Available online at www.sciencedirect.com

ScienceDirect

Mendeleev Commun., 2019, **29**, 187–189

Mendeleev Communications

Modeling of the glycine tripeptide cyclization in the Ser65Gly/Tyr66Gly mutant of green fluorescent protein

Bella L. Grigorenko,[a,b] Ekaterina D. Kots,*[a,b] Anna I. Krylov[c] and Alexander V. Nemukhin[a,b]

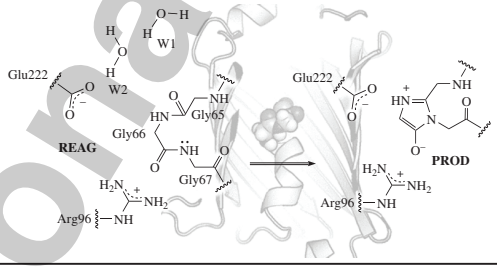
[a] *Department of Chemistry, M. V. Lomonosov Moscow State University, 119991 Moscow, Russian Federation. E-mail: kots.katya@gmail.com*

[b] *N. M. Emanuel Institute of Biochemical Physics, Russian Academy of Sciences, 119334 Moscow, Russian Federation*

[c] *Department of Chemistry, University of Southern California, Los Angeles, 90089-0482 CA, USA*

DOI: 10.1016/j.mencom.2019.03.024

The quantum mechanics/molecular mechanics approach is used to model the chain of elementary reactions involved in the backbone cyclization of the initially non-cyclized Gly65–Gly66–Gly67 tripeptide inside the protein matrix of the green fluorescent protein (GFP) Ser65Gly/Tyr66Gly mutant. The computationally characterized reaction mechanism provides support for understanding chromophore maturation in GFP-like fluorescent proteins.



The autocatalytic chromophore maturation in green fluorescent protein (GFP), a remarkable marker for *in vivo* imaging,[1–3] is a brilliant example of organic chemical reactions that occur in a specific protein environment. For the wild-type protein, a post-translational modification of the Ser65–Tyr66–Gly67 tripeptide sequence comprises a series of elementary reactions resulting in the formation of *p*-hydroxybenzylidene-imidazolinone chromophore covalently bound to the polypeptide chain. These elementary reactions, which are traditionally described as cyclization, dehydration and oxidation steps, have been investigated in multiple experimental and few computational works.[1,2] Our recent paper[4] describes these three elementary reactions of the chromophore maturation in wt-GFP using the quantum mechanics/molecular mechanics (QM/MM) approach. The new mechanism for the cyclization step has been formulated, which includes an unusual reaction intermediate of the zwitterionic character with the imidazolidine-like ring stabilized by a specific protein environment. This intermediate explains the relatively low activation barrier at the cyclization step observed in kinetic measurements.

The goal of the present work is to verify whether this mechanism is applicable for a mutant of GFP in which two of the three involved amino acid residues (Ser65 and Tyr66) are replaced by glycine ones. Previously the chromophore maturation mechanism for this mutant was investigated experimentally,[5] and the crystal structure for the protein with non-cyclized glycine tripeptide was deposited in the Protein Data Bank (PDB ID 1QYO).[6] These experimental data demonstrated that the glycine tripeptide cyclization proceeded for this mutant, but contrary to wt-GFP it resulted in the colourless product.

In this work we applied a computational protocol† similar to that used in the previous simulation for wt-GFP.[4] A model system comprised a large fraction of the protein molecule surrounding the active site where the chemical reactions occur. As a template, we used the model system created in our previous work[4] and replaced the side chains of the Ser65 and Tyr66 residues of wt-GFP by the glycine side chains. The vacancies in the protein matrix, which emerged due to a smaller size of the glycine side chains, were filled with additional water molecules. The structure was optimized in QM(DFT)/MM calculations in which the QM subsystem included the side chains of Gly65, Gly66, Gly67, Arg96, His148, Ser205 and Glu222 residues as well as seven water molecules. The presence of the charged side chains of the Arg96 and Glu222 conserved residues in the QM part was found to be crucial for a correct description of electrostatic effects at the active site.[4,5] The side chains of His148 and Ser205 residues are the important constituents of the hydrogen-bond network at the active site.[14] The water molecules assigned to QM actively participate in proton transfer in the reactions considered.[15,16]

The computationally derived structure of the model system with the non-cyclized Gly65–Gly66–Gly67 tripeptide, which we denoted as REAG, can be directly compared to the crystal structure with PDB ID 1QYO.[6] As follows from Figure 1, our model for the precyclized tripeptide corresponds well to this crystal structure.

† *Computational protocol.* To scan the sections of potential energy surfaces, the flexible effective fragment QM/MM method[7,8] was used. According to this method, the groups assigned to the MM part were represented by effective fragments contributing their one-electron potentials to the quantum Hamiltonian. Protein polypeptides were described as flexible chains of small effective fragments, and fragment–fragment interactions were computed with conventional force fields. The modified code of the GAMESS(US) software[9,10] and the TINKER software[11] was used in these QM/MM calculations. The density functional theory (DFT) approach was used in the QM subsystem with the PBE0 functional[12] and 6-31G* basis set. The AMBER force field parameters[13] were applied in MM. The estimated quantitative parameters can be improved using advanced computational approaches. First, expanding the basis set would result in a better quality of the energy profile and presumably lower reaction barriers. Second, the standard Gibbs free energy values, as obtained by proper correction from the potential energy values, would be helpful to establish correspondence with reaction rate constants.

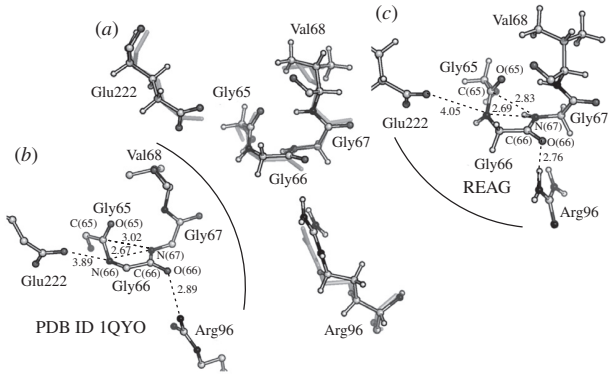


Figure 1 Structural features of the Ser65Gly/Tyr66Gly mutant of GFP: (*a*) superposition of the peptide chains from computation (balls and sticks) *vs.* peptide chains for the crystal structure PDB ID 1QYO (sticks only), (*b*) fragment of the crystal structure demonstrating some distances in Å, (*c*) fragment of the computationally derived structure demonstrating the same distances as for the crystal structure.

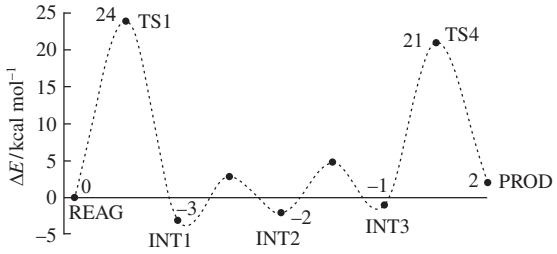


Figure 2 Computed reaction energy profile for the Gly65–Gly66–Gly67 cyclization. Stationary points are connected by a dashed curve to guide the eye.

Starting from the REAG structure, we modeled a chain of elementary reactions resulting in the structure denoted here as PROD, in which the planar imidazole ring is covalently bound to the protein. Figure 2 shows the energy profile connecting the REAG and PROD points computed by the QM/MM method. We identify three reaction intermediates, denoted as INT1, INT2, and INT3, which have common features with the corresponding intermediates in the chromophore maturation for wt-GFP.[4] Scheme 1 illustrates the corresponding chemical transformations.

The two elementary reaction steps with the largest energy barriers, namely the cyclization REAG → TS1 → INT1 and the dehydration INT3 → TS4 → PROD, deserve special consideration. The former step results in the formation of a bond between the nitrogen atom N(67) from the Gly67 residue and the carbon atom C(65) from the Gly65 residue (Figure 3). The reaction coordinate along this step is the N(67)–H distance, and its elongation results in the bond formation between N(67) and

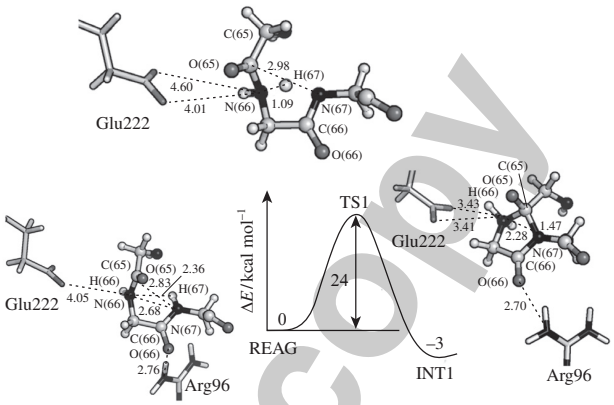
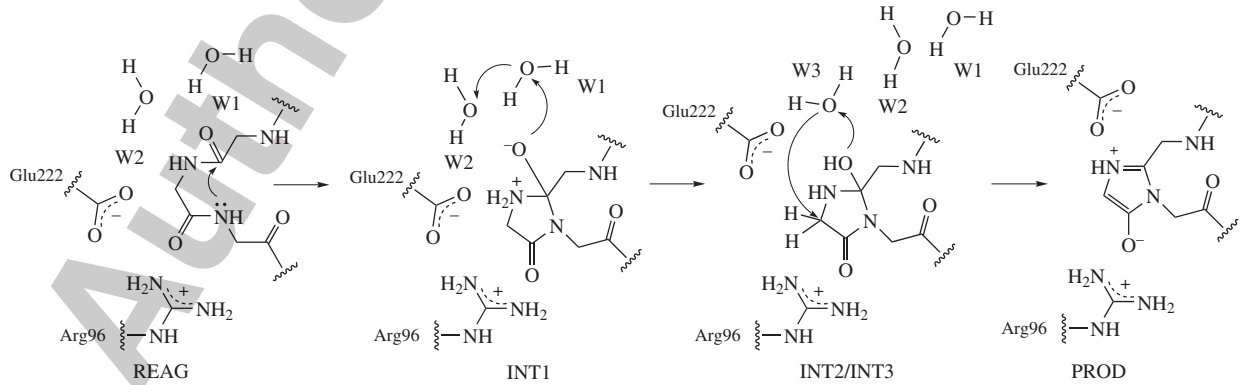


Figure 3 The first step in the Gly65–Gly66–Gly67 cyclization: REAG → TS1 → INT1. Distances between atoms are given in Å.

C(65) atoms. The structures REAG and INT1 are almost isoenergetic, and the energy barrier is 24 kcal mol^{-1}, *i.e.,* considerably higher than that for wt-GFP (13 kcal mol^{-1}).[4] The intermediate INT1 with the imidazolidine-like ring, shown in the right bottom panel in Figure 3, has H_2N^+–CO^- zwitterionic character. Stability of the structure INT1 in this specific protein environment can be rationalized as follows: the H–N(66)–H group is positively charged, and the C(65)–O(65) group with 1.33 Å interatomic distance is negatively charged. It is also important to note the hydrogen bonds occurrence for N(66) with the carboxylic oxygen atoms of the Glu222 residue as well as for O(66) with the nitrogen atom of the Arg96 residue. In the absence of such a favorable environment for structure INT1, the C(65)–N(66) bond would be unstable. Therefore, similarly to the results for wt-GFP,[4] we emphasize a crucial role of Arg96 and Glu222 residues for stabilization of structure INT1 by the electrostatic and hydrogen-bond interactions.

The latter step, namely INT3 → TS4 → PROD, results in the structure PROD with the planar imidazole ring instead of the puckered imidazolidine-like ring in the structure INT3, as illustrated in Figure 4. A hydrogen-bond chain including the C(α66)–H group, the water molecule W3 and the C(65)–O(65)H group constitute a proton transfer route conjuncted with the C(65)–O(65) bond cleavage. This reaction step corresponds to the dehydration of the intermediate INT3, which also takes place in the cyclization of wt-GFP.[4] The difference in the reaction mechanisms for the Ser65Gly/Tyr66Gly mutant considered here and wt-GFP is that for the mutant reaction the water molecules, located between the Glu222 residue and the tripeptide moiety, mediate the proton transfer, since the side chain of the Glu222 residue is farther from the reactant as compared to wt-GFP. In particular, these simulations demonstrate that the cyclization and



Scheme 1 Chemical transformations for the elementary reaction steps.

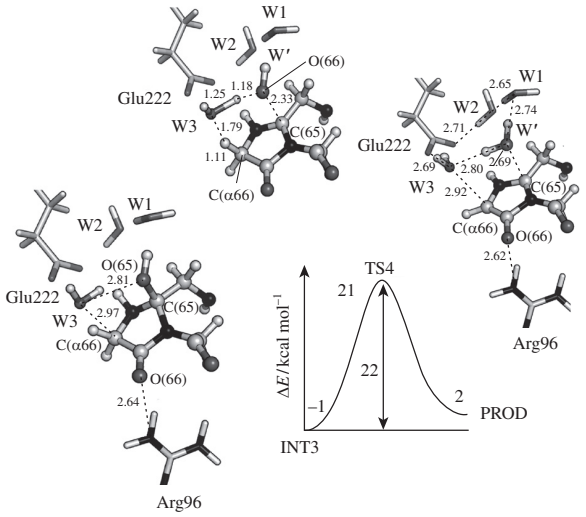


Figure 4 The dehydration step: INT3 → TS4 → PROD. Distances between atoms are given in Å.

dehydration steps can occur if the Glu222 residue side chain is replaced by another one, for example in the known Glu222Gly mutant.

In summary, in this work we have demonstrated that the Gly65–Gly66–Gly67 tripeptide moiety of the GFP mutant can be converted to the cyclized species with the planar imidazole ring. The QM/MM-computed reaction profile demonstrates the energy barriers not exceeding 24 kcal mol^{-1}, and this barrier can be overcome in reactions in proteins. As in the case of wt-GFP, the key reaction intermediate along the established pathway is the zwitterionic intermediate INT1 stabilized by interaction with the neighboring groups near the active site, namely the conserved Arg96 and Glu222 residues, plays a special role in this interaction. The application of the proposed maturation mechanism extends beyond the considered examples of wt-GFP and its Ser65Gly/Tyr66Gly mutant for the following reasons. First, all fluorescent proteins of the GFP family having the 65–67 XZG tripeptide sequence, where X is variable, Z is an aromatic amino acid and G is always glycine, undergo autocatalytic maturation, which most likely follows the specific mechanisms. Second, a related process for the Ala–Ser–Gly tripeptide moiety is known for the histidine ammonia lyase (HAL) posttranslational modification,[17] a comparison of maturation processes for GFP and HAL being performed previously.[18]

This work was supported by the Russian Foundation for Basic Research (grant no. 16-03-00078). A.I.K. acknowledges the support of the U.S. National Science Foundation (no. CHE-1566428).

The research was carried out using the equipment of the shared research facilities of HPC computing resources at M. V. Lomonosov Moscow State University. We also acknowledge the use of supercomputer resources of the Joint Supercomputer Center of the Russian Academy of Sciences.

Online Supplementary Materials

Supplementary data associated with this article can be found in the online version at doi: 10.1016/j.mencom.2019.03.024.

References

1 R. Y. Tsien, *Annu. Rev. Biochem.*, 1998, **67**, 509.
2 M. Zimmer, *Chem. Rev.*, 2002, **102**, 759.
3 A. Acharya, A. M. Bogdanov, B. L. Grigorenko, K. B. Bravaya, A. V. Nemukhin, K. A. Lukyanov and A. I. Krylov, *Chem. Rev.*, 2017, **117**, 758.
4 B. L. Grigorenko, A. I. Krylov and A. V. Nemukhin, *J. Am. Chem. Soc.*, 2017, **139**, 10239.
5 D. P. Barondeau, C. D. Putnam, C. J. Kassmann, J. A. Tainer and E. D. Getzoff, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 12111.
6 *Research Collaboratory for Structural Bioinformatics*, Protein Data Bank, www.rcsb.org.
7 B. L. Grigorenko, A. V. Nemukhin, I. A. Topol and S. K. Burt, *J. Phys. Chem. A*, 2002, **106**, 10663.
8 A. V. Nemukhin, B. L. Grigorenko, I. A. Topol and S. K. Burt, *J. Comput. Chem.*, 2003, **24**, 1410.
9 M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis and J. A. Montgomery, *J. Comput. Chem.*, 1993, **14**, 1347.
10 M. S. Gordon and M. W. Schmidt, in *Theory and Applications of Computational Chemistry. The First Fourty Years*, eds. C. E. Dykstra, G. Frenking, K. S. Kim and G. E. Scuseria, Elsevier, Amsterdam, 2005, pp. 1167–1190.
11 *Ponder lab, Tinker–Software Tools for Molecular Design*, https://dasher.wustl.edu/tinker/.
12 C. Adamo and V. Barone, *J. Chem. Phys.*, 1999, **110**, 6158.
13 W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, *J. Am. Chem. Soc.*, 1995, **117**, 5179.
14 B. L. Grigorenko, A. V. Nemukhin, I. V. Polyakov, D. I. Morozov and A. I. Krylov, *J. Am. Chem. Soc.*, 2013, **135**, 11541.
15 R. M. Wachter, *Acc. Chem. Res.*, 2007, **40**, 120.
16 D. P. Barondeau, J. A. Tainer and E. D. Getzoff, *J. Am. Chem. Soc.*, 2006, **128**, 3166.
17 T. F. Schwede, J. Rétey and G. E. Schulz, *Biochemistry*, 1999, **38**, 5355.
18 M. Donnelly, F. Fedeles, M. Wirstam, P. E. Siegbahn and M. Zimmer, *J. Am. Chem. Soc.*, 2001, **123**, 4679.

Received: 6th September 2018; Com. 18/5684