PCCP

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/pccp

Molecular Modelling of the pH Influence in the Geometry and the Absorbance Spectrum of near-Infrared TagRFP675 Fluorescent Protein.

Carlos Randino[1], Ricard Gelabert[1], Miquel Moreno[1]*, José M. Lluch[1,2] and Kiryl D. Piatkevich[3]

[1]Departament de Química, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona (Spain)

[2] Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona (Spain)

[3] Departments of Biological Engineering and Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139

[*] Corresponding author at Universitat Autònoma de Barcelona

Tel: +34935812174; E-mail address: miquel.moreno@uab.cat

Abstract

Classical molecular dynamics (MD) simulations are carried out for the recently developed TagRFP675 fluorescent protein (FP), which is specifically designed to fully absorb and emit in the near infrared (NIR) region of the electromagnetic spectrum. Since the X-ray data of TagRFP675 reveals that the chromophore exists in both the *cis* and *trans* configuration and it can also be neutral (protonated) or anionic (deprotonated) depending on the pH of the media, a total of 8 molecular dynamic simulations have been run to simulate all the possible states of the chromophore. Time-dependent DFT (TDDFT) single point calculations are performed at selected points along the simulation to theoretically mimic the absorption spectrum of the protein. Our simulations compare well (within the expected error of the computational method) with the experimental results. Our theoretical procedure allows for an analysis of the molecular orbitals involved in the lowest energy electronic excitations of the chromophore and, more interestingly, for a full analysis of the H-bond interactions between the chromophore and its surrounding residues and solvent (water) molecules. This study does not support the hypothesis, exclusively based on the analysis of X-ray data, that isomerization of nearby residues provokes the rearrangement of the hydrogen bonds in the chromophore's immediate environment leading to the observed red shift of the absorption bands at higher pHs. Instead, we attribute this shift mainly to the superposition of bands of the neutral and anionic chromophore that are expected to coexist at almost the full range of pHs experimentally analyzed. An additional factor that could contribute to this shift is the experimentally observed increase of the *cis* configuration of the chromophore at higher pHs.

Keywords

Red fluorescent proteins. Electronic absorption spectrum. Time-dependent density functional theory. Molecular dynamics simulations.

Introduction

A wide variety of fluorescent proteins (FPs) are nowadays used in bioimaging to track localization and activity of proteins, organelles and cells of living organisms in real time.[1-3] Some of them are of natural origin, mainly found in marine organisms, but the majority of FPs are artificially designed by introducing mutations in order to have a particular property, usually a specific absorption/emission wavelength. The most popular and extensively studied family of FPs is the one derived from naturally occurring green fluorescent protein (GFP) whose structure consists of a chromophore deeply buried inside a protein barrel made of beta sheets.[4, 5] Even if the colors of FPs extend now over the entire visible electromagnetic spectrum, red FPs (RFPs) have risen in the last years as the most studied ones[6-8] because the red emission spectrum has lower signal from cellular autofluorescence and light-scattering intensity drops off at longer wavelengths. Furthermore, the less energetic red light causes less damage to tissues of living beings. An even more challenging task has recently been proposed to the experts in bioengineering as hemoglobin strongly absorbs below 600 nm[9] so most of the FPs are impractical in mammalian tissues as they are opaque to almost all the wavelengths of visible light. However there exist a so called optical window between 600 and 1350 nm, where infrared absorption by water begins, in which mammalian cells are relatively transparent to light.[10] This fact has focused the recent research in engineering FPs absorbing and emitting in the near infrared part of the electromagnetic spectrum.[11, 12]

Piatkevich et al. have recently reported on a new FP called TagRFP675[13] that promises to be a major breakthrough in the search for FPs absorbing in the optical window (here and below we refer to the range of wavelengths from 600 to 1350 nm). This protein was derived from the red FP mKate by the introduction of seven point mutations. TagRFP675 absorbs at 598 nm and emits at 675 nm at physiological pH, which makes it the most red-shifted protein of the GFP family known to date. The spectroscopic study under different conditions also revealed strong pH dependence of the absorption and emission maxima. Along with the reddest emission maximum, TagRFP675 is characterized by the Stokes shift, which is 30 nm extended comparing to that of its parental protein. Quite surprisingly, the relatively large Stokes shift reported of ca. 77 nm was not ascribed to an excited state proton transfer (ESPT) reaction between the chromophore and a side chain carboxylate group of the aspartic or glutamic amino acids close by, as is the typical case for other GFP-like red FPs with a large Stokes shift, such as LSSmKates[14] and mKeima.[15, 16] An extensive network of hydrogen bond interactions between the DsRed-like chromophore and the surrounding protein matrix has been revealed through X-ray diffraction analysis of TagRFP675. Modification of this surrounding upon electronic excitation has been suggested to be responsible for the unusual TagRFP675 spectral properties.

In this paper we present a theoretical study on the TagRFP675 structure and its absorption spectra. Departing from the recently published X-ray structures of TagRFP675 (entries 4KGE and 4KGF in the Protein Data Bank for the structures at pH=4.5 and 8.0 respectively)[13] we perform MD simulations of the protein in solution at

different pHs in order to analyze how the surrounding media affects the structure of its DsRed-like chromophore and the amino acid residues close by. Given that the X-ray data indicates that the chromophore adopts a mix of *cis* and *trans* configurations (with a proportion variable as a function of the pH) we performed MD simulations for both configurations at any given pH using a protocol previously applied with success to the mKeima and LSSmKate2 proteins.[17, 18] From the full MD results we are able to theoretically reproduce the absorption spectra of TagRFP675 at three different pHs (4.5, 7.5 and 10) that roughly cover the experimental range (from 3.5 to 10.5). This study will provide a considerable insight at a molecular level into the intricate chemistry operating in these kinds of FPs. We finally propose that this study could be a corner stone in a quite ambitious project of rational design of novel FPs with fluorescence spectra fully extending within the optical window.

Method of Calculation

Our methodology is based on two main blocks: classical MD simulations and electronic excited state TD/DFT calculations.

Classical MD Simulations

In order to thoroughly reproduce the electronic absorption spectra of TagRFP675, it is necessary to explore a wide set of representative structures of the protein that simulate the thermal fluctuations that the protein undergoes in a physiological environment. Therefore, the use of classical dynamic simulations is compulsory to furnish this ensemble of representative structures. To begin with, we used the two crystallographic protein structures available on the Protein Data Bank (4KGE at pH 4.5 and 4KGF at pH 8.0) as our starting point. Although these crystallographic structures are two possible conformations of energy minima, it is only a tiny fraction of the total conformational space of the protein. Using classical MD we will also observe the dynamical stability of these structures.

Both crystal structures show two subunits, A and B. Subunit A presents only the *cis* conformation of the chromophore. However, subunit B presents both the *cis* and *trans* chromophore conformations in a 50/50 *cis/trans* ratio in the acidic structure and 60/40 ratio in the near-neutral structure.[13] For our work, we have chosen the B subunit as it contains both the *cis* and *trans* structures and have created two individual versions of TagRFP675; one with the *cis* isomer and the other with the *trans* isomer. Near the chromophore, residue Arg197 also shows two possible conformations in a complementary ratio with regard to the chromophore. We have chosen the corresponding side chain position of Arg197 according to the selected chromophore conformation to avoid nearby atoms to be placed unphysically close.

Since the low X-ray resolution does not detect hydrogen atoms, the protonation states of these amino acid residues in these systems were predicted theoretically using the PROPKA program,[19, 20] estimating the corresponding pK_a of each amino acid at any given pH (4.5, 7.5 and 10). As for the chromophore, we used a slightly different

procedure because it is not a standard amino acid and the PROPKA program does not recognize it. Therefore, its protonation state at a given pH was determined based on the protonation states of the DsRed-like chromophores in other FPs. It has to be taken into account that at pH equal to 4.5, the chromophore is protonated to a larger extent as the pK_a of the chromophore equals 5.7.[13] All the previous steps generated 8 different systems. MD simulations have been carried out for each of these 8 systems as itemized in Table 1.

To adapt the systems to physiological conditions, the CHARMM package software (c35b1 version)[21, 22] has been used together with the CHARMM-22 force field.[23-25] Special care has been paid to choose the chromophore parameters that have been utilized from the GFP parameterization done by W. Thiel and co-workers[26] and adapted by Lluch and co-workers to DS-red proteins´ chromophores.[17] The crystallographic structures have been solvated and neutralized, introducing water solvent molecules forming a cubic box of 72 Å of length of the edge and the necessary ions to cancel out the excess charge. The MD simulations of the full system consist of roughly 40,000 atoms. After solvation and neutralization, short minimizations of about 400 picoseconds (ps) using Adapted Basis Newton-Raphson algorithm were carried out to eliminate bad atom contacts between the protein and the recently added water molecules. Afterwards, the temperature of the system was increased, using the Velocity Verlet integrator algorithm with an integration step of 1 femtosecond for 120 ps, raising the temperature 25 K every 10 ps up to 300K and applying structural harmonic forces at the protein backbone. At this point, an equilibration process is carried out gradually freeing the restricted harmonic forces applied during 320 ps. 680 ps more have been run without any harmonic constraint. Finally, 10 nanoseconds of production time have been obtained for each simulation.

Electronic Excited State Calculations

The electronic excited state calculations are fundamental in the process of reproducing the electronic absorption spectrum because the method allows us to calculate the energy difference between the ground and excited states and thus, the wavelength associated with this transition.

For each complete MD simulation, a structure has been captured every 0.25 ns of the production time, gathering 400 frames in total. A sphere of 20 Å of radius has been applied to each one of these structures, taking the chromophore as the center of the sphere and reducing the number of atoms to 7,100. This highly decreases the computational cost and the system is still representative because the chromophore and its close environment remain almost unaffected. Each sphere, in turn, has been divided into two regions with different methodological approaches: the first one is the chromophore with the carbonyl group of the previous amino acid (Phe62) and the amino group of the next amino acid (Ser66). The chromophore is responsible for the radiation process, so a highly detailed treatment, namely quantum mechanics, has been applied to this part, working with the Time-Dependent/Density Functional Theory (TDDFT)

formalism, a useful and adequate method to treat this kind of systems with photo optical features. The first 15 excited electronic states have been computed for each of the 400 structures. The CAM-B3LYP functional[27] which properly describes long-range charge transfer effects[28-30] and the 6-31+G(d,p) basis set complete the set. The second region consists of the rest of the atoms within the considered sphere and it has been treated with molecular mechanics, considering all the atoms as point charges according to the CHARMM-22 force field. This adds a polarization effect to the chromophore. CHEMSHELL[31] and Gaussian 09 software packages[32] have been used to prepare all the QM/MM inputs and to compute the QM part, respectively.

The procedure of reproducing the electronic absorption spectrum is similar to that presented in previous papers.[33-36] The full set of 15 calculated electronic excited states have been sorted out with their corresponding oscillator strength and distributed according to their wavelength. All the transitions irrespective of their nature have been included. For the sake of clarity the cumulative results are presented as convoluted spectra as was already done in our previous works on other red FPs.[17, 18]

A similar procedure could be used to theoretically reproduce the fluorescence spectrum of TagRFP675. To do so a long enough MD simulation should be carried out for the excited electronic state. Unfortunately, as the classical force fields are parametrized for the ground electronic state, a QM or QM/MM scheme should be adopted and this would make the calculation of the trajectory unattainable as the computational cost of such a calculation well exceeds our present computer capabilities. In any case this is a very important issue (as the more interesting property of FPs is just that: the fluorescence!) and we are at work in order to be able to simulate fluorescence spectra in the near future.

Results and Discussion

In order to understand the subtle chemical interactions that operate in the active center of the FP, we have carried out MD simulations of different structures of TagRFP675 as explained in the previous section. The first goal of these simulations is to reproduce the absorption spectra of TagRFP675 at different pHs and compare with the spectra obtained by Piatkevich et al. at these pHs.[13] Figure 1 shows the obtained results at pH=4.5, 7.5 and 10.5 for both *cis* (left side of the Figure) and *trans* (right side of the Figure) configurations of the chromophore. As the experimental results indicate, at pH=4.5 the chromophore presents an equilibrium between its protonated (neutral) and deprotonated (anionic) forms. We have therefore performed simulations for both protonation states. At higher pH the chromophore is expected to be almost exclusively in its anionic form so at these pHs only the anionic chromophore has been considered. In order to facilitate the comparison with the experimental results previously reported,[13] Figure 1i depicts the experimental absorption spectra of TagRFP675 at different pHs. In any case it has to be remarked that a direct comparison with our simulated spectra is not feasible as the experimental results come from a mix of different conformations (*cis* and *trans*) and protonation states (neutral and anionic) of the chromophore inside the

protein. What we will do in the following discussion is to compare the position of the observed bands in each spectrum. In that sense, it is crucial to note that there is a grand total of three different absorption bands seen in the experimental spectra at the different pHs: The high energy band (360-390 nm) ascribed to the $S_0 \rightarrow S_2$ electronic transition, present in all the protonation states of the chromophore, the low energy bands of the neutral from of the chromophore (400-500 nm) and the one corresponding to the anionic form (550-650 nm). To be noted that as we are analyzing pure protonation states of the chromophore we can only find two of these bands in each simulated spectrum.

Disregarding the small absorption band originating from transitions to high excited states that appears at $\lambda < 300$ nm and is outside the wavelength zone experimentally considered, all the spectra shown in Figure 1 present two main absorption bands. In all the cases but one the most intense band is the most energetic one peaking at 390-410 nm when the chromophore is in its anionic form and 340 nm for the neutral chromophore. While the position of this band is in quite good agreement with the experimental spectra, where this band appears at 360-390 nm,[13] in our theoretical prediction it is also the more intense one whereas in the recorded absorption spectra (Figure 1i) it is, by far the more tenuous one. It is also to be noted that this band is only experimentally seen at pHs of 6 and higher as at most acidic media the very broad lowest energy band of the neutral chromophore peaking at 450 nm overlaps with this band as also seen in Figure 1i.

As for the lowest energy band, it is well known that its position depends on the protonation state of the chromophore. For GFP-like proteins the protonated (neutral) form of the chromophore absorbs at much lower wavelengths than the anionic form. In tagRFP675 the lowest energy band of the neutral chromophore peaks at 460 nm (see Figure 1i). From the results obtained for the *cis* structure at pH=4.5 (Figure 1a) this band appears largely spread between 400 and 500 nm (with the maximum at 450 nm). Similar results are found for the simulation corresponding to the *trans* structure (Figure 1b) though the band is now somewhat shifted to the blue with a maximum centered at 420 nm, so again there is a nice agreement between theory and experiment regarding the position of this band.

As just said above, the experimental spectra suggests that almost exclusively the anionic form of the chromophore exists at pH > 6. In TagRFP675 the maxima of absorption of the anionic chromophore is found around 580-602 nm (Figure 1i). Our theoretical simulations at pH=7.5 (third row in Figure 1) display a very broad band extending between 500 and 620 nm and peaked at 540 nm for the *cis* chromophore (Figure 1e). This indicates that the lowest energy absorption of the chromophore is very sensitive to the fluctuating environment. Results are similar for the *trans* structure (Figure 1f) but again slightly blue-shifted as the maximum is found at 530 nm. The larger discrepancy between theory and experiment found here is still within the expected range of error of TDDFT calculations (±0.5 eV that at this range of energy corresponds to ~ 150 nm). Results at the highest considered pH of 10 (Figure 1g and 1h

for the *cis* and *trans* configurations respectively) are almost invariant with respect to the ones just discussed at pH=7.5.

Our theoretical procedure allows for the identification and analysis of the molecular orbitals involved in the electronic excitations that are responsible for the absorption spectra. Figures 2 and 3 depict these orbitals for a set of arbitrarily chosen structures from the MD simulation of the *cis* chromophore at pH=4.5 both at its neutral (Figure 2) and anionic (Figure 3) forms. For each case two structures have been selected along the whole simulation, one having a maximum value of the oscillator strength contributing to the low-energy band of the absorption spectrum and the other one mainly contributing to the high-energy band of the corresponding absorption spectrum. Corresponding results for the *trans* chromophore at pH=4.5 are shown in Figures S1 and S2. The orientation of the chromophore used in depicting the molecular orbitals is seen in panel (a) of each Figure while panels (b) and (c) depict the shape of the orbitals mainly responsible of the low- and high-energy absorption bands respectively. Qualitatively the simulations at other pHs show no significant changes in the shape of the molecular orbitals responsible of the lowest energy electronic excitations. The electronic excitation for the low energy band of the absorption spectrum usually involves the HOMO and LUMO whereas the high energy band mainly originates from an electronic excitation from the HOMO to either the LUMO, LUMO+1 or LUMO+2, depending on the randomly chosen structure. At any rate the relative order of the virtual orbital involved is irrelevant as its position changes along the whole simulation but its shape remains qualitatively invariant.

Not surprisingly, all the involved molecular orbitals are of π type and spread over the whole chromophore. A careful analysis of these orbitals discloses that the low energy transition involves a certain degree of charge transfer from the phenolate group to the imidazolinone ring and, more specifically, to the N-acylimine group of the chromophore. This is in agreement with the (expected) effect of the N-acylimine group of extending the π system of the chromophore and so lowering the energy of the first singlet excited electronic state. As for the high-energy transition, it also has a slight charge transfer character from the phenolate group mainly to the residues directly bound to the imidazolinone ring.

A more subtle effect amenable to be theoretically analyzed is the pH-dependence of the low energy band in the absorption spectrum of the anionic chromophore. It has been experimentally observed that this band undergoes red-shift from 570 nm at pH=3.5 to 602 nm at pH=11. Our results are unable to quantitatively reproduce this shift as its range lies within the expected error of the theoretical method, as described above. On the other hand, from the comparison of our theoretical simulations the *cis* chromophores have this band displaced to slightly higher (ca. 20 nm) wavelengths at all considered pHs.

From the direct inspection of the X-ray geometries, this displacement has been hypothesized to mainly arise from two processes:[13] a) protonation-deprotonation of the

chromophore and, b) isomerization of the Asn143, Asn158 and Arg197 residues resulting in a substantial rearrangement of the hydrogen bonds near the phenolic side of the chromophore The MD simulations performed here that extend for a 10 ns time period allows for a deep analysis of the geometrical variations of the whole protein giving us an opportunity to analyze the H-bond interactions of the chromophore with the immediate protein environment for each analyzed case and so verify the rearrangement hypothesis. Generally, we found that the only residues that make stable H-bonds with the phenolic oxygen of the chromophore (and/or the hydrogen when the chromophore is neutral) are the two asparagines Asn143 and Asn158. Arg197 does not directly interact with the chromophore (though it is usually nearby) so the role of this residue has been disregarded at first instance. As noted below, there are always one or more solvent water molecules making occasional H-bonds with the phenolic side of the chromophore. Figure 4 depicts the analyzed interactions for both the *cis* and *trans* conformations being the chromophore either neutral or anionic. For all the cases a randomly chosen configuration along the whole simulation has been picked up. In the following we will analyze the distance between the donor atom and the acceptor atom of the considered H-bond interactions along the whole simulation using the criteria that a H-bond is formed when the distance between the two heavy atoms lies within a range of 2-3 Å.

The evolution of the H-bond distances between the chromophore and the Asn143 and/or Asn158 residues along the MD simulation at pH=4.5 is given in Figures 5 and 6. In these Figures, and the following ones dealing with the H-bond distances along the MD simulations, the different parts of the chromophore that interact with the surrounding residues and water molecules are indicated at the top of each Figure. The results with the protonated (neutral) chromophore in either its *cis* or *trans* conformations are analyzed in Figure 5 whereas the simulations with the anionic (deprotonated) *cis* and *trans* conformers are considered in Figure 6. The H-bond interactions with the different solvation waters are depicted in Figures S3 to S6. In all the cases the interactions involve the phenolic oxygen of the chromophore (labeled as $O_{(front)}$ in panel (a) of Figure 2,3, S1 and S2) and its bonded H-atom if the chromophore is neutral. In the neutral *cis* conformation there is a quite stable H-bond between $O_{(front)}$ and the NH_2 fragment of Asn143 (Figure 5a). This interaction is only shortly disrupted between 5.5 and 6.5 ns. The oxygen of the very same Asn143 also interacts with the phenolic hydrogen of the chromophore but this interaction (quite strong and stable at the beginning of the simulation) breaks down at around 1 ns and it is only recovered for very short periods of time along the rest of the simulation as seen in Figure 5b. Along the whole simulation the most habitual interaction of this hydrogen is with a given solvation water molecule, a molecule that changes along time as seen in Figure S3. Conversely, the simulation of the *trans* isomer only features a very stable and strong interaction between the phenolic hydrogen of the chromophore and the oxygen of Asn158 as depicted in Figure 5c. As for $O_{(front)}$, it is stabilized by different water molecules that alternate along the whole simulation (Figure S4).

When the chromophore is deprotonated, the phenolic oxygen is formally negatively charged so that it should make stronger H-bonds with the surrounding residues. For the *cis* isomer the $O_{(front)}$ has two well-formed hydrogen bonds that are seen along the whole simulation: One with a solvation water molecule that changes along time (Figure S5) and the second one with the NH_2 fragment of Asn143 (Figure 6a). As for the *trans* isomer, $O_{(front)}$ also interacts with an interchangeable water molecule (Figure S6) and with the NH_2 group of an asparagine group, now Asn158, but curiously enough, at around 1.5 ns of simulation the two hydrogen atoms of the amino group of Asn158 interchange their role as seen in Figures 6b and 6c.

Qualitatively the same picture arises when analyzing the simulations at pH=7.5 and 10 where only the anionic chromophore situation has been taken into account. Figure 7 shows that for the *cis* isomers, at both pHs the same (Cro)O···H-N(Asn143) interaction, even stronger (short O···H distance) than at pH=4.5, is clearly present along the whole simulation. Also different water molecules make short lived H-bonds with $O_{(front)}$ at these pHs (Figures S7 and S8). As for the *trans* isomers (Figure 8) there is also a (Cro)O···H-N(Asn) interaction present all along the simulations but at pH=7.5 the Asn143 residue, that is initially involved in such a H-bond, switches role with Asn158 at t~6.5 ns and this new interaction is maintained until the end of the simulation. On the other hand, only Asn158 is involved in the H-bond interaction with $O_{(front)}$ at pH=10. As usual, several solvation waters may occasionally interact with this oxygen. These interactions are seen in Figures S9 and S10.

The interactions with the surrounding media affecting other parts of the chromophore are apparently of less interest as no major differences in the H-bond network are seen in the X-ray analysis of the structures at different pH. In any case, the X-ray structures reveal that there are two carbonyl groups of the chromophore that are able to form H-bonds with the surrounding media: the carbonylic group of the acylimine end (formally the Phe62 group) and the oxygen atom directly bound to the nitrogen atom at the end the imidazolinone ring (labeled as $O_{(back)}$ in panel (a) of Figures 2,3 S1 and S2). Figure 9 depicts a set of structures randomly chosen from the corresponding MD simulations at pH=4.5 where these "back" H-bond interactions are highlighted. The time evolution of the relevant distances is shown in Figures 10 (neutral) and 11 (anionic). Analogously to what was obtained for the "front" (the phenolic group) H-bond interactions, several water molecules appear intermittently making occasional H-bonds with these two carbonyl groups. These interactions are depicted in Figures S11 to S14. Quite similar results are obtained at the other considered pHs so that for clarity purposes these additional results will not be presented here.

In the neutral *cis* configuration (Figure 10a) the H-bond interaction between Phe62 and Gln41 groups is clearly seen and is stable over all the simulation time. This interaction is also present in all the X-ray structures of the TagRFP675 chromophore resolved at different pHs. Our simulations indicate that an additional H-bond is formed between the carbonyl group of the chromophore and a water molecule: 7401 at the first half of the simulation and 8015 for the second half (Figure S11a-b). As for $O_{(back)}$, it only makes

variable H-bonds with different water molecules as indicated in Figure S11c-g (in the particular case of the snapshot captured by Figure 9a this carbonyl group interacts with water 230). For the neutral *trans* isomer (Figure 10b) the interaction of Phe62 with the Gln41 is also seen but the O-O distance is now somehow larger in average and another glutamine (Gln106) is also interacting with Phe62 in this case (Figure 10c). These two interactions compete with a third one involving a solvent water molecule (W9109). In what concerns $O_{(back)}$, it displays a quite strong and stable H-bond with Trp90 as shown in Figure 10d. Additionally, a water molecule (W266) makes a strong H-bond with $O_{(back)}$ at the first stage (3ns) of the simulation as clearly seen in Figure S12b.

As the change in the protonation state of the chromophore does not affect these two carbonyl groups, one should not expect large modifications in the interaction pattern for the anionic chromophore simulations and, in fact the strong interaction between Phe62 and Gln41 is clearly present also in the *cis* and *trans* anionic chromophore simulations (depicted in Figures 11a and d respectively). Again in the *trans* structure this interaction is weaker and Gln106 (Figure 11e) and one solvation water molecule (Figure S14a) have also noticeable interactions with the Phe62 part of the chromophore. As for the interaction of $O_{(back)}$ with the Trp90 residue, now it appears in both *cis* and *trans* isomers (Figures 11c and f respectively). Curiously enough, for the *cis* anionic chromophore the Gln106 residue also strongly interacts (at the end of the simulation time) with this carbonyl group as shown in Figure 11b.

All these results are in agreement with the analysis of the X-ray structures previously performed by Piatkevich et al.[13] Generally we found larger distances for the H-bond interactions but this is to be expected when comparing the "frozen" X-ray interatomic distances with data in solution. It is clear that the very subtle differences just noticed for the simulations at different pHs do not give explicit support to the previously suggested hypothesis that these rearrangements are the main reason behind the measured red-shift of the anionic absorption band at progressively more basic pHs. At this point some important remarks are in order:

1) The interactions of residues Asn143 and/or Asn158 with the phenolic oxygen of the anionic chromophore are not substantially modified at the different analyzed pHs (see Figures 6 to 8) so that these interactions may not be held responsible for the observed displacements in the absorption spectrum of TagRFP675 upon changes in the pH of the surrounding media.
2) At a given pH, the experimentally obtained absorption spectrum is the result of a combination of the neutral and anionic absorption bands and so, it cannot be directly compared with our simulations of "pure" protonated or deprotonated chromophore. The superposition of the two species at intermediate pHs might produce bands that are actually a combination of the pure neutral and anionic bands that overlap somehow by means of their respective queues. This overlap may produce a shift of the observed anionic band to higher wavelengths at higher pHs (as the protonated chromophore becomes less abundant). This would also explain in theoretical grounds why experimentally the red-shift of the

anionic band stops beyond pH=9 as the protonated chromophore would cease to be present at such basic solutions.

3) Invoking the superposition of the absorption bands of the different species present at a given pH, the experimentally measured slight increase of the *cis* chromophore at higher pHs could also contribute to the observed red-shift of the anionic band.

Conclusions

Classical MD simulations have been carried out for the recently developed NIR FP TagRFP675. TagRFP675 is an important step in engineering FPs that absorb and emit in the so called "optical window" (from 650 to 1350 nm). Our extensive MD simulations allows for a realistic simulation of the absorption spectra of this protein. The X-ray analysis has revealed that the chromophore of TagRFP675 may exist in both *cis* and *trans* configurations. Therefore we have performed MD simulations for both configurations. Also, as the change of the pH modifies the protonation state of the chromophore, we have had to consider both the neutral (protonated) and anionic (deprotonated) states in our simulations. Taking all this into account 8 MD runs have been obtained. Each MD simulation has been followed up to 10 ns. The first goal of these simulations was to theoretically reproduce (using a methodology previously applied to other RFPs) the absorption spectrum of TagRFP675. Results are in agreement with the actual spectra of TagRFP675 within the expected error bar of the TDDFT methodology employed to calculate the energy of the excited states.

Our theoretical calculation has allowed for a detailed analysis of the molecular orbitals involved in the lowest energy electronic excitations clearly pointing to an active role of the N-acylimine group of the chromophore in lowering the energy of the first singlet excited electronic state. A very extensive analysis of the H-bond interactions between the chromophore and the surrounding residues and solvent (water) molecules has been carried out in order to analyze the hypothesis (exclusively suggested from the analysis of the X-ray data) that isomerization of the Asn143, Asn158 and Arg197 residues provokes a substantial rearrangement of the hydrogen bonds in the phenolic region of the chromophore leading to the red shift of the absorption bands at higher pHs. Our results do not explicitly provide support to this rearrangement hypothesis, but we are able to explain the experimental results by assuming that at intermediate pHs the neutral and anionic forms of the chromophore will coexist and the measured absorption spectrum in fact will be the result of a superposition of the different bands of the two species that will overlap so that the shift of the maximum of absorption to higher wavelengths at higher pHs may mainly arise because the protonated chromophore becomes less abundant as the pH increases. A factor that additionally may contribute to this shift is the slight increase (seen experimentally) of the *cis* chromophore at higher pHs.

We believe that the quite exhaustive theoretical analysis of TagRFP675 is just the first step in a more ambitious project aiming for a fully computer based (or "in silico")

engineering of novel FPs possessing desired spectral and biochemical characteristics. One of the holy grails of mutagenesis applied to FPs is the design of new proteins fully absorbing and emitting in the NIR region of the electromagnetic spectrum. The usually long path to such an achievement can be paved down by the theoretical procedures such as the ones used in this paper so that random mutagenesis can be replaced by a kind of "theoretically assisted" mutagenesis.

Acknowledgments

This work was supported by the ''Ministerio de Economía y Competitividad'' through projects CTQ2011-24292 and CTQ2014-53144-P. Use of computational facilities at the Consorci de Serveis Científics i Acadèmics de Catalunya (CSUC) is gratefully acknowledged.

Electronic Supplementary Information (ESI) Available: Figures S1 and S2: show the shape of the molecular orbitals involved in the relevant electronic transitions of TagRFP675 for the *trans* chromophore (neutral and anionic). Figures S3 to S14 depict the relevant H-bond interactions of the chromophore with solvent water molecules at different pHs.

References

1. D. M. Chudakov, M. V. Matz, S. Lukyanov and K. A. Lukyanov, *Physiol Rev*, 2010, 90, 1103-1163.
2. O. V. Stepanenko, V. V. Verkhusha, I. M. Kuznetsova, V. N. Uversky and K. K. Turoverov, *Curr Protein Pept Sci*, 2008, 9, 338-369.
3. B. Wu, K. D. Piatkevich, T. Lionnet, R. H. Singer and V. V. Verkhusha, *Curr. Opin. Cell Biol.*, 2011, 23, 310-317.
4. M. Ormo, A. B. Cubitt, K. Kallio, L. A. Gross, R. Y. Tsien and S. J. Remington, *Science*, 1996, 273, 1392-1395.
5. F. Yang, L. G. Moss and G. N. Phillips, *Nat. Biotechnol.*, 1996, 14, 1246-1251.
6. K. D. Piatkevich, J. Hulit, O. M. Subach, B. Wu, A. Abdulla, J. E. Segall and V. V. Verkhusha, *Proc. Natl. Acad. Sci. USA*, 2010, 107, 5369-5374.
7. D. M. Shcherbakova, O. M. Subach and V. V. Verkhusha, *Angew. Chem. Int. Edit.*, 2012, 51, 10724-10738.
8. F. V. Subach and V. V. Verkhusha, *Chem. Rev.*, 2012, 112, 4308-4327.
9. G. N. Stamatas, M. Southall and N. Kollias, *J. Invest. Dermatol.*, 2006, 126, 1753-1760.
10. B. J. Tromberg, N. Shah, R. Lanning, A. Cerussi, J. Espinoza, T. Pham, L. Svaasand and J. Butler, *Neoplasia*, 2000, 2, 26-40.
11. M. Z. Lin, M. R. McKeown, H. L. Ng, T. A. Aguilera, N. C. Shaner, R. E. Campbell, S. R. Adams, L. A. Gross, W. Ma, T. Alber and R. Y. Tsien, *Chem Biol*, 2009, 16, 1169-1179.
12. D. Shcherbo, Shemiakina, II, A. V. Ryabova, K. E. Luker, B. T. Schmidt, E. A. Souslova, T. V. Gorodnicheva, L. Strukova, K. M. Shidlovskiy, O. V. Britanova, A. G. Zaraisky, K. A. Lukyanov, V. B. Loschenov, G. D. Luker and D. M. Chudakov, *Nat. Methods*, 2010, 7, 827-U1520.
13. K. D. Piatkevich, V. N. Malashkevich, K. S. Morozova, N. A. Nemkovich, S. C. Almo and V. V. Verkhusha, *Sci Rep*, 2013, 3:1847.
14. K. D. Piatkevich, V. N. Malashkevich, S. C. Almo and V. V. Verkhusha, *J. Am. Chem. Soc.*, 2010, 132, 10762-10770.

15. J. N. Henderson, M. F. Osborn, N. Koon, R. Gepshtein, D. Huppert and S. J. Remington, *J. Am. Chem. Soc.*, 2009, 131, 13212-13213.
16. S. Violot, P. Carpentier, L. Blanchoin and D. Bourgeois, *J. Am. Chem. Soc.*, 2009, 131, 10356-10357.
17. M. Nadal-Ferret, R. Gelabert, M. Moreno and J. M. Lluch, *J. Chem. Theory Comput.*, 2013, 9, 1731-1742.
18. C. Randino, M. Nadal-Ferret, R. Gelabert, M. Moreno and J. M. Lluch, *Theor Chem Account*, 2013, 132:1327.
19. T. J. Dolinsky, P. Czodrowski, H. Li, J. E. Nielsen, J. H. Jensen, G. Klebe and N. A. Baker, *Nucleic Acids Res*, 2007, 35, W522-W525.
20. T. J. Dolinsky, J. E. Nielsen, J. A. McCammon and N. A. Baker, *Nucleic Acids Res*, 2004, 32, W665-W667.
21. B. R. Brooks, C. L. Brooks, A. D. MacKerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York and M. Karplus, *J Comput Chem*, 2009, 30, 1545-1614.
22. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, *J Comput Chem*, 1983, 4, 187-217.
23. A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin and M. Karplus, *J. Phys. Chem. B*, 1998, 102, 3586-3616.
24. A. D. MacKerell, M. Feig and C. L. Brooks, *J. Am. Chem. Soc.*, 2004, 126, 698-699.
25. A. D. MacKerell Jr, B. Brooks, C. B. Brooks III, L. Nilsson, B. Roux, Y. Won and M. Karplus, in *Encyclopedia of Computational Chemistry*, ed. J. W. Sons, P. v.R.Schleyer, N.L Allinger, T. Clark, J. Gasteiger, P.A. Kollman, H.F. Schaefer III, P.R. Schreiner, Chichester, 1998, vol. 1, pp. 271-277.
26. N. Reuter, H. Lin and W. Thiel, *J. Phys. Chem. B*, 2002, 106, 6310-6321.
27. N. C. Handy, *Mol. Phys.*, 2004, 102, 2399-2409.
28. C. Randino, M. Moreno, R. Gelabert and J. M. Lluch, *J. Phys. Chem. B*, 2012, 116, 14302-14310.
29. Y. Tawada, T. Tsuneda, S. Yanagisawa, T. Yanai and K. Hirao, *J. Chem. Phys.*, 2004, 120, 8425-8433.
30. M. Y. Zhang, J. Y. Wang, C. S. Lin and W. D. Cheng, *J. Phys. Chem. B*, 2011, 115, 10750-10757.
31. P. Sherwood, A. H. de Vries, M. F. Guest, G. Schreckenbach, C. R. A. Catlow, S. A. French, A. A. Sokol, S. T. Bromley, W. Thiel, A. J. Turner, S. Billeter, F. Terstegen, S. Thiel, J. Kendrick, S. C. Rogers, J. Casci, M. Watson, F. King, E. Karlsen, M. Sjovoll, A. Fahmi, A. Schafer and C. Lennartz, *J Mol Struc-Theochem*, 2003, 632, 1-28.
32. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, N. J. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A.

D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, Gaussian, Inc., Wallingford, CT, USA, 2009.
33. P. Imhof, *J. Chem. Theory Comput.*, 2012, 8, 4828-4836.
34. C. M. Isborn, A. W. Gotz, M. A. Clark, R. C. Walker and T. J. Martinez, *J. Chem. Theory Comput.*, 2012, 8, 5092-5106.
35. E. Sanchez-Garcia, M. Doerr and W. Thiel, *J Comput Chem*, 2010, 31, 1603-1612.
36. P. Armengol, R. Gelabert, M. Moreno and J. M. Lluch, *Org. Biomol. Chem.*, 2014, 12, 9845-9852.

		cis	***trans***
pH	***4.5***	*protonated*	*protonated*
		deprotonated	*deprotonated*
	7.5	*deprotonated*	*deprotonated*
	10	*deprotonated*	*deprotonated*

Table 1: Scheme of the analyzed structures at any considered pH.

FIGURE CAPTIONS

Figure 1: Representation of the simulated absorption spectra of TagRFP675 in its *cis* (left) and *trans* (right) conformation (as indicated in the Figure) at different pHs. In all cases a convoluting Gaussian function of σ = 15 nm has been used to smooth the statistical noise. (a) and (b) in yellow background stand for the neutral chromophore at pH=4.5. The rest of the depicted spectra in green background correspond to the anionic chromophore at different pHs: 4.5 for (c) and (d); 7.5 for (e) and (f) and 10 for (g) and (h). These pHs are also indicated in the Figure. Finally panel (i) depicts the experimental absorption spectra of TagRFP675 at different pHs (the color code is indicated in the Figure).

Figure 2: Shape of the molecular orbitals involved in the relevant electronic transitions of the TagRFP675 protein as obtained from the simulation of the *cis* neutral chromophore. (a) Geometry of the chromophore. (b) Molecular orbitals mainly involved in the low energy absorption band of the absorption spectrum. (c) Molecular orbitals mainly involved in the high energy absorption band of the absorption spectrum.

Figure 3: Shape of the molecular orbitals involved in the relevant electronic transitions of the TagRFP675 protein as obtained from the simulation of the *cis* anionic chromophore. (a) Geometry of the chromophore. (b) Molecular orbitals mainly involved in the low energy absorption band of the absorption spectrum. (c) Molecular orbitals mainly involved in the high energy absorption band of the absorption spectrum.

Figure 4: Molecular scheme of the H-bond interactions between the phenolic region of the chromophore ($O_{(front)}$) and the depicted residues. (a): neutral *cis*, (b): neutral *trans*, (c): anionic *cis*, (d): anionic *trans.*

Figure 5: Evolution of the considered distances involving the hydroxy phenolic group ($O_{(front)}$) of the neutral chromophore simulations at pH=4.5. (a) and (b) are for the *cis* conformer whereas (c) corresponds to the *trans* isomer.

Figure 6: Evolution of the considered distances involving the phenolic oxygen atom ($O_{(front)}$) of the anionic chromophore simulations at pH=4.5. (a) corresponds to the *cis* conformer whereas (b) and (c) corresponds to the *trans* isomer.

Figure 7: Evolution of the considered distances involving the phenolic oxygen atom ($O_{(front)}$) of the *cis* anionic chromophore simulations at pH=7.5 (a) and pH=10 (b).

Figure 8: Evolution of the considered distances involving the phenolic oxygen atom ($O_{(front)}$) of the *trans* anionic chromophore simulations at pH=7.5 (a) and (b) and pH=10 (c).

Figure 9: Molecular scheme of the H-bond interactions between the imidazolinone ring region of the chromophore and the depicted residues. (a): neutral *cis*, (b): neutral *trans*, (c): anionic *cis*, (d): anionic *trans.*

Figure 10: Evolution of the considered distances involving the imidazolinone ring region of the chromophore ($O_{(back)}$ and Phe62) for the neutral chromophore simulations at pH=4.5. (a) stands for the *cis* conformer whereas (b) to (d) correspond to the *trans* isomer.

Figure 11: Evolution of the considered distances involving the imidazolinone ring region of the chromophore ($O_{(back)}$ and Phe62) for the anionic chromophore simulations at pH=4.5. (a) to (c) refer to the *cis* conformer whereas (d) to (f) correspond to the *trans* isomer.

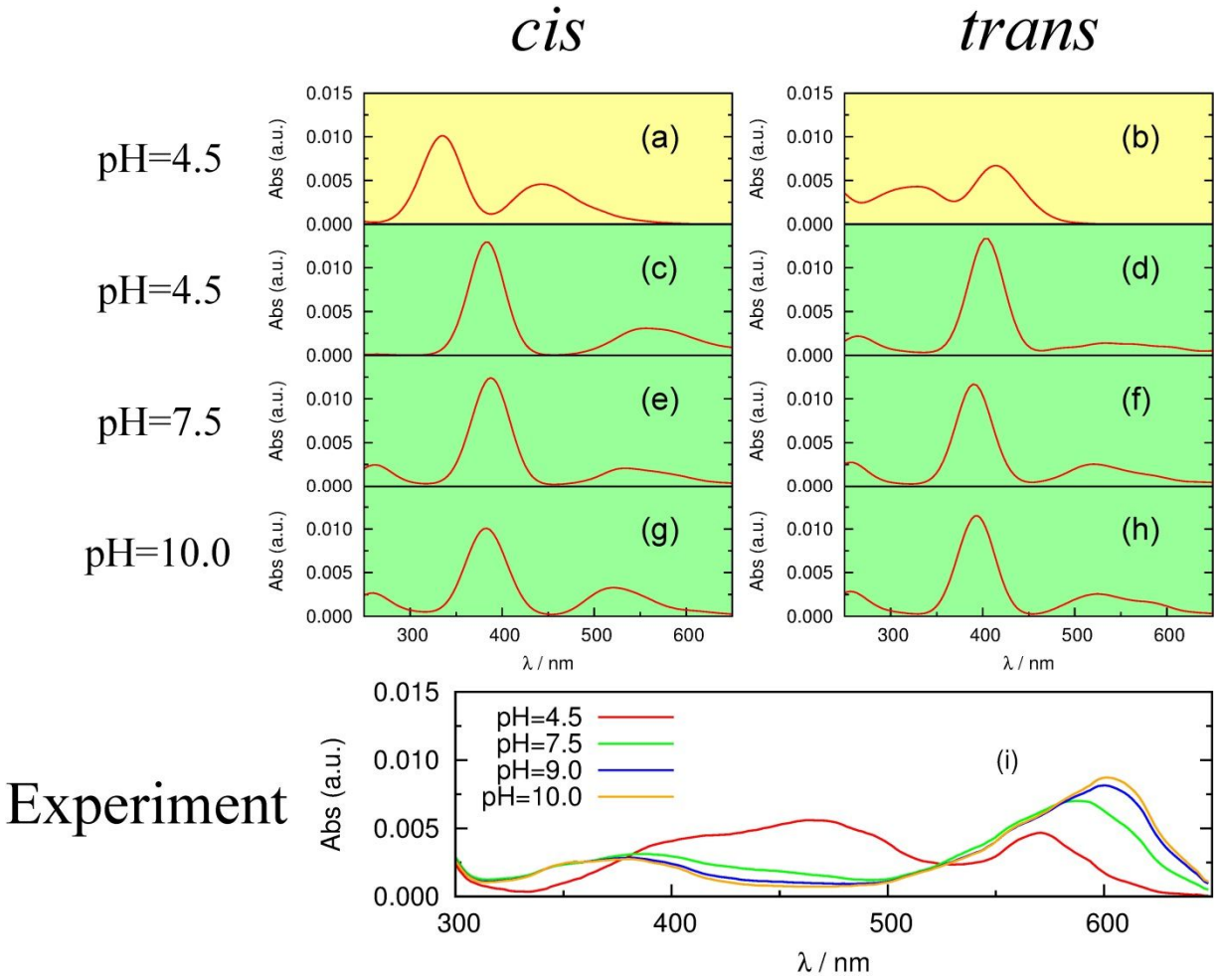
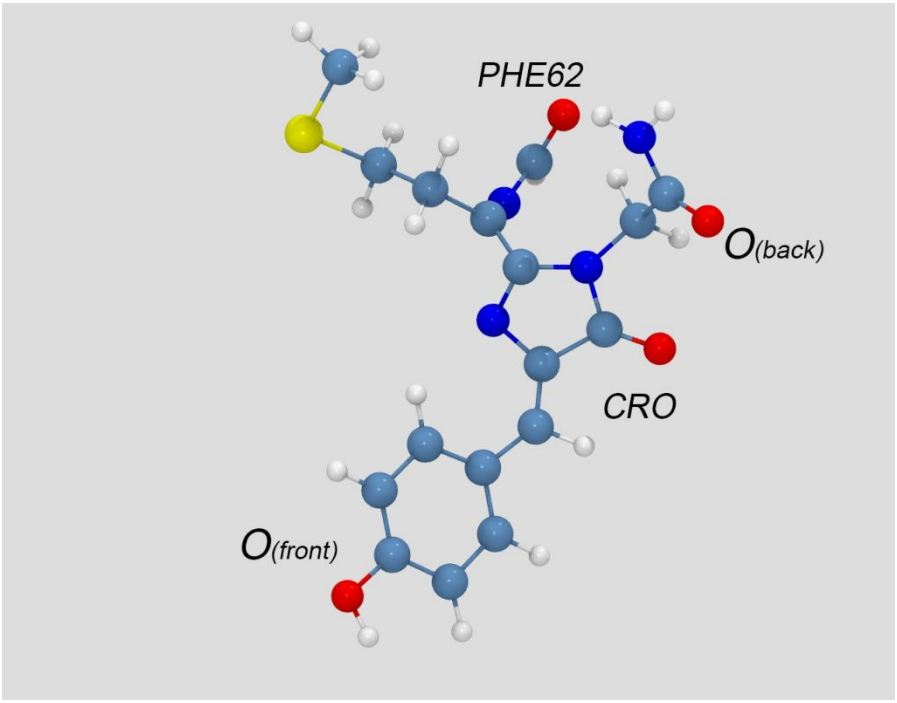
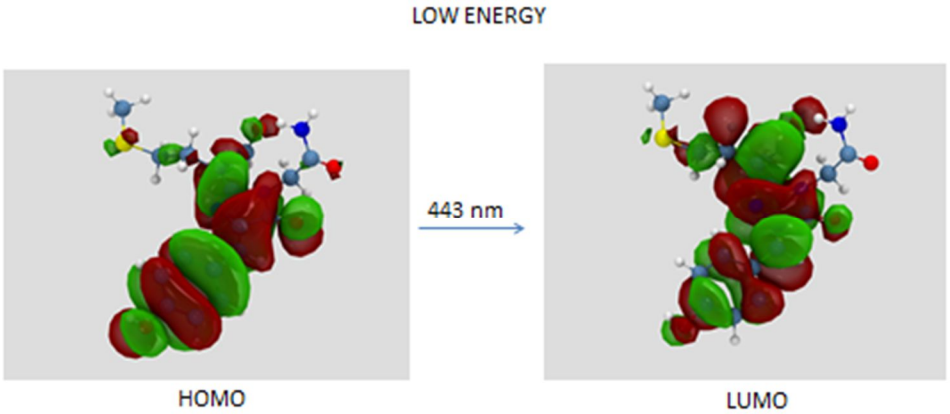


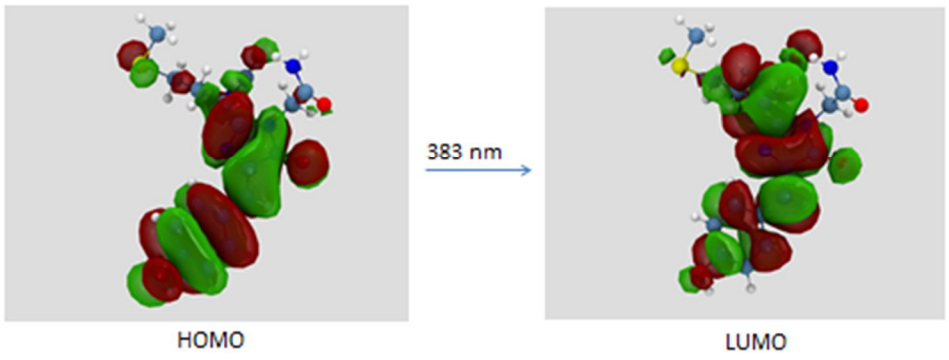
Figure 1



(a)

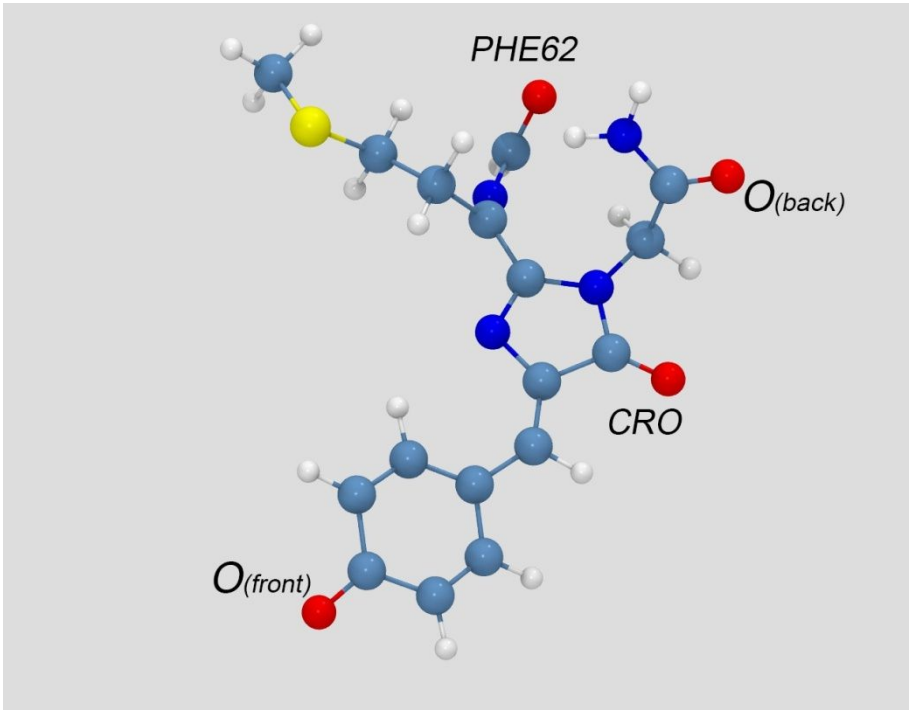


(b)

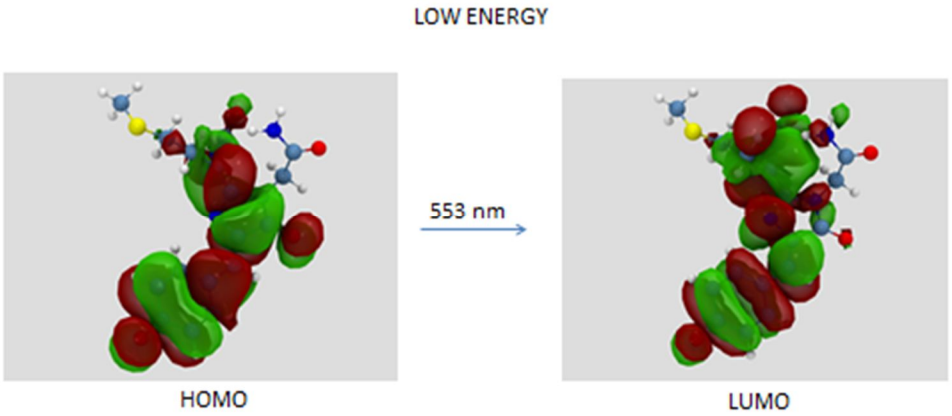


(c)

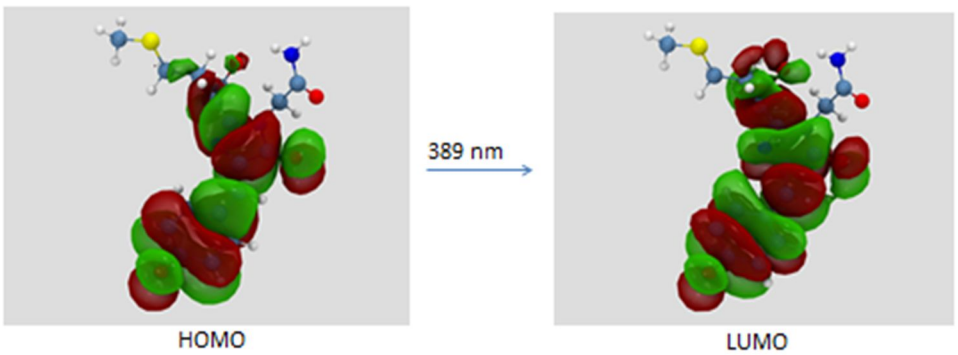
Figure 2



(a)



(b)



(c)

Figure 3.

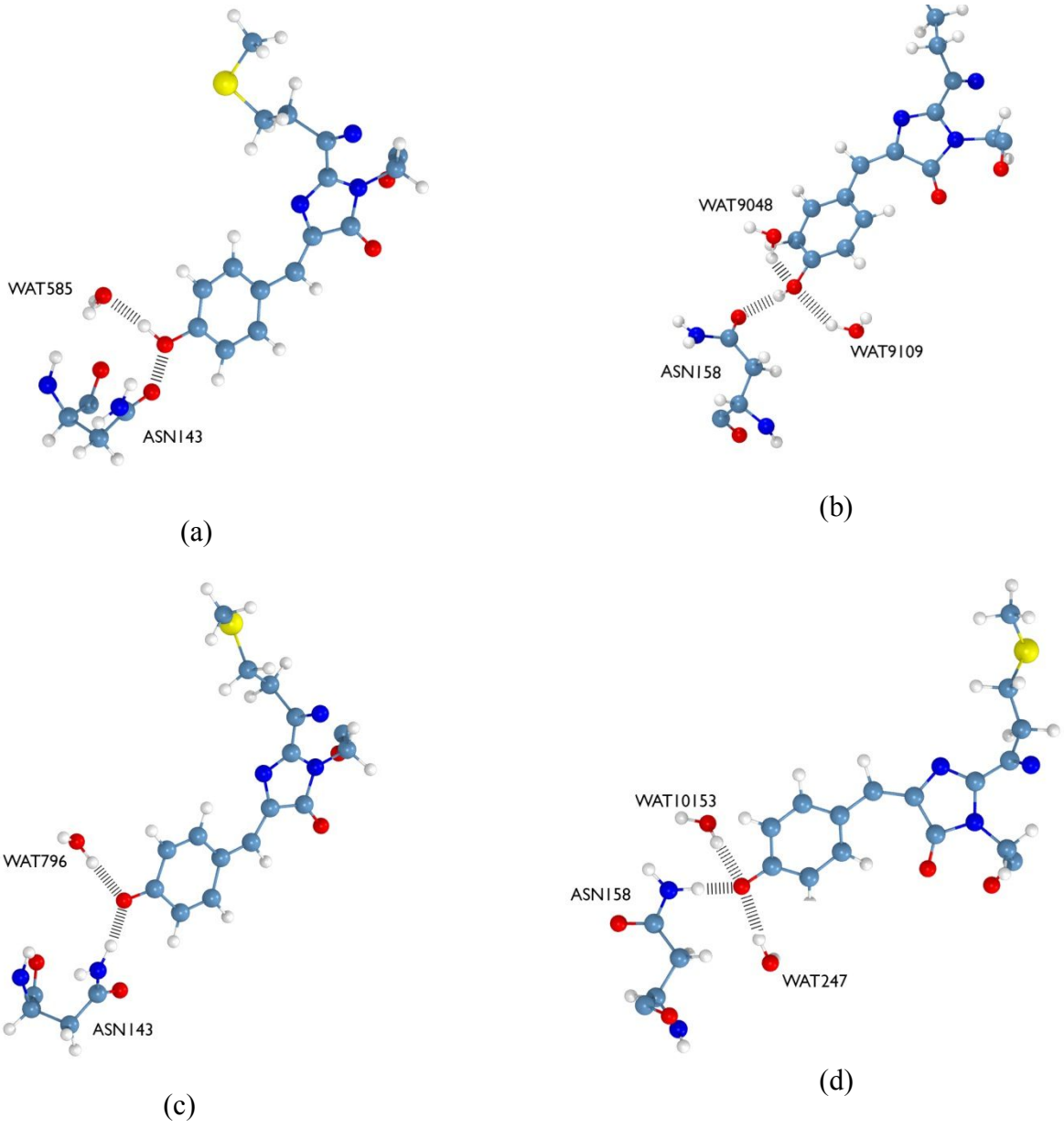
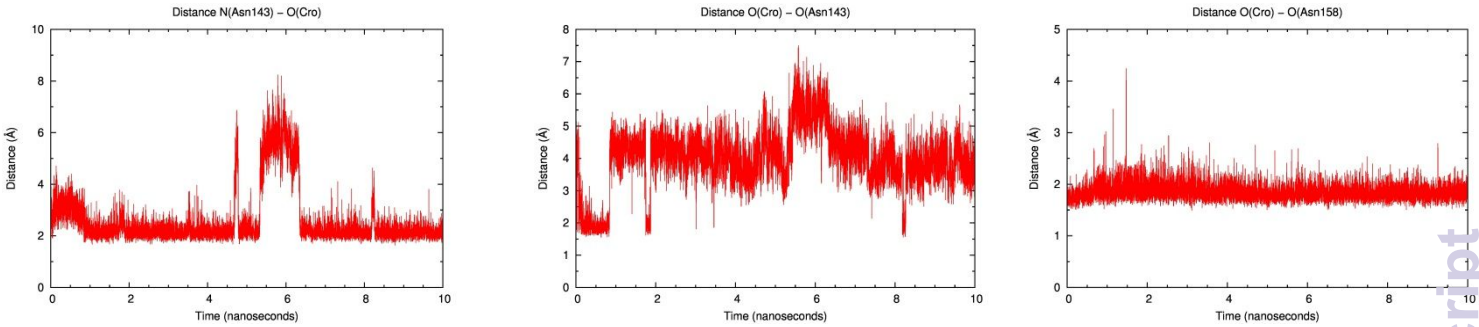
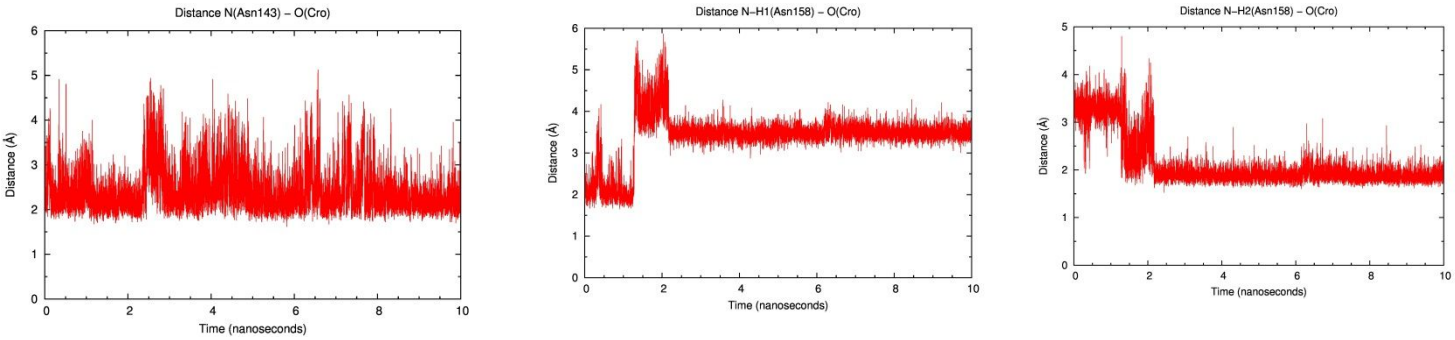


Figure 4



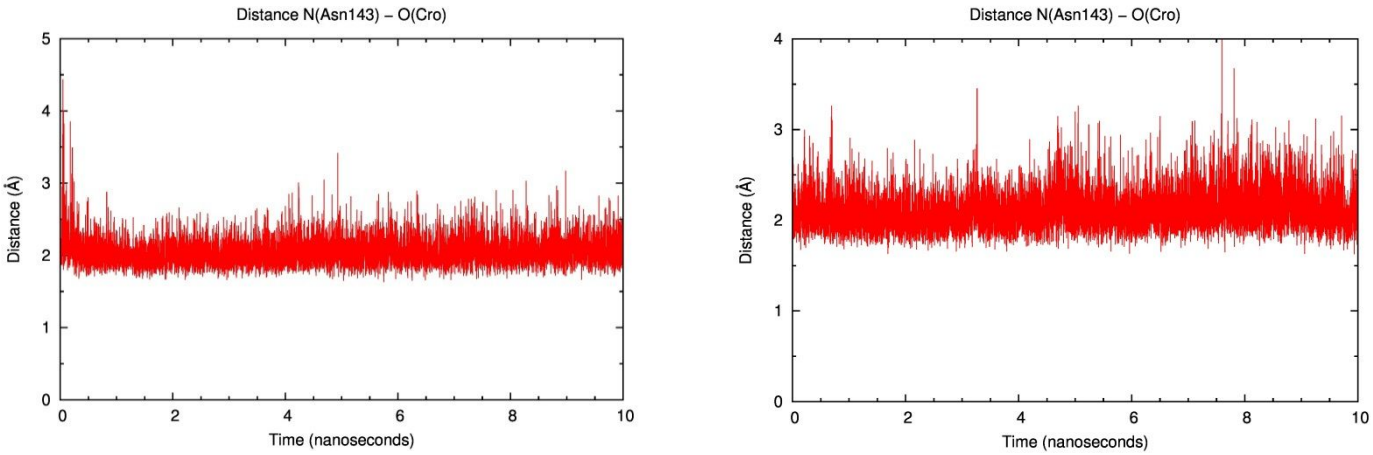
(a) (b) (c)

Figure 5



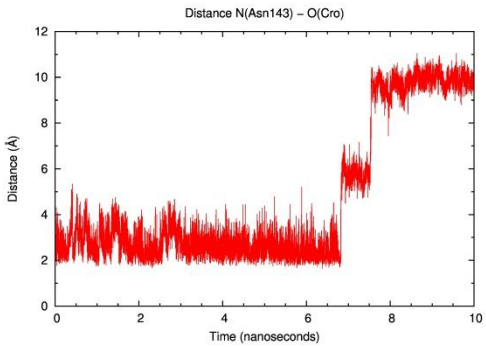
(a) (b) (c)

Figure 6

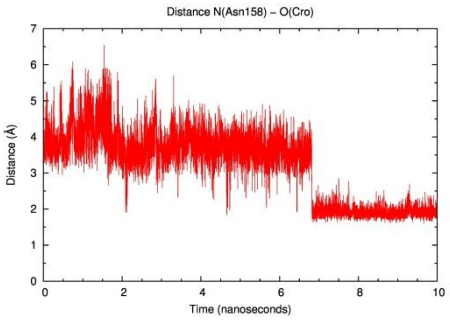


(a) (b)

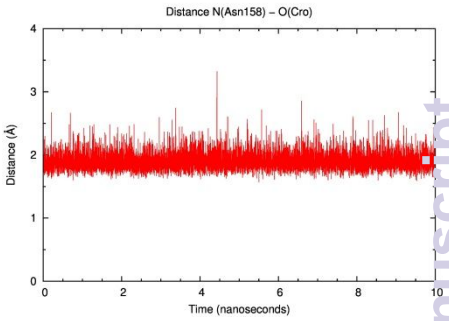
Figure 7



(a)



(b)



(c)

Figure 8

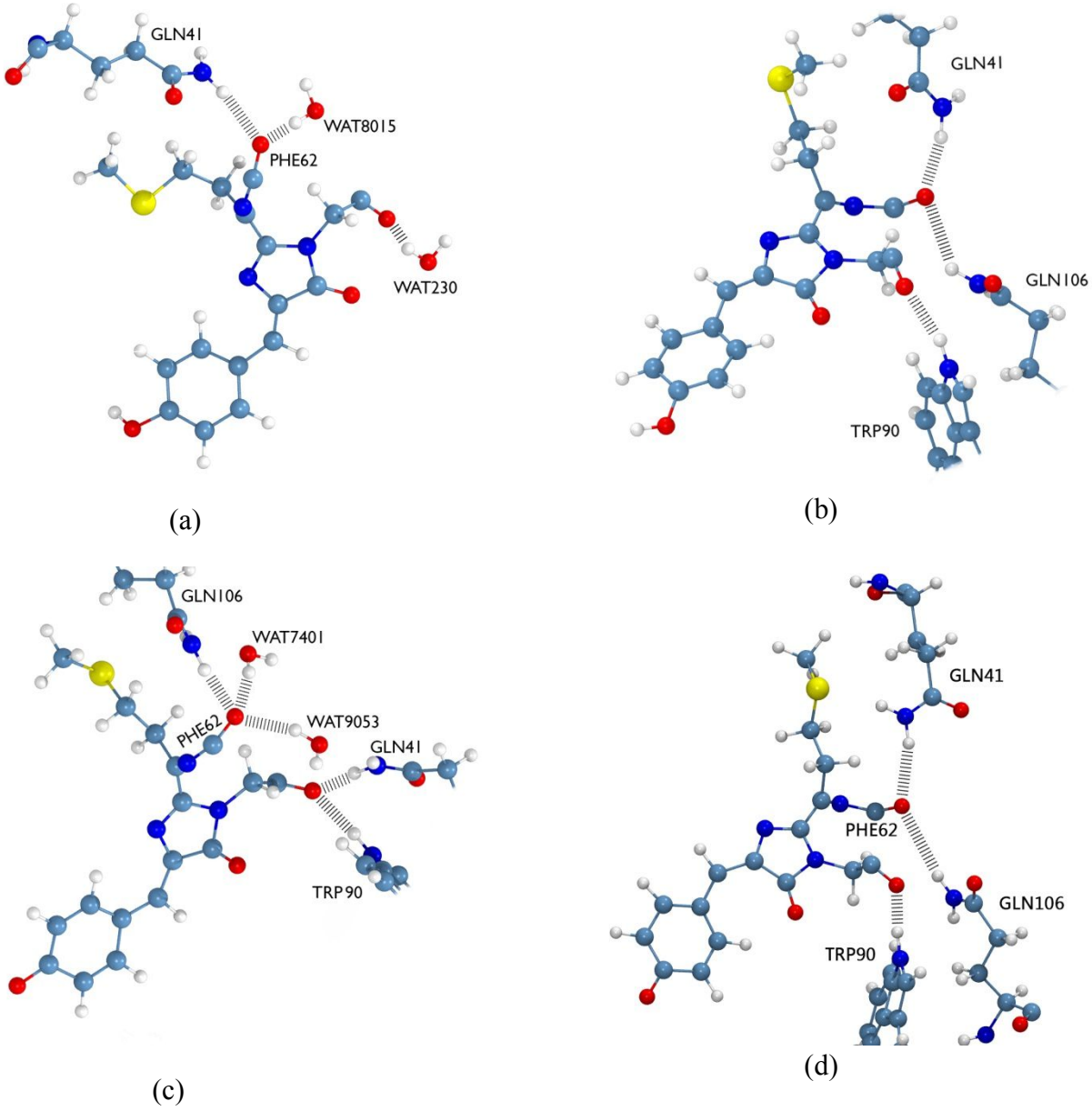
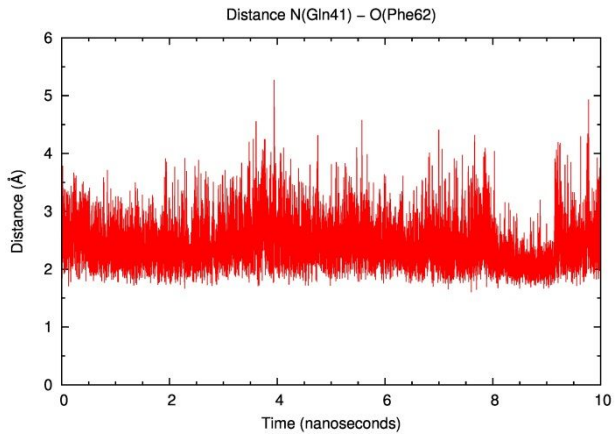
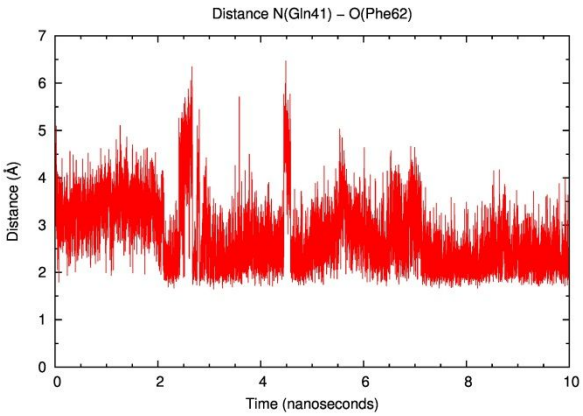


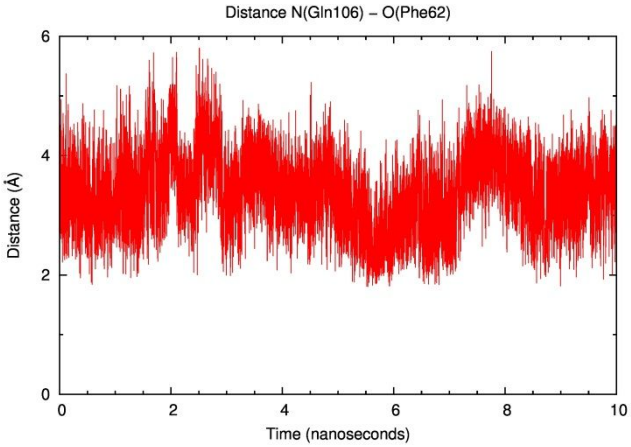
Figure 9



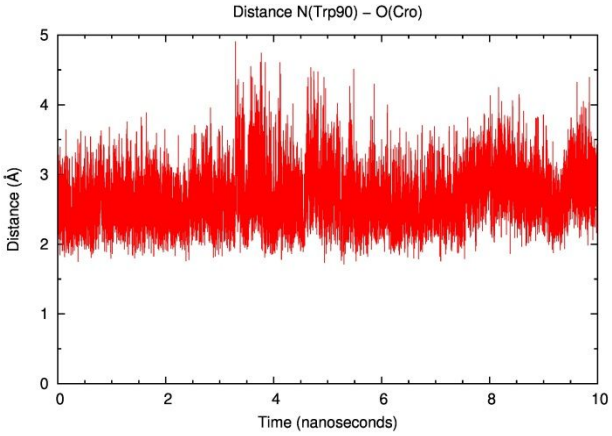
(a)



(b)

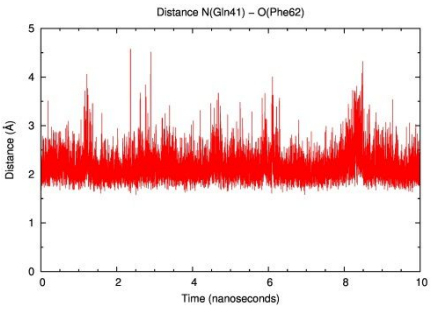


(c)

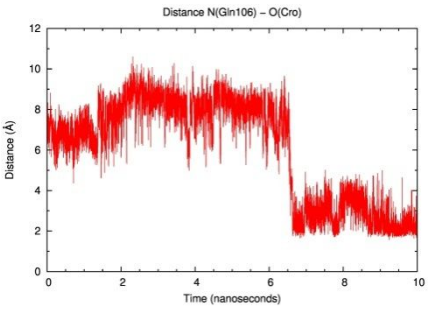


(d)

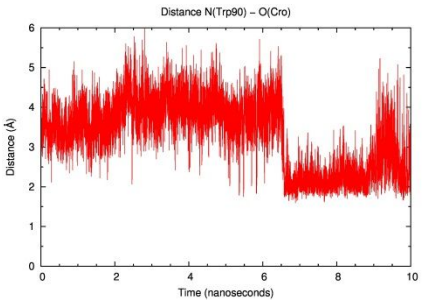
Figure 10



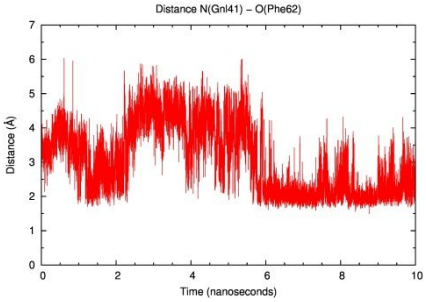
(a)



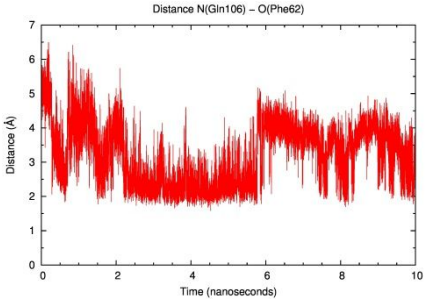
(b)



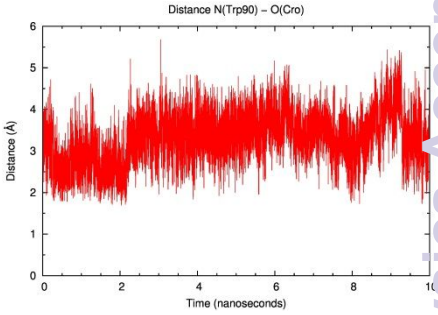
(c)



(d)



(e)



(f)

Figure 11