. Biol.* **2016**, *11* (2), 508–517. <https://doi.org/10.1021/acscchembio.5b00774>.
- (14) Chica, R. A.; Moore, M. M.; Allen, B. D.; Mayo, S. L. Generation of Longer Emission Wavelength Red Fluorescent Proteins Using Computationally Designed Libraries. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107* (47), 20257–20262. <https://doi.org/10.1073/pnas.1013910107>.
- (15) Treynor, T. P.; Vizcarra, C. L.; Nedelcu, D.; Mayo, S. L. Computationally Designed Libraries of Fluorescent Proteins Evaluated by Preservation and Diversity of Function. *Proc. Natl. Acad. Sci.* **2007**, *104* (1), 48–53. <https://doi.org/10.1073/pnas.0609647103>.
- (16) De Zitter, E.; Hugelier, S.; Duwé, S.; Vandenberg, W.; Tebo, A. G.; Van Meervelt, L.; Dedecker, P. Structure–Function Dataset Reveals Environment Effects within a Fluorescent Protein Model System. *Angew. Chemie* **2021**, *133* (18), 10161–10169. <https://doi.org/10.1002/ange.202015201>.
- (17) Pedraza-González, L.; Marín, M. D. C.; Jorge, A. N.; Ruck, T. D.; Yang, X.; Valentini, A.; Olivucci, M.; De Vico, L. Web-ARM: A Web-Based Interface for the Automatic Construction of QM/MM Models of Rhodopsins. *J. Chem. Inf. Model.* **2020**, *60* (3), 1481–1493. <https://doi.org/10.1021/acs.jcim.9b00615>.
- (18) Suellen, C.; Freitas, R. G.; Loos, P.-F.; Jacquemin, D. Cross-Comparisons between

- Experiment, TD-DFT, CC, and ADC for Transition Energies. *J. Chem. Theory Comput.* **2019**, *15* (8), 4581–4590. <https://doi.org/10.1021/acs.jctc.9b00446>.
- (19) Melaccio, F.; Ferré, N.; Olivucci, M. Quantum Chemical Modeling of Rhodopsin Mutants Displaying Switchable Colors. *Phys. Chem. Chem. Phys.* **2012**, *14* (36), 12485. <https://doi.org/10.1039/c2cp40940b>.
- (20) Schwabe, T.; Beerepoot, M. T. P.; Olsen, J. M. H.; Kongsted, J. Analysis of Computational Models for an Accurate Study of Electronic Excitations in GFP. *Phys. Chem. Chem. Phys.* **2015**, *17* (4), 2582–2588. <https://doi.org/10.1039/C4CP04524F>.
- (21) Nemukhin, A. V.; Grigorenko, B. L.; Khrenova, M. G.; Krylov, A. I. Computational Challenges in Modeling of Representative Bioimaging Proteins: GFP-Like Proteins, Flavoproteins, and Phytochromes. *J. Phys. Chem. B* **2019**, *123* (29), 6133–6149. <https://doi.org/10.1021/acs.jpcc.9b00591>.
- (22) Lischka, H.; Nachtigallová, D.; Aquino, A. J. A.; Szalay, P. G.; Plasser, F.; Machado, F. B. C.; Barbatti, M. Multireference Approaches for Excited States of Molecules. *Chem. Rev.* **2018**, *118* (15), 7293–7361. <https://doi.org/10.1021/acs.chemrev.8b00244>.
- (23) Peach, M. J. G.; Benfield, P.; Helgaker, T.; Tozer, D. J. Excitation Energies in Density Functional Theory: An Evaluation and a Diagnostic Test. *J. Chem. Phys.* **2008**, *128* (4), 044118. <https://doi.org/10.1063/1.2831900>.
- (24) Tsien, R. Y. The Green Fluorescent Protein. *Annu. Rev. Biochem.* **1998**, *67* (1), 509–544. <https://doi.org/10.1146/annurev.biochem.67.1.509>.
- (25) Wachter, R. M.; Elsliger, M. A.; Kallio, K.; Hanson, G. T.; Remington, S. J. Structural Basis of Spectral Shifts in the Yellow-Emission Variants of Green Fluorescent Protein. *Structure* **1998**, *6* (10), 1267–1277.
- (26) Nifosi, R.; Mennucci, B.; Filippi, C. The Key to the Yellow-to-Cyan Tuning in the Green Fluorescent Protein Family Is Polarisation. *Phys. Chem. Chem. Phys.* **2019**, *21* (35), 18988–18998. <https://doi.org/10.1039/C9CP03722E>.
- (27) Khrenova, M. G.; Nemukhin, A. V.; Tsirelson, V. G. Origin of the  $\pi$ -Stacking Induced Shifts in Absorption Spectral Bands of the Green Fluorescent Protein Chromophore. *Chem. Phys.* **2019**, *522*, 32–38. <https://doi.org/10.1016/j.chemphys.2019.02.010>.
- (28) Khrenova, M. G.; Mulashkin, F. D.; Bulavko, E. S.; Zakharova, T. M.; Nemukhin, A. V. Dipole Moment Variation Clears Up Electronic Excitations in the  $\pi$ -Stacked Complexes of Fluorescent Protein Chromophores. *J. Chem. Inf. Model.* **2020**, *60* (12), 6288–6297.

<https://doi.org/10.1021/acs.jcim.0c01028>.

- (29) Le Bahers, T.; Adamo, C.; Ciofini, I. A Qualitative Index of Spatial Extent in Charge-Transfer Excitations. *J. Chem. Theory Comput.* **2011**, *7* (8), 2498–2506. <https://doi.org/10.1021/ct200308m>.
- (30) Drobizhev, M.; Makarov, N. S.; Tillo, S. E.; Hughes, T. E.; Rebane, A. Describing Two-Photon Absorptivity of Fluorescent Proteins with a New Vibronic Coupling Mechanism. *J. Phys. Chem. B* **2012**, *116* (5), 1736–1744. <https://doi.org/10.1021/jp211020k>.
- (31) Drobizhev, M.; Makarov, N. S.; Tillo, S. E.; Hughes, T. E.; Rebane, A. Two-Photon Absorption Properties of Fluorescent Proteins. *Nat. Methods* **2011**, *8* (5), 393–399. <https://doi.org/10.1038/nmeth.1596>.
- (32) Drobizhev, M.; Callis, P. R.; Nifosi, R.; Wicks, G.; Stoltzfus, C.; Barnett, L.; Hughes, T. E.; Sullivan, P.; Rebane, A. Long- and Short-Range Electrostatic Fields in GFP Mutants: Implications for Spectral Tuning. *Sci. Rep.* **2015**, *5* (1), 13223. <https://doi.org/10.1038/srep13223>.
- (33) Drobizhev, M.; Tillo, S.; Makarov, N. S.; Hughes, T. E.; Rebane, A. Color Hues in Red Fluorescent Proteins Are Due to Internal Quadratic Stark Effect. *J. Phys. Chem. B* **2009**, *113* (39), 12860–12864. <https://doi.org/10.1021/jp907085p>.
- (34) Lin, C.-Y.; Romei, M. G.; Oltrogge, L. M.; Mathews, I. I.; Boxer, S. G. Unified Model for Photophysical and Electro-Optical Properties of Green Fluorescent Proteins. *J. Am. Chem. Soc.* **2019**, *141* (38), 15250–15265. <https://doi.org/10.1021/jacs.9b07152>.
- (35) Matz, M. V.; Fradkov, A. F.; Labas, Y. A.; Savitsky, A. P.; Zaraisky, A. G.; Markelov, M. L.; Lukyanov, S. A. Fluorescent Proteins from Nonbioluminescent Anthozoa Species. *Nat. Biotechnol.* **1999**, *17* (10), 969–973. <https://doi.org/10.1038/13657>.
- (36) Gross, L. A.; Baird, G. S.; Hoffman, R. C.; Baldrige, K. K.; Tsien, R. Y. The Structure of the Chromophore within DsRed, a Red Fluorescent Protein from Coral. *Proc. Natl. Acad. Sci.* **2000**, *97* (22), 11990–11995. <https://doi.org/10.1073/pnas.97.22.11990>.
- (37) Edgar, R. C. MUSCLE: A Multiple Sequence Alignment Method with Reduced Time and Space Complexity. *BMC Bioinformatics* **2004**, *5*, 113. <https://doi.org/10.1186/1471-2105-5-113>.
- (38) Shu, X.; Shaner, N. C.; Yarbrough, C. A.; Tsien, R. Y.; Remington, S. J. Novel Chromophores and Buried Charges Control Color in mFruits. *Biochemistry* **2006**, *45* (32), 9639–9647. <https://doi.org/10.1021/bi060773l>.

- (39) Henderson, J. N.; Osborn, M. F.; Koon, N.; Gepshtein, R.; Huppert, D.; Remington, S. J. Excited State Proton Transfer in the Red Fluorescent Protein MKeima. *J. Am. Chem. Soc.* **2009**, *131* (37), 13212–13213. <https://doi.org/10.1021/ja904665x>.
- (40) Pletnev, S.; Pletneva, N. V.; Souslova, E. A.; Chudakov, D. M.; Lukyanov, S.; Wlodawer, A.; Dauter, Z.; Pletnev, V. Structural Basis for Bathochromic Shift of Fluorescence in Far-Red Fluorescent Proteins EqFP650 and EqFP670. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2012**, *68* (9), 1088–1097. <https://doi.org/10.1107/S0907444912020598>.
- (41) Shen, Y.; Chen, Y.; Wu, J.; Shaner, N. C.; Campbell, R. E. Engineering of MCherry Variants with Long Stokes Shift, Red-Shifted Fluorescence, and Low Cytotoxicity. *PLoS One* **2017**, *12* (2), e0171257. <https://doi.org/10.1371/journal.pone.0171257>.
- (42) Violot, S.; Carpentier, P.; Blanchoin, L.; Bourgeois, D. Reverse PH-Dependence of Chromophore Protonation Explains the Large Stokes Shift of the Red Fluorescent Protein MKeima. *J. Am. Chem. Soc.* **2009**, *131* (30), 10356–10357. <https://doi.org/10.1021/ja903695n>.
- (43) Word, J. M.; Lovell, S. C.; Richardson, J. S.; Richardson, D. C. Asparagine and Glutamine: Using Hydrogen Atom Contacts in the Choice of Side-Chain Amide Orientation. *J. Mol. Biol.* **1999**, *285* (4), 1735–1747. <https://doi.org/10.1006/jmbi.1998.2401>.
- (44) Adamo, C.; Barone, V. Toward Reliable Density Functional Methods without Adjustable Parameters: The PBE0 Model. *J. Chem. Phys.* **1999**, *110* (13), 6158. <https://doi.org/10.1063/1.478522>.
- (45) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H–Pu. *J. Chem. Phys.* **2010**, *132* (15), 154104. <https://doi.org/10.1063/1.3382344>.
- (46) Dunning, T. H. Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron through Neon and Hydrogen. *J. Chem. Phys.* **1989**, *90* (2), 1007–1023. <https://doi.org/10.1063/1.456153>.
- (47) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117* (19), 5179–5197. <https://doi.org/10.1021/ja00124a002>.
- (48) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79* (2),

- 926–935. <https://doi.org/10.1063/1.445869>.
- (49) Valiev, M.; Bylaska, E. J.; Govind, N.; Kowalski, K.; Straatsma, T. P.; Van Dam, H. J. J.; Wang, D.; Nieplocha, J.; Apra, E.; Windus, T. L.; de Jong, W. A. NWChem: A Comprehensive and Scalable Open-Source Solution for Large Scale Molecular Simulations. *Comput. Phys. Commun.* **2010**, *181* (9), 1477–1489. <https://doi.org/10.1016/j.cpc.2010.04.018>.
- (50) Neese, F. The ORCA Program System. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**, *2* (1), 73–78. <https://doi.org/10.1002/wcms.81>.
- (51) Yanai, T.; Tew, D. P.; Handy, N. C. A New Hybrid Exchange–Correlation Functional Using the Coulomb-Attenuating Method (CAM-B3LYP). *Chem. Phys. Lett.* **2004**, *393* (1–3), 51–57. <https://doi.org/10.1016/j.cplett.2004.06.011>.
- (52) Iikura, H.; Tsuneda, T.; Yanai, T.; Hirao, K. A Long-Range Correction Scheme for Generalized-Gradient-Approximation Exchange Functionals. *J. Chem. Phys.* **2001**, *115* (8), 3540–3544. <https://doi.org/10.1063/1.1383587>.
- (53) Zhao, Y.; Truhlar, D. G. The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Noncovalent Interactions, Excited States, and Transition Elements: Two New Functionals and Systematic Testing of Four M06-Class Functionals and 12 Other Function. *Theor. Chem. Acc.* **2008**, *120* (1–3), 215–241. <https://doi.org/10.1007/s00214-007-0310-x>.
- (54) Becke, A. D. A New Mixing of Hartree–Fock and Local Density- functional Theories. *J. Chem. Phys.* **1993**, *98* (2), 1372–1377. <https://doi.org/10.1063/1.464304>.
- (55) Adamo, C.; Barone, V. Toward Reliable Adiabatic Connection Models Free from Adjustable Parameters. *Chem. Phys. Lett.* **1997**, *274* (1–3), 242–250. [https://doi.org/10.1016/S0009-2614\(97\)00651-9](https://doi.org/10.1016/S0009-2614(97)00651-9).
- (56) Mardirossian, N.; Head-Gordon, M. Mapping the Genome of Meta-Generalized Gradient Approximation Density Functionals: The Search for B97M-V. *J. Chem. Phys.* **2015**, *142* (7), 074111. <https://doi.org/10.1063/1.4907719>.
- (57) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J. Comput. Chem.* **2011**, *32* (7), 1456–1465. <https://doi.org/10.1002/jcc.21759>.
- (58) Zheng, J.; Xu, X.; Truhlar, D. G. Minimally Augmented Karlsruhe Basis Sets. *Theor. Chem. Acc.* **2011**, *128* (3), 295–305. <https://doi.org/10.1007/s00214-010-0846-z>.

- (59) Lu, T.; Chen, F. Multiwfn: A Multifunctional Wavefunction Analyzer. *J. Comput. Chem.* **2012**, *33* (5), 580–592. <https://doi.org/10.1002/jcc.22885>.
- (60) Shao, Y.; Mei, Y.; Sundholm, D.; Kaila, V. R. I. Benchmarking the Performance of Time-Dependent Density Functional Theory Methods on Biochromophores. *J. Chem. Theory Comput.* **2020**, *16* (1), 587–600. <https://doi.org/10.1021/acs.jctc.9b00823>.
- (61) Voevodin, V. V.; Antonov, A. S.; Nikitenko, D. A.; Shvets, P. A.; Sobolev, S. I.; Sidorov, I. Y.; Stefanov, K. S.; Voevodin, V. V.; Zhumatiy, S. A. Supercomputer Lomonosov-2: Large Scale, Deep Monitoring and Fine Analytics for the User Community. *Supercomput. Front. Innov.* **2019**, *6* (2), 4–11. <https://doi.org/10.14529/jsfi190201>.

# TOC Graphic

