Article

Estimating Orientation Factors in the FRET Theory of Fluorescent Proteins: The TagRFP-KFP Pair and Beyond

Maria Khrenova,[1] Igor Topol,[2,*] Jack Collins,[2] and Alexander Nemukhin[1,3]
[1]Department of Chemistry, Lomonosov Moscow State University, Moscow, Russian Federation; [2]Advanced Biomedical Computing Center, Information Systems Program, Leidos Biomedical Research, Inc., Frederick National Laboratory for Cancer Research, Frederick, Maryland; and [3]Emanuel Institute of Biochemical Physics, Moscow, Russian Federation

ABSTRACT The orientation factor κ^2, one of the key parameters defining Förster resonance energy transfer efficiency, is determined by the transition dipole moment orientations of the donor and acceptor species. Using the results of quantum chemical and quantum mechanical/molecular mechanical calculations for the chromophore-containing pockets in selected colored proteins of the green fluorescent protein family, we derived transition dipole moments corresponding to the $S_{0,min} \rightarrow S_1$ excitation for green fluorescent protein, red fluorescent protein (TagRFP), and kindling fluorescent protein, and the $S_{1,min} \rightarrow S_0$ emission for TagRFP. These data allowed us to estimate κ^2 values for the TagRFP-linker-kindling fluorescent protein tetrameric complex required for constructing novel sensors.

INTRODUCTION

Förster resonance energy transfer (FRET) is a radiationless process that occurs between a donor molecule in the excited state and an acceptor molecule in the ground state. The importance of FRET in modern biomolecular research is hard to overestimate (1–6). For this work, we concentrate on the application of FRET to the colored protein systems of the green fluorescent protein (GFP) family, well-known molecular markers used in living cells (7–9). More specifically, we are interested in theoretical evaluation of the FRET parameters in these systems, which can be used when constructing novel sensors (10–12), as well as in the interpretation of complicated spectral features of fluorescent proteins.

FRET parameters can be referred to individual properties of donor and acceptor species, namely, absorption and emission band shapes, extinction coefficients, quantum yields, and fluorescence lifetimes, as well as to quantities that are defined by donor and acceptor interactions, including R, the distance between the donor and acceptor, and the orientation factor, κ^2. The latter quantity is the only parameter defining the Förster radius R_0 (the distance at which 50% of the energy is transferred to the acceptor) that cannot be determined experimentally. In turn, the transfer efficiency E is directly related to R and R_0

$$E = 1 \Big/ \left[1 + (R/R_0)^6\right].$$

To estimate κ^2 the following formula is applied

$$\kappa^2 = [\mathbf{d} \cdot \mathbf{a} - 3(\mathbf{d} \cdot \mathbf{r}_{da})(\mathbf{a} \cdot \mathbf{r}_{da})]^2, \quad (1)$$

where $\mathbf{d}$ and $\mathbf{a}$ are transition dipole moments of the donor and acceptor, respectively, and $\mathbf{r}_{da}$ is a normalized vector connecting the two dyes. The value of κ^2 can range between 0 and 4, and a strongly simplified approach used to estimate this value is to average the donor and acceptor orientations of freely rotated dyes, leading to a mean value $\kappa^2 = 2/3$. However, this approximation may be inappropriate for the chromophores trapped inside protein barrels of fluorescent proteins.

To evaluate κ^2 accurately, using electronic structure principles, one must optimize the ground- and excited-state equilibrium geometry parameters of relevant models, and then calculate the transition dipole moment (TDM). The practical realization of this strategy requires considerable effort: large molecular model systems mimicking chromophore-containing pockets should be considered, optimizing excited-state geometry parameters is an especially difficult task, and calculating electronic transition properties to a reasonable level of accuracy is an issue of debate.

The primary goal of this study was to calculate TDMs to compute the κ^2 parameters of a specific FRET system, composed of two colored proteins in the GFP family, red fluorescent protein (TagRFP) and kindling fluorescent protein (KFP), connected by a flexible peptide linker. KFP, serving as an acceptor partner in this pair, is the Ala-143-Gly mutant of natural chromoprotein asFP595 from *Anemonia sulcata* (13–15). Brightly fluorescent TagRFP is the donor in this particular FRET pair, and is a genetic construct derived from the wild-type protein found in the sea anemone *Entacmaea quadricolor* (16,17). Employing FRET between these two proteins (connected by an appropriate peptide linker) opens new avenues for monitoring biological and physiological mechanisms. Recently, a FRET system

Submitted July 7, 2014, and accepted for publication November 6, 2014.
*Correspondence: igor.topol@fnlcr.nih.gov
Editor: Nathan Baker.

0006-3495/15/01/0126/7 $2.00
http://dx.doi.org/10.1016/j.bpj.2014.11.1859

composed of TagRFP and KFP connected by a linker built from 23 amino acids was studied experimentally as a promising novel diagnostic tool (12). Proteolic cleavage of the linker between brightly fluorescent donor TagRFP and dark acceptor KFP resulted in a substantial increase of donor fluorescence intensity and lifetime. The TagRFP-linker-KFP fusion contains the caspase-3 recognition site DEVD, which is cleaved upon caspase-3 activation.

By analyzing the results of quantum chemical (QC) and quantum mechanical/molecular mechanical (QM/MM) approaches for the corresponding chromophores within their protein environments, we deduce that the TDMs for $S_{0,min} \rightarrow S_1$ in the KFP transition and $S_{1,min} \rightarrow S_0$ in the TagRFP transition are characterized by similar orientations relative to the chromophore. Ansbacher et al. (18) have hypothesized a similar assumption by considering bare chromophores extracted from their protein matrices. Combining this fairly simple rule with knowledge of protein structure allows us to interpret the FRET efficiency for the tetrameric TagRFP-linker-KFP system and the dimeric KillerRed variant.

MATERIALS AND METHODS

Molecular models of the chromophore-containing pockets of fluorescent proteins from the GFP family were constructed for this work based on available Protein Data Bank (PDB) crystal structures. The following members were considered: GFP (the relevant crystal structure PDB ID 1EMA (19)), KFP (PDB ID 1XQM (14)), and TagRFP (PDB ID 3M22 (17)). The chemical formulae of the corresponding chromophore molecules in the anionic protonation state are presented in Fig. 1.

The GFP chromophore shown in Fig. 1 *a* represents the *cis* conformer, whereas the TagRFP and KFP chromophores (Fig. 1, *b* and *c*) are shown in the *trans* conformation, as they occur in their protein matrices.

In all cases, fairly large model systems composed of molecular groups from crystal structures were selected for equilibrium geometry parameter calculations. After addition of hydrogen atoms, assuming a customary assignment of protons to the polar residues of Asp, Glu (negatively charged), Lys and Arg (positively charged), optimization of coordinates in model systems was performed using either QM/MM or QC methods. The QM/MM approach was applied to model systems that included almost the entire protein and the QC approach to systems mimicking chromophore-containing pockets. In the latter case, the coordinates of the outermost atoms at the periphery of the clusters were kept frozen, as they were in the corresponding crystal structures.

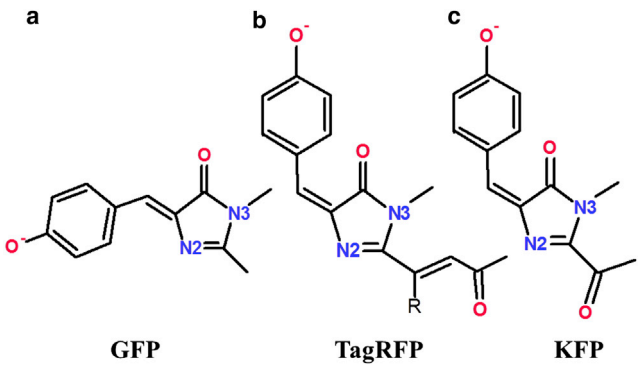

FIGURE 1 Chemical structures of the chromophores from the fluorescent proteins: (*a*) GFP, (*b*) TagRFP, (*c*) KFP. Only the oxygen and nitrogen atoms are designated. To see this figure in color, go online.

Minimum-energy, ground-state equilibrium geometry parameters in the lowest singlet electronic state S_0 were calculated using density functional theory (DFT) approaches in quantum subsystems.

In the case of TagRFP model systems, the following approaches were applied. For the QC calculations, we constructed a model cluster (called the GS-cluster) composed of the chromophore, the side chains of Leu-13, Gln-39, Ala-59 Thr-60, Ser-61, Phe-62, Ser-66, Arg-67, Arg-92, Gln-106, Tyr-117, Asn-143, Glu-145, Ser-158, Met-160, Phe-174, His-197, Leu-199, Gln-213, and Glu-215, and 11 water molecules. After adding hydrogen atoms, the system included 326 atoms in total. Optimization of geometry coordinates was performed in the B3LYP/6-31G* approximation while keeping the C_α atoms of the involved residues frozen. For QM/MM ground state geometry optimization, the protein system was subdivided into QM and MM parts. The QM subsystem included the chromophore, the nearest water molecules, and the side chains of Arg-67, Arg-92, Asn-143, Glu-145, Ser-158, His-197, and Glu-215, whereas the rest of the protein was treated as the MM subsystem. The DFT approach, with functional PBE0 and cc-pVDZ basis sets, was applied in QM, whereas AMBER force field parameters were applied in MM.

For KFP, the QM subsystem included the chromophore, the nearest water molecules, and the side chains of Lys-67, Arg-92, Glu-145, Ser-158, His-197, and Glu-215; the rest of the protein was treated as a MM subsystem. The DFT approach, with functional PBE0 and cc-pVDZ basis sets, was applied for calculations in QM, and AMBER force field parameters in MM.

The model system mimicking GFP was constructed following motifs of the crystal structure PDB ID 1EMA (19). This entry corresponds to the protein with a point mutation Ser-65-Thr compared to the wild-type GFP. The structure was augmented by hydrogen atoms, and the side chain of the residue at position 65 was set to Ser, as in the wild-type. The atomic coordinates of the model system were then optimized by the QM/MM technique, using the flexible effective fragment potential method (20,21) with PBE0/6-31G* in the QM part and the AMBER force field parameters in the MM part. The chromophore and the side chains of Glu-222, Arg-96, Ser-205, and His-148, as well as the two nearest water molecules were assigned to the quantum subsystem. This model structure refers to the so-called I-form of GFP (22).

Excited state (S_1) geometry coordinates for TagRFP models were optimized using the configuration interaction singles (CIS) approximation and the complete active space self-consistent field (CASSCF) approach. To find coordinates of the minimum energy structures in the S_1 excited state, we carried out QC calculations for molecular clusters that were constructed as follows. First, the model GS-cluster was reduced and its coordinates optimized for the second singlet S_1 state using the CIS/6-31G* approximation. We designated the corresponding 241-atom system ES-cluster1. Another model, ES-cluster2, was constructed by taking the coordinates of the QM/MM ground state structure, i.e., removing the MM part from the QM/MM model. Optimization of excited-state geometry parameters was carried out using the CASSCF procedure, keeping frozen the coordinates of the outermost atoms in the cluster. More specifically, we applied the CASSCF(6/6) method by distributing six electrons over six orbitals, and obtained a true solution for the second root in the corresponding configuration interaction expansion (in other words, the state-specific CASSCF solution was constructed). To select the active orbitals we first performed CIS calculations and choose the π-orbitals whose population changed upon the $S_0 \rightarrow S_1$ excitation. In this particular case three double occupied and three vacant orbitals in the ground state were included to the active space.

TDMs for the $S_{0,min} \rightarrow S_1$ transitions in GFP, TagRFP, and KFP, and for the $S_{1,min} \rightarrow S_0$ transition in TagRFP were computed using ab initio CIS or semiempirical ZINDO (23) approximations. We also report the ZINDO results for the vertical transition energies and the corresponding optical band maxima wavelengths. We previously showed (24–29) that such a

computational strategy, namely, estimating transition energies at the ZINDO level using a chromophore's coordinates optimized by DFT calculations, yields results that are in remarkably good agreement with experimental data.

QC calculations were carried out using Gaussian-03 (30), Firefly (31), and ORCA (32) programs. QM/MM calculations were performed using the NWChem program (33), as well as a combination of the GAMESS(US) (34) and TINKER (35) software packages, as described in (20,21). Additional details are presented below, with the results from each system.

RESULTS

Structures, absorption, and emission spectra of the TagRFP model systems

Geometry configurations of the chromophore-containing pockets in TagRFP were optimized in the ground electronic state in QC calculations for a model cluster and in QM/MM calculations for a larger model encompassing the entire protein. Remarkably, both approaches resulted in very close equilibrium configurations of the chromophore and in indistinguishable orientations of the TDM.

A view of the most important part of the model cluster, showing the chromophore and the nearest residues in the chromophore environment, is illustrated in Fig. 2. We specify several key intermolecular distances obtained from our calculations (*top and middle values* in Fig. 2) to show that a good agreement with the parent crystal structure PDB ID 3M22 (*bottom values*), solved to 2.2 Å resolution (17), is obtained. Specific equilibrium geometry parameters of the chromophore molecule in the ground electronic state are provided in Table S1 in the Supporting Material. The chromophore structure is essentially nonplanar in the protein matrix.

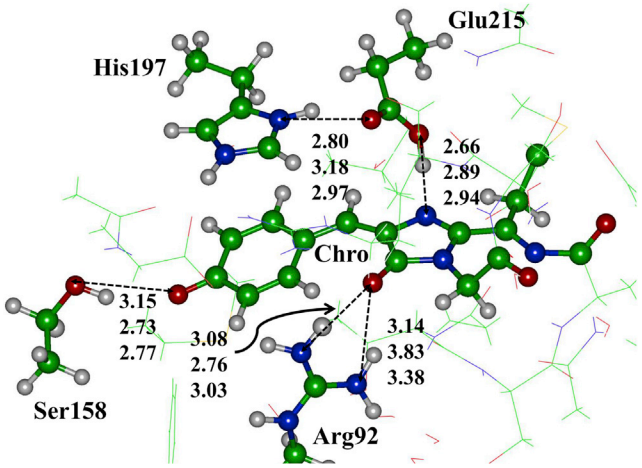

FIGURE 2 A part of the molecular cluster for the chromophore-containing pocket of TagRFP. Selected distances between heavy atoms (in Å) are arranged as follows: QC cluster model (*top*), QM/MM model (*middle*), and crystal structure PDB ID 3M22 (*bottom*). In this and forthcoming figures, carbon atoms are colored in green, oxygen in red, and nitrogen in blue. To see this figure in color, go online.

The experimental absorption band maximum for TagRFP corresponds to excitation energy of 2.23 eV or a wavelength of 555 nm (16,17). The computational results, using ZINDO at the DFT-optimized geometry for the GS-cluster model, are 2.25 eV and 551 nm; the results of the QM/MM model are 2.41 eV and 515 nm.

The equilibrium geometry parameters of the chromophore molecule in the S_1 excited state are provided in Table S2.

The experimental emission band maximum for TagRFP corresponds to an energy of 2.12 eV or a wavelength of 584 nm (16,17). Computational results (ZINDO) for the ES-cluster1 model are 2.21 eV and 559 nm, and 2.19 eV and 565 nm for the ES-cluster2 model.

Fig. 3 illustrates computed TDM directions for $S_{0,min} \rightarrow S_1$ absorption (*upper graph*) and for $S_{1,min} \rightarrow S_0$ emission (*lower graph*). Remarkably, the sets of different theoretical models used in this work (QM/MM and GS-cluster for the ground state, and ES-cluster1 and ES-cluster2 for the excited state) resulted in practically indistinguishable predictions for TDM directions. As seen in Fig. 3, absorption and emission TDMs can be viewed as vectors connecting the phenolic oxygen (O) and the imidazolinone nitrogen (N3). The absorption and emission transition dipole moments can often be considered to be colinear. In particular, similarities in absorption and emission TDM directions in fluorescent proteins were discussed in (36) by analyzing results from experiments on crystals.

A final comment in this subsection refers to the heuristic rule for guessing the TDM direction in fluorescent proteins, as formulated by Ansbacher et al. (18). Apparently, an imaginary line connecting the two farthest atoms in the

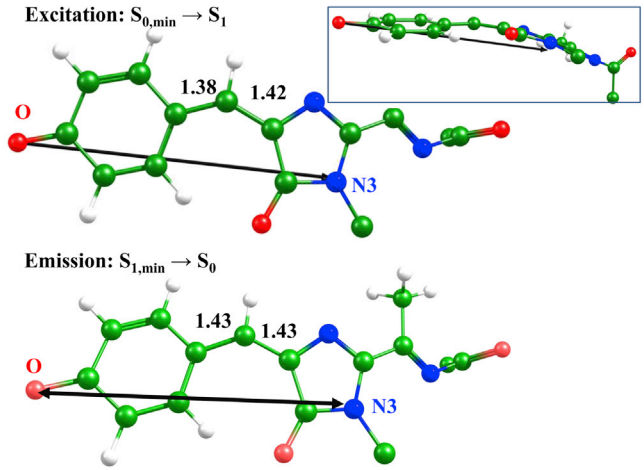

FIGURE 3 Structures of the TagRFP chromophore and TDM orientations. Upper graph refers to the S_0 minimum and the $S_{0,min} \rightarrow S_1$ transition; bottom graph refers to the S_1 minimum and the $S_{1,min} \rightarrow S_0$ transition. Distances (in Å) illustrate the bond-length changes at the methylene bridge upon excitation. TDM vectors connecting the phenolic O and N3 are indicated by double-arrow lines. The inset in the upper-right corner illustrates that the chromophore is essentially nonplanar. To see this figure in color, go online.

π-conjugated system of the molecule suggested in (18) deviates here from our prediction: in our calculations the TDM vector connects the O and N3 atomic centers, but not the two farthest oxygen atoms (Fig. 3). However, it should be noted that the authors of (18) considered mainly the *cis* form of chromophores; here, the chromophore is in the *trans* configuration.

Structure and absorption spectra of KFP and GFP model systems

A view of the most important part of the model cluster, showing the chromophore and the key residues in the chromophore environment, is illustrated in Fig. 4. Equilibrium geometry parameters of the chromophore molecule are provided in Table S3. Similar to TagRFP, the chromophore structure is essentially nonplanar in the protein matrix.

The experimental absorption band maximum in KFP corresponds to excitation energy of 2.18 eV or a wavelength of 568 nm (13). Computational results (ZINDO at the DFT-optimized geometry) for this QM/MM model are 2.31 eV or 535 nm.

The computed TDM direction for the $S_{0,min} \rightarrow S_1$ excitation in KFP is also shown in Fig. 4; apparently, it is practically the same as that calculated for TagRFP (Fig. 3).

Finally, we considered the parent member of this protein family, GFP itself, with the anionic chromophore. We illustrate this model as Fig. S1, and Table S4 lists the Cartesian coordinates of the GFP chromophore. As in all previous examples, we estimated the TDM at the equilibrium geometry configuration by using ZINDO at the DFT-optimized geometry.

In this particular case, we were able to directly compare our computational data (Fig. 5) with the TDM calculation results for the bare fluorescent protein chromophores reported in (18). The authors suggested characterizing the TDM direction by a deviation angle from the carbonyl group of the imidazolinone ring. This direction for the anionic GFP chromophore, estimated in (18) as 70°, is plotted in Fig. 5 (*dashed*, *double-arrow line*). Our TDM calculations again revealed a vector between the phenolic O and the N3 of the chromophore (*solid black line*). At a glance (Fig. 5), both theoretical approaches result in fairly consistent conclusions.

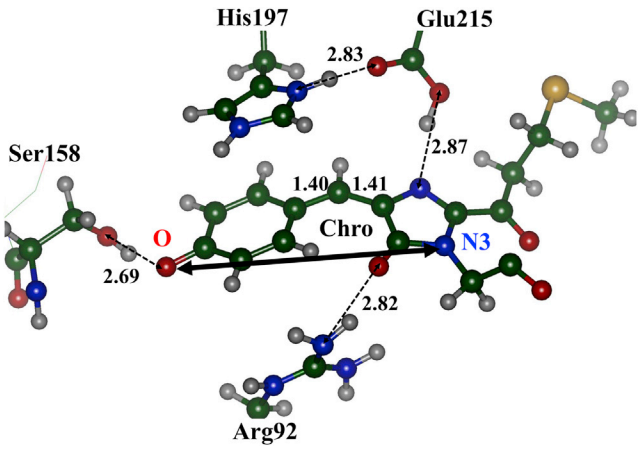

FIGURE 4 A part of the molecular cluster that includes the chromophore-containing domain of KFP. Selected distances between heavy atoms are given in Å. The TDM direction for the $S_{0,min} \rightarrow S_1$ excitation is shown as a solid, double-arrow line. To see this figure in color, go online.

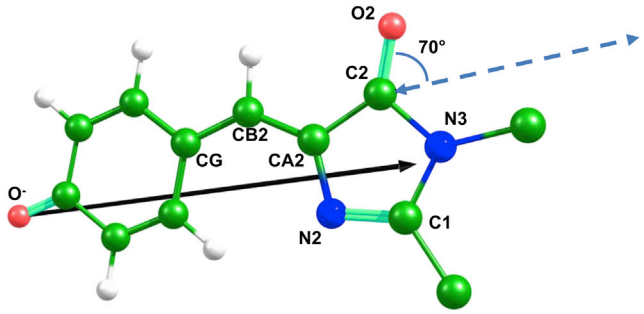

FIGURE 5 Computed TDM for GFP with the anionic chromophore. The result of our approach is shown as a solid black vector. The vector estimated in (18) is indicated by a dashed, double-arrow line. To see this figure in color, go online.

The following two subsections illustrate applications of the proposed rule to orient the TDM in fluorescent protein chromophores in the O-N3 direction for the complex TagRFP-linker-KFP and KillerRed systems.

Orientation factors for the tetrameric TagRFP-linker-KFP system

The model considered in (12) for the tetrameric TagRFP-linker-KFP system was constructed using molecular docking as a tool to arrange eight protein barrels linked by peptide chains. Analysis of the conformational flexibility of such a system is beyond the scope of this work (see (12)). Of importance, the model described in (12) did not take into account the orientation factors of donor and acceptor chromophores when estimating FRET efficiency; an averaged value of $\kappa^2 = 2/3$ was used. Here, we calculated specific orientation factors using the established rule to orient TDMs in the corresponding O-N3 directions. Because TagRFP species are located symmetrically around the core KFP tetramer complex, we considered a reduced system composed of single TagRFP species connected to a KFP tetramer by a flexible linker (Fig. 6). After photoexcitation, the TagRFP donor (D) can either fluoresce or transfer energy to one of four acceptors (A_i).

Table 1 shows distances, computed orientation factors, and FRET rate constants corresponding to all possible pairs. We can describe processes occurring after TagRFP energy absorption in terms of parallel reactions: energy can be either transferred to one of four acceptors or emitted from

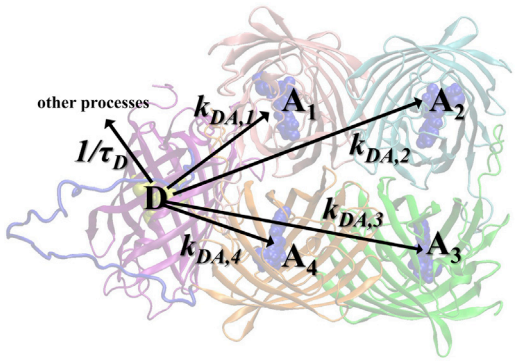

FIGURE 6 Structure of the TagRFP-linker-KFP (*tetramer*) complex. The chromophores of the donor, D (from TagRFP), and four acceptors, A_i (from KFP), are represented by the solid balls. To see this figure in color, go online.

TagRFP with the corresponding rate constants $k_{DA,i}$ ($i = 1 \div 4$) or $1/\tau_D$. The probability of occurrence of each pathway of energy transfer is proportional to the corresponding reaction rate constant; therefore probability (or efficiency) of FRET is estimated as follows:

$$E = \frac{\sum_{i=1}^{4} k_{DA,i}}{\sum_{i=1}^{4} k_{DA,i} + 1/\tau_D}.$$

Application of this formula leads to the calculated FRET efficiency for this particular system as 86%. The latter value indicates, first, that the conformation presented in (12) is not the major one for this sensor, and second, a high efficiency in this particular conformation is due to individual properties of donor and acceptor species despite the relatively low orientation factors κ^2. Therefore, a proper modification of the linker may affect a distance distribution between donor and acceptor molecules resulting in enhancement of sensor properties.

Evaluating the data shown in Table 1 immediately makes it clear that contributions to FRET efficiency are not only due to differences in distances, but also due to different orientation factors. We reiterate that an estimated κ^2 value, $2/3 = 0.67$, is typically used when evaluating FRET efficiency.

Orientation factor for the KillerRed dimer

An increased interest in one of the genetically derived red fluorescent proteins, KillerRed (37), is explained by its cytotoxicity in living cells. A series of structural and spectral studies of this protein has been completed, among which the recent work described in (38) is worth particular attention. The authors designed a Val-44-Ala mutant of KillerRed and noticed that its composition as a dimer demonstrated an intradimeric FRET effect between neutral (*green*) and anionic (*blue*) chromophores. The authors assume that the green emission peak may be explained by FRET and point to the interesting head-to-tail orientation of two chromophores in the dimer (Fig. 7).

We applied a heuristic rule to orient the TDM (from O to N3) in both chromophores in the model system, based on the crystal structure of the dimer (PDB ID 4B30), and calculated the κ^2 for such a construct. The resulting κ^2 value of 2.69 far exceeds the averaged $2/3 = 0.67$ estimate, which, in conjunction with the short distance between chromophores, explains the observed considerably FRET efficiency.

TABLE 1 Interchromophore distances (*R*), computed orientation factors (κ^2), and rate constants ($k_{DA,i}$) for the TagRFP-linker-KFP (tetramer) complex shown in Fig.6

Parameter	D-A_1	D-A_2	D-A_3	D-A_4
R, Å	38.5	63.1	57.8	37.7
κ^2	0.89	0.62	0.02	0.06
$k_{DAi,}$ ns^{-1}	2.33	0.09	<0.01	0.20

DISCUSSION

Despite the structural similarities between colored proteins in the GFP family, it is far from obvious that they should share a common TDM orientation. In fact, the chromophores of these proteins are not identical; they are characterized by varying numbers of conjugated π-bonds and noticeable differences in geometry configurations, including, in particular, the *cis* and *trans* isomers (Scheme 1). However, the calculations presently used, as well as the recent work by Ansbacher et al. (18), encourage us to assume that fairly simple rules may provide reasonably good estimates for TDM orientations within this protein family. The authors of (18) used the time-dependent (TD)-DFT method to calculate the TDM in S_0-S_1 absorption for several fluorescent protein chromophores for gas phase and solution

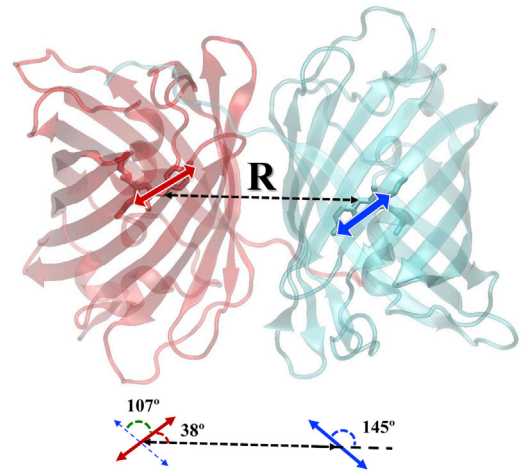

FIGURE 7 Dimeric structure of the Val-44-Ala mutant of KillerRed by motifs of PDB ID 4B30 (38). TDM directions are indicated by solid, double-arrow lines. The inset at the bottom shows the orientation angles of the TDM vectors of the crystal geometry. To see this figure in color, go online.

conditions (in the latter case, within the polarized continuum model). They borrowed the starting coordinates of the chromophore molecules from the corresponding crystal structures and reoptimized the coordinates within the DFT B3LYP/6-31G* approach in vacuo. As expected, the ground-state geometry structures of the chromophores turned out to be planar. Next, TDMs were computed by using the B3LYP/6-31+G* approach in TD-DFT. For GFP, the only system considered here and in (18), and for which direct comparison is possible, the results are consistent.

To our knowledge, novel features of our work are related to 1), modeling entire chromophore-containing pockets, not just bare chromophores (i.e., a realistic protein environment is considered, resulting in nonplanar geometry structures); 2), using other electronic structure theory methods beyond TD-DFT; and 3), calculating the S_1-S_0 transition moment directly for emission in TagRFP.

For obtaining fairly simple TDM direction estimates for absorption and emission, our recommendation is to consider the vector connecting the phenolic O and the N3 of the chromophore (Scheme 1). Applying this rule, we are able to explain the efficient FRET observed in the tetrameric TagRFP-linker-KFP system (12) and in the dimeric form of a variant of the KillerRed protein (38).

CONCLUSIONS

We report the TDM and orientation factor calculation results for TagRFP-KFP, the pair of fluorescent proteins required for construction of fusion proteins with enhanced FRET efficiency. The QM/MM and QC approaches for large molecular clusters allowed us to compute the properties of the chromophores in the protein environment. On the basis of these calculations, we deduced that TDMs for the $S_{0,min}$-S_1 and $S_{1,min}$-S_0 transitions in fluorescent proteins of the GFP family have similar orientations relative to the chromophore: the TDM direction can be estimated accurately as the vector between the phenolic O and N3 of the chromophore. Application of this simple rule to the chromophores in the tetrameric TagRFP-linker-KFP system and in the dimeric KillerRed variant allowed us to provide an interpretation of their FRET efficiencies.

SUPPORTING MATERIAL

Supporting Materials and Methods, one figure, and four tables are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(14)03071-9.

ACKNOWLEDGMENTS

We thank Prof. Alexander Savitsky for advice and discussion.

This work was supported by the Program on Molecular and Cell Biology from the Russian Academy of Sciences and by the Russian Foundation for Basic Research (project 13-03-00207). M.K. acknowledges support from the Dynasty Foundation fellowship. We acknowledge the use of supercomputer resources of the M.V. Lomonosov Moscow State University and of the Joint Supercomputer Center of the Russian Academy of Sciences. We thank the staff and administration of the Advanced Biomedical Computing Center for their support of this project. This project has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health, under contract No. HHSN261200800001E. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organization imply endorsement by the U.S. government.

REFERENCES

1. Lakowicz, J. 1999. Principles of Fluorescence Spectroscopy. Kluwer Academic and Plenum Publishers, New York.
2. Periasamy, A. 2001. Fluorescence resonance energy transfer microscopy: a mini review. *J. Biomed. Opt.* 6:287–291.
3. Margittai, M., J. Widengren, …, C. A. M. Seidel. 2003. Single-molecule fluorescence resonance energy transfer reveals a dynamic equilibrium between closed and open conformations of syntaxin 1. *Proc. Natl. Acad. Sci. USA.* 100:15516–15521.
4. Brunger, A. T., P. Strop, …, K. R. Weninger. 2011. Three-dimensional molecular modeling with single molecule FRET. *J. Struct. Biol.* 173:497–505.
5. Sindbert, S., S. Kalinin, …, C. A. M. Seidel. 2011. Accurate distance determination of nucleic acids via Förster resonance energy transfer: implications of dye linker length and rigidity. *J. Am. Chem. Soc.* 133:2463–2480.
6. Hoefling, M., N. Lima, …, H. Grubmüller. 2011. Structural heterogeneity and quantitative FRET efficiency distributions of polyprolines through a hybrid atomistic simulation and Monte Carlo approach. *PLoS ONE.* 6:e19791.
7. Zhang, J., R. E. Campbell, …, R. Y. Tsien. 2002. Creating new fluorescent probes for cell biology. *Nat. Rev. Mol. Cell Biol.* 3:906–918.
8. Zimmer, M. 2002. Green fluorescent protein (GFP): applications, structure, and related photophysical behavior. *Chem. Rev.* 102:759–781.
9. Day, R. N., and M. W. Davidson. 2009. The fluorescent protein palette: tools for cellular imaging. *Chem. Soc. Rev.* 38:2887–2921.
10. Mank, M., D. F. Reiff, …, O. Griesbeck. 2006. A FRET-based calcium biosensor with fast signal kinetics and high fluorescence change. *Biophys. J.* 90:1790–1796.
11. Rusanov, A. L., T. V. Ivashina, …, A. P. Savitsky. 2010. Lifetime imaging of FRET between red fluorescent proteins. *J. Biophotonics.* 3:774–783.
12. Savitsky, A. P., A. L. Rusanov, …, A. V. Nemukhin. 2012. FLIM-FRET imaging of caspase-3 activity in live cells using pair of red fluorescent proteins. *Theranostics.* 2:215–226.
13. Chudakov, D. M., V. V. Belousov, …, K. A. Lukyanov. 2003. Kindling fluorescent proteins for precise in vivo photolabeling. *Nat. Biotechnol.* 21:191–194.
14. Quillin, M. L., D. M. Anstrom, …, S. J. Remington. 2005. Kindling fluorescent protein from *Anemonia sulcata*: dark-state structure at 1.38 A resolution. *Biochemistry.* 44:5774–5787.
15. Henderson, J. N., and S. J. Remington. 2006. The kindling fluorescent protein: a transient photoswitchable marker. *Physiology (Bethesda).* 21:162–170.
16. Merzlyak, E. M., J. Goedhart, …, D. M. Chudakov. 2007. Bright monomeric red fluorescent protein with an extended fluorescence lifetime. *Nat. Methods.* 4:555–557.
17. Subach, O. M., V. N. Malashkevich, …, V. V. Verkhusha. 2010. Structural characterization of acylimine-containing blue and red chromophores in mTagBFP and TagRFP fluorescent proteins. *Chem. Biol.* 17:333–341.

18. Ansbacher, T., H. K. Srivastava, …, A. Shurki. 2012. Calculation of transition dipole moment in fluorescent proteins—towards efficient energy transfer. *Phys. Chem. Chem. Phys.* 14:4109–4117.

19. Ormö, M., A. B. Cubitt, …, S. J. Remington. 1996. Crystal structure of the *Aequorea victoria* green fluorescent protein. *Science.* 273:1392–1395.

20. Grigorenko, B. L., A. V. Nemukhin, …, S. K. Burt. 2002. 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.* 106:10663–10672.

21. Nemukhin, A. V., B. L. Grigorenko, …, S. K. Burt. 2003. Flexible effective fragment QM/MM method: validation through the challenging tests. *J. Comput. Chem.* 24:1410–1420.

22. Grigorenko, B. L., A. V. Nemukhin, …, A. I. Krylov. 2013. 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.* 135:11541–11549.

23. Zerner, M. C. 1991. *In* Reviews in Computational Chemistry, *Vol. 2*. K. B. Lipkowitz, and D. B. Boyd, editors. VCH Publishing, New York, pp. 313–366.

24. Topol, I., J. Collins, …, A. Nemukhin. 2009. On photoabsorption of the neutral form of the green fluorescent protein chromophore. *Biophys. Chem.* 145:1–6.

25. Topol, I., J. Collins, and A. Nemukhin. 2010. Modeling spectral tuning in monomeric teal fluorescent protein mTFP1. *Biophys. Chem.* 149:78–82.

26. Topol, I., J. Collins, …, A. Nemukhin. 2011. Computational strategy for tuning spectral properties of red fluorescent proteins. *Biophys. Chem.* 158:91–95.

27. Topol, I., J. Collins, …, A. Nemukhin. 2012. Modeling absorption of the kindling fluorescent protein with the neutral form of the chromophore. *Int. J. Quantum Chem.* 112:2947–2951.

28. Topol, I., J. Collins, and A. Nemukhin. 2012. Modeling structures and spectra of fluorescent proteins in the coordinate-locking cluster approach: application to the photoswitchable protein asFP595. *Computational Molecular Bioscience.* 2:83–91.

29. Nemukhin, A., I. Topol, …, M. Khrenova. 2013. Quantum chemistry in studies of fluorescent and photosensing proteins. *Int. J. Quantum Chem.* 113:1828–1832.

30. Frisch, M. J., G. W. Trucks, …, J. A. Pople. 2004. Gaussian 03, Revision C.02. Gaussian, Wallingford, CT.

31. Granovsky, A. A. Firefly version 8.0. http://classic.chem.msu.su/gran/firefly/index.html. Accessed September 15, 2014.

32. Neese, F. 2012. The ORCA program system. *WIREs Comput. Mol. Sci.* 2:73–78.

33. Valiev, M., E. J. Bylaska, …, W. A. de Jong. 2010. NWChem: a comprehensive and scalable open-source solution for large scale molecular simulations. *Comput. Phys. Commun.* 181:1477–1489.

34. Schmidt, M. W., K. K. Baldridge, …, J. A. Montgomery. 1993. General atomic and molecular electronic structure system. *J. Comput. Chem.* 14:1347–1363.

35. Ponder, J. W., and F. M. Richards. 1987. An efficient Newton-like method for molecular mechanics energy minimization of large molecules. *J. Comput. Chem.* 8:1016–1024.

36. Rosell, F. I., and S. G. Boxer. 2003. Polarized absorption spectra of green fluorescent protein single crystals: transition dipole moment directions. *Biochemistry.* 42:177–183.

37. Bulina, M. E., D. M. Chudakov, …, K. A. Lukyanov. 2006. A genetically encoded photosensitizer. *Nat. Biotechnol.* 24:95–99.

38. de Rosny, E., and P. Carpentier. 2012. GFP-like phototransformation mechanisms in the cytotoxic fluorescent protein KillerRed unraveled by structural and spectroscopic investigations. *J. Am. Chem. Soc.* 134:18015–18021.

Estimating Orientation Factors in the FRET Theory of Fluorescent Proteins: the TagRFP-KFP Pair and Beyond

Maria Khrenova, [†] Igor Topol, [‡*] Jack Collins, [‡] and Alexander Nemukhin [†§]

[†] Department of Chemistry, Lomonosov Moscow State University, Moscow, Russian Federation

[‡] Advanced Biomedical Computing Center, Frederick National Laboratory for Cancer Research, Frederick, MD, USA

[§] Emanuel Institute of Biochemical Physics, Moscow, Russian Federation

Supporting Materials

Table S1. Cartesian coordinates (Å) of the TagRFP chromophore in the ground S_0 state optimized in the QM(PBE0/cc-pVDZ)/MM(AMBER) calculations.

N -10.9220 -38.5530 23.4360
H -10.0180 -38.8090 23.8200
C -12.0720 -38.3710 24.2910
H -12.9480 -38.4840 23.6370
C -12.1340 -39.4140 25.3580
H -11.4130 -39.1850 26.1470
H -13.1240 -39.3710 25.8250
C -11.8930 -40.8450 24.9110
C -10.6120 -41.4100 25.0480
H -9.8000 -40.8310 25.4590
C -10.3790 -42.7370 24.6530
H -9.3950 -43.1670 24.7740
C -11.4180 -43.5050 24.1050
H -11.2330 -44.5260 23.8010
C -12.9410 -41.6280 24.3910
H -13.9370 -41.2270 24.3120
C -12.7020 -42.9500 23.9770
H -13.5090 -43.5460 23.5740
C -12.2100 -36.9490 24.8680
O -13.1650 -36.6420 25.5520
N -11.2690 -36.0180 24.4430
C -10.2060 -35.7350 25.1140
C -9.2560 -34.7660 24.5910
N -8.0900 -34.5630 25.1640
N -9.3540 -34.0350 23.4140
C -8.1570 -33.3480 23.2200
O -7.9780 -32.5690 22.2520
C -7.3460 -33.7510 24.3420

C -5.9850 -33.6020 24.7040
H -5.7930 -34.0550 25.6820
C -4.8650 -33.1080 24.0710
C -3.6040 -33.2090 24.7630
H -3.5940 -33.6840 25.7490
C -4.8610 -32.5120 22.7630
H -5.8070 -32.4600 22.2270
C -2.4510 -32.7300 24.2360
H -1.4970 -32.7850 24.7730
C -3.7200 -32.0200 22.2180
H -3.7130 -31.5340 21.2400
C -2.4580 -32.0700 22.9420
O -1.4270 -31.5490 22.4750
C -9.8180 -36.3830 26.4170
H -10.1580 -37.4260 26.4080
H -8.7260 -36.3790 26.5040
C -10.4150 -35.6870 27.6020
H -9.9780 -34.6880 27.6970
H -11.4920 -35.5620 27.4330
S -10.2130 -36.5660 29.1710
C -8.4210 -36.5560 29.3290
H -7.9530 -37.0910 28.5070
H -8.0810 -35.5330 29.2910
H -8.1280 -36.9980 30.2800
C -10.3980 -34.0560 22.4250
H -10.0270 -33.4760 21.5650
H -10.6180 -35.0780 22.0950
C -11.7070 -33.4450 22.8870
O -12.7590 -33.6740 22.2970

Table S2. Cartesian coordinates (Å) of the TagRFP chromophore in the excited S_1 state optimized in the QC CASSCF(6/6)/cc-pVDZ calculations.

C -11.1370 -38.3310 22.2310
O -11.9140 -37.5490 21.7820
N -11.0780 -38.7040 23.5300
H -10.5300 -39.5110 23.7410
C -12.1690 -38.3900 24.4150
H -13.1330 -38.5960 23.9480
C -12.2510 -36.9570 24.9370
O -13.1400 -36.6520 25.6590
N -11.2840 -36.0580 24.4960
C -10.2590 -35.7810 25.1740
C -9.2520 -34.8710 24.6070
N -8.1030 -34.7100 25.1770
N -9.3340 -34.1800 23.4080
C -8.1330 -33.5470 23.1990
O -7.9170 -32.7970 22.2420
C -7.3020 -33.9690 24.3340
C -5.9230 -33.8760 24.6890
H -5.7150 -34.3760 25.6240

C -4.7800 -33.3690 23.9900
C -3.4770 -33.4810 24.6180
H -3.4010 -34.0250 25.5520
C -4.8230 -32.7220 22.7230
H -5.7700 -32.6600 22.2240
C -2.3770 -32.8950 24.1010
H -1.4330 -32.9360 24.6240
C -3.7020 -32.1250 22.1780
H -3.7410 -31.5770 21.2480
C -2.4430 -32.0990 22.8880
O -1.4910 -31.4060 22.4990
C -9.8940 -36.3900 26.5070
H -10.6920 -37.0300 26.8710
H -8.9720 -36.9630 26.4120
C -10.4340 -34.1000 22.4550
H -10.0250 -33.7180 21.5260
H -10.8510 -35.0860 22.2820
C -11.6060 -33.1910 22.8000
O -12.4490 -33.0260 21.9690
N -11.6520 -32.6500 24.0340
H -10.8300 -32.5410 24.5920
H -10.3610 -38.8110 21.6160
H -12.0960 -39.0240 25.2970
H -9.7200 -35.6130 27.2500
H -12.4210 -32.0360 24.2120

Table S3. Cartesian coordinates (Å) of the KFP chromophore in the ground S_0 state optimized in the QM(PBE0/cc-pVDZ)/MM(AMBER) calculations.

C 58.8120 71.7570 -1.2300
N 58.2300 72.7920 -0.6630
N 58.0190 71.2120 -2.2300
C 56.8590 71.9640 -2.3280
O 55.9470 71.6870 -3.1250
C 57.0390 73.0110 -1.3340
C 56.3650 74.2300 -1.1350
H 56.8460 74.8190 -0.3500
C 55.3220 74.8800 -1.8040
C 55.1170 76.2800 -1.5500
H 55.7200 76.7610 -0.7760
C 54.4840 74.2810 -2.8010
H 54.5880 73.2110 -2.9780
C 54.2360 77.0370 -2.2670
H 54.1120 78.1070 -2.0850
C 53.5900 75.0210 -3.5140
H 52.9460 74.5780 -4.2730
C 53.4450 76.4440 -3.3220
O 52.6610 77.1180 -4.0470
O 60.6430 70.4190 -1.7150
C 60.1710 71.2710 -0.9600

C 60.9520 71.8280 0.1900
H 61.9940 71.9140 -0.1460
H 60.5830 72.8200 0.4740
C 60.9010 70.9240 1.3860
H 59.8580 70.8210 1.7000
H 61.2700 69.9320 1.1070
S 61.8520 71.5120 2.8100
C 63.4980 70.9220 2.3600
H 63.4950 69.8340 2.3060
H 63.7990 71.3280 1.3940
H 64.2120 71.2360 3.1210
C 58.2880 70.1330 -3.1470
H 57.4400 70.1130 -3.8490
H 59.2080 70.3100 -3.7130
C 58.3980 68.7610 -2.5080
O 58.9100 67.8280 -3.1110

Table S4. Cartesian coordinates (Å) of the GFP chromophore in the ground S_0 state optimized in the QM(PBE0/6-31G*)/MM(AMBER) calculations.

C 25.1520 26.6260 36.4550
C 25.9820 27.0850 37.6110
N 27.2210 27.4630 37.5280
C 27.6300 27.7280 38.8320
C 28.7850 28.2990 39.2850
H 28.8120 28.3760 40.3730
C 29.9090 28.8900 38.6190
C 30.0890 28.9620 37.2190
H 29.3710 28.4700 36.5710
C 30.8930 29.5230 39.4150
H 30.7740 29.4870 40.4940
C 31.1740 29.6210 36.6700
H 31.2980 29.6370 35.5900
C 31.9730 30.1820 38.8760
H 32.7210 30.6340 39.5210
C 32.1750 30.2520 37.4710
O 33.2340 30.8190 36.9700
C 26.5120 27.4500 39.7380
O 26.3960 27.5710 40.9630
N 25.4940 27.0520 38.9090
C 24.1340 26.9580 39.3910
H 23.9440 27.7550 40.1000
H 23.4330 27.0920 38.5770
C 23.8020 25.6450 40.0500
O 22.9580 25.5520 40.9440
N 24.4760 24.6220 39.5180

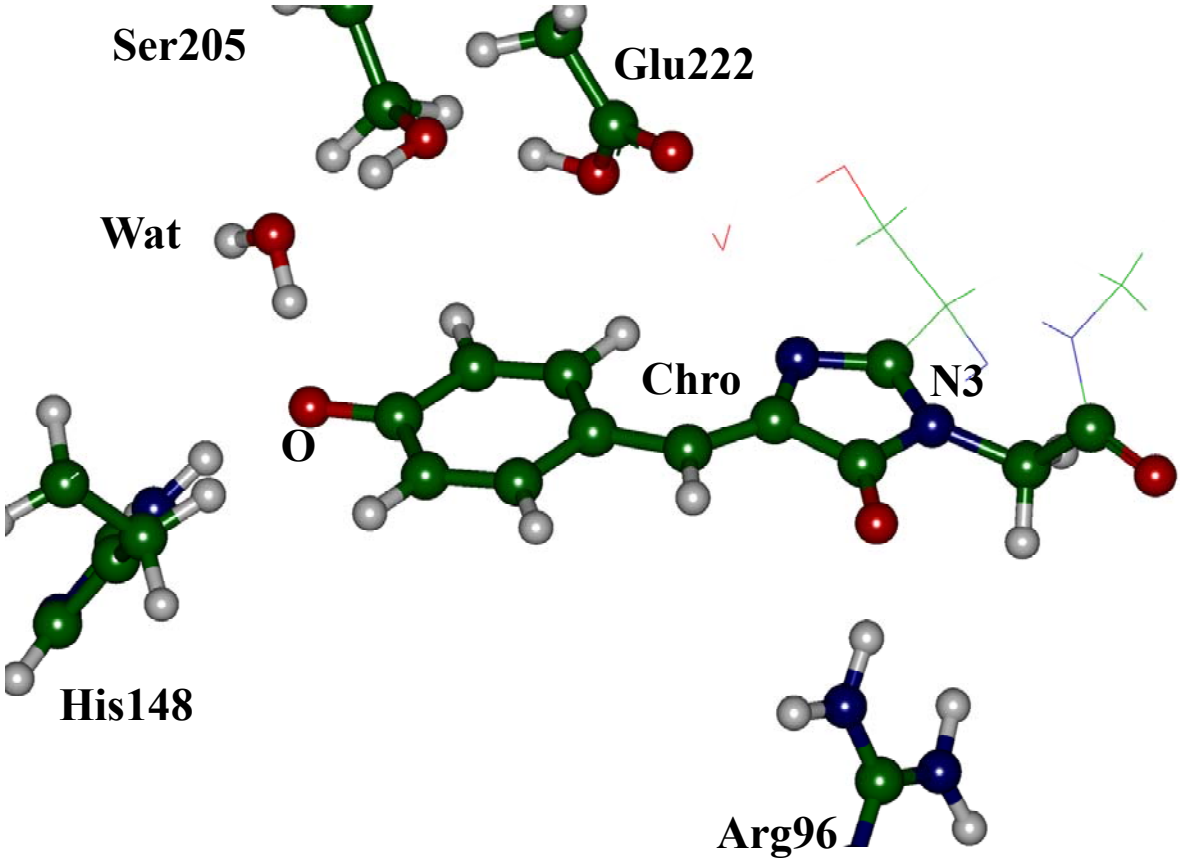

Figure S1. A part of the model system for the chromophore containing pocket of GFP.