

Computational Challenges in Modeling of Representative Bioimaging Proteins: GFP-Like Proteins, Flavoproteins, and Phytochromes

Alexander V. Nemukhin,^{*,†,‡,§} Bella L. Grigorenko,^{†,‡} Maria G. Khrenova,^{†,§} and Anna I. Krylov^{||}

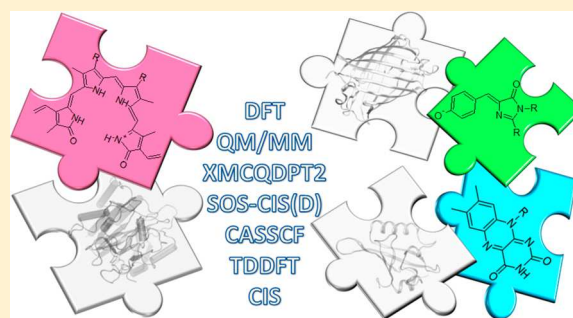
[†]Department of Chemistry, Lomonosov Moscow State University, Moscow 119991, Russia

[‡]Emanuel Institute of Biochemical Physics, Russian Academy of Sciences, Moscow 119334, Russia

[§]Federal Research Center of Biotechnology, Bach Institute of Biochemistry, Russian Academy of Sciences, Moscow 119071, Russian

^{||}Department of Chemistry, University of Southern California, Los Angeles, California 90089-0482, United States

ABSTRACT: Remarkable success in engineering novel efficient biomarkers based on fluorescent and photoactive proteins provokes a question of whether computational modeling of their properties can contribute to this important field. In this Feature Article, we analyze selected papers devoted to computer simulations of three types of photoactive systems: the green fluorescent protein and its derivatives, the flavin-binding proteins, and the phytochrome domains. The main emphasis is on structures, optical spectra, and chemical reactions in the chromophore-containing pockets. Quantum chemistry, quantum mechanics/molecular mechanics, and molecular dynamics methods are effective tools in these simulations. We highlight both the success stories and the persisting challenges, discussing the ways of elevating theoretical approaches to the level of testable predictions.



■ INTRODUCTION

This Feature Article was inspired by a recent review from 2017 by Roger Tsien and co-workers, “The Growing and Glowing Toolbox of Fluorescent and Photoactive Proteins”,¹ which describes the ongoing experimental efforts to engineer efficient markers for multicolor imaging in living cells. The review presents a comprehensive summary of the progress achieved during the past two decades aimed to “to improve and modify the fundamental properties of fluorescent proteins with the goal of adapting them for a fantastic range of applications” in biomedicine and biotechnology. This progress and the richness of the field are illustrated visually by a multicolor mosaic in Figure 1 (created following the motifs of Figure 2 in ref 1).

Recognizing a remarkable success of the experimental research in this area, a question arises of whether theoretical approaches based on computational modeling of these systems can contribute to this important and ever-growing field. There are a great number of original papers reporting the results of computer simulations of fluorescent and photoactive proteins and their chromophores, carried out at various theoretical levels, as well as several review articles.^{2–5} Here, we analyze selected papers devoted to the computer simulations of the systems described ref 1: the green fluorescent protein (GFP) and its derivatives, the flavin-binding proteins, and the phytochrome domains. Our main focus is how to elevate the existing theoretical approaches to the next level of accuracy and robustness.

The basic motifs of the chromophores from these three types of proteins are shown in Figure 2. Those are the *p*-hydroxybenzylidene-imidazolinone species for GFP-like proteins (Figure 2a), the isoalloxazine ring of flavin-based moieties (Figure 2b), and the linear tetrapyrrole (bilin) chromophore of phytochrome domains (Figure 2c).

The significance of GFP and its homologues and derivatives as fluorescent tags cannot be overstated.^{6–11} It is impossible to imagine modern molecular biology, medicine, or biotechnology without these *in vivo* reporters. The chromophore of the parent member of the family, the wild-type GFP, comprises phenolate and imidazolinone moieties connected by a methine bridge. It is covalently bound to the peptide chains, as shown in Figure 2a. The formation of this π -conjugated system occurs upon protein maturation, in a series of post-translational modifications of the natural tripeptide sequence Ser-Tyr-Gly. In other family members, Ser and Tyr can be replaced by other residues.

The flavin-based blue-light photoreceptors are the foundational members of protein families called BLUF and LOV domains, which have been extensively studied in recent years as fluorescent and photoactive species.¹² Both domains incorporate non-covalently bound cofactors: flavin adenine dinucleotide (FAD) or flavin mononucleotide (FMN), the

Received: January 20, 2019

Revised: April 21, 2019

Published: May 1, 2019

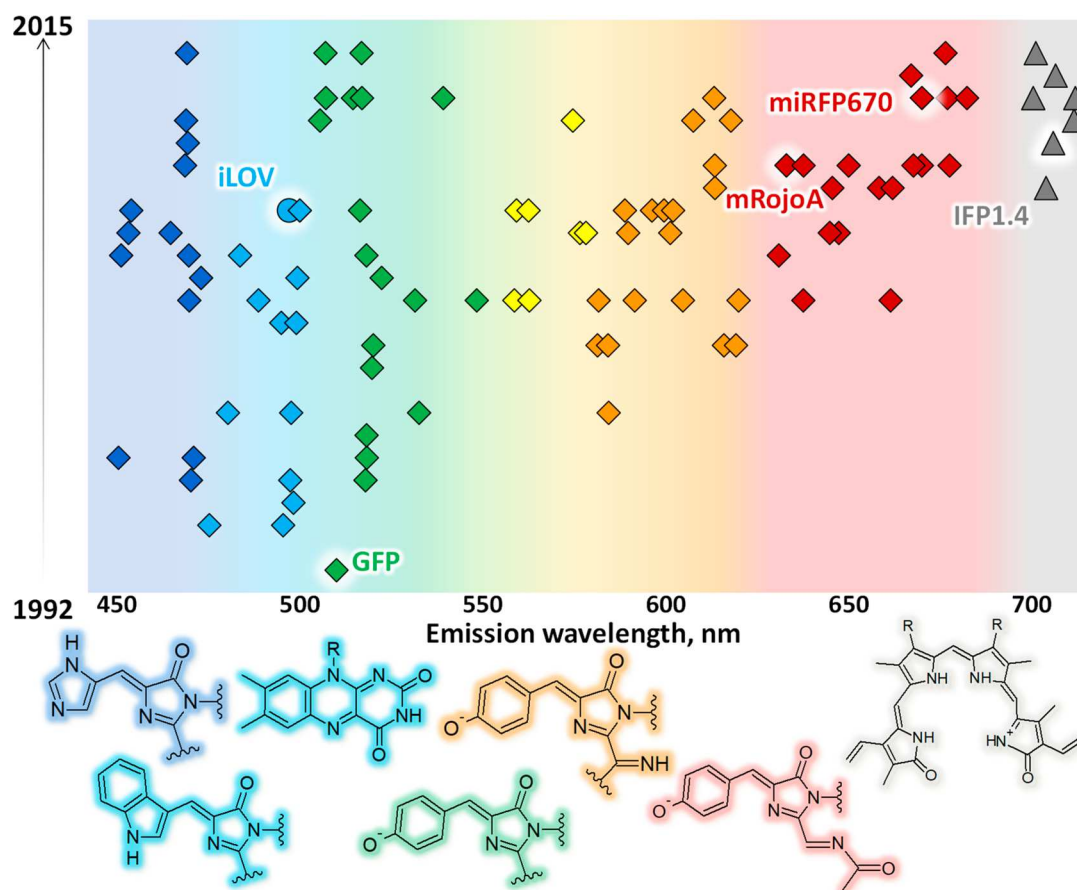


Figure 1. Chart constructed by motifs of Figure 2 in ref 1, showing the timeline of the creation of fluorescent proteins and their emission band maxima. Each colored diamond denotes a particular protein; the names are shown for the species that are discussed in the present article.

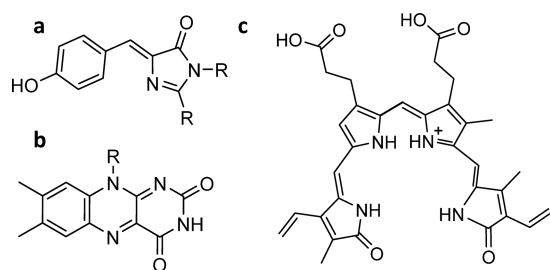


Figure 2. Chromophores from GFP-like proteins (a), flavin-binding proteins (b), and phytochrome domains (c).

light-sensing part of which is the isoalloxazine ring (Figure 2b). The name BLUF is derived from the blue light using EAD combination, while LOV stands for the light, oxygen, voltage. The latter abbreviation gave rise to several charming titles of LOV-related scientific papers, e.g., “A LOV Story...”,¹³ “LOV to BLUF...”,¹⁴ and “LOVely Enzymes...”.¹⁵ BLUF and LOV domains are small, compact photosensing modules of approximately 100 amino acid side chains. In recent years, the flavin-based fluorescent proteins (FbFPs)^{12,16–19} drew on an enhanced attention as biomarkers. In particular, an improved LOV variant called iLOV^{20,21} and its mutants are well suited for *in vivo* imaging due to the increased photostability.

The phytochrome-based protein domains with the linear tetrapyrrole chromophores constitute an emerging class of biomarkers.^{22–24} The extended π -system of these chromo-

phores accounts for smaller gaps between the ground and excited electronic states than in the GFP-like proteins or in flavin-containing domains. This prompted researchers to turn to bacterial phytochromes as templates for engineering infrared and near-infrared fluorescent proteins. In these constructs, the chromophore is a biliverdin molecule (Figure 2c), covalently bound to a cysteine residue in the protein domains via a thioether bond.

Biological aspects of imaging markers based on the GFP-like proteins, flavoproteins, and phytochromes are highlighted in the review article.¹ In brief, optical properties of these three types of proteins with the chromophores shown in Scheme 1 cover overlapping yet distinct spectral ranges (see Figure 1). FbFPs offer certain advantages compared to the GFP family of reporters, owing to their small size and oxygen-independent fluorescence. Phytochromes are promising reporters in the red/far-red spectral range where light penetrates the furthest through animal tissues.

In this Feature Article, we analyze selected computational studies of these three types and describe the applied models and methods, with the goal to discuss the advantages and drawbacks of the existing theoretical methods applicable to this type of systems. The computational approaches used in these simulations are based on quantum chemistry (QM), molecular mechanics (MM), and molecular dynamics methods. The modeling relies on available structures from the Protein Data Bank (PDB).²⁵ Numbering of residues is given according to the corresponding PDB structures. The three letter code for amino acids is used throughout the paper, except for the

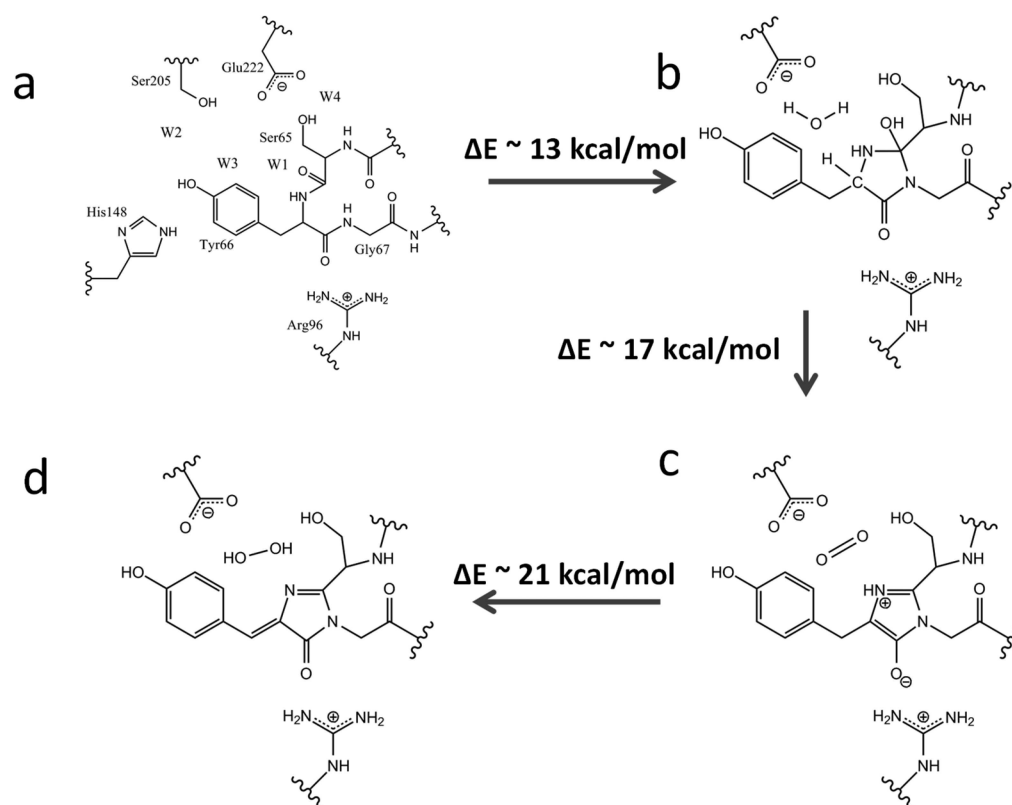


Figure 3. Chromophore maturation mechanism in GFP according to the simulations in ref 38. Panel a shows the initial structure with the Ser65-Tyr66-Gly67 tripeptide; panel b illustrates the cyclized but non-dehydrated intermediate; panel c shows the dehydrated intermediate with the added oxygen molecules; panel d shows the mature chromophore with the formed hydrogen peroxide molecule.

protein mutants that are named using a one letter code, e.g., GFP-S65T.

SIMULATIONS OF GFP-LIKE PROTEINS

Chromophore Maturation in GFP. Mechanistic understanding of the chromophore's maturation in GFP is an obviously important issue, because it provides a solid basis for predictions of maturation mechanisms in other fluorescent proteins. A unique feature of GFP is that its chromophore (shown in Figure 2a) is formed autocatalytically from the initially noncyclized tripeptide sequence, Ser65-Tyr66-Gly67 in the wild-type (wt) GFP, under aerobic conditions inside the protein molecule. A tentative chain of reactions including the cyclization, dehydration, and oxidation steps has been outlined in the earliest experimental studies by Tsien et al.,²⁶ which were reproduced in subsequent reviews^{7,8} and research papers. However, despite considerable efforts, the molecular-level mechanism of the chromophore maturation is still debated. For example, a long-standing disputed issue is the sequence of the reaction steps: cyclization–dehydration–oxidation (the so-called Getzoff mechanism^{27–32}) or cyclization–oxidation–dehydration (the Wachter mechanism^{33–35}).

Experimental approaches for elucidating the reaction mechanism of the GFP chromophore maturation include kinetics^{31,33,35,36} and crystallography.^{27–32,37} The latter studies attempt trapping the reaction intermediates by mutations that slow down a particular step. All approaches agree that the cyclization is the first reaction step, whereas different interpretations of the experimental evidence provide support to either Getzoff's or Wachter's mechanisms.⁸

From the theoretical side, the first complete computational description of all elementary reaction steps resulting in the chromophore's maturation in GFP was reported in ref 38. Other computational studies attempted to characterize the cyclization^{39–41} and some aspects of dehydration⁴² and oxidation⁴³ steps; however, the reported quantum chemical calculations yielded the reaction barriers that were too high and inconsistent with the experimental kinetics data.

In ref 38, a model system with the noncyclized tripeptide, Ser65-Tyr66-Gly67, was constructed and the reaction pathway from the initial structure to the structure with the fully formed mature protein-bound chromophore was computed using the QM/MM approach. According to these simulations, which fully characterized the chain of elementary reactions connecting the tripeptide with the mature chromophore, the sequence of steps is cyclization–dehydration–oxidation (Figure 3).

This conclusion was supported by a successful comparison of the computed properties³⁸ with those extracted from the experimental data. First, the computationally derived structures of the reaction product and several reaction intermediates agreed well with the relevant crystal structures. Especially important was a favorable comparison of the final structure with the mature chromophore (Figure 3d) with the high-resolution crystal structure PDB ID: 2WUR.⁴⁴ Also, the model system with the dehydrated intermediate (Figure 3c) was consistent with the trapped structures of mutated GFP variants PDB ID: 2FZU³¹ and PDB ID: 3LVC.³⁷ Second, the computed barrier heights (see Figure 3) on the potential energy surface were compared to the activation barriers estimated from the measured reaction rates^{31,33,36} by using

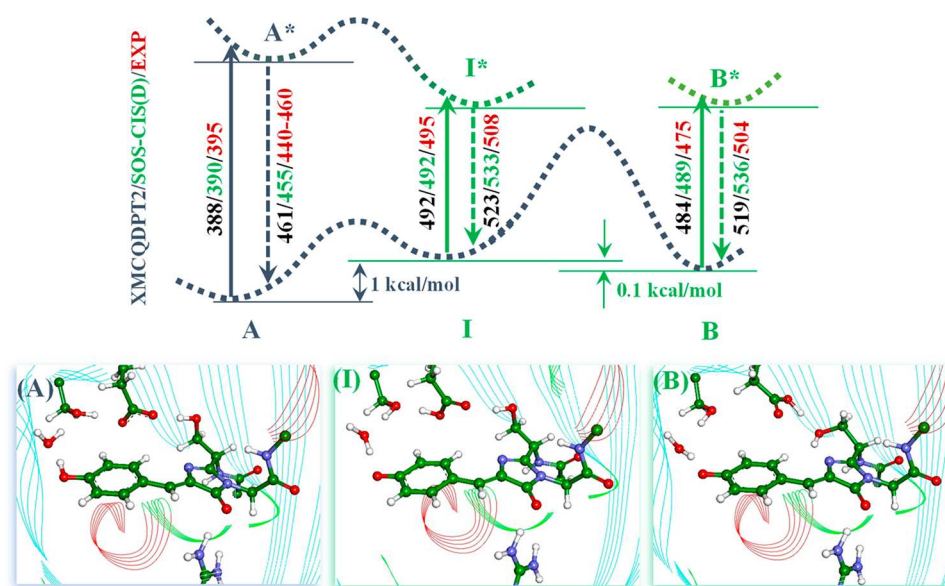


Figure 4. Computational modeling of the A → I → B route in wt-GFP.

the transition-state theory: the highest computed energy barriers at the cyclization–dehydration (17 kcal/mol) and oxidation (21 kcal/mol) steps agreed reasonably well with the values derived from the kinetics measurements (20.7 and 22.7 kcal/mol, respectively). The main conclusions of this study were reinforced by follow-up calculations of the glycine tripeptide in the S65G/Y66G mutant of GFP, which forms only a partially cyclized chromophore.^{27,45}

The most recent experimental study,⁴⁶ which has been published after the computational paper,³⁸ confirmed the cyclization–dehydration–oxidation sequence in the chromophore's formation in GFP. This conclusion was based on the mass-spectrometric analysis of the peptides derived from the engineered nonfluorescent mutant GFP-S65T/G67A.

These computational studies represent a success of the theory: the computed structures on the ground-state potential energy surface and the corresponding energetics were consistent with the experimental observations. The discrepancies for activation barriers as large as 4 kcal/mol are not unexpected for the method used in these studies, QM(PBE0/6-31G*)/MM(AMBER).

Simulations Related to the “Life and Death” of GFP.

After its “birth”, GFP lives an eventful life, being engaged in a variety of biophysical processes, described in numerous original reports and comprehensive reviews.^{7,9,47,48} Along with the chromophore formation mechanism, a complete molecular-level understanding of these processes is the cornerstone for the rational design of novel efficient biomarkers. The simulations of these diverse and complex phenomena require the consideration of the evolution of the system on the excited-state potential energy surfaces and transitions between different electronic states, which is a harder task than scanning the ground-state potential energy surface.

As a first step in this direction, we discuss the computational characterization of the energy landscape and optical spectra along the A → I → B proton transfer route in the canonical three-state GFP model,⁴⁹ which explains the observed photophysical properties of the wt-GFP. The reaction scheme for excited-state proton transfer, including the protein forms with the conventional names A, B, and I, was clearly presented

in Figure 4 in ref 9 and reproduced in multiple subsequent papers. Figure 4 in the present article shows computationally derived molecular models of the chromophore-containing pockets in the A, B, and I forms.⁵⁰

The relevant experimental evidence can be summarized as follows. Form A, with the absorption band maximum at ~395 nm (3.14 eV), consists of the neutral chromophore and deprotonated Glu222 (a critical residue near the chromophore). Form B absorbs at ~475 nm (2.61 eV) and emits at ~508 nm (2.44 eV); it features the anionic chromophore (deprotonated at the phenolic oxygen) and protonated Glu222. Under physiological conditions, the population of form B is approximately one-sixth that of form A, meaning that in the ground state form A is lower in energy. Transformations between A and B occur through an intermediate form termed I. Upon excitation of the neutral chromophore, series of proton transfer steps occur via the hydrogen-bond network that includes the phenolic ring of the chromophore, the nearby water molecule, and the Ser205 and Glu222 side chains, ultimately resulting in the anionic chromophore and protonated Glu222. The step from A to I is simply a protonation change, whereas the transition from I to B includes a slow conformational change of the protein.

The results of computational modeling⁵⁰ summarized in Figure 4 illustrate the success of the theory in reproducing the bulk of the experimental data. The technical details of computational protocols are discussed below. In short, a model system mimicking form A of the wt-GFP was constructed from the relevant crystal structure; the structural parameters and the respective energies on the ground state (S_0) potential energy surface were computed using the QM(PBE0/6-31G*)/MM(AMBER) approach; minimum-energy points on the excited-state (S_1) potential energy surface were optimized using the configuration interaction singles method; the vertical $S_0 \rightarrow S_1$ and $S_1 \rightarrow S_0$ transition energies were computed by a variety of methods including the XMCQDPT2/SAX-CASSCF(N/M) and SOS-CIS(D) approaches. In the former, the string SAX-CASSCF(N/M) refers to an X-state-averaged SA-CASSCF calculation with N electrons distributed over M orbitals, XMCQDPT2 designates

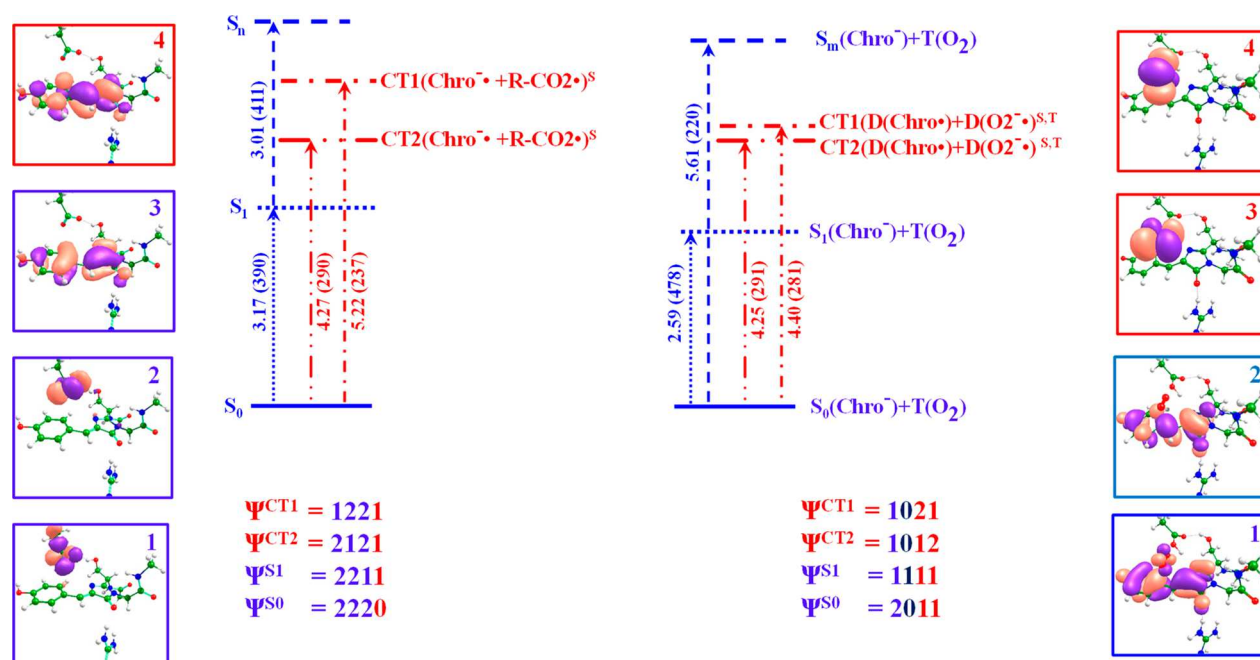


Figure 5. Excited states in GFP according to the calculations from refs 69 and 70. The panels showing molecular orbitals with variable occupancies in the wave functions Ψ are placed at the left and right sides. The energy diagrams and the wave function composition in the ground state (Ψ^{S0}), the first excited state (Ψ^{S1}), and the states of charge-transfer character (Ψ^{CT1} and Ψ^{CT2}) are shown in the center. The left part of the figure shows the system with the neutral chromophore and the side chain of Glu222. The right part of the figure shows the system with the anionic chromophore and molecular oxygen.

the extended multiconfiguration quasi-degenerate perturbation theory,⁵¹ and SOS-CIS(D) is the approach with scaled-opposite-spin corrections to second-order contributions to configuration interaction singles.^{52,53}

A series of molecular models⁵⁰ describing the photoinduced intermediates for the wt-GFP and the GFP-S65T mutant provided a concrete theoretical support to the three-state model originally based on purely experimental evidence. According to the calculation results, the structure representing form A in the wild type GFP is the lowest in energy, whereas form B is about 1 kcal/mol higher, in accord with the experimentally derived population analysis of GFP conformations. Form I is energetically slightly higher than B, and a high energy barrier separating I and B is consistent with the slow $I \rightarrow B$ transition. The most notable change near the chromophore at this reaction segment is the disruption of the Ser205–Glu222 hydrogen bond due to the *anti* $\rightarrow$ *syn* flip of Glu222 (Figure 4). In accord with the experimental evidence, in the GFP-S65T mutant (which has no spectral signatures of the A form⁵⁴), the structures with the anionic chromophore are significantly lower in energy than their counterparts with the neutral chromophore. The calculations⁵⁰ led to the reexamination of the role of the nearby amino acid residues in the chromophore-containing pocket. On the basis of the analysis of the crystal structures, it is commonly posited that the Thr203 residue in the chromophore-containing pocket plays a critical role by stabilizing the intermediates with the anionic chromophore,⁹ but the simulations did not support this proposal, emphasizing instead the hydrogen bonding of the chromophore with the His148 residue.

A subsequent computational study of the proton translocation along the transient hydrogen-bond network in GFP examined large segments of the potential energy surface in the ground and excited states⁵⁵ and reported the structures and

energetics that were close to those presented in ref 50. Given that different protocols were employed in the two studies, this is an important result validating the robustness of the theory. The most recent work on modeling proton wire dynamics in GFP⁵⁶ also reported the results generally consistent with the findings of ref 50.

Thus, these studies illustrate the success of the employed theoretical protocols in describing structures, energies, and chemical transformations in the chromophore-binding pocket in the ground state. The computed wavelengths of the vertical $S_0 \rightarrow S_1$ and $S_1 \rightarrow S_0$ transitions summarized in Figure 4 show an excellent agreement between the theoretical and experimental values. Experimental positions of the absorption band maxima assigned to the A, I, and B forms are 395, 495, and 475 nm, respectively. The emission band maximum of the forms with the anionic chromophore lies at 508 nm; the values corresponding to the A form with the neutral chromophore are within 440–460 nm. We conclude that the assignment of the reaction intermediates in the GFP photocycle, which was the main goal of the computational study,⁵⁰ was performed successfully.

The accuracy of various protocols used in calculations of optical spectra for the systems like GFP with fairly large conjugated chromophores calls for discussion. The literature features numerous studies presenting the calculations of the electronic spectra of the protein-bound GFP chromophore.^{57–65} Olivucci et al.^{57,58} used the CASSCF method in QM to scan the potential energy surfaces and the multi-reference perturbation theory (CASPT2//CASSCF) to calculate the positions of the $S_0 \rightarrow S_1$ and $S_1 \rightarrow S_0$ transitions in GFP. Optical spectra of GFP and several mutated variants with different protonation states of the chromophore were calculated by Hasegawa et al.⁶⁰ using a configuration-interaction based method for electronic energy differences

between the ground and excited states and accounted for the effect of the protein matrix by QM/MM. Filippi et al.^{63,64} used a variety of electronic structure methods, including TD-DFT, multireference perturbation theory (e.g., CASPT2), quantum Monte Carlo in combination with static point charges, DFT embedding, and classical polarizable embedding, to study coupling of the photoexcited GFP chromophore with the protein environment. Schwabe et al.⁶⁵ employed a high level quantum chemistry method (second-order approximate coupled-cluster approach with the resolution of identity technique (RI-CC2)) to compare results of a hybrid RI-CC2/polarizable imbedding model and cluster approaches to compute accurate excitation energies in GFP. All of these studies emphasize the importance of the effect of the protein environment on the spectral properties of protein-bound chromophores as well as the need for reliable quantum chemistry methods for describing the electronic structure of the chromophore. The accuracy in the excitation-energy calculations in these photoactive proteins using advanced computational protocols can be illustrated by the results of modeling several families of photoreceptor proteins, as summarized, e.g., in Figure 1 in ref 66. The experimental peak maxima can be reproduced within approximately 20–30 nm in this spectral range, which corresponds to the 0.10–0.15 eV error bars in transition energies.

Eventually, upon repeated irradiation, all photoactive proteins gradually lose their optical output and become bleached, i.e., optically dead. The “death” of GFP is related to irreversible changes either in the chromophore or in the nearby molecular groups. From a practical point of view, photobleaching is often a parasitic process limiting imaging applications. However, several techniques exploit photobleaching, such as super-resolution imaging⁶⁷ and different methods for tracing protein dynamics by fluorescence loss and recovery.⁶⁸ More broadly, photoinduced chemical changes of the chromophore or nearby protein residues can lead to changes in color, which can be exploited in optical high-lighting, superresolution, and other applications.

The results of simulations of photoinduced decarboxylation⁶⁹ and chromophore decomposition⁷⁰ in the wt-GFP are discussed at length in the review article.³ For the goals of this paper, we stress that a careful consideration of highly excited electronic states, including those of the charge-transfer (CT) character, is required in these simulations. For example, the key step in modeling^{69,71} the decarboxylation of the Glu222 side chain upon irradiation of the A form of GFP by intense violet or UV light^{72,73} entailed identification of the CT states with sufficient oscillator strength and with excitation energy corresponding to the electron transfer from deprotonated Glu222 to the chromophore (Chro); these states, identified by the population analysis of their wave functions, are shown in the left part in Figure 5. The key molecular orbitals with variable occupancies in the wave function are shown in the square panels in Figure 5.

The first step in modeling irreversible bleaching of GFP in the fluorescent state (with the anionic chromophore, Chro[−]) upon photoexcitation in the presence of molecular oxygen was the identification of the CT states corresponding to the electron transfer Chro[−] → O₂ (the right part in Figure 5).⁷⁰ The calculations^{69,70} revealed that, once the CT states are reached, the system can undergo series of low-barrier transformations leading to either Glu222 side chain decarboxylation or the chromophore destruction.

Simulations of these processes begin with the optimization of the initial structures, e.g., the GFP structures in the A or B forms, either without or with molecular oxygen in the chromophore-containing pocket. This task can be efficiently performed at the QM(DFT)/MM level. Then, the estimates of excited state energies at the key stationary points should be performed by using advanced quantum-chemical approaches (e.g., see the recent review⁷⁴), followed by a careful analysis of the orbitals involved in different electronic states. The CASSCF-based wave functions are capable of describing the CT nature of the target states and of simulating the evolution of the electronically excited system up to the crossing with the ground-state surface. At this point, the computational strategy can switch back to the QM(DFT)/MM methods to follow the reaction pathway to reaction products. We note that the existing computational protocols cannot cover all relevant reaction segments with a uniformly high quantitative accuracy. Even more importantly, these calculations involve system-dependent parametrization (such as constructing model systems, devising QM/MM separation, choosing active space and state-averaging protocol in CASSCF, etc.) and cannot be applied in an automated black-box manner. Nevertheless, the insights derived from such calculations are very valuable, as they provide a molecular-level picture of very complex phenomena involving photoactive proteins.

The Design of the Triple-Decker Motif in GFP-Like Proteins. The attempts to engineer fluorescent proteins with red-shifted spectra and a less flexible chromophore have stimulated an extensive search of prospective chromoproteins from other natural organisms, paralleled by the efforts to modify the *p*-hydroxybenzylidene-imidazolinone chromophore (Figure 2a) and its immediate environment. The latter research proceeded along several directions.⁷⁵ In the rational design of proteins, point mutations is an important strategy. In the context of GFP, its early success is exemplified by the creation of the GFP-S65T variant⁷⁶ from wt-GFP. Red-shifted spectra can be realized by increasing the size of the conjugated π -system in the chromophore. A large family of red fluorescent proteins with such chromophores is widely employed.^{1,10,75} Another design idea entails the π -stacking between the conjugated chromophore and the neighboring aromatic groups. Development and refinement of such structure–property based hypotheses provides a basis for the rational design of novel variants with even more red-shifted optical bands and increased photostability.^{77–79}

In the GFP-derived systems, the π -stacking of the *p*-hydroxybenzylidene-imidazolinone chromophore with the tyrosine side chain was first exploited experimentally, resulting in the creation of the yellow fluorescent protein. To shift the absorption and emission bands by ~20 nm to the red, a few mutations were introduced in GFP, with the most critical one being T203Y.^{80,81} The crystal structures of the mutant showed nearly coplanar π -stacking between the chromophore and the introduced Tyr203. The same idea, the replacement of Thr by Tyr near the chromophore of the red fluorescent protein mCherry along with accompanying point mutations, resulted in the protein variant called mRojoA.⁷⁷ In ref 77, a combined computational and experimental approach was used to prescreen prospective mutants of mCherry *in silico* and to generate combinatorial libraries.

We used a different computational strategy to propose a construct called the triple-decker motif.⁸² On the basis of QM/MM calculations, we predicted a new structural pattern of the

conventional anionic GFP chromophore sandwiched between two tyrosine residues. The minimal set of four mutations T62Y/Y145S/I167G/T203Y was required to accommodate the Tyr side chain at position 62 instead of Thr, in the same manner as the replacement of Thr203 by Tyr in the yellow fluorescent protein.^{80,81} According to the calculations, the absorption/emission wavelengths of this triple-decker construct are shifted by ~ 40 nm (absorption) and ~ 25 nm (emission) relative to the parent species, namely, the I form of wt-GFP. The stationary points on the ground-state potential energy surface were located using the QM(PBE0/6-31G*)/MM(AMBER) method. Optimization of the S_1 excited-state geometry was performed by using the QM(CASSCF(10/9)/6-31G*)/MM(AMBER) method. Calculations of the vertical $S_0 \rightarrow S_1$ and $S_1 \rightarrow S_0$ energy differences were performed using the XMCQDPT2 and SOS-CIS(D) approaches.

This theoretical proposal⁸² has stimulated experimental efforts⁷⁹ to design a variant of the previously engineered mRojoA protein,⁷⁷ in which the chromophore was surrounded by two tyrosine side chains: Tyr63–chromophore–Tyr197 (numbering of residues is given by the structure of mCherry PDB ID: 2H5Q). Unfortunately, the desired coplanar arrangement of the three aromatic rings was not achieved, as clearly seen in the crystal structure of this variant (PDB ID: 5H89).⁷⁹ Rather, only the chromophore–Tyr197 pair is π -stacked in the protein, whereas the second tyrosine (Tyr63) is oriented perpendicularly, edge-to-face with respect to the chromophore.

On the basis of careful analysis of the protein structure of mRojoA, we suggested⁸³ a series of mutations capable of fixing Tyr63 in a coplanar orientation with respect to the chromophore, along with the previously achieved π -stacking of the chromophore with Tyr197 (Figure 6). Each replacement had a specific purpose. These mutations introduced molecular groups to form hydrogen bonds with the hydroxyl group of Tyr63 and to keep the benzene ring of Tyr63 parallel to the chromophore. The minimum-energy structures in the ground state were optimized with QM(DFT)/MM. Importantly, the calculations verified that such a triple-decker construct remained stable in the course of protein dynamics. To this end, we carried out molecular dynamics simulations with the QM/MM potentials. The QM parts comprised the chromophore, the side chains of Tyr63, Arg95, Glu215, Ser146, and Tyr197, and two water molecules. The computational protocol PBE-D3/TZV2P-MOLOPT-GTH was used in QM, and the

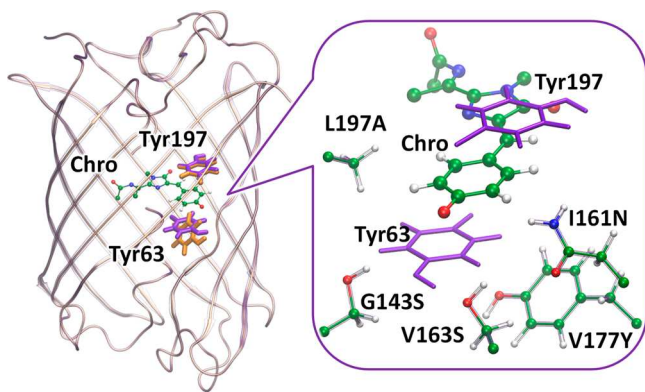


Figure 6. Improvement of the triple-decker motif in mRojoA, as suggested in ref 83.

AMBER force field, in MM. The simulations showed that different starting conformations converged to similar structures with the triple-decker motif (see Figure 6 in ref 83).

Computational protocols for characterizing the absorption and emission energies were essentially the same as those applied for the GFP-like triple-decker construct. Namely, the XMCQDPT2 and SOS-CIS(D) approaches were applied, using the ground-state structures optimized by QM(DFT)/MM and excited-state structures optimized by CASSCF. The results showed that this chromophore unit in the new protein variant had lower excitation energies than the parent mCherry and mRojoA species.

To conclude this section, we note that the computationally predicted triple-decker motif exploiting the π -stacking of the chromophore's phenolate group with tyrosine or other aromatic residues suggests a promising strategy toward the design of fluorescent proteins with the longer absorption/emission wavelengths and, possibly, increased fluorescence quantum yield.

■ SIMULATIONS OF FLAVIN-BINDING PROTEIN DOMAINS

In this section, we describe two examples of computational modeling of flavin-based photoreceptors: the search for better variants of iLOV and the characterization of molecular mechanisms of photoinduced transformations in the BLUF domain.

iLOV Story: Computational Predictions of iLOV Variants. As mentioned in the Introduction, an engineered protein iLOV^{20,21} is presently among the most popular imaging biomarkers because of its increased photostability, small size, and oxygen-independent fluorescence. However, its absorption (447 nm, 2.77 eV) and emission (497 nm, 2.49 eV) are far from the biotransparent window (650–900 nm) in which light easily penetrates through mammalian tissues. This shortcoming motivated research efforts aiming to design variants with optical bands shifted to longer wavelengths.

A computational study⁸⁴ proposed to replace Gln489 in the protein structure PDB ID: 4EET²¹ (Figure 7) by lysine. The rationale for this mutation is that the positively charged amino group (Lys) near the flavin chromophore can stabilize its π -electron system in the excited state. For example, the charged arginine side chain is involved in the π -stacking with the

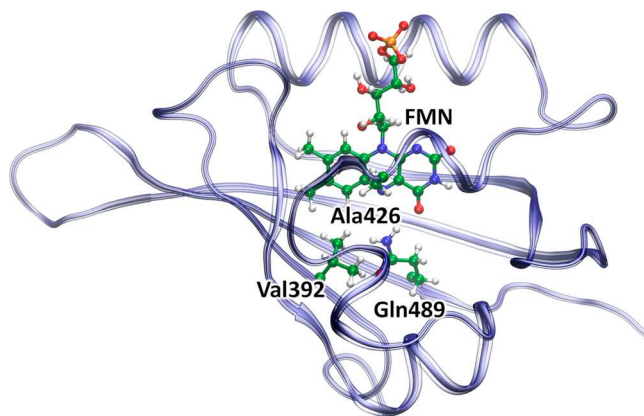


Figure 7. iLOV model constructed from the crystal structure PDB ID: 4EET showing the FMN cofactor and several amino acid side chains near FMN.

chromophore in the eqFP650 and eqFP670 proteins.⁸⁵ The effect of the electrostatic fields due to the charged groups on the spectra has also been discussed for rhodopsins (see, e.g., ref 86) and, more recently, for several other systems.⁸⁷

Indeed, the calculations described in ref 84 show that the mutant iLOV-Q489K should have red-shifted absorption and emission bands, if the side chain of lysine can be accommodated inside the chromophore-binding pocket. However, the later work of Davari et al.⁸⁸ demonstrated that, in the dominant population of iLOV-Q489K conformers, the amino group of Lys489 points away from the chromophore, whereas the population with the charged amino group near the chromophore is minor. Consequently, in the experimentally engineered⁸⁸ variant iLOV-Q489K, the optical bands were only slightly shifted to the shorter wavelengths relative to iLOV.

Importantly, despite distinctly different computational protocols (XMCQDPT2/SA-CASSCF(12/12) in ref 84 and TD-DFT (B3LYP/SVP) in ref 88), both papers^{84,88} agree that *in silico* the variant iLOV-Q489K should feature red-shifted bands in the case of proper orientation of Lys toward the chromophore.

The lesson learned from these studies^{84,88} is that a careful analysis of protein conformations based on extensive molecular dynamics simulations is a necessary step for a reliable prediction of protein properties. Equipped with this experience, we carried out⁸⁹ the next attempt to predict red-shifted iLOV mutants. The idea was to lock the lysine side chain in a favorable position with the amino group near the isoalloxazine ring, preventing its flip away from the chromophore. Unfortunately, as was found in ref 88 (and confirmed in a later work⁸⁹), an attempt to introduce the charged amino group of Lys close to the chromophore by single mutation Q489K was not successful. To achieve this goal, more mutations are necessary. Figure 7 shows positions in the structure of iLOV, which played key roles in the computational design of the variants capable of accommodating the Lys amino group near the isoalloxazine ring. A comprehensive discussion of the computational strategy is given in the original paper;⁸⁹ only the most salient findings are discussed here.

Figure 8 shows the molecular groups in the chromophore-containing pocket of two prospective iLOV mutants: iLOV-Q489K/L470T (left panel) and iLOV-V392K/F410 V/A426S (right panel). In the first case, Gln489 is replaced by Lys, but additionally, Leu470 is replaced by Thr. Importantly, the QM/MM dynamical simulations show that the iLOV-Q489K/

L470T construct is stable. Another promising place for lysine is a position occupied by Val392 in the initial structure (Figure 7). The variant iLOV-V392K/F410 V/A426S required three replacements; the MD simulations confirmed that the structure with the charged amino group near the chromophore is stable.

Calculations of the electronic energy differences between the two lowest singlet states were carried out at different theoretical levels, including XMCQDPT2 and SOS-CIS(D), and showed that both mutants should have absorption and emission band maxima significantly red-shifted with respect to the parent iLOV (Table 1).

Exploring a slightly different direction, a very recent computational study⁹⁰ proposed other variants of iLOV. This study reported series of model systems promising red-shifted spectra derived by the substitution of FMN by its analogues, 8-amino-FMN, 8-methylamino-FMN, and 1-deaza-FMN, coupled with additional point mutations.

Mechanisms of Photoinduced Transformations in the BLUF Domain. Unlike many other types of photoreceptors in which the chromophore undergoes changes upon irradiation, the activation mechanism of BLUF is accompanied by a reconstruction of the hydrogen-bond patterns near the isoalloxazine ring.^{91–93} The mechanism of formation of the light-signaling state as well as its back recovery in BLUF proteins is beyond the scope of this article; detailed discussion on these topics can be found in the recent papers.^{94–98} Instead, here we focus on the computationally characterized reactions in the chromophore-binding pocket, which suggest an interesting transformation in the side chain of the glutamine residue located near flavin.

In 2007, on the basis of FTIR studies of the BLUF protein, a tautomerization in the glutamine side chain in the course of the photoinduced reactions was tentatively proposed.⁹⁹ A year later, two independent computational papers described the reaction mechanisms in two different BLUF proteins (AppA-BLUF¹⁰⁰ and BlrB¹⁰¹), supporting the amide–imide tautomerization in the glutamine side chain. Figure 9, which is based on the data from ref 100, illustrates the main steps of the photoreaction.

The computationally derived mechanism for AppA-BLUF is consistent with the available crystal structures in the chromophore-containing pocket PDB ID: 2LYG and PDB ID: 1YRX; it also explains other experimental observations, including 10–15 nm shift in the absorption band maximum, 20 cm^{−1} shift of the characteristic vibrational band, and vanishing of the NMR peak corresponding to the amide group of glutamine upon the formation of the light state. Nevertheless, alternative mechanisms of BLUF photoreactions, without glutamine tautomerization, have also been discussed in the literature. Recently, further evidence for Gln tautomerization in BLUF domains was provided by experiments⁹⁵ using the isotope ¹⁵N labeling at the ϵ -position of the glutamine followed by the FTIR measurements and by additional calculations. The QM(PBE0/cc-pVDZ)/MM-(AMBER) calculations¹⁰² were carried out to optimize the structures of the BlrB-BLUF including the amide form (dark state) and imide form (light state) of Gln and to compute vibrational spectra of the corresponding species. Comparison of the computed and observed patterns in the double difference spectrum for the unlabeled and labeled bands in BlrB-BLUF provides a strong support to the model assuming glutamine tautomerization.

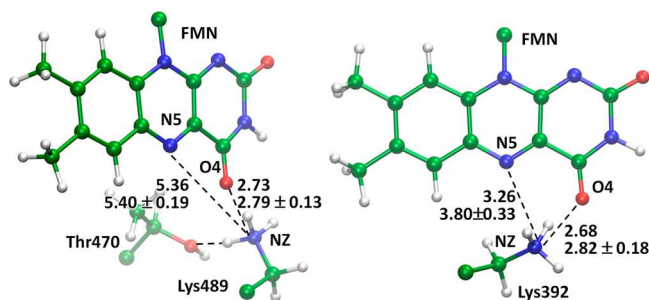
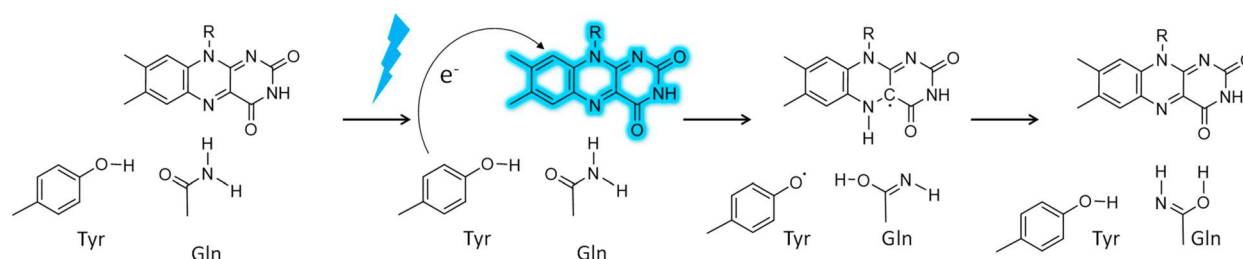
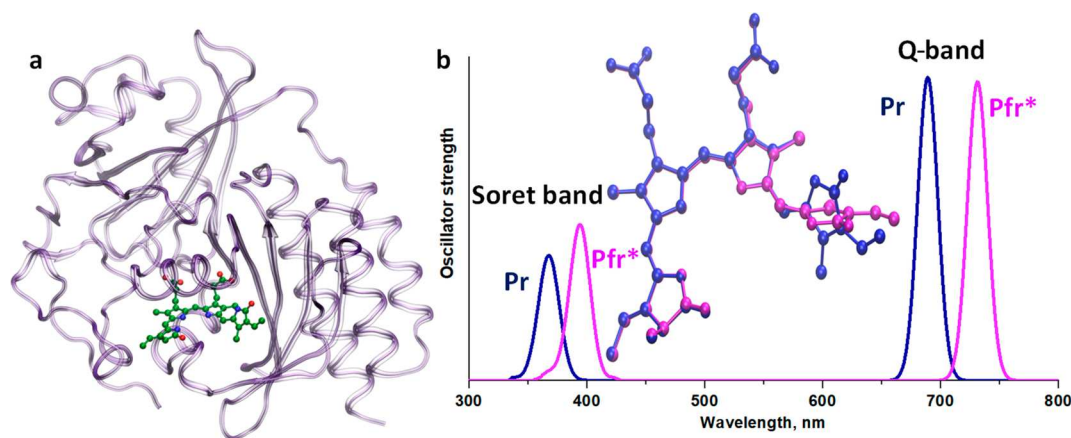


Figure 8. Selected molecular groups in the iLOV mutants: iLOV-Q489K/L470T (left) and iLOV-V392K/F410V/A426S (right). Distances (in Å) are from the QM(PBE0/cc-pVDZ)/MM(AMBER) optimized structures and from the QM/MM MD trajectories (average values with standard deviations in parentheses).

Table 1. Wavelengths (nm) and Transition Energies (eV) in Parentheses Corresponding to the Vertical Electronic Transitions between S_0 and S_1 in iLOV and Prospective Mutants Computed at the XMCQDPT2 and SOS-CIS(D) Levels

system	XMCQDPT2		SOS-CIS(D)	
	absorption	emission	absorption	emission
iLOV	441 (2.81)	491 (2.52)	394 (3.14)	480 (2.58)
iLOV-Q489K/L470T	469 (2.64)	533 (2.32)	416 (2.98)	518 (2.39)
iLOV-V392K/F410V/A426S	492 (2.52)	529 (2.34)	429 (2.89)	534 (2.32)

**Figure 9.** Photoreaction in BLUF domains involving the flavin chromophore, the glutamine, and the tyrosine side chains.**Figure 10.** (a): view on the chromophore-binding domain in IFP1.4. (b): computed spectra of the model system showing positions of the Q-band and Soret band in the biliverdin conformation corresponding to the Pr form (blue) and in the conformation with the twisted D-ring as expected in the Pfr form (red).

Illustrating the success of the theory, the most recent review⁹³ on BLUF proteins, devoted mostly to experimental studies, supports the model involving glutamine tautomerization in the chromophore-binding pocket.

SIMULATIONS OF PHYTOCHROMES

The infrared and near-infrared fluorescent proteins engineered from natural phytochrome domains are important for *in vivo* imaging because their absorption and emission bands fall into the optical transparency window of biological materials. The photocycle of natural phytochromes, carrying a covalently bound linear tetrapyrrole chromophore, usually switches between two stable forms, the red-adsorbing form (Pr) and the far-red-adsorbing form (Pfr). This mechanism is widely employed in plants, bacteria, and fungi to regulate a variety of functions including circadian rhythms and adjustments to the light spectrum. Photoexcitation of the Pr form initiates the double-bond isomerization at the methine bridge between the pyrrole rings C and D (Figure 2c), ultimately leading to the Pfr form. The absorption spectra of phytochromes with the bilin chromophores feature two bands: the main absorption peak known as the Q-band (660–770 nm) and the less intense Soret band (380–420 nm).

Computational studies of phytochromes are relatively scarce, as compared to other classes of photoreceptors. First of all, the large size of the chromophore (Figure 2c) precludes the use of many advanced excited-state methods. Because of that, the time-dependent density functional theory (TD-DFT) is the most popular approach in simulations of phytochrome excitation spectra,^{103–108} although for the gas-phase chromophore calculations using the symmetry-adapted cluster-configuration interaction (SAC-CI) method have been reported.¹⁰⁹

We focus here on two simulations of the infrared fluorescent protein IFP1.4¹¹⁰ constructed from a natural chromophore-binding domain (CBD) of the bacteriophytochrome from *Deinococcus radiodurans* (DrCBD) through a series of point mutations. Experimentally, it has been characterized by crystallography and spectroscopy studies. This domain carries a linear tetrapyrrole biliverdin IX α as a chromophore.

The authors of ref 106 constructed their model system from the crystal structure PDB ID: 2O9C of the protein reported at 1.45 Å resolution.¹¹¹ Optimization of the ground-state structure of the Pr form and calculations of the vertical excitation energy were performed with the ONIOM version of QM/MM by using TDDFT with a variety of functionals. On

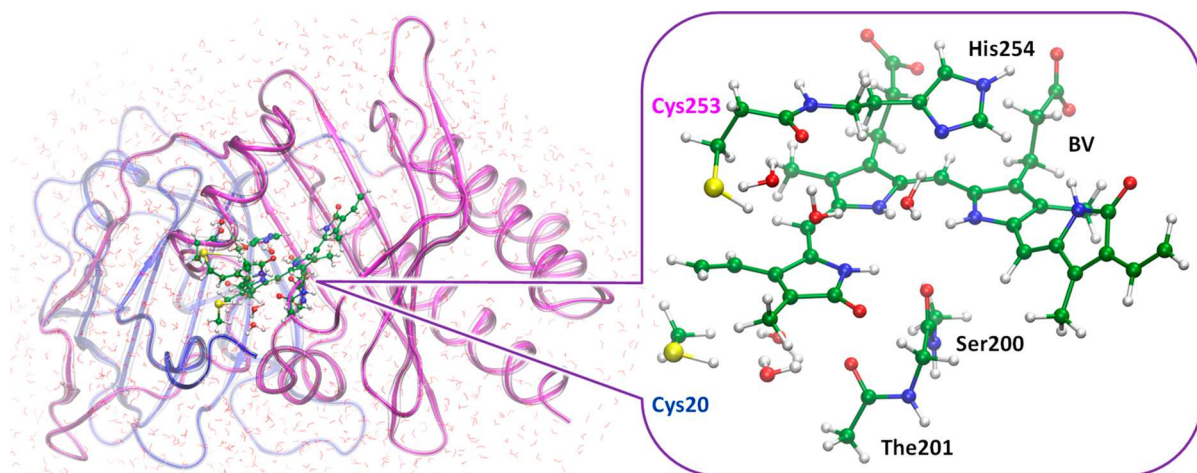


Figure 11. Illustration of BV binding to the phytochrome domains. The model system used in QM/MM calculations. The PAS domain is colored in blue, and the GAF domain is colored in magenta (left). The QM subsystem (right) is shown in ball and stick representations.

the basis of the computed excitation energies, the functionals were graded with respect to the experimental Q-band maximum. The authors concluded that the position of the Q-band was better reproduced using the chromophore's geometry from the crystal structure than with the QM/MM relaxed geometry parameters. Overall, the errors of more than 50 nm in absorption bands were reported. We also point out that using excitation energy alone for calibration purposes is not sufficient: often, the functionals showing the best performance for excitation energies produce qualitatively wrong wave functions.¹¹²

The authors of ref 113 relied on another protein crystal structure PDB ID: 5AJG obtained at 1.11 Å resolution¹¹⁰ and on an alternative computational protocol, involving QM-(PBE0/6-31G*)/MM optimization of the ground-state structure representing the Pr and Pfr forms. To design the latter structure, the D-ring of biliverdin was rotated, and the parameters were reoptimized in QM/MM calculations. Vertical excitation energies and corresponding oscillator strengths were calculated using the TD-DFT and XMCQDPT2/SA10-CASSCF(16/12)/cc-pVDZ calculations. The main findings are illustrated in Figure 10.

This work¹¹³ basically confirmed the conclusions formulated in refs 105 and 106 obtained with TD-DFT, that the excitation energies of the Q-band in the Pr form are overestimated by 0.2–0.3 eV. Calculations using XMCQDPT2 were more successful: the deviations of the computed energies from the experimental values were 0.05 eV for the Q-band (or 20 nm) and 0.12 eV for the Soret band. The theory reproduced the red-shift from Pr to Pfr form (0.1 eV, to be compared with 0.13 eV in experiments). Finally, it was demonstrated that there is no need to borrow structural parameters of the chromophore from crystal structures to simulate the absorption spectra. Consistent treatment of all parameters (both geometry and energy) solely by computations forms a firm basis for a systematic improvement of the predictive power of modeling such complex systems.

Another family of bacteriophytochrome-based infrared and near-infrared fluorescent proteins^{24,114,115} employs the same chromophore, biliverdin IX α (Figure 2c). In particular, Verkhusha et al. has successfully engineered a bright monomeric variant, miRFP670.^{24,115} To clarify its features, we note that most phytochromes share a common multi-

domain architecture consisting of two coupled photosensory domains called PAS and GAF which are connected to a phytochrome-specific domain. The bilin chromophore is accommodated in the GAF domain; however, it may be covalently bound to the cysteine residue in either the GAF or PAS domain.²³ The crystallography studies of miRFP670 revealed two different covalent adducts of the biliverdin–protein complexes; the chromophore molecule forms either a single thioether C–S bond with the cysteine residue from the GAF domain or with two cysteines, one from PAS and one from GAF. From experimental studies, the second reaction channel appears to be dominant.

The computational work¹¹⁶ clarified the mechanism of competing chemical reactions of protein assembly in miRFP670 leading to both products. A model structure of the non-covalently bound biliverdin molecule inside the protein cleft of miRFP670 was constructed from the crystal structure PDB ID: 5VIV.²⁴ The thioether covalent bonds between biliverdin and cysteines found in the crystal structure were manually cleaved, and the planar structure of the A-ring and vinyl group of biliverdin was restored. The QM/MM methods were employed to characterize the structures, energetics, and dynamics of reacting species along two reaction channels by binding the biliverdin molecule either to Cys253 from the GAF domain only or to both cysteines, Cys20 (PAS) and Cys253 (GAF) (Figure 11).

The results show that the nucleophilic attack of Cys253 from the GAF domain leads to a single C–S bond formation with an activation energy of 16 kcal/mol. Another pathway, leading to the biliverdin adduct with two C–S bonds, is characterized by lower energy barriers, less than 11 kcal/mol. The competition between these reaction pathways explains the experimentally observed mixture of both adducts. The kinetic scheme for these competing reactions derived in simulations is fully consistent with the experimental observations. This first simulation¹¹⁶ of phytochrome assembly represents a success of computational modeling.

DISCUSSION OF MODELS AND METHODS USED IN SIMULATIONS

As the first step in all simulations, relevant structures from the Protein Data Bank²⁵ are used to construct initial three-dimensional molecular models. Although the structures from

the PDB contain all structural parameters of a protein, the use of crystallographic models requires a certain care, especially for mobile regions of a protein.¹¹⁷ For example, conformations of the glutamine side chain in AppA-BLUF from the relevant PDB structures were discussed in ref 100 with respect to the position of nitrogen and oxygen atoms. With few exceptions, PDB structures contain only the coordinates of heavy (non-hydrogen) atoms; thus, the addition of hydrogen atoms is an important step. Usually, the negatively charged species for Asp and Glu and positively charged species for Lys and Arg are assumed. Additional help can be provided by PropKa,¹¹⁸ but every system must be considered with care. For instance, in GFP studies, Glu222 is negatively charged only in the A form with the neutral chromophore and it is protonated in forms I and B. Protonation states of His residues are selected manually, upon the inspection of all neighboring molecular groups. To validate the choice of protonation states, the comparison of the averaged structural parameters with those from the crystal structure is often used.^{119–121} Comparison of the experimental and computed IR and UV–vis absorption spectra is also helpful. Thus, obtained model systems are solvated in water (i.e., surrounded by a large amount of explicit water molecules) and neutralized. The number and initial positions of counterions are very important both for soluble and for membrane proteins. For example, one should add a sufficient number of counterions to neutralize all surface groups of the protein, and not only the overall excess charge.^{62,122}

Molecular dynamics simulations are then carried out, to relax the coordinates of the initial all-atom models. Conventional force fields (AMBER, CHARMM, etc.) are usually employed, but those need to be augmented by specially derived parameters for nonstandard residues, such as the chromophore molecules.^{120,123} The NAMD,¹²⁴ Gromacs,¹²⁵ and VMD¹²⁶ software packages are very useful for preparing models for subsequent quantum-based simulations.

The quantum mechanics/molecular mechanics (QM/MM) methods^{127,128} are essential tools for studying the properties of fluorescent and photoactive proteins. The chromophore moieties and neighboring molecular groups are assigned to QM subsystems, and the rest of the protein is treated at the MM level. Algorithms of QM/MM calculations implemented in many modern software packages enable calculations with fairly large QM subsystems, which reduces the sensitivity of the results to the QM–MM boundary. Molecular models shown in Figures 4, 6, and 11 illustrate the QM choices typical for the systems discussed in this article. For example, inclusion of the chromophore properly capped from the peptide chains to which it is covalently bound, the side chains of the critical glutamine and arginine residues (Glu222 and Arg96 in GFP), the side chains of residues (Ser205, Thr203, and His148 in GFP), and water molecules involved in the hydrogen-bond network near the chromophore appears to be necessary for simulations of GFP and related proteins. Such selections are guided by chemical intuition and play a critical role in obtaining qualitatively correct results. To obtain more robust quantitative results, rational and automated strategies to determine QM/MM partitioning are being developed.^{129,130} Once a molecular model and a QM/MM setup are selected, methods to compute energies and forces in the QM and MM must be specified. We refer to the recent review article¹³¹ discussing the application of QM/MM in complex environments for computational spectroscopy, which is also important for the goals of the present article.

With minor exceptions, a choice of force field parameters for MM is not critical, in contrast to the choice of quantum chemical methods. The Kohn–Sham DFT is almost exclusively used in QM in calculations of the structures and chemistry in the chromophore-binding pockets in the ground electronic state. Selection of a specific functional for these calculations is an important step in simulations; see, e.g., the review article¹³² benchmarking a total of 200 functionals. The majority of studies of GFP-related systems and flavin-containing and phytochrome protein domains mentioned in the present article employed the popular functional PBE0,¹³³ often augmented with the dispersion correction D3.¹³⁴ Several studies^{132,135} praise the PBE0 functional as one of the most reliable DFT options. According to the conclusions formulated in the review article,¹³² the structural parameters computed with PBE0-D3 are reliable, but the errors about 3 kcal/mol in computing barrier heights on the reaction profiles are to be expected. The analysis performed in ref 132 shows that it is unrealistic to expect errors in barrier heights less than 1.7 kcal/mol with any DFT functional. Improvement in DFT is a matter of current research, and success along this direction would elevate computational predictions to a higher level.¹³⁶

The lessons learned from modeling the systems with the triple-decker motif and with iLOV mutants clearly demonstrated the importance of extensive equilibrium sampling of protein conformations. We emphasize here the significance of carrying out dynamical simulations with the QM/MM potentials, because setting up classical MD with conventional force fields involves significant efforts for deriving parameters for the chromophores. If executed with poorly chosen parameters, such simulations damage the structures of the chromophore-binding pocket. In this regard, the extension of an effective fragment potential method to biomolecules shows much promise.^{137–139}

Finally, we comment on calculations of energy differences between electronic states. All proteins considered in the present paper carry the organic chromophores sharing a common feature: a π -conjugated system. Their electronic structure is deceptively simple and can be understood within the Hückel model.² The challenge is to calculate the corresponding transition energies with errors less than, say, 0.3 eV typical for many quantum chemistry approaches. Once converted to the wavelength scale in the corresponding spectral range, such errors (up to 50 nm in positions of spectral peaks) may appear too large from the point of view of the experimentalists.

As illustrated by the examples discussed in this article, the XMCQDPT2 and SOS-CIS(D) methods often produce excitation within 10 nm from the experiment. However, in both cases, careful validation for each new system is required. As any other second-order method, SOS-CIS(D) can fail unexpectedly, especially when describing electronic states of different nature. Furthermore, it is not applicable for computing global potential energy surfaces and bond-breaking. Thus, there is a strong motivation for developing algorithms and computer codes that can reduce computational cost and scaling of the more robust EOM-CCSD method¹³⁹ and improve its accuracy by including triples corrections. The XMCQDPT2 method can be tweaked to deliver accurate results by system-specific choice of the active states and state-averaging procedure but is quite expensive. This level of theory is useful for well-thought-out projects involving fluorescent and photoactive proteins.

Nevertheless, despite these limitations, the theory is capable of describing general trends when comparing different proteins of the same family and different mutants quite reliably (see, e.g., ref 129). Even more importantly, it provides a vehicle for visualization and hypothesis testing. The calculations deliver important insights into the intricate details of underlying mechanisms. We hope that vigorous research aiming to improve the electronic structure methodology may deliver robust and inexpensive methods, suitable for large-scale calculations and more extensive screening. For example, improving the predictive power of the TD-DFT method^{140,141} would have considerable impact on this field.

Another important challenge is the complexity of the workflows commonly employed to tackle these types of projects. As discussed above, for a complete description of a new system, one needs to create an all-atom model, which involves several critically important steps. The calculations combine molecular dynamics and QM/MM calculations using a variety of methods. Often, several software packages are used within the same project. Thus, improving interoperability of molecular modeling codes and developing more robust software tools that could help to automate the key steps and facilitate the creation of easily customizable workflows combining multiple algorithms and multiple software stacks is highly desirable.^{129,142}

CONCLUSION

In this Feature Article, we analyzed several projects devoted to computational simulations of structure and properties of fluorescent and photoactive proteins of three different types, which are used as imaging biomarkers in living cells. We considered the well-known GFP and its derivatives, the flavin-based proteins, and the protein domains from bacteriophytochromes. Although quite different in their chemical origin, the organic chromophores of these proteins share a common feature: an extensive π -conjugated system. Therefore, one may expect that similar computational protocols can be employed to model these systems. The analysis shows that the computational characterization of structures of the chromophore-binding pockets and chemical transformations occurring on the ground electronic state energy surfaces leads to reliable results consistent with experimental data, particularly with available crystal structures and kinetic studies. This point is illustrated by the simulations of GFP maturation, GFP reaction intermediates, modeling transformations in the BLUF domain, and phytochrome assembly. The QM(DFT)/MM approach is an adequate tool for such applications, although errors in calculating barrier heights along reaction profiles as large as 3–4 kcal/mol¹³² may be expected with the standard functionals. Moreover, the cancellation of errors, which is common in the QM/MM models, makes such predicted accuracy increase uncertain. With improved functionals and improved schemes for the QM/MM partitioning,¹³⁰ the accuracy should increase, furthering the progress in computer modeling of the processes in photoactive and fluorescent proteins. Computational design of novel protein mutants, as illustrated by the examples of the triple-decker motif and iLOV variants, demonstrates a strong need to employ QM/MM dynamical simulations for equilibrium sampling and for the verification of the stability of various protein conformations. Development of methods for the dynamical QM/MM simulations is important for multiple applications of computer simulations, including the field of photoreceptors.¹⁴³ The accuracy and robustness of calculations

of excited electronic states are still to be improved. Treatments of large organic chromophores (such as the ones in Figure 2) embedded into proteins using conventional quantum chemistry methods within the QM/MM scheme cannot at present claim the accuracy of a few nm even for the lowest electronic transition. As illustrated by the examples here, a properly configured XMCQDPT2 method can describe electronic transitions within ~ 10 nm from the experimental values. The accuracy of less expensive TD-DFT methods does not approach that level, although they are very useful to recognize the trends, say, in series of mutated protein species. Modeling structures and dynamics of excited states still presents a big challenge to the theory, motivating further research. On an optimistic note, we stress that the ability to visualize, at the atomic level, processes in such complicated systems as fluorescent and photoactive proteins is the greatest achievement of computer simulations. We believe that this field will continue to expand and include more complex and, consequently, more interesting objects. In particular, we expect that modeling of novel systems, such as components of optogenetic modules and photoactivated enzymes,^{15,144–146} will expand the scope of quantum-based computational photobiology.

AUTHOR INFORMATION

Corresponding Author

*E-mail: anemukhin@yahoo.com; anem@lcc.chem.msu.ru.

ORCID

Alexander V. Nemukhin: 0000-0002-4992-6029

Maria G. Khrenova: 0000-0001-7117-3089

Anna I. Krylov: 0000-0001-6788-5016

Notes

The authors declare no competing financial interest.

Biographies



Alexander V. Nemukhin is currently Professor of Physical Chemistry at the Chemistry Department of the M.V. Lomonosov Moscow State University (Russia) and the Director of Laboratory for Computer Modeling of Biomolecules and Nanomaterials at the N.M. Emanuel Institute of Biochemical Physics of Russian Academy of Sciences. He is holding a title of Distinguished Professor of the M.V. Lomonosov Moscow State University. Prof. Nemukhin received his Ph.D. in 1975 and Doctoral degree in 1989, both from the Moscow State University. The theses were focused on quantum-chemical studies of molecular properties. His early work was devoted, in part, to modeling properties of matrix-isolated species at low temperatures. Since 2000, his main research interest has been development and

application of molecular modeling methods to study biomolecular systems.



Bella L. Grigorenko received her Ph.D. at the Physics Department of the M.V. Lomonosov Moscow State University (Russia) in the field of Solid State Physics. Her Doctoral degree devoted to molecular modeling (2004) is from the Department of Chemistry at the Moscow State University. Dr. Grigorenko is currently a Leading Researcher at the Physical Chemistry Division of the Moscow State University and a Researcher at the N.M. Emanuel Institute of Biochemical Physics of Russian Academy of Sciences. Dr. Grigorenko's main interests are the development and application of molecular modeling methods to study biomolecular systems. Grigorenko's contributions to this field have been recognized by the award from Russian Corporation of Nanotechnologies (2010) and the award from the Russian Ministry of Science and High Education (2006).



Maria G. Khrenova is currently a Leading Researcher at the M.V. Lomonosov Moscow State University (Russia) and the Head of the Group of Molecular Modeling at the "Fundamentals of Biotechnology" Research Center of the Russian Academy of Sciences. She received her Ph.D. in 2011 and Doctoral degree in 2016, both from the Moscow State University. Her research interests are focused on the application of advanced computational chemistry methods to study processes relevant to biochemistry, biomedicine, and biotechnology. Dr. Khrenova also contributes to promoting and popularizing science through technical seminars, public lectures, and mass media. She regularly participates in the "Science Festival", which takes place in all major Russian cities every fall. She has delivered lectures inspiring teenagers about Chemistry and its applications. In an effort of popularizing science, she has given many interviews to Russian newspapers and TV programs. Dr. Khrenova has been recognized by several awards including the Stipend of President of the Russian Federation, the Moscow Government Award, and the

Shuvalov Award. Her research was supported by the L'Oréal-UNESCO for Women in Science program in 2014.



Anna I. Krylov received her M.Sc. (with honors) in Chemistry from the Moscow State University (Russia) in 1990 and her Ph.D. (cum laude) in Physical Chemistry from the Hebrew University of Jerusalem (Israel) in 1996. After postdoctoral training in Prof. M. Head-Gordon's group at UC Berkeley, she started her research in electronic structure theory at the Department of Chemistry of the University of Southern California in Los Angeles, where she is currently Gabilan Distinguished Professor in Science and Engineering and Professor of Chemistry. Krylov's research is focused on theoretical modeling of open-shell and electronically excited species. Krylov's vision is to develop and devise computational tools for treatment of complicated open-shell electronic structures ranging from bound and unbound excited states to complicated polyradical species in the gas phase and in complex environments (solutions, molecular solids, and proteins). She develops robust black-box methods to describe complicated multiconfigurational wave functions in single-reference formalisms, such as coupled-cluster and equation-of-motion approaches. Using the tools of computational chemistry, and in collaboration with experimental groups, Krylov also investigates the role that radicals and electronically excited species play in such diverse areas as combustion, gas- and condensed-phase chemistry, astrochemistry, solar energy, quantum information storage, bioimaging, and light-induced processes in biology. Krylov's contributions to electronic structure method development (in particular, the spin-flip method) have been recognized by several awards including the WATOC's Dirac medal (2007), the Bessel Prize from the Humboldt Foundation (2011), the Theoretical Chemistry Award from the Physical Chemistry Division of ACS (2012), and Mainz Guestprofessorship and Mildred Dresselhaus Award from University of Hamburg (2018). She is an elected member of the International Academy of Quantum Molecular Science, a Fellow of the American Physical Society, the American Chemical Society, and the American Association of Advancement of Science, and 2018 Simons Fellow in Theoretical Physics.

■ ACKNOWLEDGMENTS

This work was supported by the Russian Science Foundation (Project No. 17-13-01051) to B.L.G. and M.G.K. and by the U.S. National Science Foundation (No. CHE-1566428) to A.I.K. The research is carried out using the equipment of the shared research facilities of HPC computing resources at Lomonosov Moscow State University supported by the project RFMEFI62117X0011. The use of supercomputer resources of the Joint Supercomputer Center of the Russian Academy of Sciences is also acknowledged.

REFERENCES

- (1) Rodriguez, E. A.; Campbell, R. E.; Lin, J. Y.; Lin, M. Z.; Miyawaki, A.; Palmer, A. E.; Shu, X.; Zhang, J.; Tsien, R. Y. The Growing and Glowing Toolbox of Fluorescent and Photoactive Proteins. *Trends Biochem. Sci.* **2017**, *42* (2), 111–129.
- (2) Bravaya, K. B.; Grigorenko, B. L.; Nemukhin, A. V.; Krylov, A. I. Quantum Chemistry Behind Bioimaging: Insights from Ab Initio Studies of Fluorescent Proteins and Their Chromophores. *Acc. Chem. Res.* **2012**, *45* (2), 265–275.
- (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.
- (4) Gozem, S.; Luk, H. L.; Schapiro, I.; Olivucci, M. Theory and Simulation of the Ultrafast Double-Bond Isomerization of Biological Chromophores. *Chem. Rev.* **2017**, *117* (22), 13502–13565.
- (5) Schwabe, T. Review of Proteinochromism Modeling in Computational Chemistry. *Curr. Org. Chem.* **2017**, *21* (10), 856–871.
- (6) Tsien, R. Y. The Green Fluorescent Protein. *Annu. Rev. Biochem.* **1998**, *67* (1), 509–544.
- (7) Zimmer, M. Green Fluorescent Protein (GFP): Applications, Structure, and Related Photophysical Behavior. *Chem. Rev.* **2002**, *102* (3), 759–782.
- (8) Craggs, T. D. Green Fluorescent Protein: Structure, Folding and Chromophore Maturation. *Chem. Soc. Rev.* **2009**, *38* (10), 2865–2875.
- (9) Remington, S. J. Green Fluorescent Protein: A Perspective. *Protein Sci.* **2011**, *20* (9), 1509–1519.
- (10) Subach, F. V.; Verkhusha, V. V. Chromophore Transformations in Red Fluorescent Proteins. *Chem. Rev.* **2012**, *112* (7), 4308–4327.
- (11) Nienhaus, K.; Ulrich Nienhaus, G. Fluorescent Proteins for Live-Cell Imaging with Super-Resolution. *Chem. Soc. Rev.* **2014**, *43* (4), 1088–1106.
- (12) Losi, A.; Gärtner, W. Old Chromophores, New Photoactivation Paradigms, Trendy Applications: Flavins in Blue Light-Sensing Photoreceptors. *Photochem. Photobiol.* **2011**, *87* (3), 491–510.
- (13) Chen, E.; Swartz, T. E.; Bogomolni, R. A.; Kliger, D. S. A LOV Story: The Signaling State of the Phot1 LOV2 Photocycle Involves Chromophore-Triggered Protein Structure Relaxation, As Probed by Far-UV Time-Resolved Optical Rotatory Dispersion Spectroscopy. *Biochemistry* **2007**, *46* (15), 4619–4624.
- (14) Christie, J. M.; Gawthorne, J.; Young, G.; Fraser, N. J.; Roe, A. J. LOV to BLUF: Flavoprotein Contributions to the Optogenetic Toolkit. *Mol. Plant* **2012**, *5* (3), 533–544.
- (15) Krauss, U.; Lee, J.; Benkovic, S. J.; Jaeger, K.-E. LOVely Enzymes - towards Engineering Light-Controllable Biocatalysts. *Microb. Biotechnol.* **2010**, *3* (1), 15–23.
- (16) Mukherjee, A.; Walker, J.; Weyant, K. B.; Schroeder, C. M. Characterization of Flavin-Based Fluorescent Proteins: An Emerging Class of Fluorescent Reporters. *PLoS One* **2013**, *8* (5), e64753.
- (17) Wingen, M.; Potzkei, J.; Endres, S.; Casini, G.; Rupprecht, C.; Fahlke, C.; Krauss, U.; Jaeger, K.-E.; Drepper, T.; Gensch, T. The Photophysics of LOV-Based Fluorescent Proteins – New Tools for Cell Biology. *Photochem. Photobiol. Sci.* **2014**, *13* (6), 875–883.
- (18) Mukherjee, A.; Schroeder, C. M. Flavin-Based Fluorescent Proteins: Emerging Paradigms in Biological Imaging. *Curr. Opin. Biotechnol.* **2015**, *31*, 16–23.
- (19) Buckley, A. M.; Petersen, J.; Roe, A. J.; Douce, G. R.; Christie, J. M. LOV-Based Reporters for Fluorescence Imaging. *Curr. Opin. Chem. Biol.* **2015**, *27*, 39–45.
- (20) Chapman, S.; Faulkner, C.; Kaiserli, E.; Garcia-Mata, C.; Savenkov, E. I.; Roberts, A. G.; Oparka, K. J.; Christie, J. M. The Photoreversible Fluorescent Protein iLOV Outperforms GFP as a Reporter of Plant Virus Infection. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (50), 20038–20043.
- (21) Christie, J. M.; Hitomi, K.; Arvai, A. S.; Hartfield, K. A.; Mettlen, M.; Pratt, A. J.; Tainer, J. A.; Getzoff, E. D. Structural Tuning of the Fluorescent Protein iLOV for Improved Photostability. *J. Biol. Chem.* **2012**, *287* (26), 22295–22304.
- (22) Shcherbakova, D. M.; Balaban, M.; Verkhusha, V. V. Near-Infrared Fluorescent Proteins Engineered from Bacterial Phytochromes. *Curr. Opin. Chem. Biol.* **2015**, *27*, 52–63.
- (23) 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.
- (24) Balaban, M.; Shcherbakova, D. M.; Pletnev, S.; Pletnev, V. Z.; Lagarias, J. C.; Verkhusha, V. V. Designing Brighter Near-Infrared Fluorescent Proteins: Insights from Structural and Biochemical Studies. *Chem. Sci.* **2017**, *8* (6), 4546–4557.
- (25) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28* (1), 235–242.
- (26) Heim, R.; Prasher, D. C.; Tsien, R. Y. Wavelength Mutations and Posttranslational Autooxidation of Green Fluorescent Protein. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91* (26), 12501–12504.
- (27) Barondeau, D. P.; Putnam, C. D.; Kassmann, C. J.; Tainer, J. A.; Getzoff, E. D. Mechanism and Energetics of Green Fluorescent Protein Chromophore Synthesis Revealed by Trapped Intermediate Structures. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (21), 12111–12116.
- (28) Barondeau, D. P.; Kassmann, C. J.; Tainer, J. A.; Getzoff, E. D. Understanding GFP Chromophore Biosynthesis: Controlling Backbone Cyclization and Modifying Post-Translational Chemistry. *Biochemistry* **2005**, *44* (6), 1960–1970.
- (29) Wood, T. I.; Barondeau, D. P.; Hitomi, C.; Kassmann, C. J.; Tainer, J. A.; Getzoff, E. D. Defining the Role of Arginine 96 in Green Fluorescent Protein Fluorophore Biosynthesis. *Biochemistry* **2005**, *44* (49), 16211–16220.
- (30) Barondeau, D. P.; Kassmann, C. J.; Tainer, J. A.; Getzoff, E. D. Understanding GFP Posttranslational Chemistry: Structures of Designed Variants That Achieve Backbone Fragmentation, Hydrolysis, and Decarboxylation. *J. Am. Chem. Soc.* **2006**, *128* (14), 4685–4693.
- (31) Barondeau, D. P.; Tainer, J. A.; Getzoff, E. D. Structural Evidence for an Enolate Intermediate in GFP Fluorophore Biosynthesis. *J. Am. Chem. Soc.* **2006**, *128* (10), 3166–3168.
- (32) Barondeau, D. P.; Kassmann, C. J.; Tainer, J. A.; Getzoff, E. D. The Case of the Missing Ring: Radical Cleavage of a Carbon–Carbon Bond and Implications for GFP Chromophore Biosynthesis. *J. Am. Chem. Soc.* **2007**, *129* (11), 3118–3126.
- (33) Zhang, L.; Patel, H. N.; Lappe, J. W.; Wachter, R. M. Reaction Progress of Chromophore Biogenesis in Green Fluorescent Protein. *J. Am. Chem. Soc.* **2006**, *128* (14), 4766–4772.
- (34) Wachter, R. M. Chromogenic Cross-Link Formation in Green Fluorescent Protein. *Acc. Chem. Res.* **2007**, *40* (2), 120–127.
- (35) Pouwels, L. J.; Zhang, L.; Chan, N. H.; Dorrestein, P. C.; Wachter, R. M. Kinetic Isotope Effect Studies on the de Novo Rate of Chromophore Formation in Fast- and Slow-Maturing GFP Variants. *Biochemistry* **2008**, *47* (38), 10111–10122.
- (36) Reid, B. G.; Flynn, G. C. Chromophore Formation in Green Fluorescent Protein. *Biochemistry* **1997**, *36* (22), 6786–6791.
- (37) Pletneva, N. V.; Pletnev, V. Z.; Lukyanov, K. A.; Gurskaya, N. G.; Goryacheva, E. A.; Martynov, V. I.; Wlodawer, A.; Dauter, Z.; Pletnev, S. Structural Evidence for a Dehydrated Intermediate in Green Fluorescent Protein Chromophore Biosynthesis. *J. Biol. Chem.* **2010**, *285* (21), 15978–15984.
- (38) Grigorenko, B. L.; Krylov, A. I.; Nemukhin, A. V. Molecular Modeling Clarifies the Mechanism of Chromophore Maturation in the Green Fluorescent Protein. *J. Am. Chem. Soc.* **2017**, *139* (30), 10239–10249.
- (39) Siegbahn, P. E. M.; Wirstam, M.; Zimmer, M. Theoretical Study of the Mechanism of Peptide Ring Formation in Green Fluorescent Protein. *Int. J. Quantum Chem.* **2001**, *81* (2), 169–186.
- (40) Ma, Y.; Sun, Q.; Li, Z.; Yu, J.-G.; Smith, S. C. Theoretical Studies of Chromophore Maturation in the Wild-Type Green Fluorescent Protein: ONIOM(DFT:MM) Investigation of the

Mechanism of Cyclization. *J. Phys. Chem. B* **2012**, *116* (4), 1426–1436.

(41) Ma, Y.; Zhang, H.; Sun, Q.; Smith, S. C. New Insights on the Mechanism of Cyclization in Chromophore Maturation of Wild-Type Green Fluorescence Protein: A Computational Study. *J. Phys. Chem. B* **2016**, *120* (24), 5386–5394.

(42) Ma, Y.; Yu, J.-G.; Sun, Q.; Li, Z.; Smith, S. C. The Mechanism of Dehydration in Chromophore Maturation of Wild-Type Green Fluorescent Protein: A Theoretical Study. *Chem. Phys. Lett.* **2015**, *631–632*, 42–46.

(43) Ma, Y.; Sun, Q.; Smith, S. C. The Mechanism of Oxidation in Chromophore Maturation of Wild-Type Green Fluorescent Protein: A Theoretical Study. *Phys. Chem. Chem. Phys.* **2017**, *19* (20), 12942–12952.

(44) Shinobu, A.; Palm, G. J.; Schierbeek, A. J.; Agmon, N. Visualizing Proton Antenna in a High-Resolution Green Fluorescent Protein Structure. *J. Am. Chem. Soc.* **2010**, *132* (32), 11093–11102.

(45) Grigorenko, B. L.; Kots, E. D.; Krylov, A. I.; Nemukhin, A. V. Modeling of the Glycine Tripeptide Cyclization in the Ser65Gly/Tyr66Gly Mutant of Green Fluorescent Protein. *Mendeleev Commun.* **2019**, *29*, 187–189.

(46) Bartkiewicz, M.; Kazazić, S.; Krasowska, J.; Clark, P. L.; Wielgus-Kutrowska, B.; Bzowska, A. Non-Fluorescent Mutant of Green Fluorescent Protein Sheds Light on the Mechanism of Chromophore Formation. *FEBS Lett.* **2018**, *592* (9), 1516–1523.

(47) van Thor, J. J. Photoreactions and Dynamics of the Green Fluorescent Protein. *Chem. Soc. Rev.* **2009**, *38* (10), 2935.

(48) Meech, S. R. Excited State Reactions in Fluorescent Proteins. *Chem. Soc. Rev.* **2009**, *38* (10), 2922.

(49) Chattoraj, M.; King, B. A.; Bublitz, G. U.; Boxer, S. G. Ultra-Fast Excited State Dynamics in Green Fluorescent Protein: Multiple States and Proton Transfer. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93* (16), 8362–8367.

(50) Grigorenko, B. L.; Nemukhin, A. V.; Polyakov, I. V.; Morozov, D. I.; Krylov, A. I. First-Principles Characterization of the Energy Landscape and Optical Spectra of Green Fluorescent Protein along the A→I→B Proton Transfer Route. *J. Am. Chem. Soc.* **2013**, *135* (31), 11541–11549.

(51) Granovsky, A. A. Extended Multi-Configuration Quasi-Degenerate Perturbation Theory: The New Approach to Multi-State Multi-Reference Perturbation Theory. *J. Chem. Phys.* **2011**, *134* (21), 214113.

(52) Grimme, S. Improved Second-Order Møller–Plesset Perturbation Theory by Separate Scaling of Parallel- and Antiparallel-Spin Pair Correlation Energies. *J. Chem. Phys.* **2003**, *118* (20), 9095–9102.

(53) Rhee, Y. M.; Head-Gordon, M. Scaled Second-Order Perturbation Corrections to Configuration Interaction Singles: Efficient and Reliable Excitation Energy Methods. *J. Phys. Chem. A* **2007**, *111* (24), 5314–5326.

(54) Brejc, K.; Sixma, T. K.; Kitts, P. A.; Kain, S. R.; Tsien, R. Y.; Ormo, M.; Remington, S. J. Structural Basis for Dual Excitation and Photoisomerization of the Aequorea Victoria Green Fluorescent Protein. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94* (6), 2306–2311.

(55) Nadal-Ferret, M.; Gelabert, R.; Moreno, M.; Lluch, J. M. Transient Low-Barrier Hydrogen Bond in the Photoactive State of Green Fluorescent Protein. *Phys. Chem. Chem. Phys.* **2015**, *17* (46), 30876–30888.

(56) Shinobu, A.; Agmon, N. Proton Wire Dynamics in the Green Fluorescent Protein. *J. Chem. Theory Comput.* **2017**, *13* (1), 353–369.

(57) Martin, M. E.; Negri, F.; Olivucci, M. Origin, Nature, and Fate of the Fluorescent State of the Green Fluorescent Protein Chromophore at the CASPT2//CASSCF Resolution. *J. Am. Chem. Soc.* **2004**, *126* (17), 5452–5464.

(58) Sinicropi, A.; Andruniow, T.; Ferré, N.; Basosi, R.; Olivucci, M. Properties of the Emitting State of the Green Fluorescent Protein Resolved at the CASPT2//CASSCF/CHARMM Level. *J. Am. Chem. Soc.* **2005**, *127* (33), 11534–11535.

(59) Laino, T.; Nifosi, R.; Tozzini, V. Relationship between Structure and Optical Properties in Green Fluorescent Proteins: A

Quantum Mechanical Study of the Chromophore Environment. *Chem. Phys.* **2004**, *298* (1–3), 17–28.

(60) Hasegawa, J.-Y.; Fujimoto, K.; Swerts, B.; Miyahara, T.; Nakatsuji, H. Excited States of GFP Chromophore and Active Site Studied by the SAC-CI Method: Effect of Protein-Environment and Mutations. *J. Comput. Chem.* **2007**, *28* (15), 2443–2452.

(61) Topol, I.; Collins, J.; Polyakov, I.; Grigorenko, B.; Nemukhin, A. On Photoabsorption of the Neutral Form of the Green Fluorescent Protein Chromophore. *Biophys. Chem.* **2009**, *145* (1), 1–6.

(62) Bravaya, K. B.; Khrenova, M. G.; Grigorenko, B. L.; Nemukhin, A. V.; Krylov, A. I. Effect of Protein Environment on Electronically Excited and Ionized States of the Green Fluorescent Protein Chromophore. *J. Phys. Chem. B* **2011**, *115* (25), 8296–8303.

(63) Filippi, C.; Buda, F.; Guidoni, L.; Sinicropi, A. Bathochromic Shift in Green Fluorescent Protein: A Puzzle for QM/MM Approaches. *J. Chem. Theory Comput.* **2012**, *8* (1), 112–124.

(64) Daday, C.; Curutchet, C.; Sinicropi, A.; Mennucci, B.; Filippi, C. Chromophore–Protein Coupling beyond Nonpolarizable Models: Understanding Absorption in Green Fluorescent Protein. *J. Chem. Theory Comput.* **2015**, *11* (10), 4825–4839.

(65) 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.

(66) Melaccio, F.; Ferré, N.; Olivucci, M. Quantum Chemical Modeling of Rhodopsin Mutants Displaying Switchable Colors. *Phys. Chem. Chem. Phys.* **2012**, *14* (36), 12485.

(67) Burnette, D. T.; Sengupta, P.; Dai, Y.; Lippincott-Schwartz, J.; Kachar, B. Bleaching/Blinking Assisted Localization Microscopy for Superresolution Imaging Using Standard Fluorescent Molecules. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (52), 21081–21086.

(68) van Drogen, F.; Peter, M. Revealing Protein Dynamics by Photobleaching Techniques. *Methods Mol. Biol.* **2004**, *284*, 287–306.

(69) Grigorenko, B. L.; Nemukhin, A. V.; Morozov, D. I.; Polyakov, I. V.; Bravaya, K. B.; Krylov, A. I. Toward Molecular-Level Characterization of Photoinduced Decarboxylation of the Green Fluorescent Protein: Accessibility of the Charge-Transfer States. *J. Chem. Theory Comput.* **2012**, *8* (6), 1912–1920.

(70) Grigorenko, B. L.; Nemukhin, A. V.; Polyakov, I. V.; Khrenova, M. G.; Krylov, A. I. A Light-Induced Reaction with Oxygen Leads to Chromophore Decomposition and Irreversible Photobleaching in GFP-Type Proteins. *J. Phys. Chem. B* **2015**, *119* (17), 5444–5452.

(71) Ding, L.; Chung, L. W.; Morokuma, K. Reaction Mechanism of Photoinduced Decarboxylation of the Photoactivatable Green Fluorescent Protein: An ONIOM(QM:MM) Study. *J. Phys. Chem. B* **2013**, *117* (4), 1075–1084.

(72) van Thor, J. J.; Gensch, T.; Hellingwerf, K. J.; Johnson, L. N. Phototransformation of Green Fluorescent Protein with UV and Visible Light Leads to Decarboxylation of Glutamate 222. *Nat. Struct. Biol.* **2002**, *9* (1), 37–41.

(73) Bell, A. F.; Stoner-Ma, D.; Wachter, R. M.; Tonge, P. J. Light-Driven Decarboxylation of Wild-Type Green Fluorescent Protein. *J. Am. Chem. Soc.* **2003**, *125* (23), 6919–6926.

(74) 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.

(75) Day, R. N.; Davidson, M. W. The Fluorescent Protein Palette: Tools for Cellular Imaging. *Chem. Soc. Rev.* **2009**, *38* (10), 2887–2921.

(76) Heim, R.; Cubitt, A. B.; Tsien, R. Y. Improved Green Fluorescence. *Nature* **1995**, *373* (6516), 663–664.

(77) 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.

(78) 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.

(79) 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.

(80) 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.

(81) Griesbeck, O.; Baird, G. S.; Campbell, R. E.; Zacharias, D. A.; Tsien, R. Y. Reducing the Environmental Sensitivity of Yellow Fluorescent Protein. *J. Biol. Chem.* **2001**, *276* (31), 29188–29194.

(82) Grigorenko, B. L.; Nemukhin, A. V.; Polyakov, I. V.; Krylov, A. I. Triple-Decker Motif for Red-Shifted Fluorescent Protein Mutants. *J. Phys. Chem. Lett.* **2013**, *4* (10), 1743–1747.

(83) Khrenova, M. G.; Polyakov, I. V.; Grigorenko, B. L.; Krylov, A. I.; Nemukhin, A. V. Improving the Design of the Triple-Decker Motif in Red Fluorescent Proteins. *J. Phys. Chem. B* **2017**, *121* (47), 10602–10609.

(84) Khrenova, M. G.; Nemukhin, A. V.; Domratcheva, T. Theoretical Characterization of the Flavon-Based Fluorescent Protein iLOV and Its Q489K Mutant. *J. Phys. Chem. B* **2015**, *119* (16), 5176–5183.

(85) 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 EFP650 and EFP670. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2012**, *68* (9), 1088–1097.

(86) Melaccio, F.; Ferré, N.; Olivucci, M. Quantum Chemical Modeling of Rhodopsin Mutants Displaying Switchable Colors. *Phys. Chem. Chem. Phys.* **2012**, *14*, 12485–12495.

(87) Orozco-Gonzalez, Y.; Kabir, M. P.; Gozem, S. Electrostatic Spectral Tuning Maps for Biological Chromophores. *J. Phys. Chem. B* **2019**, DOI: 10.1021/acs.jpcc.9b00489.

(88) Davari, M. D.; Kopka, B.; Wingen, M.; Bocola, M.; Drepper, T.; Jaeger, K.-E.; Schwaneberg, U.; Krauss, U. Photophysics of the LOV-Based Fluorescent Protein Variant iLOV-Q489K Determined by Simulation and Experiment. *J. Phys. Chem. B* **2016**, *120* (13), 3344–3352.

(89) Khrenova, M. G.; Meteleshko, Y. I.; Nemukhin, A. V. Mutants of the Flavoprotein iLOV as Prospective Red-Shifted Fluorescent Markers. *J. Phys. Chem. B* **2017**, *121* (43), 10018–10025.

(90) Meteleshko, Y. I.; Nemukhin, A.; Khrenova, M. Novel Flavon-Based Fluorescent Proteins with Red-Shifted Emission Bands: A Computational Study. *Photochem. Photobiol. Sci.* **2019**, *18*, 177–189.

(91) Mathes, T.; Götze, J. P. A Proposal for a Dipole-Generated BLUF Domain Mechanism. *Front. Mol. Biosci.* **2015**, *2*, 62.

(92) Park, S.-Y.; Tame, J. R. H. Seeing the Light with BLUF Proteins. *Biophys. Rev.* **2017**, *9* (2), 169–176.

(93) Fujisawa, T.; Masuda, S. Light-Induced Chromophore and Protein Responses and Mechanical Signal Transduction of BLUF Proteins. *Biophys. Rev.* **2018**, *10* (2), 327–337.

(94) Khrenova, M. G.; Nemukhin, A. V.; Grigorenko, B. L.; Krylov, A. I.; Domratcheva, T. M. Quantum Chemistry Calculations Provide Support to the Mechanism of the Light-Induced Structural Changes in the Flavon-Binding Photoreceptor Proteins. *J. Chem. Theory Comput.* **2010**, *6* (8), 2293–2302.

(95) Domratcheva, T.; Hartmann, E.; Schlichting, I.; Kottke, T. Evidence for Tautomerisation of Glutamine in BLUF Blue Light Receptors by Vibrational Spectroscopy and Computational Chemistry. *Sci. Rep.* **2016**, *6* (1), 22669.

(96) Khrenova, M. G.; Domratcheva, T.; Nemukhin, A. V. Molecular Mechanism of the Dark-State Recovery in BLUF Photoreceptors. *Chem. Phys. Lett.* **2017**, *676*, 25–31.

(97) Goyal, P.; Hammes-Schiffer, S. Role of Active Site Conformational Changes in Photocycle Activation of the AppA BLUF Photoreceptor. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (7), 1480–1485.

(98) Goings, J. J.; Reinhardt, C. R.; Hammes-Schiffer, S. Propensity for Proton Relay and Electrostatic Impact of Protein Reorganization in Slr1694 BLUF Photoreceptor. *J. Am. Chem. Soc.* **2018**, *140* (45), 15241–15251.

(99) Stelling, A. L.; Ronayne, K. L.; Nappa, J.; Tonge, P. J.; Meech, S. R. Ultrafast Structural Dynamics in BLUF Domains: Transient Infrared Spectroscopy of AppA and Its Mutants. *J. Am. Chem. Soc.* **2007**, *129* (50), 15556–15564.

(100) Domratcheva, T.; Grigorenko, B. L.; Schlichting, I.; Nemukhin, A. V. Molecular Models Predict Light-Induced Glutamine Tautomerization in BLUF Photoreceptors. *Biophys. J.* **2008**, *94* (10), 3872–3879.

(101) Sadeghian, K.; Bocola, M.; Schütz, M. A Conclusive Mechanism of the Photoinduced Reaction Cascade in Blue Light Using Flavon Photoreceptors. *J. Am. Chem. Soc.* **2008**, *130* (37), 12501–12513.

(102) Grigorenko, B. L.; Khrenova, M. G.; Nemukhin, A. V. Amide–imide Tautomerization in the Glutamine Side Chain in Enzymatic and Photochemical Reactions in Proteins. *Phys. Chem. Chem. Phys.* **2018**, *20* (37), 23827–23836.

(103) Durbeej, B.; Borg, O. A.; Eriksson, L. A. Computational Evidence in Favor of a Protonated Chromophore in the Photoactivation of Phytochrome. *Chem. Phys. Lett.* **2005**, *416* (1–3), 83–88.

(104) Durbeej, B. On the Primary Event of Phytochrome: Quantum Chemical Comparison of Photoreactions at C4, C10 and C15. *Phys. Chem. Chem. Phys.* **2009**, *11* (9), 1354.

(105) Matute, R. A.; Contreras, R.; González, L. Time-Dependent DFT on Phytochrome Chromophores: A Way to the Right Conformer. *J. Phys. Chem. Lett.* **2010**, *1* (4), 796–801.

(106) Falklöf, O.; Durbeej, B. Modeling of Phytochrome Absorption Spectra. *J. Comput. Chem.* **2013**, *34* (16), 1363–1374.

(107) Falklöf, O.; Durbeej, B. Steric Effects Govern the Photoactivation of Phytochromes. *ChemPhysChem* **2016**, *17* (7), 954–957.

(108) Falklöf, O.; Durbeej, B. Computational Identification of Pyrrole Ring C as the Preferred Donor for Excited-State Proton Transfer in Bacteriophytochromes. *ChemPhotoChem* **2018**, *2* (6), 453–457.

(109) Hasegawa, J.; Isshiki, M.; Fujimoto, K.; Nakatsuji, H. Structure of Phytochromobilin in the Pr and Pfr Forms: SAC-CI Theoretical Study. *Chem. Phys. Lett.* **2005**, *410* (1–3), 90–93.

(110) Feliks, M.; Lafaye, C.; Shu, X.; Royant, A.; Field, M. Structural Determinants of Improved Fluorescence in a Family of Bacteriophytochrome-Based Infrared Fluorescent Proteins: Insights from Continuum Electrostatic Calculations and Molecular Dynamics Simulations. *Biochemistry* **2016**, *55* (31), 4263–4274.

(111) Wagner, J. R.; Zhang, J.; Brunzelle, J. S.; Vierstra, R. D.; Forest, K. T. High Resolution Structure of Deinococcus Bacteriophytochrome Yields New Insights into Phytochrome Architecture and Evolution. *J. Biol. Chem.* **2007**, *282* (16), 12298–12309.

(112) Mewes, S. A.; Plasser, F.; Krylov, A.; Dreuw, A. Benchmarking Excited-State Calculations Using Exciton Properties. *J. Chem. Theory Comput.* **2018**, *14* (2), 710–725.

(113) Polyakov, I. V.; Grigorenko, B. L.; Mironov, V. A.; Nemukhin, A. V. Modeling Structure and Excitation of Biliverdin-Binding Domains in Infrared Fluorescent Proteins. *Chem. Phys. Lett.* **2018**, *710*, 59–63.

(114) Rumyantsev, K. A.; Turoverov, K. K.; Verkhusha, V. V. Near-Infrared Bioluminescent Proteins for Two-Color Multimodal Imaging. *Sci. Rep.* **2016**, *6*, 36588.

(115) Shcherbakova, D. M.; Balaban, M.; Emelyanov, A. V.; Brenowitz, M.; Guo, P.; Verkhusha, V. V. Bright Monomeric Near-Infrared Fluorescent Proteins as Tags and Biosensors for Multiscale Imaging. *Nat. Commun.* **2016**, *7* (1), 12405.

(116) Khrenova, M. G.; Kulakova, A. M.; Nemukhin, A. V. Competition between Two Cysteines in Covalent Binding of Biliverdin to Phytochrome Domains. *Org. Biomol. Chem.* **2018**, *16* (40), 7518–7529.

- (117) Lamb, A. L.; Kappock, T. J.; Silvaggi, N. R. You Are Lost without a Map: Navigating the Sea of Protein Structures. *Biochim. Biophys. Acta, Proteins Proteomics* **2015**, *1854* (4), 258–268.
- (118) Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem. Theory Comput.* **2011**, *7* (2), 525–537.
- (119) Laurent, A. D.; Mironov, V. A.; Chapagain, P. P.; Nemukhin, A. V.; Krylov, A. I. Exploring Structural and Optical Properties of Fluorescent Proteins by Squeezing: Modeling High-Pressure Effects on the MStrawberry and MCherry Red Fluorescent Proteins. *J. Phys. Chem. B* **2012**, *116* (41), 12426–12440.
- (120) Faraji, S.; Krylov, A. I. On the Nature of an Extended Stokes Shift in the mPlum Fluorescent Protein. *J. Phys. Chem. B* **2015**, *119* (41), 13052–13062.
- (121) Bogdanov, A. M.; Acharya, A.; Titelmayer, A. V.; Mamontova, A. V.; Bravaya, K. B.; Kolomeisky, A. B.; Lukyanov, K. A.; Krylov, A. I. Turning On and Off Photoinduced Electron Transfer in Fluorescent Proteins by π -Stacking, Halide Binding, and Tyr145 Mutations. *J. Am. Chem. Soc.* **2016**, *138* (14), 4807–4817.
- (122) Vegh, R. B.; Bravaya, K. B.; Bloch, D. A.; Bommarius, A. S.; Tolbert, L. M.; Verkhovsky, M.; Krylov, A. I.; Solntsev, K. M. Chromophore Photoreduction in Red Fluorescent Proteins Is Responsible for Bleaching and Phototoxicity. *J. Phys. Chem. B* **2014**, *118* (17), 4527–4534.
- (123) Reuter, N.; Lin, H.; Thiel, W. Green Fluorescent Proteins: Empirical Force Field for the Neutral and Deprotonated Forms of the Chromophore. Molecular Dynamics Simulations of the Wild Type and S65T Mutant. *J. Phys. Chem. B* **2002**, *106* (24), 6310–6321.
- (124) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005**, *26* (16), 1781–1802.
- (125) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4* (3), 435–447.
- (126) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14* (1), 33–38.
- (127) Warshel, A.; Levitt, M. Theoretical Studies of Enzymic Reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme. *J. Mol. Biol.* **1976**, *103* (2), 227–249.
- (128) Senn, H. M.; Thiel, W. QM/MM Methods for Biomolecular Systems. *Angew. Chem., Int. Ed.* **2009**, *48* (7), 1198–1229.
- (129) Melaccio, F.; del Carmen Marín, M.; Valentini, A.; Montisci, F.; Rinaldi, S.; Cherubini, M.; Yang, X.; Kato, Y.; Stenrup, M.; Orozco-Gonzalez, Y.; et al. Toward Automatic Rhodopsin Modeling as a Tool for High-Throughput Computational Photobiology. *J. Chem. Theory Comput.* **2016**, *12* (12), 6020–6034.
- (130) Karelina, M.; Kulik, H. J. Systematic Quantum Mechanical Region Determination in QM/MM Simulation. *J. Chem. Theory Comput.* **2017**, *13* (2), 563–576.
- (131) Morzan, U. N.; Alonso de Armiño, D. J.; Foglia, N. O.; Ramírez, F.; González Lebrero, M. C.; Scherlis, D. A.; Estrin, D. A. Spectroscopy in Complex Environments from QM–MM Simulations. *Chem. Rev.* **2018**, *118* (7), 4071–4113.
- (132) Mardirossian, N.; Head-Gordon, M. Thirty Years of Density Functional Theory in Computational Chemistry: An Overview and Extensive Assessment of 200 Density Functionals. *Mol. Phys.* **2017**, *115* (19), 2315–2372.
- (133) Adamo, C.; Barone, V. Toward Reliable Density Functional Methods without Adjustable Parameters: The PBE0Model. *J. Chem. Phys.* **1999**, *110* (13), 6158.
- (134) 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.
- (135) Medvedev, M. G.; Bushmarinov, I. S.; Sun, J.; Perdew, J. P.; Lyssenko, K. A. Density Functional Theory Is Straying from the Path toward the Exact Functional. *Science* **2017**, *355* (6320), 49–52.
- (136) Rupp, M.; von Lilienfeld, O. A.; Burke, K. Guest Editorial: Special Topic on Data-Enabled Theoretical Chemistry. *J. Chem. Phys.* **2018**, *148* (24), 241401.
- (137) Grigorenko, B. L.; Nemukhin, A. V.; Topol, I. A.; Burt, S. K. Modeling of Biomolecular Systems with the Quantum Mechanical and Molecular Mechanical Method Based on the Effective Fragment Potential Technique: Proposal of Flexible Fragments. *J. Phys. Chem. A* **2002**, *106* (44), 10663–10672.
- (138) Nemukhin, A. V.; Grigorenko, B. L.; Topol, I. A.; Burt, S. K. Flexible Effective Fragment QM/MM Method: Validation through the Challenging Tests. *J. Comput. Chem.* **2003**, *24* (12), 1410–1420.
- (139) Gurunathan, P. K.; Acharya, A.; Ghosh, D.; Kosenkov, D.; Kaliman, I.; Shao, Y.; Krylov, A. I.; Slipchenko, L. V. Extension of the Effective Fragment Potential Method to Macromolecules. *J. Phys. Chem. B* **2016**, *120* (27), 6562–6574.
- (140) Jacquemin, D.; Wathelet, V.; Perpète, E. A.; Adamo, C. Extensive TD-DFT Benchmark: Singlet-Excited States of Organic Molecules. *J. Chem. Theory Comput.* **2009**, *5* (9), 2420–2435.
- (141) Charaf-Eddin, A.; Planchat, A.; Mennucci, B.; Adamo, C.; Jacquemin, D. Choosing a Functional for Computing Absorption and Fluorescence Band Shapes with TD-DFT. *J. Chem. Theory Comput.* **2013**, *9* (6), 2749–2760.
- (142) Krylov, A.; Windus, T. L.; Barnes, T.; Marin-Rimoldi, E.; Nash, J. A.; Pritchard, B.; Smith, D. G. A.; Altarawy, D.; Saxe, P.; Clementi, C.; et al. Perspective: Computational Chemistry Software and Its Advancement as Illustrated through Three Grand Challenge Cases for Molecular Science. *J. Chem. Phys.* **2018**, *149* (18), 180901.
- (143) Schnedermann, C.; Yang, X.; Liebel, M.; Spillane, K. M.; Lugtenburg, J.; Fernández, I.; Valentini, A.; Schapiro, I.; Olivucci, M.; Kukura, P.; Mathies, R. A. Evidence for a Vibrational Phase-Dependent Isotope Effect on the Photochemistry of Vision. *Nat. Chem.* **2018**, *10*, 449–455.
- (144) Seifert, S.; Brakmann, S. LOV Domains in the Design of Photoresponsive Enzymes. *ACS Chem. Biol.* **2018**, *13* (8), 1914–1920.
- (145) Losi, A.; Gardner, K. H.; Möglich, A. Blue-Light Receptors for Optogenetics. *Chem. Rev.* **2018**, *118* (21), 10659–10709.
- (146) Marin, M. C.; Agathangelou, D.; Orozco-Gonzalez, Y.; Valentini, A.; Kato, Y.; Abe-Yoshizumi, R.; Kandori, H.; Choi, A.; Jung, K.-H.; Haacke, S.; Olivucci, M. Fluorescence Enhancement of a Microbial Rhodopsin via Electronic Reprogramming. *J. Am. Chem. Soc.* **2019**, *141* (1), 262–271.