ARTICLE IN PRESS

Available online at www.sciencedirect.com
SCIENCE DIRECT®

Chemical Physics Letters 419 (2005) 48–54

CHEMICAL PHYSICS LETTERS

www.elsevier.com/locate/cplett

How do insertions affect green fluorescent protein?

Murat Cetinkaya [a], Ahmet Zeytun [b], Jorge Sofo [c,d], Melik C. Demirel [a,d,*]

[a] College of Engineering, Pennsylvania State University, 212 EES Bldg., Office 205A, University Park, PA 16802 , USA
[b] Bioscience Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA
[c] Department of Physics, Pennsylvania State University, University Park, PA 16802, USA
[d] Materials Research Institute, Pennsylvania State University, University Park, PA 16802, USA

Received 28 March 2005; in final form 10 November 2005

Abstract

A computer-based modeling and bench-top experiments are combined to understand the fluorescence of GFP. Random octapeptides are inserted into individual loops of the GFP. Amino acid sequences and fluorescence levels of clones from each loop are determined. The effect of peptide insertions into the loop regions of GFP are studied computationally using quantum mechanics and molecular dynamics calculations. Our experimental and computational results show that the random peptide insertions into the loop regions change the absorption intensity of the GFP but not the absorption wavelength.

1. Introduction

The green fluorescent protein (GFP) is an intrinsically fluorescent protein extracted from the jellyfish *Aequorea victoria* [1]. Wild type (WT) GFP is a monomeric protein with a molecular weight of 28 kDa [2]. GFP has a β-can structure with 11 antiparallel β-strands and three α-helices. Moreover, there are loops at each end improving the water-proof structure of the β-can so that the GFP chromophore is thoroughly protected. The GFP chromophore is formed by the modification of residues 65–66–67 (Ser–Tyr–Gly) into a chemical structure referred to as 4-(*p*-hydroxy-benzylidene)-imidazolidin-5-one [3]. Several mechanisms have been proposed for the chromophore formation, but the currently accepted mechanism is the rapid cyclization between residues Ser65 and Gly67 resulting in an imidazolin-5-one structure proceeded by dehydration of Ser65 carbonyl oxygen, and oxygenation of Tyr66 at a much slower rate [4–6]. The β-can topology also creates the environment for the side chain interactions occurring at the chromophore proximity (i.e., residues Thr203, Ser205 and Glu222).

Excited-state dynamics of the GFP have been studied and the two distinct excitation spectra have been resolved [5–7]. The 475 nm peak is the result of the deprotonated (anionic) chromophore. On the other hand, the 395 nm peak arises from the protonated (neutral) chromophore. Creemers et al. [8,9] and Volkmer et al. [10] have investigated the chromophore chemistry with high resolution optical spectroscopy at low temperatures. Patterson et al. [11] determined the quantum yield of the wt-GFP as 0.79.

The GFP chromophore formation and its mutants have been extensively studied (see a review by Miyawaki et al. [12]). Peptide insertions have been studied experimentally by two groups [13,14]. Abedi et al. [14] selected 10 candidate sites for peptide insertion (Leu–Glu–Glu–Phe–Gly–Ser). They showed that the location of the peptide insertion affects the fluorescence levels of the GFP. However, a precise understanding of fluorescence loss is still lacking. Particularly, our aim is to create GFP molecules carrying random amino acids insertions and understand the effect of these insertions (e.g. by change in excitation and emission) experimentally and computationally.

* Corresponding author. Fax: +1 814 865 9974.
E-mail addresses: mcd18@psu.edu, mdemirel@engr.psu.edu (M.C. Demirel).

0009-2614/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.cplett.2005.11.046

In this Letter, we combined loop structure modeling and molecular dynamics modeling to simulate dynamics of the GFP and explain the reasons of absorption intensity change by semi-empirical quantum mechanical methods. The results are compared with our experimental measurements. Our calculations show that the random peptide insertions change the absorption intensity of the GFP but not the absorption wavelength. The computational results are in good agreement with experiments.

2. Methods

2.1. Molecular dynamics

All molecular dynamics (MD) simulations are performed with the AMBER force-field and simulation package [15]. The GFP structure (PDB code 1EMB) is explicitly solvated with TIP3P water [16] with an 8 Å buffer (~9000 water molecules) around itself. The electrical charge of the environment (all negative) is also neutralized with Na^+ ions in the solvent. The systems are simulated with periodic boundary conditions. Particle mesh Ewald (PME) method and 10 Å cut-off non-bonded interactions are selected for electrostatics treatment. Additionally, the tail part of the GFP molecule (residues 230–238) is truncated since it causes an excessive free energy due to local disorder [17]. Infact, it is experimentally demonstrated that truncation of the GFP tail does not affect the fluorescence characteristics of the protein [18]. MD simulations are performed after the geometry optimization of GFP structures with steepest descent and conjugate gradient algorithms. Each simulation starts with a 15 ps constant volume MD in which the system temperature is increased gradually to 300 K, while constraining the solute with increased force constants. Then, the simulation continues with a 1 ns constant pressure (1 atm) and temperature (300 K) MD run with 1 fs time steps.

2.2. Loop modeling

The protein data bank (PDB) structure 1EMB is chosen as a template structure for peptide insertions. The insertion sequence is placed into one of the 9 sites defined in Table 1. Fig. 1a shows the loop modeling procedure in a schematic way. All preparation steps are performed using the Swiss-Pdb Viewer software package [19]. The insertions are modeled with the loop search method and the de novo generation technique [20]. The loop search method scans the protein fragments in the PDB and finds the structures satisfying the user-specified geometrical constraints. The loop search method finds structures resembling existing protein conformations. The insert does not fit to the original structure for certain insertions because of N and C terminus overlaps. The de novo generation technique solves this problem by reconstructing the N and C terminus of the insertion region. Random dihedral angles are assigned for the N and C terminus atoms of insertion sequence and the possibilities are refined. All insertions modeling procedures are finalized with energy minimization techniques. In order to test the accuracy of initial models, several loop structures are used in each set of MD simulations, and the results showed agreement with each other. Here, only one set of result is presented for each GFP insertion.

2.3. Quantum mechanical calculations

The chromophore of the GFP is modeled by a heterocyclic ring composed of three amino acid residues 65–66–67 (Ser–Tyr–Gly). The backbone group of Ser and Gly are replaced by hydrogen atoms (Fig. 2b). The four states of GFP (i.e., anionic, neutral, cationic and zwitterion) are considered for calculations. The average geometry of the chromophore structure is obtained from molecular dynamics calculations. The absorption spectra of the model systems are calculated with the spectroscopic INDO-CI method [21,22]. Oscillator strengths (absorption intensities) are determined in the one-center approximation to the dipole length operator. We note that ZINDO is a semi-empirical method and we used this method for qualitative comparison of GFP insertions. A detailed quantitative understanding of the chromophore structure and energetics using time-dependent density functional was studied by Laino et al. [23]. However, the semi-empirical method provides a reasonable qualitative description of the effect of different insertions. It should be noted that the accuracy of the excitation energy varies (~0.7 eV) depending on

Table 1
Ten insertion sequences are selected from Fig. 1

Insertion site	Sequence	Location	RMS deviation	Relative excitation intensity
1	GFP-control	N/A	0.794	66844
2	WWHSGVDG	172–173	0.797	66558
3	RRGRHRRM	23–24	0.802	56024
4	ARLEHGGF	213–214	0.809	55643
5	RDQTGGLA	173–174	0.794	59637
6	FGGLAGRC	173–174	0.816	58131
7	APTICTHC	173–174	0.799	60910
8	VMPMVHES	173–174	0.803	79637
9	WRTMDPES	172–173	0.796	80302
10	PVGGRVDG	173–174	0.799	95283

Measured excitation intensities, insertion locations and RMS deviation (1 ns average) are shown for these sites.

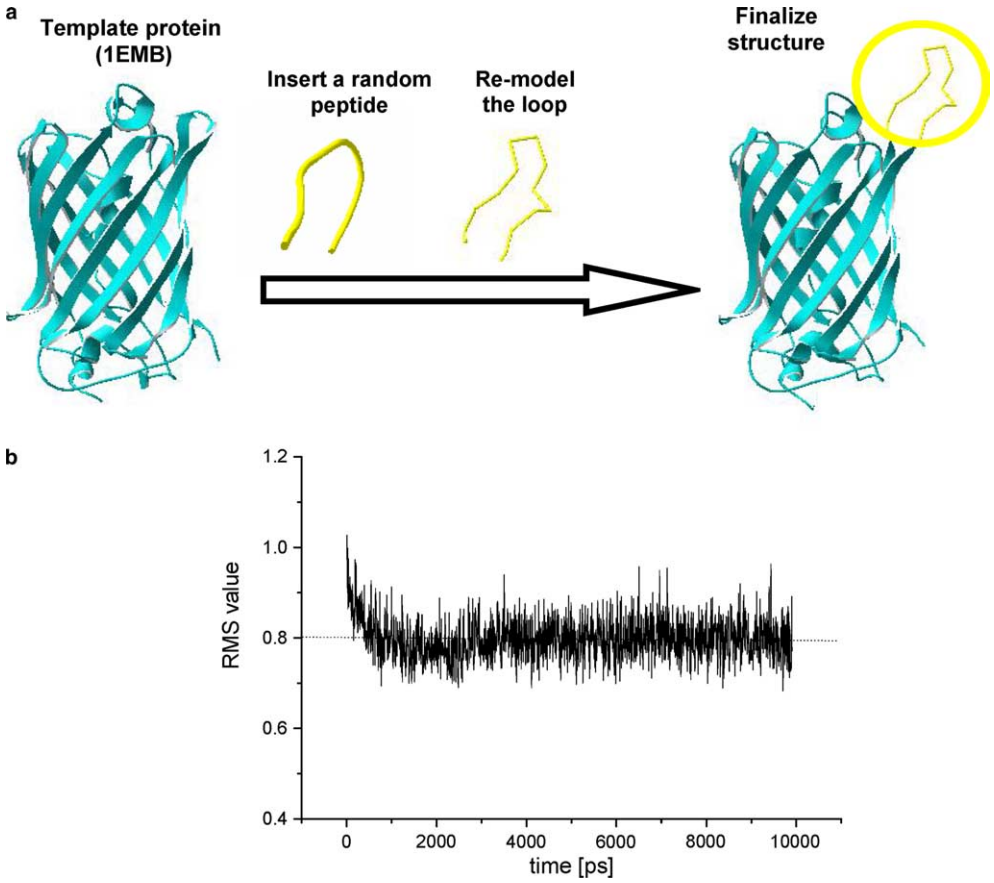

Fig. 1. Schematic representation for loop modeling is shown (a). PDB code 1EMB is used as a template structure. The RMS value for the MD simulation (10 ns) is shown for sequence # 2. (b) The RMS value converges after 0.5 ns simulation.

the calculation method (e.g, SA6CAS, SA2CAS, TD-DFT and EOM-CCSD) for the same chromophore structure (see [24, p. 50] and references therein).

2.4. Protein expression, purification and characterization

A new variant of GFP called superfolder GFP [25] is used for insertion experiments. The superfolder GFP is more permissive for introducing exogenous sequences into the loop regions, compared to all available GFPs. GFP insertions are prepared by single loop insertions using random peptides as $(NNK)_8$ (i.e., N = G, A, T, or C, and K = T or G). The random peptide is introduced into GFP with several primers (see Table 2) using PCR according to standard protocols [26]. Enzymes and buffers are bought from New England Biolabs (NEB) and all PCR reactions are done using the proofreading Vent-polymerase (NEB). The GFP carrying random peptide is transferred into a phagemid vector (pDAN5) and expressed in DH5αF cells. Single clones are picked for expression and sequencing. Cells are grown at 37 °C, induced with 1 mM isopropylthiogalactoside (IPTG) and then incubated for another 6 h. The cell pellets are harvested by centrifugation at 4 °C at 3000 rpm for 20 min. The His_6-tagged GFP is purified by metal affinity chromatography and gel filtration. Fluorescence intensities are determined after purification using a fluorescence spectrometer (Perkin–Elmer) at 2 μg/ml concentration in PBS.

3. Results

Random octapeptides $(NNK)_8$ are inserted into four different loops of the GFP (Fig. 2a). The results from insertions demonstrate that the number of green colonies varies from 0–10% (loop 2 amino acids 102–103) to 80–90% (loop 3 amino acids 172–173; loop 1 amino acids 22–24). Experimental data shows that loop region insertions dramatically change the emission and excitation intensity of the GFP depending on the peptide sequence that is inserted to the loop. We have selected 10 highly fluorescent sequences for the computational studies. These 10 sequences, insertion locations and the corresponding absorption intensities are listed in Table 1. The emission and excitation spectra for the GFP are measured and shown in Fig. 2c. The wavelength of excitation (480 nm) and emission (509 nm) do not depend on the insertion sequence. On the other hand, the intensity of emission and excitation are sensitive to the insertion sequence (Table 1).

Fluorescence emission intensities for several insertions are shown in Fig. 3 (i.e., sequence number 1 is the GFP, 2 is the background, and 3–60 are different sequences that are inserted to loops of GFP). The data is collected after

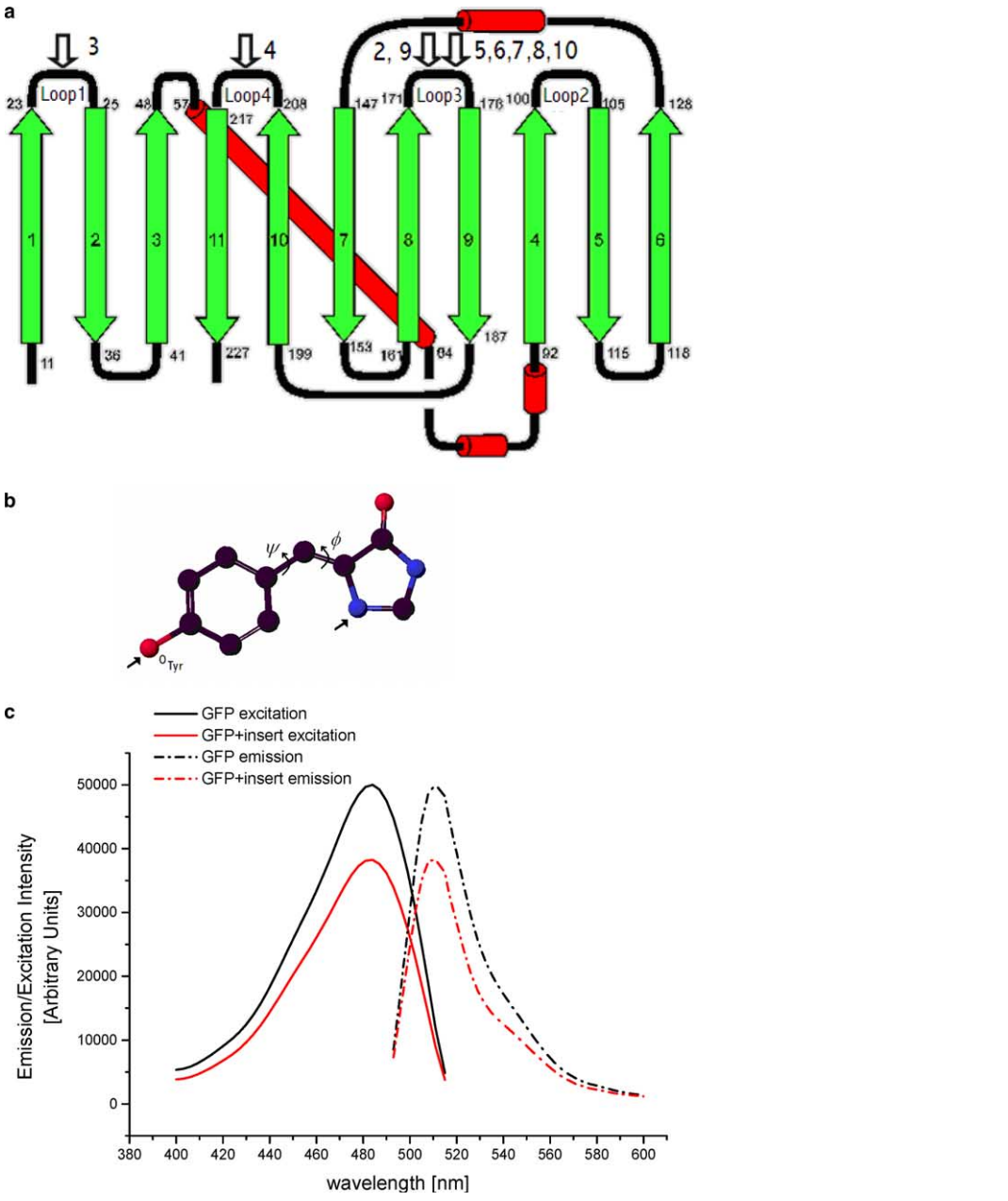

Fig. 2. Ten candidate sites for peptide insertions are shown. (a) Arrows and the numbers (bold) show the insertion sites and the locations listed in Table 1. The model chromophore structure is shown in the anionic state (b). The protonation sites of the anion are indicated by arrows in the model chromophore structure. Emission and excitation spectra for the GFP are measured with and without an insertion (c). The excitation and emission intensity change but the wavelength is constant.

the protein purification using a His_6-tag metal affinity column and the intensity is normalized with respect to concentration of GFP. We assumed that the fluorescence emission intensities lower than 10% of the GFP is non-fluorescent. In Fig. 3b, we showed the percentage of amino acids composition found in loop regions. Expected NNK values are also shown. Arginine and glycine are the most common amino acids observed in sequencing. However, they are commonly observed in both fluorescent and non-fluorescent GFP clones (Fig. 3c).

Ten highly fluorescent sequences (Table 1) are simulated using AMBER force-field for 1 ns at 300 K with explicit water. The overall RMS value (0.80 ± 0.01) is calculated on the backbone atoms. The RMS value is large; because the insertions cause local conformational changes along the backbone. A 10 ns simulation is also performed to ensure convergence (Fig. 1b).

The average structure of the chromophore is obtained from MD results for each insertion and the ZINDO method is used for absorption calculations. We calculated the absorption wavelength for the 10 insertions (Table 1). The results show that the excitation energy is 2.8 ± 0.1 eV for the anionic state. Standard deviation denotes the variation of excitation energy between

Table 2
The primers for random peptide insertions

Loop 1 (23)														
E	L	D	G	D	V	N	8X	G	H	K	F	S	V	R	
GAA	TTA	GAT	GGT	GAT	GTT	AAT	8NNK	GGG	CAC	AAA	TTT	TCT	GTC	AGA	GG
Loop 2 (102)														
R	T	I	S	F	K	D	18X	D	G	T	Y	K	T	R	
CGC	ACT	ATA	TCT	TTC	AAA	GAT	8NNK	GAC	GGG	ACC	TAC	AAG	ACG	CGT	G
Loop 3 (173)														
I	R	H	N	V	E	D	8X	G	S	V	Q	L	A	D	
ATT	CGC	CAC	AAC	GTT	GAA	GAT	8NNK	GGT	TCC	GTT	CAA	CTA	GCA	GAC	C
Loop 4 (213)														
L	S	K	D	P	N	E	8X	K	R	D	H	M	V	L	
CTT	TCG	AAA	GAT	CCC	AAC	GAA	8NNK	AAG	CGT	GAC	CAC	ATG	GTC	CTT	

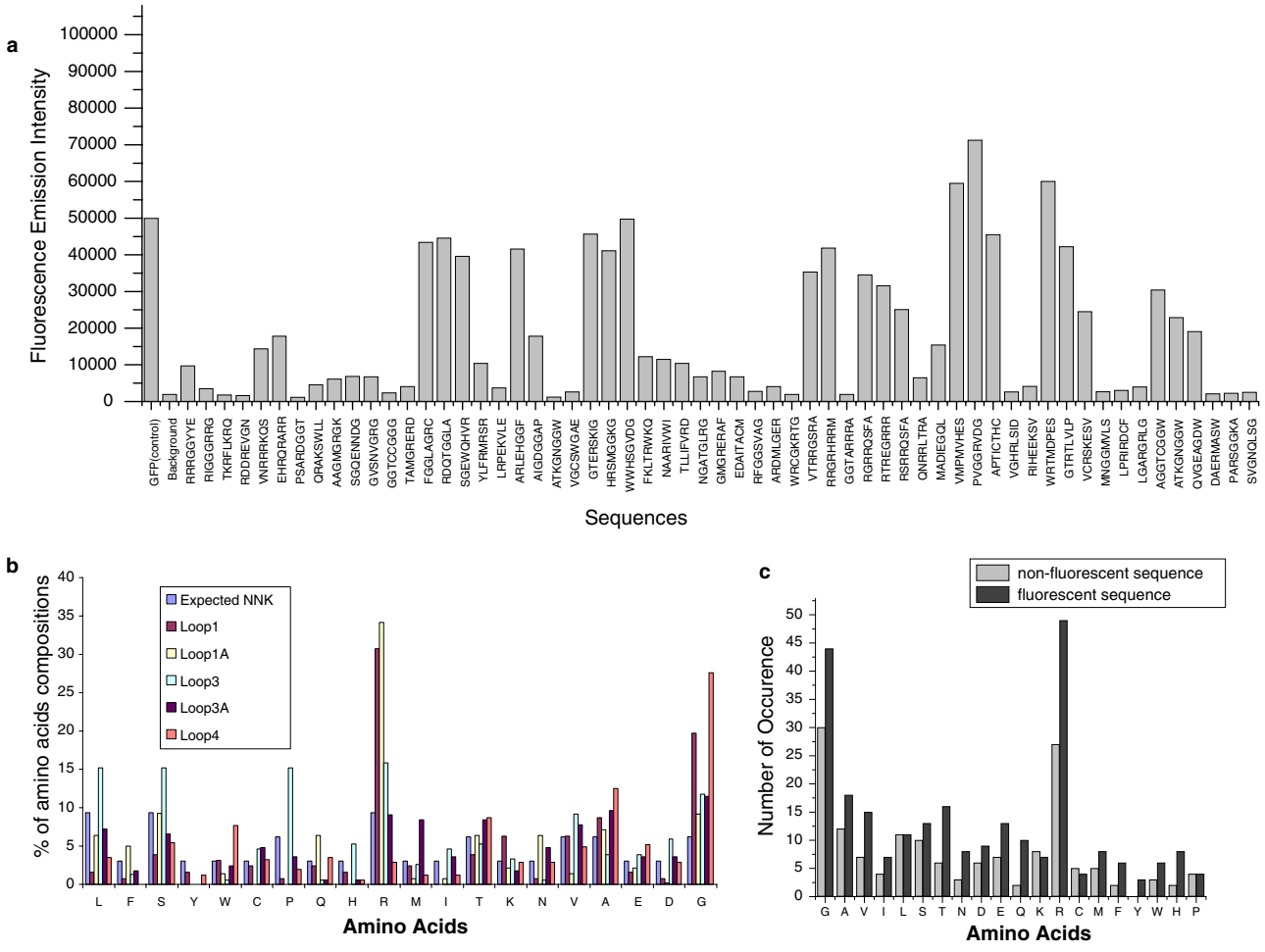

Fig. 3. Fluorescence emission intensity data are shown in (a): sequence number 1 is the GFP, 2 is the negative control, and 3–60 are different sequences that are inserted to the loops of GFP. Percentage of amino acids compositions in loop regions are plotted (b). The same percentage compositions are grouped into two sets: fluorescent and non-fluorescent (c). Arginine and glycine have the highest peaks.

insertions. The excitation wavelength shifts to red by 20 nm depending on the protonation of phenolic oxygen (66 Tyr) of the chromophore [27]. In addition, both computational and experimental results show that the absorption intensity varies with insertion sequence. The absorptions of the chromophore are calculated for the cation (3.1 ± 0.1 eV) and the zwitterion (2.4 ± 0.1 eV) states. It should be noted, that the accuracy of the excitation energy varies ($\sim$0.7 eV) based on the calculation method [24].

We also compared the dynamics of chromophore in the folded and unfolded conformation. It is known that

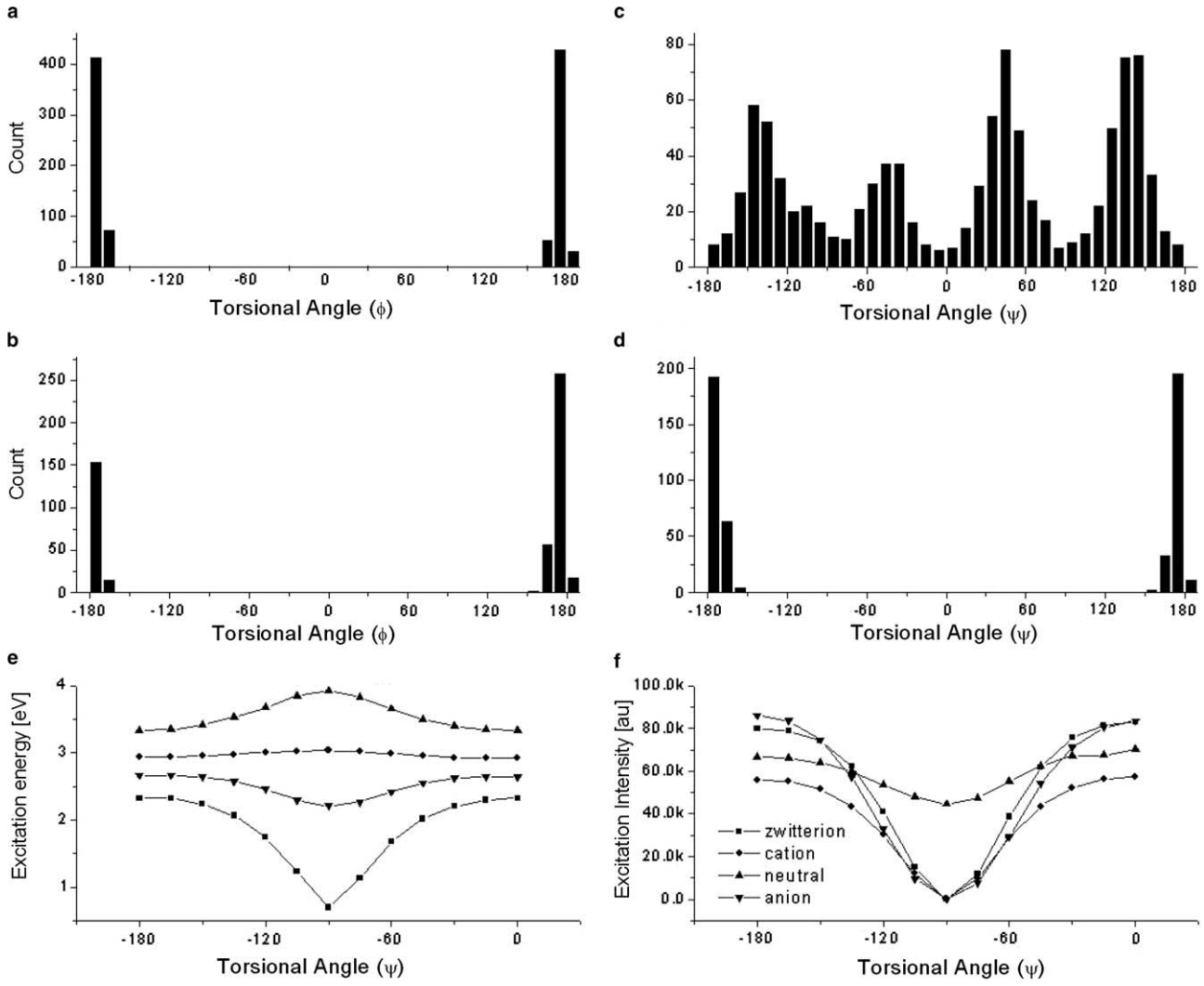

Fig. 4. The histogram for the torsional angles ϕ and ψ are shown for the unfolded-GFP (a,c) and the folded-GFP (b,d) respectively. Excitation energies (e) and intensities (f) are plotted for all possible rotations of the torsional angle ψ. Radiationless deactivation in GFP is favorable in the unfolded configuration for all four states (i.e. neutral, anion, cation and zwitterion).

unfolded GFP is non-fluorescent which indicates the importance of the β-barrel structure. Model compounds identical to the chromophore have been synthesized, but none fluoresced in solution [4]. Fig. 4 shows the histogram of the torsional angle ϕ and ψ (rotations about the two cyclic bonds connecting the two rings of the chromophore). In the folded case, both angles are constrained around their planar configuration (Fig. 4b,d). However, in the unfolded case, the torsional angle ψ visits all possible conformations (Fig. 4c). Both experimental and computational results have suggested that the torsional isomerization lead to a radiationless deactivation in GFP chromophore [28,29]. We have calculated the absorption energy and corresponding intensity (oscillator strength) for the torsional angle ψ (Fig. 4e,f). The intensity of absorption changes drastically with rotation. Our results indicate that the radiationless deactivation in GFP is favorable in the unfolded configuration for all four states (i.e. neutral, anion, cation and zwitterion) studied here.

4. Conclusion

We have studied emission and excitation spectrum of the GFP by inserting random peptide fragments into four loop sites. Conformation of the chromophore structure is obtained using molecular dynamics simulations and absorption characteristics are calculated by semi-empirical quantum mechanical methods. Fig. 5 shows a direct correlation of the excitation intensities between experimentally and computationally obtained results for GFP insertions. Both computational and experimental results show that random peptide insertions change the excitation and emission intensity of GFP but the wavelength of excitation and emission do not change based on the insertion. We plan to work on the understanding and characterization of non-fluorescent and fluorescent GFPs by experimental studies, such as CD spectra, isothermal titration and by computational studies, such as time dependent density functional theory.

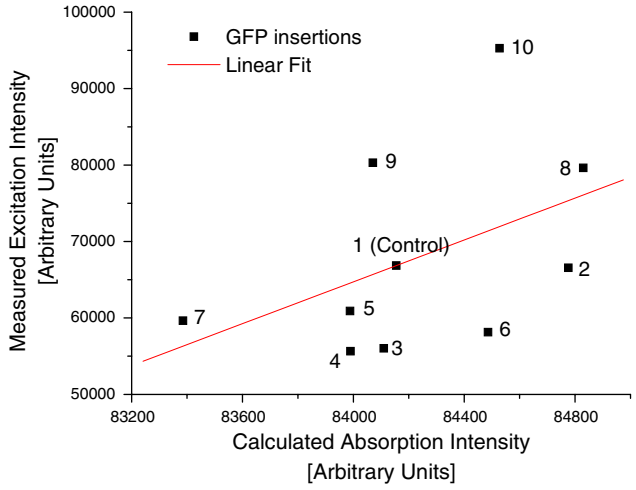

Fig. 5. Experimental and computational absorption intensities are plotted. Both results show that the random peptide insertions change the absorption intensity of the GFP.

Acknowledgments

This research is supported by Materials Simulation Center, a Penn State Center for Nanoscale Science (MRSEC-NSF) and MRI facility, Institute for Complex Adaptive Matter fellowship and Alexander von Humboldt fellowship (to M.C.D.). We thank Dr. Thomas Jovin at the Max Planck Institute for Biophysical Chemistry for discussions.

References

[1] R.Y. Tsien, Annu. Rev. Biochem. 67 (1998) 509.
[2] F.G. Prendergast, K.G. Mann, Biochemistry 17 (1978) 3448.
[3] C.W. Cody, D.C. Prasher, W.M. Westler, F.G. Prendergast, W.W. Ward, Biochemistry 32 (1993) 1212.
[4] H. Niwa, S. Inouye, T. Hirano, T. Matsuno, S. Kojima, M. Kubota, M. Ohashi, F.I. Tsuji, Proc. Natl. Acad. Sci. USA 93 (1996) 13617.
[5] A.B. Cubitt, R. Heim, S.R. Adams, A.E. Boyd, L.A. Gross, R.Y. Tsien, Trends Biochem. Sci. 20 (1995) 448.
[6] R. Heim, D.C. Prasher, R.Y. Tsien, Proc. Natl. Acad. Sci. USA 91 (1994) 12501.
[7] R. Nifosi, V. Tozzini, Proteins – Struct. Function Genet. 51 (2003) 378.
[8] T.M.H. Creemers, A.J. Lock, V. Subramaniam, T.M. Jovin, S. Volker, Chem. Phys. 275 (2002) 109.
[9] T.M.H. Creemers, A.J. Lock, V. Subramaniam, T.M. Jovin, S. Volker, Nature Struct. Biol. 6 (1999) 557.
[10] A. Volkmer, V. Subramaniam, D.J.S. Birch, T.M. Jovin, Biophys. J. 78 (2000) 1589.
[11] G.H. Patterson, S.M. Knobel, W.D. Sharif, S.R. Kain, D.W. Piston, Biophys. J. 73 (1997) 2782.
[12] A. Miyawaki, N.H. Takeharu, M. Hideaki, Curr. Opin. Chem. Biol. 7 (2003) 557.
[13] B. Peelle, J. Lorens, W.Q. Li, J. Bogenberger, D.G. Payan, D.C. Anderson, Chem. Biol. 8 (2001) 521.
[14] M.R. Abedi, G. Caponigro, A. Kamb, Nucleic Acids Res. 26 (1998) 623.
[15] D.A. Pearlman, D.A. Case, J.W. Caldwell, W.S. Ross, T.E. Cheatham, S. Debolt, D. Ferguson, G. Seibel, P. Kollman, Comput. Phys. Commun. 91 (1995) 1.
[16] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, J. Chem. Phys. 79 (1983) 926.
[17] M. Ormo, A.B. Cubitt, K. Kallio, L.A. Gross, R.Y. Tsien, S.J. Remington, Science 273 (1996) 1392.
[18] J. Dopf, T.M. Horiagon, Gene 173 (1996) 39.
[19] N. Guex, M.C. Peitsch, Electrophoresis 18 (1997) 2714.
[20] H.D. Höltje, W. Sippl, D. Rognan, G. Folkers, Molecular Modeling: Basic Principles and Applications, Wiley, 2003.
[21] M.C. Zemer, G.H. Loew, R.F. Kirchner, U.T. Miiller-Westerhof, J. Am. Chem. Soc. 102 (1980) 589.
[22] J.E. Ridley, M.C. Zemer, Theor. Chim. Acta 32 (1973) 111.
[23] T. Laino, R. Nifosi, V. Tozzini, Chem. Phys. 298 (2004) 17.
[24] S.C. Olsen, The Electronic Excited States of Green Fluorscent Protein Chromophore Models, University of Illinois, Urbana-Champaign, 2004.
[25] G.S. Waldo, B.M. Standish, J. Berendzen, T.C. Terwilliger, Nature Biotechnol. 17 (1999) 691.
[26] J. Sambrook, E.F. Fritsch, T. Maniatis, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press, New York, 2001.
[27] A.A. Voityuk, M.E. MichelBeyerle, N. Rosch, Chem. Phys. Lett. 272 (1997) 162.
[28] W. Weber, V. Helms, J.A. McCammon, P.W. Langhoff, Proc. Natl. Acad. Sci. USA 96 (1999) 6177.
[29] A.A. Voityuk, M.-E. Michel-Beyerle, N. Rosch, Chem. Phys. Lett. 296 (1998) 269.