Chapter 19

Exciting Green Flourescent Protein

Volkhard Helms[1,4], Erik F. Y. Hom[1], T. P. Straatsma[2], J. Andrew McCammon[1], and Peter Langhoff[3]

[1]Department of Chemistry and Biochemistry, University of California at San Diego, La Jolla, CA 92093–0365
[2]Pacific Northwest Laboratory, Richland, WA
[3]Department of Chemistry, Indiana University, Bloomington, IN 47405

Dynamical and luminescent properties of the Green Fluorescent Protein (GFP) are being explored via molecular dynamics (MD) simulations and quantum chemistry calculations with the aim of facilitating the rational development of GFP as a probe for cellular functions. Results from an MD simulation of wild type GFP demonstrate the rigidity of the structural framework of GFP, and a very stable hydrogen bond network around the chromophore. Furthermore, excited state calculations have been performed on the chromophore in vacuum, and we report about our work in progress here.

Interfacing Quantum Mechanical and Molecular Mechanics methods has become a popular device to study properties of a small quantum regime embedded in a solvent or protein environment that is described classically. Here, we discuss preliminary results of parallel quantum and classical calculations. Not only do we want to understand the energetic and dynamic features of an organic chromophore and its surroundings in detail, we would also like to use its absorption and fluorescent properties as a detector to study structural and dynamic behaviour of its protein environment.

Green Fluorescent Protein

Green Fluorescent Protein (GFP) is a spontaneously fluorescent protein isolated from the Pacific Northwest jellyfish *Aequorea victoria.* Its apparent role in the jellyfish is to transduce, via fluorescence resonance energy transfer, the blue chemiluminescence of another protein, aequorin, as green fluorescent light (1). Since its first identification, cell biologists have recognized the biotechnological potential of GFP, and have very successfully used GFP as a noninvasive marker in living cells, e.g. as a reporter of gene expression and protein localization (2), and in the detection of protein-protein interactions (3). Recently, three-dimensional crystal structures of the wild type protein, of a Ser65 → Thr mutant, and of a number of other mutants, were determined (4–7).

GFP consists of a single chain of 238 residues that forms an 11-stranded $\beta-$barrel, wrapped around a central helix. The barrel is a nearly perfect cylinder 42 $\AA$ long and 24 $\AA$ in diameter. At the center of this $\beta-$barrel 'canister' lies the central helix and the chromophore that is responsible for GFP's luminescence, see Figure 1. This unusual *p*-hydroxylbenzylideneimidazolidinone chromophore results from the autocatalytic cyclization of the polypetide backbone between residues Ser65 and Gly67, followed by oxidation of the $\alpha - \beta$ bond of Tyr66. It is completely protected from bulk solvent, although there are a number of bound water molecules in the protein interior.

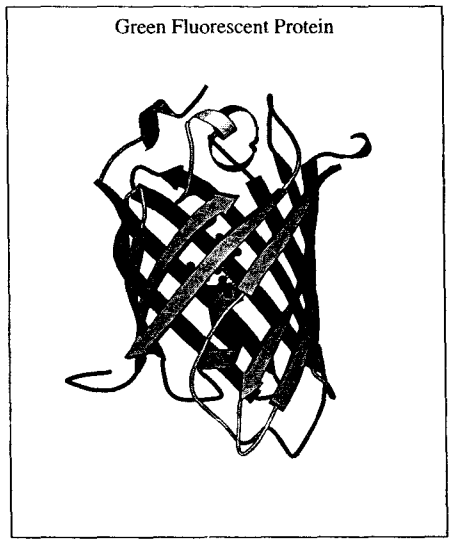

Figure 1 Schematic diagram of the cylindrical shape of GFP. Only elements of secondary structure and the chromophore (in black) are shown.

Wild type GFP has two absorption maxima at 395 and 475 nm that are believed to be due to a neutral and anionic form of the chromophore, respectively (8). A ca. 80% neutral : 20% anionic equilibrium appears to be governed by a hydrogen

[4]email: vhelms@ucsd.edu.

bonding network that permits proton transfer between the chromophore and neighboring side chains (6). Figure 2 shows simplified drawings of the proposed hydrogen bonding networks around the two ionization states of the chromophore (4, 7).

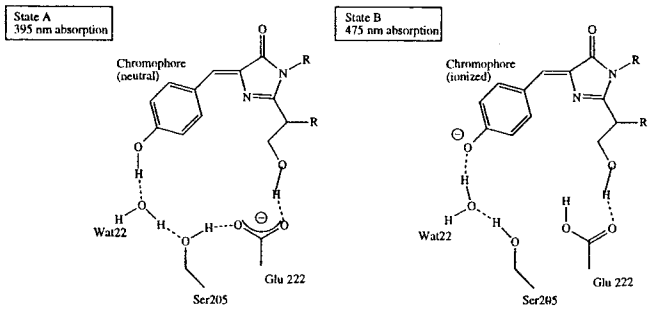

Figure 2 Hydrogen bond networks close to neutral and anionic chromophore.

Mutant versions of GFP with significantly different spectral properties have been constructed, and permit, in principle, simultaneous tagging and localization of proteins and the monitoring of association events *in vivo*. Our present work in progress aims at elucidating the structural, dynamical, and electronic factors responsible for the remarkable properties of GFP.

In spectroscopic single molecule experiments, two mutants exhibited a fast on/off blinking behaviour under constant illumination, and a conversion into a long-lived dark state (9). From here, the molecule could only be switched back to the original emissive state by continuosly irradiating at 405 nm. The authors noted that, presumably, the fast blinking occurs with an anionic chromophore, while the slow conversion reflects the transition to a form with a neutral chromophore. Certainly, these mutants contain different solvent networks in the chromophore surrounding, and the results are not directly transferable to the present study, but they give an indication about the slow timescales involved with these transitions.

Relation to Other Work

This work follows previous theoretical work on the photoisomerization of retinal in the membrane protein bacteriorhodopsin (10) and on metal-containing enzymes such as plastocyanin (11) involved in electron transport. The study of GFP is facilitated compared to these proteins by the availability of excellent crystallographic

data with well resolved internal water positions and by the fact that GFP does not contain complex co-factors such as a heme-moiety or metal atoms.

Based on INDO/S calculations of different forms of a model chromophore *in vacuo* and in solvent, it was suggested (12) that the bare nitrogen in the imidazolidinone ring is protonated. This was based on the fact that the results for the protonated chromophore were in best agreement with the experimental absorption frequencies. In our work we do not account for this possibility, because we feel that protonation of the ring nitrogen would cause polar and charged protein groups in its vicinity to move closer to this site than is observed in the crystal structures.

Molecular Dynamics Simulation of GFP

An MD simulation of wild type GFP with a neutral chromophore was performed with the parallel NWChem program (13). The simulation system consisted of one GFP molecule solvated in a cubic, periodic solvent box with dimensions of 69 $\AA$. The AMBER94 all-atom force field (14) was employed, resulting in 3594 solute atoms and 9381 solvent molecules, with a total of 31726 dynamic atoms. Six Na^+-ions were positioned in low energy positions on the protein surface to make the system overall neutral, and long-range electrostatic interactions were treated by the particle-mesh-Ewald technique (15). The partial atomic charges of the ring atoms of the chromophore were obtained from an electrostatic potential fit with the CHELPG method (16) to the HF/6-31G(d) optimized electron density. The partial atomic charges of the atoms of the chromophore's Serine side chain and those of the connecting peptide bonds were taken from the AMBER94 force field. The remaining charge of 0.2 e was assigned to atom CA3. Parameters for covalent and Lennard-Jones interactions were also adapted from AMBER94. The positions of aromatic hydrogens and hydrogens of crystal water molecules were optimized with the WHAT IF program (17). Starting from the crystal structure (6), the system was heated and equilibrated by 100 steps of steepest descent energy minimization of the whole system, by subsequent 20 ps MD equilibration of the solvent only at 300 K, and by short MD simulations of 5 ps length each of the whole system at 50 K, 100 K, ..., 300 K under constant volume and constant energy conditions. A 1 ns MD simulation was then performed at a constant temperature of 300 K and at constant pressure. The individual energy components were well equilibrated after 100 ps, and the last 900 ps were considered for analysis.

The Molecular Dynamics Simulation was Performed to Characterize :

- *How rigid the GFP β-barrel structure is*
  After the crystal structure of GFP became known, it was obvious that the encapsulation of the chromophore is probably responsible for the high quantum yield of fluorescence, for the inability of O_2 to quench the excited state, and for the resistance of the chromophore to titration of the external pH (4, 18).

We investigated the dynamical behavior of the protein during the simulation by an essential dynamics analysis (19) that is implemented in the WHAT IF program (17). Figure 3 shows maximum and minimum projections of the trajectory onto the eigenvector with the largest eigenvalue.

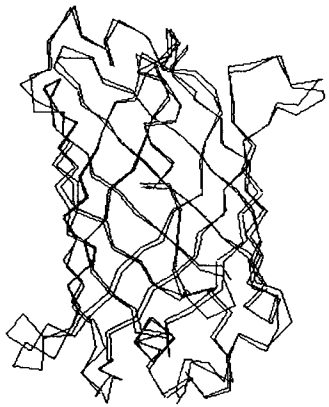

Figure 3 Projections of the MD simulation of GFP wild type onto the eigenvector with the largest eigenvalue.

Detectable motions can only be seen in the loop segments, and the β-barrel is almost unaffected. Apparently, its unusual cylindrical architecture makes GFP a rock-solid protein compared to other proteins folds. Also, the RMS-deviation of the C_α atoms from the energy minimized crystal structure remained at around 0.1 nm during the production phase. This is a very low value since proteins often show RMS-deviations of 0.2 - 0.3 nm during MD simulations of this length.

- *The permeability of the space inside the β-barrel to bulk solvent molecules.*

No water molecules were observed to penetrate the interior of GFP during the simulation.

- *The hydrogen bonding pattern central to the conversion between the neutral and anionic chromophore forms.*

Occupancies were calculated for the hydrogen bond network in the chromophore surrounding. All important hydrogen bonds had occupancies between 80 and 100 % during the simulation. The six water molecules in the immediate vicinity of the chromophore behave like in ice.

A more detailed analysis of the simulations will be presented elsewhere. Not surprisingly, the GFP structure provides a very stable framework for the enclosed chromophore. It is noteworthy that this rigidity extends to the immediate surrounding of the chromophore where a very stable arrangement of hydrogen bonds and water positions was found. The cylindrical β-barrel thus provides a natural protection of the fluorescent signal against environmental factors such as pH, O_2, and metal ions.

Contrary to our original plans that included a statistical sampling of the protein dynamics by performing quantum calculations on a large number of snapshots from the molecular dynamics trajectory, performing single quantum calculations on a small number of selected protein conformations seems perfectly justified.

Quantum Calculations of the GFP Chromophore

The *ab initio* QM calculations aim at characterizing the chromophore absorption spectrum under the influence of the protein environment and thermal fluctuations. Although good quantitative agreement with experimental data is desirable, we are initially most interested in identifying key protein groups which could be modified to shift the wavelength of excitation, improve absorption intensity, and/or narrow the bandwidth of absorption. These calculations will serve as an important base to derive simplified models for predicting spectral changes due to mutations in GFP.

In Vacuo Calculations

The geometries of neutral and anionic model chromophore have been optimized at Hartree Fock 6-31G(d) level (264 basis functions) with the GAMESS program (20). Fifty seven molecular orbitals are occupied in each case. Planar symmetry was employed since the chromophore is found in almost planar conformations in the various crystal structures. Vertical singlet excitations into the 10 lowest virtual IVO orbitals were performed separately for A" $\rightarrow$ A", A' $\rightarrow$ A", A' $\rightarrow$ A', and for A" $\rightarrow$ A' transitions. The transition dipole moments for mixed A' $\rightarrow$ A" and A" $\rightarrow$ A' excitations are close to zero, and all A' $\rightarrow$ A' excitations are very high energy (> 12 eV). To understand the low-lying spectrum of the chromophore, it is therefore sufficient to focus on A" $\rightarrow$ A" excitations. The results for IVO calculations for the A" $\rightarrow$ A" excitations are shown in table I.

Table I. Vertical excitation energies (eV) for A" $\rightarrow$ A" excitations into virtual IVO orbitals.

from MO # to	state 2	state 3	state 4	state 5	state 6
57	5.534	7.411	8.830	9.654	12.093
56	7.305	8.716	9.990	12.415	13.959
55	7.572	9.613	10.664	11.652	13.658
53	8.889	11.783	12.119	12.763	14.683
51	10.916	13.455	14.235	14.855	17.032
49	11.195	13.489	14.348	15.696	17.962

Only a few vertical excitations have energies below 10 eV. Accordingly, the active space for CASSCF-type calculations may be chosen rather small unless very accurate results are desired. The result from an MCSCF calculation (21) with an active space including the 3 highest occupied π-orbitals and the 3 lowest unoccupied π^*-orbitals is 5.628 eV, in good agreement with the IVO result for the HOMO $\rightarrow$ LUMO excitation. The excitation energies for the anionic chromophore were ca. 1 eV lower, in agreement with the experimentally observed red-shift for the anionic form. Still, the excitation energies are rather high compared to the expected values of ca. 3 eV. We expect a significant drop of these energies by also optimizing the excited state geometries and by including the effects of electron correlation.

Inclusion of the Protein Environment

An important challenge and difficulty arises from the representation of the protein environment. Work on bacteriorhodopsin by the groups of Schulten, Warshel, and others has shown that it is essential to incorporate the protein surroundings for a correct description of the electronic structure of the chromophore. Experimental spectra of a model GFP chromophore in water and in various organic solvents (22) were blue-shifted by 30 - 50 nm compared to spectra in the protein environment. Yet, these shifts are rather small and demonstrate that the chromophore environment in GFP is not of unusual nature.

We plan to represent the protein environment by effective fragment potentials (23) that are being developed in combination with the GAMESS program.

Acknowledgments

We would like to acknowledge the help of Kim Baldridge with using the GAMESS program. Carl Winstead was of critical help with the single excitation calculations. We also thank Roger Tsien for valuable discussions. NWChem Version 3.0, as developed and distributed by Pacific Northwest National Laboratory, P. O. Box 999, Richland, Washington 99352 USA, and funded by the U. S. Department of Energy, was used to obtain some of these results. This work was supported by a postdoctoral NATO fellowship to VH by the Deutscher Akademischer Austauschdienst, and by grants from NSF, the San Diego Supercomputing Center and the NSF Supercomputer Centers MetaCenter program. VH is also a fellow of the Program in Mathematics and Molecular Biology and of the La Jolla Interfaces in Sciences Program.

Literature Cited

(1) Ward, W.; Cody, C.; Hart, R.; Cormier, M. *Photochem. Photobiol.* **1980**, *31*, 611 - 615.

(2) Cubitt, A.; Heim, R.; Adams, S.; Boyd, A.; Gross, L.; Tsien, R. *Trends Biochem. Sci.* **1995**, *20*, 448–455.

(3) Rizzuto, R.; Brini, M.; Giorgi, F. D.; Rossi, R.; Heim, R.; Tsien, R.; Pozzan, T. *Curr. Biol.* **1996**, *6*, 183–188.

(4) Ormö, M.; Cubitt, A.; Kallio, K.; Gross, L.; Tsien, R.; Remington, S. *Science* **1996**, *273*, 1392–1395.

(5) Yang, F.; Moss, G.; Jr, G. P. *Nat. Biotech.* **1996**, *14*, 1246–1251.

(6) Brejc, K.; Sixma, T.; Kitts, P.; Kain, S.; Tsien, R.; Ormö, M.; Remington, S. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 2306–2311.

(7) Palm, G.; Zdanov, A.; Gaitanaris, G.; Stauber, R.; Pavlakis, G.; Wlodawer, A. *Nat. Struct. Biol.* **1997**, *4*, 361–365.

(8) Chattoraj, M.; King, B.; Bublitz, G.; Boxer, S. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 8362–8367.

(9) Dickson, R.; Cubitt, A.; Tsien, R.; Moerner, W. *Nature* **1997**, *388*, 355–358.

(10) Logunov, I.; Schulten, K. *J. Am. Chem. Soc.* **1996**, *118*, 9727–9735.

(11) Pierloot, K.; Kerpel, J. D.; Ryde, U.; Roos, B. *J. Am. Chem. Soc.* **1997**, *119*, 218–226.

(12) Voityuk, A.; Michel-Beyerle, M.; Rösch, N. *Chem. Phys. Lett.* **1997**, *272*, 162–167.

(13) *NWChem, A Computational Chemistry Package for Parallel Computers, Version 3.0.* High Performance Computational Chemistry Group, Pacific Northwest National Laboratory, Richland, Washington 99352, USA, 1997.

(14) Cornell, W.; Cieplak, P.; Bayly, C.; Gould, I.; Jr., K. M.; Ferguson, D.; Spellmeyer, D.; Fox, T.; Caldwell, J.; Kollman, P. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.

(15) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **98**, *98*, 10089–10093.

(16) Breneman, C.; Wiberg, K. *J. Comput. Chem.* **1990**, *11*, 361–373.

(17) Vriend, G. *J. Mol. Graph.* **1990**, *8*, 52–56.

(18) Phillips, G. *Curr. Op. Struct. Biol.* **1997**, *7*, 821–827.

(19) Amadei, A.; Linssen, A.; Berendsen, H. *Proteins* **1993**, *17*, 412–425.

(20) Schmidt, M.; Baldridge, K.; Boatz, J.; Elbert, S.; Gordon, M.; Koseki, J. J. S.; Matsunaga, N.; Nguyen, K.; Su, S.; Windus, T. *J. Comp. Chem.* **1993**, *14*, 1347–1363.

(21) Roos, B.; Taylor, P.; Siegbahn, P. *Chem. Phys.* **1980**, *48*, 157–173.

(22) Niwa, H.; Inouye, S.; Hirano, T.; Matsuno, T.; Kojima, S.; Kubota, M.; Ohashi, M.; Tsuji, F. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 13617–13622.

(23) Day, P.; Jensen, J.; Gordon, M.; Webb, S.; Stevens, W.; Krauss, J.; Cohen, D. G. H. B. D. *J. Chem. Phys.* **1996**, *105*, 1968–1986.