Structural forms of green fluorescent protein by quantum mechanics/molecular mechanics calculations

B. L. Grigorenko,[a] I. V. Polyakov,[a] A. P. Savitskii,[b] and A. V. Nemukhin[a,c]*

[a]Department of Chemistry, M. V. Lomonosov Moscow State University, 1 Leninskie Gory, 119992 Moscow, Russian Federation. Fax: +7 (495) 932 8846. E-mail: polyakoviv@gmail.com
[b]A. N. Bakh Institute of Biochemistry, Russian Academy of Sciences, 33 Leninsky prosp., 119991 Moscow, Russian Federation.
[c]N. M. Emanuel Institute of Biochemical Physics, Russian Academy of Sciences, 4 ul. Kosygina, 119334 Moscow, Russian Federation

The molecular modeling of structural forms of the green fluorescent protein (GFP) with the Ser65Thr single-site mutation was performed by the quantum mechanics/molecular mechanics (QM/MM) method. Two model systems were constructed based on the crystallographic structure from the Protein Data Bank (PDB entry code 1EMA.) The model systems differ in the initial protonation state of the side chain of the amino acid residue Glu222 near the chromophore. The atomic coordinates of the protein macromolecule corresponding to the equilibrium geometric configurations were determined by total energy minimization using the QM/MM method within the density functional theory approximation PBE0/cc-pVDZ for the quantum subsystem that consists of the chromophore, a water molecule, and the side chains of Arg96, Glu222, and Ser205, and with the parameters of the AMBER force field for the molecular mechanics subsystem. In the analysis of the results, particular attention was given to the hydrogen bond redistribution in the chromophore-containing region of the protein caused by a change in the protonation state of the chromophore. The results obtained from the model containing the initially protonated side chain of Glu222 suggest a new interpretation of the photophysical processes in the green fluorescent protein.

Key words: green fluorescent protein, molecular modeling, hybrid quantum mechanics/molecular mechanics methods.

The discovery and the use of color fluorescent proteins of the green fluorescent protein (GFP) family have stimulated great interest in these proteins because of the possibility of labeling cell clones, thus allowing literally the visual observation of the course of intracellular events.[1—3] These proteins are crystallographically characterized in sufficient detail[4,5] and data for many of them are deposited in the Protein Data Bank (PDB).[6] The proteins of the GFP family have a barrel structure consisting of tightly packed β-sheets that efficiently protect the (*p*-hydroxybenzylidene)imidazolinone chromophore derived from the tripeptide Ser65—Tyr66—Gly67 from the bulk solvent. The transformations of the chromophore that occur in macromolecules under irradiation with light of a particular wavelength are responsible for the photophysical properties of fluorescent proteins. However, in spite of numerous publications on the structure and dynamics of GFP, the knowledge on the functional states of the protein is far from complete.

The absorption spectrum of GFP shows a band with the absorption maximum at approximately 400 nm, which is assigned to the so-called A form of the protein, and a band of lower intensity with the maximum at 480 nm assigned to the B form.[7,8] The positions of the absorption bands and the intensity ratio depend on a number of external factors, as well as on the presence of mutations in the protein molecules. It is assumed that the band of the A form belongs to the absorption of the chromophore in the neutral state, and the band of the B form corresponds to the anionic state containing the deprotonated hydroxyl group of the tyrosine residue (Tyr66) involved in the chromophore (hereinafter, this fragment is denoted Cro(Tyr66)). The excitation at both wavelengths gives rise to the fluorescence with a peak at about 510 nm. Hence, it was suggested that the photoexcitation of the A form is accompanied by the proton transfer to give the B form. Figure 1 shows a part of the environment of the chromophore in the protonated form in the Ser65Thr mutant of GFP.[9]

Published in Russian in *Izvestiya Akademii Nauk. Seriya Khimicheskaya,* No. 1, pp. 62—66, January, 2010.

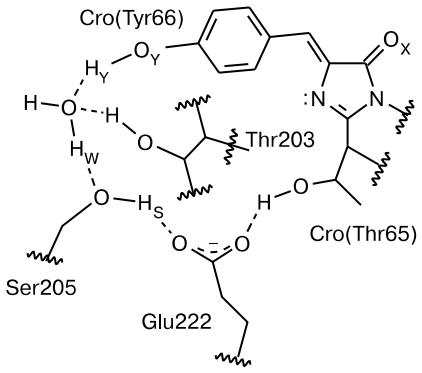

Fig. 1. Chromophore environment in the protein GFP(Ser65Thr).

Upon the excitation of the neutral form of the A chromophore characterized by a short-wavelength absorption band (HO_Y, N, O_x), the proton H_Y is transferred from the fragment Cro(Tyr66) to the nearest water molecule W, another protons of the water molecule, H_W, is transferred to the side chain of the residue Ser205, and the transfer of the proton H_S from Ser205 to the initially unprotonated side chain of Glu222 is the final event in the sequence of relatively fast transformations of the A form to the intermediate I form. Slower conformational transformations with the involvement of amino acid residues in the chromophore-containing region, which result in the transformation from the I form to the B form, stabilize the protein structure containing the chromophore in the anionic configuration responsible for the fluorescence. They are characterized by a long-wavelength absorption band (O_Y, N, O_X) and are in particular, accompanied by a substantial reorientation of the side chain of Thr203. These hypotheses were drawn from the X-ray diffraction data and spectroscopic studies of the proteins and chromophore molecules in the crystals and in solution.

Alternative hypotheses concern the involvement of different protonation states of the chromophore in photophysical transformations of GFP. The A form was referred to the cation of the chromophore $(HO_Y, HN, O_X)^+$, and the B form was attributed to the complex of the zwitterion (O_Y^-, HN^+, O_X) with the nearest molecular groups.[10,11] In a subsequent study by this research team,[12] it was suggested to consider the chromophore structures (O_Y, N, HO_X), $(HO_Y, N, HO_X)^+$, $(O_Y, HN, HO_X)^+$, and $(HO_Y, HN, HO_X)^{2+}$ in addition to the above-mentioned structures.

Of no less significance are the results of the modeling using quantum theory methods. In particular, the positions of the bands in the optical absorption spectra of model chromophores were calculated by the INDO/S semiempirical quantum chemical method.[10—12] The constants of acidity of the O_Y, N, and O_X atoms of the GFP chromophore were determined by quantum chemical methods.[13,14] Based on these constants, suggestions were made concerning the protonation states of the chromophore in solution and in the protein environment. The properties of small molecular clusters modeling the structures of chromophore-containing peptides of the protein GFP were calculated by the density functional theory methods.[15,16]

The molecular modeling of protein systems with the use of hybrid quantum mechanics/molecular mechanics (QM/MM) methods[17] would be expected to give the most significant results. In these methods, the properties of the most important fragment of the macromolecule are considered at the quantum mechanical level, whereas the major part of the protein matrix surrounding the quantum subsystem is described by molecular mechanics force fields. Evidently, the quantum subsystem should include the chromophore, and the boundary between the QM and MM parts should be such that all conjugated bonds responsible for the light absorption and emission are described by quantum equations. It is also reasonable to include the amino-acid side chains, which are located near the chromophore and can participate in the proton transfer involving the chromophore group, in the QM subsystem. An example of the structure modeling for another protein of the GFP family, *viz.*, the kindling fluorescent protein asFP595, was described.[18,19]

The results of the modeling of the structural forms of GFP in the chromophore-containing region by the QM/MM methods are scarcely covered in the literature. A simple and less reliable procedure for the determination of the structural parameters of a model protein system by the AM1 semiempirical quantum chemical method was employed.[20] Only the neutral form of the chromophore (HO_Y, N, O_X) was considered with the inclusion of the side chains of Arg96, His148, Ser205, and Glu222, the chromophore, and a water molecule in the quantum subsystem. The starting coordinates of the heavy atoms were estimated from the structure deposited in the PDB with the entry code 1GFL.[5] This is the structure of the wild-type protein GFP containing the residue Ser65. The main goal of the modeling was to theoretically estimate the positions of the bands in the absorption spectra using the TD-DFT approach. The positions of the bands in the absorption and emission spectra of GFP were calculated[21] by the QM/MM method at the CASSCF level of theory for the quantum subsystem and using the CHARMM force field for the MM subsystem. Only the neutral form of the chromophore (HO_Y, N, O_X) in the protein matrix, which corresponds to the structure deposited in the PDB with the entry code 1GFL,[5] was considered. The quantum subsystem included the side chains of Arg96, His148, Ser205, and Glu222, the chromophore, and a water molecule. The calculated positions of the bands are in satisfactory agreement with the experimental data (the deviations are 20—30 nm). In a series of calculations for the successively

extended quantum subsystem, the influence of the charged amino acid residue Arg96 located near the chromophore on the electronic spectrum was investigated. The optical spectra of the native form of GFP were calculated also[22] using the SAC/CI configuration interaction method to estimate the energy differences between the ground and excited states in the quantum subsystem (the chromophore in the neutral form (HO_Y, N, O_X), a water molecule, and the side chains of Ser205 and Glu222) and the point charge model for the other part of the protein matrix.

Therefore, in the above-mentioned studies, the chromophore of the native protein GFP(Ser65) was considered based on the structure with the PDB entry 1GFL[5] assuming the neutral form of the chromophore (HO_Y, N, O_X) and the anionic state of the side chain of Glu222. In the present study, we report new results of the structure modeling of GFP by the QM/MM method, with attention given to the following problems. For practical applications of GFP, the structure of GFP(Ser65Thr) deposited in the PDB with the entry code 1EMA[4] is more interesting because of the predominance of the form with the long-wavelength absorption band. Model systems of GFP with the initially protonated side chain of Glu222 have not been hitherto considered. At least, the conserved amino acid residue Glu in the analogous position in the protein asFP595 is assumed to be protonated in the dark form. Let us emphasize that the positions of protons in the molecular groups of macromolecules cannot be unambiguously determined by experimental methods. Hence, the results of modeling are helpful in studying the protein structure and dynamics.

Models and methods

In the present study, we considered two models for calculations by the QM/MM method of equilibrium geometric configurations of the GFP(Ser65Thr) structures in the chromophore-containing region of the protein. In the model 1, the side chain of the amino acid residue Glu222 was initially assumed to be unprotonated with a charge of −1. In the model 2, a proton was added to the system and the side chain of Glu222 was assumed to be protonated with a charge of 0. The initial state of the chromophore was assumed to be neutral (HO_Y, N, O_X). The initial coordinates of the heavy atoms for both model systems were taken from the structure with the PDB entry 1EMA.[4]

The modeling was carried out by the flexible effective fragment QM/MM method.[23,24] This method has advantages over other QM/MM schemes in that the quantum and molecular mechanical subsystems are rather adequately related, allowing the consideration of the contributions of the molecular groups of the MM subsystem (electron penetration) to the quantum Hamiltonian, as well as the influence of the changes in the geometric parameters of the QM subsystem on the conformations of the MM groups. In this approach, the quantum subsystem included the chromophore. the side chains of Arg96, Glu222, and Ser205 and the water molecule W. The energies and forces in the QM subsystem were calculated by the density functional theory method with the PBE functional[25] and the cc-pVDZ basis set.[26] The other atoms of the model system (a total of 85 atoms in the model 1 and 86 atoms in the model 2) were grouped into 642 effective fragments whose positions (except for the fragments at the periphery of the protein macromolecule) were optimized along with the atoms of the quantum subsystem against the total energy minimum. The energies and forces in the MM subsystem were calculated using the AMBER force field parameters.[27]

Results and Discussion

In the model 1, we assumed that the side chain of Glu222 was initially unprotonated, and the chromophore exists in the neutral form with the proton at the O_Y center. However, the geometry optimization yielded a minimum on the potential energy surface for the ground state (Fig. 2, *a*), which is characterized by the proton transfer from the initial structure along the hydrogen bond network O_Y—W—Ser205—Glu222. As a result, a structure containing the chromophore in the anionic state and the protonated side chain of Glu222 was obtained without substantial conformational changes of the protein macromolecule. The latter structure corresponds to the structure in the excited state in the scheme reported earlier.[9]

Figure 2 gives the calculated distances between the heavy atoms and the distances measured in the crystallographic structures 1EMA and 1GFL. It is unreasonable to include the oxygen centers of the CO_2 group of Glu222 in this comparison. Hence, the results of measurements from the C_δ center of the amino acid residue Glu222 to the oxygen atom of Cro(Thr65) and the nitrogen atom of the chromophore are given. There is also little point in comparing the distances to the oxygen atom of the hydroxyl group of Cro(Thr65) with the corresponding distances in the structure 1GFL, because the latter is the structure of the protein containing the Cro(Ser65) fragment.

The results obtained for the model 2 (see Fig. 2, *b*) are very convincing. Thus, all molecular groups are well saturated with hydrogen bonds. Except for the positions of the protons and the oxygen atoms of the CO_2 group of Glu222, the equilibrium geometric configurations in the models 1 and 2 have no substantial differences.

A comparison of the equilibrium geometric parameters of the model systems calculated by the QM/MM method with the crystallographic structures of GFP shows that the coordinates of the heavy atoms are in reasonable agreement. In spite of the fact that the coordinates of the structure 1EMA were used for the optimization, the calculated atomic distances are even in better agreements with the data for the structure 1GFL (see Fig. 2).

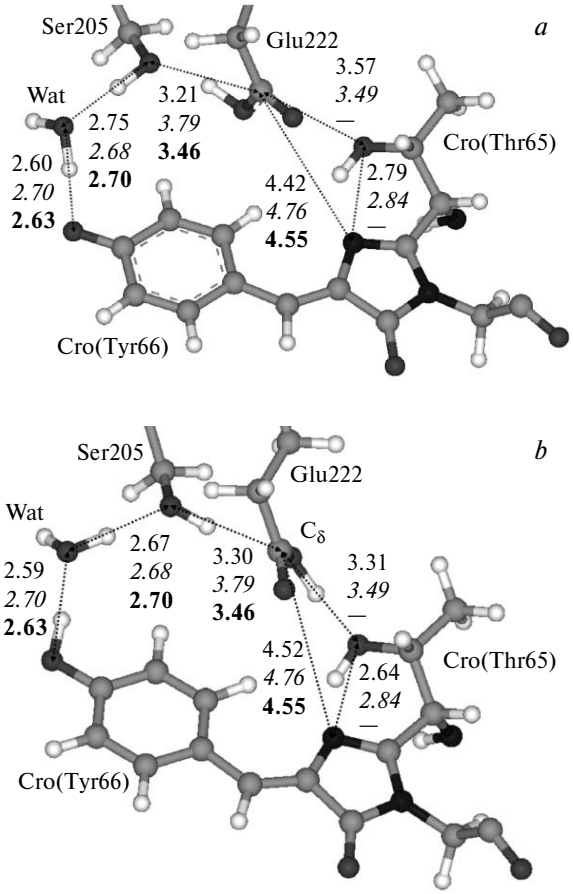

Fig. 2. Structure of the chromophore-containing region in the model containing the initially unprotonated residue Glu222 (model 1) (*a*) and in the model containing the protonated residue Glu222 (model 2) (*b*). The distances between the heavy atoms (Å) were determined from the results of calculations (upper line), the crystal structure with the PDB entry code 1EMA (middle line), and the crystal structure with the PDB entry code 1GFL containing the residue Ser65 (lower line).

Thus, an important result obtained based on the model 1 (see Fig. 2) is that, even on the assumption of the neutral state of the chromophore (HO_Y, N, O_X) and the unprotonated state of the side chain of Glu222, this system in the ground state is characterized by a minimum on the potential energy surface corresponding to the proton transfer along the chromophore—water—Ser205—Glu222 hydrogen bond system. In a recent review,[28] it was emphasized that the results obtained by molecular dynamics calculations,[29–33] in which the potential surfaces have minima for the structures with a proton located at the chromophore in the ground state and at Glu222 in the excited state, are very helpful in the accepted scheme of photoinduced transitions between the A form (the neutral state of the chromophore) and the B form (the anionic state of the chromophore) of GFP.[9] According to our data, the proton can be located at Glu222, resulting in the anionic form of the chromophore $(O_Y, N, O_X)^-$ in the ground state as well.

The results obtained with the use of the model 2 (see Fig. 2, *b*) are more important. The perfect hydrogen bond network obtained in the calculations is involved in the linking of two potential protonation centers of the chromophore (O_Y and N) in the protein: O_Y—W—Ser205——Glu222—Cro(Thr65)—N. Therefore, we cannot rule out the hypothesis[10–12] about the involvement of not only the neutral and anionic forms, but also of the cationic form of the chromophore $(HO_Y, HN, O_X)^+$ in the phototransformations of GFP. In the latter case, it is sufficient that the proton be located at the O_Y center and the proton be transferred along the chain from Glu222 through CroThr65 to the N center. As an alternative, the proton can be transferred along the total chain O_Y—W—Ser205—Glu222—— Cro(Thr65)—N, resulting in the formation of the zwitterionic form of the chromophore (O_Y^-, HN^+, O_X). The other structures discussed[12] where the proton is located at the O_X center, *viz.*, (O_Y, N, HO_X), $(HO_Y, N, HO_X)^+$, are $(O_Y, HN, HO_X)^+$, are unlikely in the protein due to the close proximity of O_X and the positively charged residue Arg96, but these structures can play a role in solution.

Hence, the scheme of phototransformations in GFP that assumes the light absorption by the chromophore in the neutral or anionic form and the fluorescence in the anionic form involving the necessary proton transfer in the excited state from the neutral form to Glu222[9] is probably not the only one. The equilibrium between the cationic and zwitterionic forms of the chromophore in the protein can be considered as equally probable with the equilibrium between the neutral and anionic forms, thus explaining the photodynamics of GFP.

This study was partially financially supported by the Russian Foundation for Basic Research (Project Nos 07-03-00059 and 08-03-91104-AFGIR) and the Russian Academy of Sciences (Program of the Presidium of the Russian Academy of Sciences "Molecular and Cell Biology").

We thank the Belarusian—Russian Union State (SKIF-GRID research-technical program), the Interregional Computer Center of the Tomsk State University, the Scientific-Research Computer Center of the M. V. Lomonosov Moscow State University, and the Interdepartmental Supercomputer Center of the Russian Academy of Sciences for providing the computer time on SKIF-Cyberia, SKIF-MGU, and MVS-100K supercomputers.

References

1. R. Y. Tsien, *Ann. Rev. Biochem.*, 1998, **67**, 509.
2. M. Zimmer, *Chem. Rev.*, 2002, **102**, 759.
3. Yu. A. Labas, A. V. Gordeeva, A. F. Fradkov, *Priroda* [*Nature*], 2003, **3**, 33 (in Russian).

4. M. Ormö, A. B. Cubitt, K. Kallio, L. A. Gross, R. Y. Tsien, S. J. Remington, *Science*, 1996, **273**, 1392.
5. F. Yang, L. G. Moss, G. N. Phillips, *Nat. Biotech.*, 1996, **14**, 1246.
6. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, *Nucleic Acids Res.*, 2000, **28**, 235.
7. M. Chattoraj, B. A. King, G. U. Bublitz, S. G. Boxer, *Proc. Nat. Acad. Sci.*, 1996, **93**, 8362.
8. H. Lossau, A. Kummer, R. Heinecke, F. Pollinger-Dammer, C. Kompa, G. Bieser, T. Jonsson, C. M. Silva, M. M. Yang, D. C. Youvan, M. E. Michel-Beyerle, *Chem. Phys.*, 1996, **213**, 1.
9. K. Brejc, T. K. Sixma, P. A. Kitts, S. R. Kain, R. Y. Tsien, M. Ormo, S. J. Remington, *Proc. Nat. Acad. Sci.*, 1997, **94**, 2306.
10. A. A. Voityuk, M. E. Michel-Beyerle, N. Rösch, *Chem. Phys. Lett.*, 1997, **272**,162.
11. A. A. Voityuk, M. E. Michel-Beyerle, N. Rösch, *Chem. Phys.*, 1998, **231**, 13.
12. A. A. Voityuk, A. D. Kummer, M. E. Michel-Beyerle, N. Rösch, *Chem. Phys.*, 2001, **269**, 83.
13. J. El. Yazal, F. G. Prendergast, D. A. Shaw, Y. P. Pang, *J. Am. Chem. Soc.*, 2000, **122**, 11411.
14. C. Scharnagl, R. A. Raupp-Kossmann, *J. Phys. Chem. B*, 2004, **108**, 477.
15. T. Laino, R. Nifosм, V. Tozzini, *Chem. Phys.*, 2004, **298**, 17.
16. R. B. Zhang, M. T. Nguyen, A. Ceulemans, *Chem. Phys. Lett.*, 2005, **404**, 250.
17. A. Warshel, M. Levitt, *J. Mol. Biol.*, 1976, **109**, 227.
18. B. Grigorenko, A. Savitsky, I. Topol, S. Burt, A. Nemukhin, *Chem. Phys. Lett.*, 2006, **424**, 184.
19. B. Grigorenko, A. Savitsky, I. Topol, S. Burt, A. Nemukhin, *J. Phys. Chem. B*, 2006, **110**, 18635.
20. M. A. L. Marques, X. Lopez, D. Varsano, A. Castro, A. Rubio, *Phys. Rev. Lett.*, 2003, **90**, 258101.
21. A. Sinicropi, T. Andruniow, N. Ferre, R. Basosi, M. Olivucci, *J. Am. Chem. Soc.*, 2005, **127**, 11534.
22. J. Hasegawa, K. Fujimoto, B. Swerts, T. Miyahara, H. Nakatsuji, *J. Comput. Chem.*, 2007, **28**, 2443.
23. B. L. Grigorenko, A. V. Nemukhin, I. A. Topol, S. K. Burt, *J. Phys. Chem. A.*, 2002, **106**, 10663.
24. A. V. Nemukhin, B. L. Grigorenko, I. A. Topol, S. K. Burt, *J. Comput. Chem.*, 2003, **24**, 1410.
25. J. P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. B*, 1996, **77**, 3865.
26. T. H. Dunning, *J. Chem. Phys.*, 1989, **90**, 1007.
27. W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, Jr., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman, *J. Am. Chem. Soc.*, **117**, 5179.
28. P. J. Tonge, S. R. Meech, *J. Photochem. Photobiol. A. Chem.*, 2009, **205**, 1.
29. O. Vendrell, R. Gelabert, M. Moreno, J. M. Lluch, *Chem. Phys. Lett.*, 2004, **396**, 202.
30. O. Vendrell, R. Gelabert, M. Moreno, J. M. Lluch, *J. Am. Chem. Soc.*, 2006, **128** , 3564.
31. O. Vendrell, R. Gelabert, M. Moreno, J. M. Lluch, *J. Chem. Theor. Comput.*, 2008, **4**, 1138.
32. O. Vendrell, R. Gelabert, M. Moreno, J. M. Lluch, *J. Phys. Chem. B*, 2008, **112**, 13443.
33. O. Vendrell, R. Gelabert, M. Moreno, J. M. Lluch, *J. Phys. Chem. B*, 2008, **112**, 5500.

Received July 13, 2009