Molecular Modeling Clarifies the Mechanism of Chromophore Maturation in the Green Fluorescent Protein

Bella L. Grigorenko,†,‡ Anna I. Krylov,*,§ and Alexander V. Nemukhin*,†,‡,§

†Department of Chemistry, Lomonosov Moscow State University, Moscow 119991, Russia
‡Emanuel Institute of Biochemical Physics, Russian Academy of Sciences, Moscow 119991, Russia
§Department of Chemistry, University of Southern California, Los Angeles, California 90089-0482, United States

Supporting Information

ABSTRACT: We report the first complete theoretical description of the chain of elementary reactions occurring in chromophore maturation in the green fluorescent protein (GFP). All reaction steps including cyclization, dehydration, and oxidation are characterized at the quantum mechanics/molecular mechanics (QM/MM) computational level using density functional theory in quantum subsystems. Starting from a structure of the wild-type protein with the noncyclized Ser65-Tyr66-Gly67 tripeptide, we modeled cyclization and dehydration reactions. We then added molecular oxygen to the system and modeled the oxidation reaction resulting in the mature protein-bound chromophore. Computationally derived structures of the reaction product and several reaction intermediates agree well with the relevant crystal structures, validating the computational protocol. The highest computed energy barriers at the cyclization–dehydration (17 kcal/mol) and oxidation (21 kcal/mol) steps agree well with the values derived from the kinetics measurements (20.7 and 22.7 kcal/mol, respectively). The simulations provide strong support to the mechanism involving the cyclization–dehydration–oxidation sequence of the chromophore’s maturation reactions. The results also establish a solid basis for predictions of maturation mechanisms in other fluorescent proteins.

1. INTRODUCTION

Fluorescent proteins (FPs) have transformed imaging capabilities in life sciences by enabling in vivo observation of biological processes with an unprecedented level of detail.4–6 FPs are exploited in super-resolution microscopy, which affords a spatial resolution of ~10 nm, about 20 times smaller than the diffraction limit.7,8 Unique properties of FPs have also inspired their applications in biotechnology, including nanobiophotonic devices,9,10 optical data storage,11,12 and even novel laser designs.13

The utility of GFP in biology and medicine is grounded in the autocatalytic chromophore formation, which occurs upon protein folding via post-translational modification of the three amino-acid residues. The chromophore formation does not require any cofactors or natural enzymes, meaning that it is the sequence of amino acids that needs to be encoded in the genetic material of a cell for a fluorescent tag to be produced by the cell’s machinery. Post-translational modifications are rather ubiquitous in biology and are contributing to natural protein diversity.14 For example, modifications of their side chains expand the scope of natural amino acids by more than an order of magnitude, from 20 to 23 naturally occurring amino acids to more than 350 modified species.15–17

The chromophore is formed by Ser65–Gly67 XZG tripeptide sequence, where X is variable, Z is a tyrosine in all naturally occurring FPs and is aromatic (e.g., histidine in blue FP) in all fluorescent laboratory-derived mutants, and G is always glycine. In wt-GFP, X is serine and Z is tyrosine; in enhanced GFP X is threonine. The tripeptide sequence is located in the α-helix, buried in the center of the GFP’s β-barrel. The post-translational chemistry of the 65–67 tripeptide might be driven by the strained structure of the α-helix, which features ~80° kink in the middle. Mechanistic understanding of chromophore maturation is obviously important for successful protein engineering. It is also of a fundamental significance, because similar chemical transformations are operational in other biological systems, e.g., cyclization in histidine ammonia lyase enzyme.18 Yet, despite considerable efforts, the molecular-level mechanism of the chromophore’s maturation is not fully elucidated. A tentative chain of chemical transformations leading to the mature chromophore in GFP, which was outlined in the early 1990s by the pioneers of the GFP research19,20 and then reproduced in multiple papers and reviews,21–23 is shown in Figure 1.

The mechanism involves three main steps (or reaction stages): cyclization, dehydration, and oxidation. Peptide backbone cyclization is initiated by formation of a covalent bond between the Gly67 amide nitrogen and the Ser65 carbonyl carbon (structure III in Figure 1). Dehydration leads to structure IV, followed by oxidation of the Tyr66 C=O bond.
producing fully conjugated chromophore (structure V). Oxidation has been identified as a rate-determining step.\textsuperscript{24}

Several elegant experimental studies\textsuperscript{14,25–29} have provided strong support for this mechanism and fleshed out additional details, including the role of important nearby residues (Arg96 and Glu222). A common approach for elucidating a mechanism is trapping the reaction intermediates by mutations that slow down a particular step.\textsuperscript{25}–\textsuperscript{29} Using this technique, Getzoff and co-workers showed that Arg96Ala mutation slows the maturation rate from minutes to months, which allowed them to obtain the crystal structure of the colorless immature precyclized intermediate (structure III in Figure 1).\textsuperscript{25} Importantly, despite the slow rate, the chromophore eventually matures (as confirmed by the fluorescence properties and the crystal structure), suggesting that the reaction pathway in the mutant is representative of that in the parent FP. Later, by structural characterization of FP folded at anaerobic conditions and reduced forms of the mature chromophores, Getzoff’s group established\textsuperscript{27} that the dehydrated intermediate (structure IV) exits in the enolate form stabilized by Arg96. In a more recent study, Pletneva et al. designed a mutant that forms a colorless intermediate (structure III) in wt-GFP to the fully formed mature chromophore. This study provides additional support in favor of oxidation being the last step of the maturation process.

This mechanism has been challenged by Wachter and co-workers.\textsuperscript{30,31} Using kinetic studies and mass spectrometry, they confirmed production of hydrogen peroxide in a stoichiometric 1:1 ratio with the mature chromophore and found that hydrogen peroxide production precedes the onset of the GFP fluorescence, which led to the conclusion that the last step of the mechanism is dehydration. A different magnitude of the kinetic isotope effect for the oxidation and dehydration steps (small for the former and large for the latter)\textsuperscript{31} has provided further support for dehydration being the last step in chromophore’s formation. These findings underscore the complexity of post-translational modifications in GFP and suggest that there might be more than one pathway of chromophore formation and that the competition between different pathways might depend on the concrete structure of the protein. For example, mutations introduced to modify the energy profile of the chromophore formation reaction might switch the reaction mechanism to an entirely different pathway.

Given the controversial experimental evidence and the complexity of the system, the theoretical modeling of the chromophore maturation reaction can provide decisive evidence in favor of a particular mechanistic hypothesis. Here, we present a detailed computational study of the full reaction pathway, from the original tripeptide sequence (Ser65-Tyr66-Gly67) in wt-GFP to the fully formed mature chromophore. We employ high-level electronic structure methods and the QM/MM (quantum mechanics/molecular mechanics) approach to model the protein environment. This is the first calculation of the entire reaction pathway, including the rate-determining oxidation step. Previous theoretical studies have only attempted to characterize the cyclization\textsuperscript{32–37} and some aspects of the dehydration\textsuperscript{38} steps; however, the corresponding quantum chemical calculations yielded reaction barriers that were too high and inconsistent with the experimental kinetics. As shown below, our simulations lead to results consistent with the available experimental data. Our results support the cyclization–dehydration–oxidation sequence and provide insight into the nature of the key intermediates in the maturation process. Following tradition, below we refer to the two GFP maturation schemes as Getzoff’s mechanism (cyclization–dehydration–oxidation) and Wachter’s mechanism (cyclization–oxidation–dehydration).

2. MODELS AND METHODS

A suitable structure from the Protein Data Bank\textsuperscript{39} usually serves as a source of coordinates of heavy atoms of a model system, which provides a starting point for simulating reactions inside a protein. However, for the wt-GFP (in which the chromophore is formed from the initial Ser65-Tyr66-Gly67 sequence), no PDB structures with this tripeptide in a precyclized form are available. Only for GFP mutants there are several PDB entries containing natural amino acid residues at positions 65–66–67. For instance, PDB ID 2AWJ\textsuperscript{36} contains the
noncyclized Thr65-Tyr66-Gly67 moiety, with Thr65 instead of Ser65 and the critical Arg96 residue replaced by methionine. As reported by Getzoff and co-workers,40 this structure shows dramatic distortions around Tyr66 compared to wt-GFP. Consequently, building a model system from such mutants would entail tremendous efforts toward restoring computationally a polypeptide structure of the GFP precursor. We choose another strategy to prepare a precylized structure of the protein corresponding to initial point I in Figure 1. A model system mimicking wt-GFP suitable for QM/MM calculations of reaction profiles was carefully prepared and validated in our previous study.41 It was developed on the basis of the crystal structure PDB ID 1EMA42 using conventional treatment with the molecular modeling tools, and it was shown to yield accurate spectroscopic properties of different structures involved in the GFP photocycle.43 We also point out a perfect agreement between this structure and the high-resolution (0.9 Å) crystal structure PDB ID 2WUR of wt-GFP reported recently.44 Thus, this model system is a suitable starting point for simulations of chemical transformations inside the chromophore-containing pocket. According to an important note from ref 25, no distortions of the protein structure outside the region occupied by the residues at positions 65–66–67 were seen in any X-ray crystallography studies of the chromophore maturation process.

Stationary points corresponding to the minimum-energy points on the ground electronic state potential energy surface (PES) of wt-GFP in our previous study41 were located using an extension of the QM/MM scheme based on the effective fragment potential approach46,47 that allows the effective fragments to be flexible.46,47 In this scheme, the groups assigned to the MM part are represented by effective fragments contributing their one-electron potentials to the quantum Hamiltonian,46 the peptide chains of the protein are described as flexible chains of small effective fragments,46,47 and fragment–fragment interactions are computed with conventional force fields. In the present work we started from the lowest-energy structure of the model system with the neutral chromophore corresponding to form A of wt-GFP and assigned a nearly identical set of molecular groups as before41 to the QM subsystem. To prepare a polypeptide chain corresponding to the initial structure (structure I in Figure 1) we manually retro-engineered the Ser65-Tyr66-Gly67 sequence from the chromophore and used the same QM/MM computational protocol as previously41 to locate the equilibrium geometry of the structure designated REAG below. In the latter, we assigned tripeptide Ser65-Tyr66-Gly67, the side chains of Arg96, His148, Ser205, and Glu222, and four water molecules to QM (see Figure 2). Thus, the QM subsystem comprises not only the residues that undergo chemical transformations (Ser65, Tyr66, Gly67) but also almost all nearby hydrophilic groups. His148 and Ser205 were included in the QM because they form hydrogen bonds with the active-site residues.

Further details of the QM selection are given in the Supporting Information (Figure S1). We note that the QM/EFP method (in which the environment is modeled by polarizable effective fragments) describes electrostatic and polarization interactions between the QM and the MM parts with accuracy approaching a full quantum description of the entire system.46,47 To model the oxidation stage in wt-GFP chromophore maturation one of the water molecules in QM was replaced by O2.

To scan the sections along the ground-state PES of the system at the cyclization and dehydration stages (without molecular oxygen), we used the QM(PBE0/6-31G*)/MM(AMBER) protocol.41 At the oxidation stage either restricted (for singlet surface) or unrestricted (for triplet surface) DFT(PBE0/6-31G*) calculations were carried out. The PBE0 functional46,47 was chosen based on its overall robust performance and because this functional is based on solid theoretical grounds and does not include adjustable parameters.47 On the basis of our previous studies of enzyme-catalyzed reactions (see, for example, ref 51), we believe that the main mechanistic conclusions are not sensitive to the choice of the functional. To further support this point, we carried out additional calculations of selected stationary points with another functional, B3LYP, but did not find considerable differences. As previously,41 a modified code of the GAMESS(US)52,53 program and TINKER54 were used in the QM/MM calculations.

3. RESULTS AND DISCUSSION

3.1. Cyclization–Dehydration Reaction in wt-GFP.

Figure 2 shows molecular groups assigned to the QM subsystem in QM/MM simulations of the cyclization and dehydration steps. Figure 3 shows our molecular model at the equilibrium geometry of REAG including the groups directly involved in chemical transformations upon chromophore formation. Designation of atoms used throughout the paper is also given in Figure 3.

We begin by comparing several structural parameters in REAG with those referring to precyclized tripeptide Thr65-Tyr66-Gly67 from the crystal structure PDB ID 2AWJ of the GFP mutant.40 In this variant, Arg96Met mutation was introduced (along with Ser65Thr), which might lead to significant distortions of the active site. However, the comparison of the tripeptide conformations is still instructive.

Figure 3 and Table 1 show the relevant structural parameters of the active site in REAG and the crystal (PDB ID 2AWJ) structure, focusing on the atoms that ultimately form the chromophore. As one can see, the distances between the heavy atoms and the respective valence angles in REAG agree very
assumed a direct attack of N67 on C65 and a simultaneous (or multistep) transfer of proton (H65) from N67 to O65. In this work, we found another reaction route for cyclization with a reasonable energy barrier. As illustrated in Figure 3, in the tight conformation of REAG the distance from H65 to N66 (2.29 Å) is much shorter than that to O65 (2.70 Å), so that one can expect a lower energy barrier for the transfer of H65 to N66. The REAG and INT1 structures are nearly isoenergetic, and the energy of TS1 is 13 kcal/mol (see Figure S3 in Supporting Information).

The conversion of INT1 into the structure with the five-membered imidazolidine-like ring, i.e., structure III in Figure 1 and INT3 in Figure 4, proceeds via a route with low energy barriers (less than 5 kcal/mol). The corresponding transformations are described in the Supporting Information (Figure S4). We note that the local energy minima, INT2a and INT2b (Figure 4), do not necessarily give rise to the true reaction intermediates: the reaction flow on the free energy surface may skip these intermediates separated by low barriers.

Analysis of the dehydration elementary step from INT3 to INT4 reveals that proton transfer via TS4 is assisted by a water molecule in the active site (W1 in Figure 2). A hydrogen-bond chain, including the Cα66-H, W1, and O66-H groups, provides a route for proton transfer coupled with Cα66-H and C65-O65 bond cleavage (see Figure S5 in Supporting Information). The minimum-energy pathway shows cleavage of the bond between Cα66 and its proton, Hα66 and elongation and breaking of the C65-O65 bond. The energy of the corresponding transition state TS4 is 17 kcal/mol above the intermediate INT3.

Removal of a proton from Cα66 is considered in both Getzoff’s and Wachter’s mechanisms. It has been suggested that Arg96 acidifies Cα66, facilitating further transformations of the intermediates. Wachter’s mechanism assumes a proton transfer from Cα66 to the side chain of Glu222 mediated by a water molecule when starting from the cyclized intermediate (INT3 in our terms). Presumably, this step is required to generate a negatively charged intermediate capable of donating an electron to the oxygen molecule and to initiate oxidation prior to dehydration. We attempted to locate such an intermediate, but we could not find a minimum-energy structure with a proton attached to the Glu222 side chain. We observed that either the proton shuttled back to Cα66 or the system advanced to the dehydrated intermediate (INT4 in our scheme). Getzoff’s mechanism assumes a water elimination from INT3 at the expense of proton removal from N66 and its recombination with the hydroxyl from C65. In our simulations of the minimum energy pathway, elongation of the N66-H bond is, in fact, coupled with cleavage of the Cα66-O65 bond. However, simultaneously the Cα66-H bond is cleaved, the same chain of proton transfer reactions as in our proposal for dehydration occurs, and the N66-H bond is finally restored, giving rise to the same reaction intermediate INT4. Therefore, attempts to move along the reaction coordinate defined as the N66-H distance led to the same product but with higher energy barriers. On the basis of these observations, we advocate here for the mechanism illustrated in Figure 4 (see also Figure S5 in Supporting Information). We note that the dehydrated zwitterionic intermediate in our scheme (INT4) is a good candidate to donate an electron to the oxygen molecule (Figure 5), which provides additional support to our proposal.

**3.2. Oxidation Reaction.** To model the oxidation step of the chromophore maturation reaction, the model system was

---

**Table 1. Structural Parameters of REAG**

<table>
<thead>
<tr>
<th>parameter</th>
<th>REAG</th>
<th>X-ray</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cα65−C65</td>
<td>1.52</td>
<td>1.50</td>
</tr>
<tr>
<td>C65−O65</td>
<td>1.23</td>
<td>1.23</td>
</tr>
<tr>
<td>C65−N66</td>
<td>1.36</td>
<td>1.34</td>
</tr>
<tr>
<td>N66−Cα66</td>
<td>1.45</td>
<td>1.48</td>
</tr>
<tr>
<td>Cα66−C66</td>
<td>1.53</td>
<td>1.51</td>
</tr>
<tr>
<td>Cα66−O66</td>
<td>1.24</td>
<td>1.23</td>
</tr>
<tr>
<td>Cα66−N67</td>
<td>1.33</td>
<td>1.33</td>
</tr>
<tr>
<td>O65−C65−N66</td>
<td>122</td>
<td>121</td>
</tr>
<tr>
<td>C65−N66−Cα66</td>
<td>115</td>
<td>110</td>
</tr>
<tr>
<td>N66−Cα66−C66</td>
<td>110</td>
<td>107</td>
</tr>
<tr>
<td>Cα66−C66−N67</td>
<td>117</td>
<td>115</td>
</tr>
<tr>
<td>N66−N67</td>
<td>2.67</td>
<td>2.53</td>
</tr>
<tr>
<td>C65−N67</td>
<td>2.93</td>
<td>3.03</td>
</tr>
<tr>
<td>O65−N67</td>
<td>3.06</td>
<td>3.20</td>
</tr>
</tbody>
</table>

See Figure 3 for atom-numbering scheme. The distances are in Angstroms; angles are in degrees. †PDB ID 2AWJ, ref 40.
modified as follows. We manually introduced dioxygen molecule into INT4WT by replacing a water molecule and reoptimized the structure with QM(UDFT)/MM (note that by introducing O2, the system assumes triplet multiplicity).

The resulting structure (denoted oxINT1) is shown in the left panel in Figure 5; this is the initial structure for the oxidation step, and the relative energies of the subsequent elementary steps are given with respect to this structure. The oxygen molecule resides near the chromophore and the Glu222 side chain with distances of \( \sim 2.7 \) Å between its oxygen atoms and carbon atoms of the chromophore, C66 and C65.

We found that the dehydrated chromophore precursor in \(^{\text{as}}\)INT1, denoted by Chro', shows no reactivity toward oxygen; however, reactive species could be obtained after proton transfer from N66 of Chro' to Glu222. The ensuing transformations proceed on the triplet PES from \(^{\text{as}}\)INT1 to \(^{\text{as}}\)INT2 via the transition state \(^{\text{as}}\)TS1 at 10 kcal/mol. The chromophore precursor in \(^{\text{as}}\)INT2 is denoted Chro'.

The structure of \(^{\text{as}}\)INT2 can be described as the charge-transfer complex. First, elongation of the O=O bond in dioxygen from the initial value of 1.22 Å to 1.26 Å (see right panels in Figure 5) supports an increasing anionic character, from O2 to O2\(^-\); second, comparison of the electronic structures of \(^{\text{as}}\)INT1 and \(^{\text{as}}\)INT2 supports electron transfer from Chro' to dioxygen. (In Figure S6 of the Supporting Information we present the results of electronic structure

Figure 4. Energy profile and chemical structures of the key intermediates (in insets) along the reaction segment from the precyclized tripeptide to the dehydrated intermediate.

Figure 5. First elementary step in oxidation from \(^{\text{as}}\)INT1 to reaction intermediate \(^{\text{as}}\)INT2 on the triplet-state PES.
calculations at the \(^{1\alpha}\)INT1 and \(^{1\alpha}\)INT2 geometries showing molecular orbitals and the composition of the multiconfigurational wave functions.)

Importantly, energies of the triplet and singlet states are nearly degenerate at the \(^{1\alpha}\)INT2 geometry. Therefore, we assume that starting from this point, the reaction pathway follows the singlet PES, and the same computational protocol as at the cyclization and dehydration steps can be applied. The computed full QM(PBE0/6-31G*)/MM(AMBER) energy diagram for the oxidation reaction, which can be subdivided into three segments, is shown in Figure 6. First, unconstrained QM/MM minimization from \(^{1\alpha}\)INT2 in the singlet state leads to the minimum energy structure, \(^{1\alpha}\)INT3, in which the covalent bond between the oxygen atom of O\(_2\)\(^{-}\) and C\(_{65}\) from Chro’ is formed. A small barrier (less than 5 kcal/mol) separates the intermediate \(^{1\alpha}\)INT3 from the species in which both oxygen atoms of dioxygen are bound to carbon atoms C\(_{66}\) and C\(_{65}\). To accomplish this elementary step, the proton from Glu222 should be transformed to N\(_{66}\). The energy of the resulting reaction intermediate, \(^{1\alpha}\)INT4, is \(\sim16.5\) kcal/mol below \(^{1\alpha}\)INT1. We describe the reaction segment \(^{1\alpha}\)INT2 \(\rightarrow^{1\alpha}\)INT4 in the Supporting Information (Figures S7 and S8).

Figure 6. Energy profile and structures of the key intermediates (in insets) for the oxidation reaction. Fragment shown in red corresponds to the triplet state; dark blue curve corresponds to the singlet PES.

Figure 7. Elementary step \(^{1\alpha}\)INT4 \(\rightarrow^{1\alpha}\)TS3 \(\rightarrow^{1\alpha}\)INT5 in which a proton is detached from C\(_{66}\), initiating a proton transfer route via a water molecule to dioxygen species. Top center panel illustrates the vibrational mode with the imaginary frequency in the transition state \(^{1\alpha}\)TS3. Kinetic isotope effect due to replacement of both protium atoms by deuterium atoms at C\(_{66}\) is estimated for this step.
the top center panel (\textsuperscript{\textit{ox}TS3}), accounts for the formation of the intermediate \textsuperscript{\textit{ox}INT5} in which the C\textsubscript{66}O\textsubscript{1} bond is broken.

At this step, kinetic isotope effect (KIE) due to dideuterium substitution at C\textsubscript{66} can be estimated theoretically. Wachter and co-workers studied\textsuperscript{31} KIEs in the chromophore maturation reaction for the isotope-enriched protein bearing Tyr residues deuterated at C\textsubscript{66} The rate constants derived from kinetics curves for unlabeled and deuterium-labeled protein allowed the authors to obtain a KIE of 5.89 ± 2.81. The vibrational mode with imaginary frequency (both for H and D) depicted in the top center panel in Figure 7 shows the involvement of an atom detached from C\textsubscript{66}. Calculations of zero-point energies of \textsuperscript{\textit{ox}INT4} and \textsuperscript{\textit{ox}TS3} followed by the estimates of the rate constants ratio resulted in a KIE value of 5.45, in agreement with the experimental data. It should be noted that, contrary to our model, the reactions involving proton transfer from C\textsubscript{66} were assigned in ref 31 to the reaction stage preceding oxidation.

The final elementary step, \textsuperscript{\textit{ox}INT5} \rightarrow \textsuperscript{\textit{ox}TS4} \rightarrow PROD, illustrated in Figure 6 and Figure S9 in the Supporting Information, corresponds to the highest barrier (21.3 kcal/mol) at the oxidation step (and the entire reaction of the chromophore’s formation). At this step, again, a concerted proton transfer takes place, utilizing the hydrogen-bond network segments: N\textsubscript{66}H\cdotsO\textsubscript{2} (Glu222) and O\textsubscript{2} (Glu222)\cdotsH\cdotsO\textsubscript{2} (Glu222)\cdotsO\textsubscript{2} (Glu222)\cdotsO\textsubscript{2} (Glu222). As a result, hydrogen peroxide molecule \textsubscript{H}\textsubscript{2}O\textsubscript{2} and the mature chromophore are formed. Although the covalent bonds and the conjugated \pi system of the chromophore are fully formed at this stage, it is possible that the appearance of the fluorescence may be delayed because of the structural relaxation of the protein adjusting to the mature chromophore (i.e., changes in the conformations and protonation states of the nearby residues). In such a case, the appearance of the fluorescence would be delayed relative to the peroxide production, as observed by Wachter and co-workers.\textsuperscript{31}

3.3. Closing the Cycle: Validation of the Structures and Connection to the Kinetics Data. 3.3.1. Comparison to the Crystal Structures. We initiated our simulations from the model system mimicking wt-GFP with an unmodified tripeptide sequence (Figure 3), REAG. Then we modeled a chain of chemical reactions, which resulted in the final structure, PROD, with the mature, fully formed chromophore. Comparison of the chromophore-containing pocket in the computationally derived structure PROD to that in the crystal structure PDB ID 2WUR\textsuperscript{43} of wt-GFP solved with high resolution (0.9 Å) showing hydrogens presents a stringent test for our simulations. Figure S10 of the Supporting Information shows a superposition of theoretical (PROD) and crystal (PDB ID 2WUR) structures. Given that the coordinates of the experimental structure PDB ID 2WUR have not been used as input data in our calculations, the agreement is excellent, providing strong support to the simulation results.

A comparison of model structures of the reaction intermediates with the relevant structures from the PDB is also important. The product of the cyclization step, structure INT3 (structure III in Figure 1), contains an imidazolidine-like ring that should be further transformed into the imidazoline-like ring. The Protein Data Bank contains an entry PDB ID 2QRF (the cyclized-only intermediate of GFP chromophore maturation in the Gln183Glu/Ser65Thr mutant) deposited in 2009 by Getzoff’s group, but this result has not been published. It is possible that replacement of Gln183 (located near Arg96) by Glu has resulted in trapping the reaction intermediate. An inspection of the crystal structure shows that the five-membered imidazolidine-like ring is strongly nonplanar, as it is in our intermediates, structures INT1\textsuperscript{−}INT3.

This structure should correspond to the intermediate preceding dehydration, since the dehydrated species contain a planar imidazoline-like ring. Figure S11 in the Supporting Information superimposes the crystal structure PDB ID 2QRF and our INT3, which closely resembles the imidazolidine fragment of the partially formed chromophore. This comparison also provides support to our simulation results.

Two papers reported crystal structures of GFP variants with the dehydrated but nonoxidized chromophores.\textsuperscript{27,29} Reference 27 reported the PDB ID 2FZU structure obtained for chemically reduced GFPsol (the solubility-optimized GFP variant F64L/S65T/F99S/M153T/V163A). Two best-defined conformers with 62% and 38% occupancies were reported. The structure was tentatively assigned to the enolate intermediate preceding oxidation.\textsuperscript{27}

Pletneva et al.\textsuperscript{29} characterized a crystal structure (PDB ID 3LVC) of the aceGFP-Gly222Glu protein. Initially, aceGFP is a bright green fluorescent protein, differing from wt-GFP by several mutations, including the substitution of Glu222 by Gly. In ref 29, the authors introduced a single reverse mutation, Gly222Glu, and obtained a colorless nonfluorescent mutant, called aceGFP-Gly222Glu, with an immature chromophore. The authors\textsuperscript{29} demonstrated that the maturation of the chromophore can be induced photochemically via photo-oxidation. They described the immature chromophore as the product of the cyclization and dehydration steps.

In Table 2, we compare the structure of INT4 with the two reported structures of the nonoxidized intermediate.\textsuperscript{27,29} We observe that all immature chromophores contain a planar imidazoline-like ring but differ in the degree of nonplanarity of the rings. The agreement between the INT4 structure and the crystal structures of cyclized and dehydrated intermediates is

<table>
<thead>
<tr>
<th>geometry parameter</th>
<th>computed INT4</th>
<th>crystal PDB ID 3LVC</th>
<th>crystal PDB ID 2QRF</th>
<th>crystal PDB ID 2FZU</th>
</tr>
</thead>
<tbody>
<tr>
<td>C\textsubscript{66}−C\textsubscript{65}</td>
<td>1.51</td>
<td>1.52</td>
<td>1.38</td>
<td>1.39</td>
</tr>
<tr>
<td>C\textsubscript{66}−C\textsubscript{65}</td>
<td>1.49</td>
<td>1.48</td>
<td>1.54</td>
<td>1.55</td>
</tr>
<tr>
<td>C\textsubscript{66}−C\textsubscript{65}−C\textsubscript{65}</td>
<td>1.25</td>
<td>1.22</td>
<td>1.23</td>
<td>1.15</td>
</tr>
<tr>
<td>C\textsubscript{66}−C\textsubscript{65}−C\textsubscript{65}−N\textsubscript{66}</td>
<td>−1.52</td>
<td>1.78</td>
<td>1.78</td>
<td>−1.17</td>
</tr>
<tr>
<td>C\textsubscript{66}−C\textsubscript{65}−C\textsubscript{65}−N\textsubscript{66}</td>
<td>34</td>
<td>−53</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>O\textsubscript{66}−O\textsubscript{65}</td>
<td>1.39</td>
<td>1.49</td>
<td>1.41</td>
<td>1.38</td>
</tr>
<tr>
<td>O\textsubscript{66}−O\textsubscript{65}</td>
<td>1.26</td>
<td>1.22</td>
<td>1.25</td>
<td>1.26</td>
</tr>
<tr>
<td>O\textsubscript{66}−N\textsubscript{67}</td>
<td>1.41</td>
<td>1.37</td>
<td>1.40</td>
<td>1.41</td>
</tr>
<tr>
<td>N\textsubscript{67}−C\textsubscript{67}</td>
<td>1.44</td>
<td>1.47</td>
<td>1.45</td>
<td>1.46</td>
</tr>
<tr>
<td>N\textsubscript{67}−C\textsubscript{67}</td>
<td>1.34</td>
<td>1.38</td>
<td>1.36</td>
<td>1.36</td>
</tr>
<tr>
<td>C\textsubscript{67}−C\textsubscript{68}</td>
<td>1.49</td>
<td>1.51</td>
<td>1.49</td>
<td>1.49</td>
</tr>
<tr>
<td>C\textsubscript{67}−N\textsubscript{68}</td>
<td>1.34</td>
<td>1.27</td>
<td>1.32</td>
<td>1.33</td>
</tr>
<tr>
<td>N\textsubscript{68}−C\textsubscript{68}</td>
<td>1.39</td>
<td>1.46</td>
<td>1.37</td>
<td>1.38</td>
</tr>
<tr>
<td>N\textsubscript{68}−C\textsubscript{68}</td>
<td>2.16</td>
<td>2.24</td>
<td>2.21</td>
<td>2.18</td>
</tr>
<tr>
<td>N\textsubscript{68}−C\textsubscript{68}</td>
<td>2.23</td>
<td>2.34</td>
<td>2.29</td>
<td>2.26</td>
</tr>
<tr>
<td>O\textsubscript{68}−O(Glu222)</td>
<td>3.01</td>
<td>3.54</td>
<td>4.02</td>
<td>3.89</td>
</tr>
<tr>
<td>O\textsubscript{68}−O(Thr96)</td>
<td>2.58</td>
<td>2.85</td>
<td>2.78</td>
<td>2.86</td>
</tr>
</tbody>
</table>

\textsuperscript{a}See Figure 3 for atom-numbering scheme. Distances are in Angstroms; angles are in degrees. \textsuperscript{b}Conformer 62%. \textsuperscript{c}Conformer 38%.
fairly good, again providing support to the conclusions drawn from our simulations.

3.3.2. Comparison with the Experimental Kinetics Data. Table 3 summarizes experimentally derived and computed energy barriers. Reid and Flynn, 24 who investigated the Ser65Thr-GFP chromophore maturation in vitro, proposed three distinct kinetic steps: first, protein folding precedes chromophore modification and occurs fairly slowly ($k_f = 2.44 \times 10^{-3} \text{s}^{-1}$); second, cyclization of the tripeptide with a rate constant of $3.8 \times 10^{-3} \text{s}^{-1}$; and third, oxidation of the cyclized chromophore ($k_{ox} = 1.51 \times 10^{-4} \text{s}^{-1}$).

Within the transition-state theory, the rate constant $0.0038 \text{s}^{-1}$ corresponds to a barrier of 20.7 kcal/mol. This barrier corresponds to the combined cyclization–dehydration steps. Our calculations give the highest barrier for cyclization–dehydration step of $\sim 17$ kcal/mol, which is in semiquantitative agreement with the experimentally derived value.

The rate constant for oxidation $0.000151 \text{s}^{-1}$ corresponds to a barrier of 22.7 kcal/mol. Kinetics of the oxidation step has also been studied by Getzoff and co-workers 27 and by Wachter and co-workers. 30 They reported reaction rates, which can be converted to the activation energies 22.3 27 and 22.0 kcal/mol. 30 Our calculations give the highest barrier for oxidation of $\sim 21.3$ kcal/mol, which is in excellent agreement with the reported experimental values.

Table 3. Comparison of Experimentally Derived and Computed Energy Barriers

<table>
<thead>
<tr>
<th>reaction step</th>
<th>experimental</th>
<th>calculated</th>
</tr>
</thead>
<tbody>
<tr>
<td>cyclization–dehydration</td>
<td>20.7\textsuperscript{a}</td>
<td>17</td>
</tr>
<tr>
<td>oxidation</td>
<td>22.3\textsuperscript{a} \textsuperscript{b} 22.0\textsuperscript{c}</td>
<td>21</td>
</tr>
</tbody>
</table>

\textsuperscript{a} Reference 24, \textsuperscript{b} Reference 27, \textsuperscript{c} Reference 30.

3.4. Mechanistic Implications and Connection with Previous Studies. Here, we computed the complete cycle of chemical reactions leading to chromophore formation in wt-GFP, from the initial tripeptide (Ser65-Tyr66-Gly67) to the mature GFP chromophore. The simulations provide strong support to the cyclization–dehydration–oxidation mechanism. The computed structures of the key intermediates are validated against the available experimental crystallographic data. The simulations provide clear molecular-level visualization of the elementary steps along the reaction pathway.

The two mechanistic interpretations providing a detailed picture of chemical reactions shown in Figure 1 are commonly referred to as “Getzoff’s mechanism” and “Wachter’s mechanism” (see, for example, ref 23), both formulated on the basis of the experimental studies. Getzoff’s mechanism assumes a sequence of stages illustrated in Figure 1, cyclization–dehydration–oxidation, while Wachter’s mechanism suggests an alternative reaction flow, cyclization–oxidation–dehydration. Our simulations are consistent with the cyclization–dehydration–oxidation mechanism but flesh out additional details and offer new insights into the nature of the reaction intermediates. To illustrate the key mechanistic issues we present schematically in Figure 8 the segment of the reaction pathway from the cyclized intermediate (INT3), which is common for all three approaches. The upper pathway is Getzoff’s mechanism, the bottom one is Wachter’s mechanism, and the middle one is the mechanism derived on the basis of the present simulations. After formation of the cyclized intermediate the reaction pathways of three mechanisms follow different routes, finally arriving to the same reaction product, the protein with the mature chromophore.

The dehydration stage in Getzoff’s mechanism (the top panel in Figure 8) assumes a formal removal of a proton from N66 and a hydroxyl from C65. The detailed transformations, which...
have not been disclosed, may include several elementary steps, e.g., involving the side chain of Glu222 and water molecule(s). Wachter’s mechanism assumes a proton transfer from Cα66 to the side chain of Glu222 mediated by a water molecule when starting from the cyclized intermediate. Presumably, this step is required to generate a negatively charged intermediate (see the bottom panel in Figure 8) capable of donating an electron to the oxygen molecule and to initiate oxidation prior to dehydration. Our mechanism (center panels in Figure 8) is consistent with removal of a proton from Cα66 (unlike Getzoff’s mechanism) and a hydroxyl from Cα66 thus resulting in the preoxidized intermediate (denoted INT4). All proposed reaction mechanisms of the oxidation step share a common feature that the preoxidized chromophore donates an electron to O2, thus activating the reactions of the corresponding intermediates with molecular oxygen. Molecular simulations provide direct evidence for this step; our intermediate INT2 (see Figure 5) corresponds to the point where an electron transfer to the oxygen molecule occurs and where the triplet and singlet PESs cross. Formation of peroxy compounds at the oxidation step (INT3, INT4, INT5) is also a common feature of the proposed reaction pathways.

However, there are several notable differences in the corresponding chemical transformations (indicated by notations pathway-1, pathway-2, and pathway-3 in Figure 8) as described in this work. We emphasize that our work presents consistent with removal of a proton from Cα66 ( unlike Getzoff’s mechanism) and a hydroxyl from Cα66 thus resulting in the preoxidized intermediate (denoted INT4). All proposed reaction mechanisms of the oxidation step share a common feature that the preoxidized chromophore donates an electron to O2, thus activating the reactions of the corresponding intermediates with molecular oxygen. Molecular simulations provide direct evidence for this step; our intermediate INT2 (see Figure 5) corresponds to the point where an electron transfer to the oxygen molecule occurs and where the triplet and singlet PESs cross. Formation of peroxy compounds at the oxidation step (INT3, INT4, INT5) is also a common feature of the proposed reaction pathways.

The key experimental results, which support the cyclization—oxidation—dehydration mechanism, are based on the observation that fluorescence of mature chromophore upon reactions with the mGFPsol variant follows the production of hydrogen peroxide.31 The delayed fluorescence of the mature chromophore might be explained by possible structural relaxation of the protein accommodating the newly formed chromophore (the relaxation might involve conformational changes and/or changes in protonation states of nearby residues). The local structure around the chromophore and its fluorescent ability may also be affected by the presence of hydrogen peroxide. Moreover, we cannot exclude the possibility of multiple competing pathways including those in which H2O2 is formed prior to chromophore formation (e.g., in which the chromophore appears first as a zwitterionic intermediate). We agree with the comment by Pletneva et al.29 (advocating for the cyclization—dehydration—oxidation mechanism) that “this might indicate the complexity of chromophore biosynthesis, suggesting more than one possible reaction pathway” of the chromophore’s maturation.

3.4.1. Cyclization Reaction: A Long-Standing Problem of Computationally Derived Energetics. Several computational studies attempted to model the mechanism of the ring formation in GFP, i.e., to simulate elementary reactions from a model system mimicking a protein with the tripeptide sequence Ser65-Tyr66-Gly67 (structure I in Figure 1 or REAG in our model) to a required intermediate with the imidazolidine-like ring (structure III in Figure 1 or INT3 in our model). Siegbahn et al. reported32 the results of DFT calculations for a series of molecular clusters representing the active site. The structure of reactants was graphically restored from the configuration in the crystal structure (PDB ID 1GFL) with the mature chromophore to the polypeptide, as it would be prior to the autocatalytic cyclization. The authors clearly stated that they were not satisfied with the calculated energetics for the mechanism assuming a direct attack of Nα67 on Cα66 as depicted in Figure 1 (for atom numbering, see also Figure 3). Prompted by the calculation results they put forward an alternative suggestion, that dehydration of Tyr66 to dehydrotyrosine (removal of a proton from Cα66) occurs prior to cyclization, but this hypothesis was not supported by the experiments. Ma et al. explored33,34,37 several options to model the cyclization step for wt-GFP: they used the approach of a quantum-chemical cluster and the QM/MM ONIOM method.35 Molecular models were constructed33 on the basis of crystal structure PDB ID 2AWJ, the original wt-GFP residues were restored, and the structure was refined by using molecular dynamics and quantum-chemical calculations. In particular, the authors considered the mechanisms studied by Siegbahn et al.32 as well as the original suggestions, most of which were rejected by the authors due to the computed high-energy barriers. The most recent paper36 proposed a two-step mechanism according to which, first, a proton on the amide nitrogen of Gly67 (Hα67 in our designation, see Figure 3) transfers to O65; second, the amide nitrogen (Nα67) attacks the carbonyl carbon of Ser65 (Cα65). The computed energy profile (calculated with the ONIOM(B3LYP/6-31G**/AMBER) electronic-embedding scheme) shows barriers as high as 33 kcal/mol, which is significantly higher than the experimentally derived activation energy of 21 kcal/mol.24,30

In contrast to previous theoretical studies, we were able to find a pathway from the initial tripeptide to the cyclized intermediate, with energy barriers consistent with the kinetic studies. The highest barrier at the cyclization stage in our simulations, 13 kcal/mol, does not contradict the experimental kinetic studies.

A novel feature in our mechanism is the zwitterionic intermediate, INT1. The stability of this structure is easily explained: the H−Nα66−H group is positively charged, and the Cα65−Oα65 group with an 1.32 Å interatomic distance is negatively charged. Additional stabilization of this intermediate is provided by the hydrogen bonds of O65 with Nα66 and Oα66 as well as the hydrogen bonds of N65 with O65 (Glu222) and O66 with N(Arg96). In the absence of such favorable environment for INT1, the Cα65−Nα66 bond would be unstable.

Our simulation of the cyclization step is validated against the experimental structural and kinetics data. First, our model with the mature chromophore agrees well with high-resolution crystal structure PDB ID 2WUR (Figure S10), the structure of REAG is consistent with crystal structure PDB ID 2AWJ (Figure 3 and Table 1), and the structure of INT3 is consistent with crystal structure PDB ID 2QRF (Figure S11). Second, the computed energy profile of the cyclization stage for wt-GFP (Figure 4) lies within reasonable limits consistent with the kinetic studies.

3.4.2. Dehydration Reaction. The QM/MM-computed minimum energy pathways (pathway-2 in Figure 8 for wt-GFP) suggest that dehydration is assisted by a water molecule in the active site. A hydrogen-bond chain including the Cα66H, water, and Cα65−Oα65H groups provides a route for proton transfer coupled with the Cα65−Oα65 bond cleavage. The dehydration mechanism for wt-GFP gains support from the crystallography studies. Table 2 compares the structure of the
computationally derived dehydrated intermediate INT4 with two crystal structures (PDB ID 3LVC and PDB ID 2FZU) obtained in two different experiments. Pletneva et al. employed mutagenesis to create the aceGFP-Gly222Glu variant, which exhibited no absorption/fluorescence in the visible region. Barondeau et al. chemically reduced the GFPsol variant and also obtained colorless species. Both papers support formation of the dehydrated intermediate prior to the oxidation stage. The comparison presented in Table 2 demonstrates that our INT4 matches these two crystal structures PDB ID 3LVC and PDB ID 2FZU. The energy from the measured rate constant by Reid and Flynn. We believe that our computed value, 17 kcal/mol, provides an estimate from below. First, the conformation of the Tyr66 side chain was maintained as such in the protein with the mature chromophore; however, it should be different in the reagent structure with the tripeptide Ser65-Tyr66-Gly67. An amount of energy of the order of few kcal/mol, which is not taken into account in simulating this step, would be required to change the conformation of Tyr66.

### 3.4.3. Oxidation Reaction.

No attempts to model the oxidation step in the GFP chromophore maturation have been reported prior to this work. The papers from Getzoff’s and Wachter’s groups tentatively describe chemical transformations followed by the decisive step, the transfer of an electron from a relevant intermediate to dioxygen, generating a caged radical pair. This is a well-known way by which spin-forbidden reactions with molecular oxygen was demonstrated and co-workers in crystallography studies.

We found a route which in this particular protein environment led to cleavage of both oxygen–carbon bonds (Cα–Ox1 and C65–Ox2) and formation of hydrogen peroxide. We successfully reached the structure with the mature chromophore in wt-GFP. Comparison with a high-resolution crystal structure shows that all chemical transformations involved in the chromophore’s maturation are well localized, an aspect noted by Zimmer et al.

We believe that our computed value, 17 kcal/mol, provides an estimate from below. First, the conformation of the Tyr66 side chain was maintained as such in the protein with the mature chromophore; however, it should be different in the reagent structure with the tripeptide Ser65-Tyr66-Gly67. An amount of energy of the order of few kcal/mol, which is not taken into account in simulating this step, would be required to change the conformation of Tyr66.

### 4. CONCLUSION

We described the sequence of chemical reactions at an atomic scale from a precyclded polypeptide (REAG) to the reaction products (PROD) clarifying the mechanistic details of the chromophore’s maturation reaction in wt-GFP. This is the first computational study of the entire maturation reaction. The calculations of all steps are performed at the uniform theoretical level, QM(DFT)/MM. No calculations of the oxidation of the dehydrated intermediate have been reported prior to our study. The proposed mechanism is consistent with the available experimental structural and kinetic data. Molecular modeling provides a detailed visualization of the elementary steps in the maturation process and clarifies several important aspects of the corresponding reactions. Here, the reactions occur outside the protein barrel, and the presence of the charged groups (Arg96, Glu222) near the active site as well as a hydrogen-bond network are instrumental for stabilization of the key reaction intermediates, INT2, INT1, and INT4. The results of simulations establish a solid basis for predictions of maturation mechanisms in other fluorescent proteins.

**ASSOCIATED CONTENT**

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.7b00676.

Additional energy profiles and detailed structures along the maturation pathway, molecular-orbital analysis of the Chro−O2 charge-transfer complex (PDF)

**AUTHOR INFORMATION**

**Corresponding Authors**

*akrylov@usc.edu*

*anemukhin@yahoo.com; anem@lcc.chem.msu.ru*

**ORCID**

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

**Notes**

The authors declare no competing financial interest.

**ACKNOWLEDGMENTS**

B.L.G. and A.V.N. thank Prof. Alexander Majouga for helpful discussions. This work was supported by the U.S. National Science Foundation (CHE-1566428, A.I.K.) and the Russian Foundation for Basic Research (project no. 16-03-00078, A.V.N.). The authors acknowledge the use of supercomputer resources of the Lomonosov Moscow State University and of the Joint Supercomputer Center of the Russian Academy of Sciences.

**REFERENCES**
