Computational modeling structure and spectra of biological chromophores

Jack R. Collins, Igor A. Topol, Alexander V. Nemukhin, Alexander P. Savitsky

a Advanced Biomedical Computing Center, Advanced Technology Program, SAIC- Frederick Inc., NCI-Frederick, Frederick, MD, 21702-1201, USA
b Department of Chemistry, M.V. Lomonosov Moscow State University, Moscow, 119991, Russian Federation
c A.N. Bach Institute of Biochemistry, Russian Academy of Sciences, Leninsky prospekt, 33, Moscow, 119071, Russian Federation

ABSTRACT

Modern computational approaches based on quantum mechanical methods to characterize structures and optical spectra of biological chromophores in the gas phase, in solutions and proteins are discussed. Primary attention is paid to the chromophores from the family of the green fluorescent protein (GFP) widely used as a biomarker in living cells. Beyond GFP, photophysical properties of the monomeric teal fluorescent protein (mTFP1) and the kindling fluorescent protein asFP595 are simulated. We apply modern quantum chemical approaches for high level calculations of the structures of the chromophore binding pockets and to estimate spectral bands corresponding to the S0-S1 optical transitions. A special attention is paid to evaluate effects of point mutations in the vicinity of the chromophore group. Theoretical data provide important information on the chromophore properties aiming to interpret the results of experimental studies of fluorescent proteins.

Keywords: biological chromophores, GFP, mTFP1, quantum modeling

INTRODUCTION

Photosensing proteins which contain organic chromophores with extended π-conjugated systems play a crucial role in living organisms [1]. In particular, they include the light converters which emit light as a consequence of the primary photoexcitation as in Green Fluorescent Protein (GFP) like proteins, widely used in biology and medicine as biomarkers in vivo [2,3]. One of the most intriguing members from the GFP family is the so-called kindling protein asFP595 [4,5], which is initially non-fluorescent, but in response to intense green light irradiation it becomes brightly fluorescent. Photoswitching properties of this kindling fluorescent protein may be useful in many applications of modern nano- and biotechnology. The recently characterized monomeric teal fluorescent protein from Clavularia coral (mTFP1) [6,7] promises a unique bright and photostable species.

Modern spectroscopy tools as well as molecular modeling tools are intensively used to gain the knowledge of events occurring upon photoexcitation of organic chromophores in the gas-phase, in solutions and in protein matrices. This contribution is devoted to computational approaches to describe properties of the chromophores in various media aiming to interpret the results of experimental studies of fluorescent proteins. We rely on ab initio and semiempirical methods of quantum chemistry to calculate equilibrium geometry configurations of the chromophore molecules in the gas phase and in the immediate environment of the molecular groups from protein matrices in the ground electronic state. We analyze possible protonation states of the chromophore molecules and calculate the corresponding pKa values. To estimate the vertical excitation energies corresponding to the transitions between lowest singlet states of the model systems we consider different techniques which are capable of providing an accuracy comparable to that of experimental measurements. Special attention is paid to evaluate effects of point mutations in the vicinity of the chromophore group.

*anemukhin@yahoo.com; phone 7 495 939-1096; fax 7 495 939-0283; lcc.chem.msu.ru
Fig. 1 illustrates the chemical structure of the denatured GFP chromophore, 4'-hydroxybenzylidene-2,3-dimethylimidazolinone (HBDI) in its so-called neutral form in its cis-conformation. According to both experimental studies and quantum chemical calculations of the pKa values in aqueous solutions performed, in particular, in our recent paper [8], the phenolic oxygen atom O1 can be easily deprotonated, and the anionic structure of the GFP chromophore is often considered as the reference species for an analysis of its properties. The same chromophore in the anionic form is present in the monomeric teal FP, however, the absorbance and fluorescence emission maxima in mTFP1 are blue-shifted compared to those in the enhanced GFP variant due to differences in amino acid residues in the binding pocket.

![Chemical structure of the GFP chromophore.](image1)

The chromophore of the asFP595 kindling protein in the trans-conformation of the neutral form is depicted in Fig. 2. Compared to HBDI the latter possesses an enlarged π-conjugated system and correspondingly is characterized by longer absorption and emission spectral bands.

![Chemical structure of the asFP595 chromophore.](image2)

Fig. 3 illustrates a structure of the model system which comprises the chromophore and a fairly large fraction of the protein asFP595. To obtain such a system for an analysis of the fluorescent protein properties we apply the combined quantum mechanical – molecular mechanical (QM/MM) technique and compute energies and forces in the selected subsystem of primary interest (the QM-part) by using quantum chemistry approaches and energies and forces in the remaining subsystem (MM-part) by using empirical force field parameters. In Fig. 3 the selected QM-part is depicted by balls and sticks. As starting coordinates of heavy atoms we use those from the structures deposited to the Protein Data Bank (PDB). Then hydrogen atoms are added to the model system with the help of modern molecular modeling tools, and the ground state geometry parameters are carefully optimized in QM/MM energy minimization. Generally, several
schemes to assign specific protonation states of the chromophore group as well as of the nearby amino acid residues are considered in simulations.

For example, an analysis of protonation of His197 and Glu215 (Fig.3) plays a crucial role in modeling structure, spectra and dynamics of the asFP595 protein [9].

When considering the structure of this particular model system we used two different QM/MM schemes. The first one was the mechanical embedding QM/MM method [10] and the second one was the effective fragment potential (EFP) based QM/MM approach originally developed by Gordon and co-authors [11]. In the mechanical embedding method the QM and MM interaction terms $V_{QM/MM}$ are specified by an explicit analytic formula, while in the EFP-based theory the QM and MM coupling is treated through the contributions from the molecular groups of the MM-part (effective fragments) to the one-electron integrals of the quantum Hamiltonian.

We found that both QM/MM approaches produced fairly consistent minimum energy geometry configurations. An agreement to the reference crystal structure [12] is also reasonable in both cases. The greatest difference between the results of two QM/MM approaches in this case refers to the description of the Glu215 side chain (Fig.3). In the EFP-based QM/MM optimized model structure the location of the proton shared by the oxygen atom of Glu215 and the nitrogen atom of the chromophore is on the nitrogen center providing the zwitterionic form of the chromophore. In the mechanical embedding QM/MM structure the chromophore is in the anionic form, and Glu215 corresponds to the protonated species. Consequently, the distances between Glu215 and chromophore are noticeably different: 2.6 Å vs 2.9 Å. In this respect the EFP-based QM/MM structure is better consistent with the crystal structure with the distance 2.7 Å [12].
Accurate calculation of the electronic excitation energies of organic chromophores poses a challenge in modeling spectra of fluorescent proteins. A variety of quantum chemical methods applied for the GFP-like chromophores ranges from simple semiempirical estimates to fairly complicated and consuming \emph{ab initio} techniques. In publication [13] we considered different versions of the time-dependent density functional theory (TDDFT) method, which is now one of the most popular modeling tools, to compute the vertical excitation energy of the $S_0$-$S_1$ optical transition for the GFP anionic chromophore.

Table 1. The excitation energies, the corresponding wavelengths and oscillator strengths for the cis-anionic form of HBDI (Fig.1) calculated in different quantum chemical approximations.

<table>
<thead>
<tr>
<th>Calculation method</th>
<th>$\Delta E$, eV</th>
<th>$\lambda$(max), nm</th>
<th>Oscillator strength</th>
</tr>
</thead>
<tbody>
<tr>
<td>TDDFT(BP86)//B3LYP/6-31+G**</td>
<td>2.94</td>
<td>422</td>
<td>0.86</td>
</tr>
<tr>
<td>CIS//B3LYP/6-31+G**</td>
<td>3.75</td>
<td>330</td>
<td>1.55</td>
</tr>
<tr>
<td>ZINDO//B3LYP/6-31+G**</td>
<td>2.59</td>
<td>479</td>
<td>1.22</td>
</tr>
<tr>
<td>aug-MCQDPT2/PBE0/aug-cc-pVDZ [15]</td>
<td>2.54</td>
<td>489</td>
<td></td>
</tr>
<tr>
<td>Experiment [14]</td>
<td>2.59</td>
<td>479</td>
<td></td>
</tr>
</tbody>
</table>

In the second row of Table 1 we show one more result of the TDDFT approach with the use of the BP86 functional at the ground state geometry configuration optimized in the B3LYP/6-31+G** approximation. The calculated wavelength of the electronic transition, 422 nm, differs by about 50 nm from the experimental absorption band maximum [14], typical for the TDDFT applications. The result obtained in the \emph{ab initio} configuration interaction method with single excitations (CIS) shown in the third row of Table 1 is much worse, 330 nm. An achievement of a highly developed \emph{ab initio} technique is illustrated here by showing the result of the augmented version of the multiconfigurational quasidegenerate perturbation theory in the second order (aug-MCQDPT2) for the DFT optimized geometry configuration [15]. Such an approach, based on the state averaged complete active space SCF wavefunctions, allows one to obtain positions of the spectral bands within 10-15 nm from the corresponding experimental values for several types of biological chromophores [15-17], however, it requires considerable computational efforts.

The fourth row of Table 1 shows an application of the well-known semiempirical ZINDO technique [18] to estimate the excitation energies of organic chromophores. Its application here for the DFT optimized ground state equilibrium geometry parameters gives a nearly perfect result. Therefore we apply this strategy (ZINDO//DFT) for modeling optical spectra of the systems modeling active sites of fluorescent proteins.

**MODELING MUTATIONS IN MONOMERIC TEAL FLUORESCENT PROTEIN**

To model optical spectra of the monomeric teal fluorescent protein (mTFP1) we started from the coordinates of heavy atoms in the crystal structure PDBID:2HQK. The extended cluster model was constructed by including into the system the chromophore unit, the side chains of 10 nearest amino acid residues (Asn69, Arg70, Arg95, Asp144, Ser146, Ghu148, His163, His197, Ile199, Ghu215) and 4 water molecules. Equilibrium geometry parameters of this initial structure were optimized in the B3LYP/6-31G approximation, and vertical excitation energies were computed by using the ZINDO technique. We considered all possible variants of protonation/deprotonation status of the His residues by comparing the computed and experimental band positions. The best agreement was achieved when both His163 and His197 side chains were assumed to be protonated (the calculated wavelength is 465 nm versus experimental 462 nm). Next we carried out series of calculations by mutating the chromophore molecule and selected nearby residues. Experimental studies were performed for the engineered proteins in which the Tyr67 part of the chromophore was mutated by His and Trp [19] showing that all these versions retained their blue-shifted emission relative to the enhanced GFP counterpart. The results of ZINDO//B3LYP calculations for both mutants Y67H and Y67W well agree with the experimental findings. Simulations allowed us to predict spectral shifts upon mutations of His163 and His197 by Met, of Asp144 by Ala, of Arg75 by Leu, and of Ile199 by Lys. Qualitative conclusions on the role of these mutations in the spectral properties of mTFP1 are in accord with the experimental observations.
CONCLUSION

This contribution describes the practical means how to apply modern tools of quantum chemistry for accurate calculations of structures and optical band positions in the spectra of biological chromophores in different media. We underline three major problems of such calculations: (i) optimization of equilibrium geometry parameters, (ii) computation of energy differences between the excited and ground state energies, (iii) estimates of contributions from the environmental molecular groups of solvent or protein. The last one presents the major challenge. Small variations in positions of charged or polarized environmental molecular groups may cause large shifts in optical band positions far exceeding effects of geometry changes or details of electronic structure theory used in vertical energy calculations. Even from the qualitative side, molecular modeling provides an essential support to experimental studies in spite of the specified difficulties in obtaining precise values for the molecular parameters. Formulation of possible mechanisms of photoexcitation, identification of the functional states of the chromophores, elucidation the role of point mutations in the photoreceptor proteins may be considerably facilitated with the help of theoretical estimates.

ACKNOWLEDGMENTS

This project has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health, under contract number 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 organizations imply endorsement by the U.S. Government. This work was partly supported by the Science and Innovation Federal Russian Agency (project # 02.522.11.2002), by the Program for Molecular and Cell Biology from Russian Academy of Sciences, by the joint grant from U.S. Civilian Research and Development Foundation (project #RUC1-2913-MO-07) and the Russian Foundation for Basic Research (project # 08-03-91104-AFGIR).

REFERENCES