Angewandte Chemie
Eine Zeitschrift der Gesellschaft Deutscher Chemiker

Supporting Information

69451 Weinheim, Germany

Photoswitching of the Fluorescent Protein asFP595: Mechanism, Proton Pathways, and Absorption Spectra (Supporting Information)

Lars V. Schäfer, Gerrit Groenhof, Astrid R. Klingen, G. Matthias Ullmann, Martial Boggio-Pasqua, Michael A. Robb, and Helmut Grubmüller

I. METHODS

A. MD Simulations

All simulations were performed in a rectangular periodic box of about $8 \times 8 \times 8\,\mathrm{nm}^3$. Each system contained in total about 15,000 TIP4P water molecules, including 340 crystallographic water molecules. After assigning the protonation pattern of the chromophore pocket, all other polar, aromatic, and aliphatic hydrogens were added to the protein with the HB2MAK[1] routine of WHATIF.[2] To each of the systems, sodium and chloride ions at physiological concentration were added to compensate for the net positive charge of the protein. The actual number of ions used depended on the total charge of the protein, which differed for the five chosen protonation patterns of the chromophore cavity. The final systems contained around 68,000 atoms. Prior to the simulations, the systems were energy minimized for 1000 steps using a steepest descent algorithm. Subsequently, a 200 ps MD simulation was performed with harmonic position restraints on all heavy protein atoms (force constant $1000\,\mathrm{kJ\,mol^{-1}nm^{-2}}$) to equilibrate the water and the ions.

All simulations were run at constant temperature and pressure by coupling to external baths[3] ($\tau_\mathrm{T} = 0.1\,\mathrm{ps}$, $\tau_\mathrm{p} = 1\,\mathrm{ps}$). LINCS[4] was used to constrain bond lengths, thus allowing a time step of 2 fs for the force field simulations. SETTLE[5] was applied to constrain the internal degrees of freedom of the water molecules. A twin-range cut-off method was used for the Lennard-Jones interactions. Interactions within 1.0 nm were updated every time step, whereas Lennard-Jones interactions between 1.0 and 1.6 nm were updated every ten steps. Coulomb interactions within 1.0 nm were computed each step as well, beyond this cut-off, the particle mesh Ewald method[6] was used, with a reciprocal grid spacing of $0.12\,\mathrm{nm}^{-1}$. All simulations were performed using the Gromacs simulation package[7] together with the OPLS all-atom force field.[8] Force field parameters for the chromophore were taken from Ref. 9.

B. UV-vis spectra

To calculate the optical absorption spectra of the asFP595 chromophore, we extracted sufficiently large ensembles (100 structures) from each force field trajectory and computed the excitation energy for each of these structures. From each MD trajectory, 100 equidistant frames were selected, each frame separated by 75 ps of simulation time. Prior to the calculation of the excited states, very short QM/MM geometry optimizations were performed on every structure in the

ensemble, using the ONIOM mechanical embedding scheme.[10] The chromophore was described quantum mechanically (Fig. 1A) at the RHF/3-21G* ab-initio level during geometry optimization. The rest of the protein, water and ions, were described by the OPLS all-atom force field. A maximum of 50 steepest descent steps were computed to correct for force field inaccuracies. The value of 50 steps was small enough to leave the thermodynamic ensemble nearly unperturbed, yielding improved configurations for the calculation of the excitation spectra.

For the calculation of the TDDFT UV-vis spectra, the chromophore was described quantum mechanically. The remainder of the system, consisting of the apo-protein, water molecules, and ions in the periodic box, was modeled with the OPLS forcefield. The $C_\beta - C_\gamma$ bond of the Met63 sidechain and the $N - C_\alpha$ bond of the Gly65 backbone, which connect the QM and MM subsystems, were replaced by constraints, and the QM part was capped with hydrogen link atoms (Fig. 1A). The force on the link atom was distributed over both atoms of the respective bonds according to the lever rule. The QM subsystem experienced the Coulomb field of all MM atoms within a sphere of 1.6 nm radius, and Lennard-Jones interactions between QM and MM atoms were also included. To avoid over-polarization of the QM subsystem by the MM charges, the partial charges of the MM atoms next to the link atoms as well as of the hydrogen atoms bound to these atoms were set to zero. To maintain an overall integer charge, the charge of the sulphur atom in the Met63 side chain was reduced by a factor of two. In addition, all charges in the MM subsystem were replaced by gaussian charge distributions during the TDDFT calculations. These gaussian distributions were centered on the MM atom, and had a width of $\sigma = 0.3\,\mathrm{nm}$, a value that has been shown to yield improved results.[11] All QM/MM TDDFT calculations were performed with Gromacs 3.3[7] and its QM/MM interface[12] to Gaussian03.[13] Modifications were made in the one-electron integral routines of Gaussian03 for the TDDFT computations of the chromophore polarized by smeared MM charges.

For all TDDFT calculations, the Vosko-Wilk-Nusair local density functional[14] was used together with the Becke exchange functional[15] and the Perdew density gradient correction.[16] A double-zeta basis set of gaussian type orbitals was used with one added polarization function (6-31G*). This density functional method yielded accurate excitation energies for various systems ranging from atoms and small molecules[17] to large bio-organic systems like the free base porphin.[18] In addition, a recent TDDFT study comparing a number of density functionals showed that this functional yields the best excitation energies for the GFP chromophore in the gas phase.[19] To estimate convergence with respect to the basis set size, the excitation energies of representa-

tive structures taken from each of the MD ensembles were re-evaluated using the more complete 6-31+G* and 6-311G* basis sets.

For the calculation of the ZINDO spectra, the entire chromophore and the side chains of residues Lys67, Arg92, Glu145, Ser158, His197, Glu215, and the crystallographic water molecule W233 were described at the quantum level. To this end, the following bonds were cut and the open valences were capped with hydrogen atoms: the $C_\gamma - C_\beta$ bonds of residues Glu145, His197, and Glu215; the $C_\delta - C_\gamma$ bond of Arg92, the $C_\beta - C_\alpha$ bond of Ser158; the $C_\varepsilon - C_\delta$ bond of Lys67; and the $N - C_\alpha$ bond of the Gly65 backbone.

C. Poisson-Boltzmann electrostatics calculations

To quantify the population of the different protonation states of the chromophore pocket, we have performed Poisson-Boltzmann electrostatics (PBE) calculations[20] on wt-asFP595 (pdb entry 2A50[9]) and a mutant structure, which shows density for both cis and trans chromophores, and thus allowed to evaluate the effect of trans-cis photoisomerization on the protonation states of the chromophore pocket.

To prepare the crystal structures for the PBE calculations, a monomer was extracted from each structure, and hydrogen atoms were added with WHATIF.[1,2] Subsequently, the positions of all hydrogen atoms and of the chromophore atoms were optimized using GROMACS and the OPLS force field (l-bfgs, 200 steps). Protonation probabilities were then calculated for all protonatable sites in the protein using the multiflex routine of the MEAD program package[21] and a Metropolis Monte Carlo (MC) algorithm to sample the protonation state energies.[22]

In the PBE calculations, partial charges from the CHARMM27 parameter set[23] were used for the protein. Partial charges for the chromophore were adopted from Ref. 9. Bondi radii[24] were used except for hydrogen $\left(R_\mathrm{H} = 1.0\,\text{Å}\right)$. The dielectric constants were set to $\varepsilon = 4$ for the protein and $\varepsilon = 80$ for the solvent, the ionic strength to $I = 0.1\,\mathrm{M}$, and the temperature to 300 K. For characterizing the titration behavior of the different sites in the protein, pK_A-values of appropriate model compounds in aqueous solution were used as reference values. Standard model compound pK_A-values[20] were used for the sidechains of Asp, Glu, Lys, Arg, Tyr and Cys. Histidines were treated as described by Bashford *et al.*[25] The pK_A-value of the chromophore imidazolinone (N1, Fig. 1B) was estimated according to the thermodynamic cycle 1 described in Ref. 26 at the B3LYP/6-31G* level of theory using the polarizable continuum water model.[27–31] The calculated

values strongly depend on the chromophore conformation and are 9.1 and 4.7 for trans and cis, respectively. Because the pK_A of the chromophore phenolate group (O12, Fig. 1B) does not critically depend on the conformation, we used the value of 8.2 measured for the structurally similar GFP chromophore.[32]

The MC calculations involved 500 equilibration scans and 20,000 production scans for every pH-step. Each MC scan comprised N steps, with N being the total number of protonatable sites. In each MC step, the protonation state of one group was changed, and the change in energy was evaluated. In double (triple) MC steps, the protonation state of two (three) groups was changed simultaneously. Double (triple) MC steps were applied to groups with an interaction energy larger than 2 (3) pK-units.

From the calculated protonation probabilities at a given pH, a subset of N residues likely to change their protonation state were identified in each structure: these residues have a protonation probability $\langle x \rangle(\mathrm{pH}) \in [0.1, 0.9]$. For residues in these subsets, we have calculated the relative energies of all 2^N possible protonation states directly from the MEAD output.[20] The protonation forms of all other residues were kept fixed (residues with $\langle x \rangle(\mathrm{pH}) < 0.1$ were considered deprotonated, residues with $\langle x \rangle(\mathrm{pH}) > 0.9$ were considered protonated). The chromophore and the potentially relevant residues Glu215 and His197 were always included into the subset of N residues for which all protonation state energies were computed. The probability of state n was calculated from its relative energy g_n according to

$$p_n(\mathrm{pH}) = \frac{\exp\left(-g_n(\mathrm{pH})/RT\right)}{\sum_n^{2^N}\left[\exp\left(-g_n(\mathrm{pH})/RT\right)\right]}.$$

With N varying between 10 and 11 for the different calculations, the list of state energies and populations is too large to be shown. To illustrate that the considered protonation states have similar energies and that many of them contribute to the protonation probability of individual residues, Table IA) lists states that have a population larger than 1 % in the wild type protein at $pH = 7$. In this case, the chromophore, Glu215 and His197, as well as His1, His24, His105, His169, His173, His175, Glu203 and His231 were included into the subset of $N = 11$ residues. More than 50,000 state energies and corresponding populations were in fact calculated, the 15 lowest-energy states of which are listed. Table IB) shows the corresponding data for the cis mutant structure. Here, the 16 lowest-energy states are listed (population > 1 % at $pH = 7$). For the cis mutant, Glu203 was found to be always deprotonated and not included into the subset of N

residues.

Table II shows the population of all forms of all residues that were included in the subset of *N* residues for the state energy calculations for one or more structures. The populations for the individual residues are calculated as the sum over the populations of all states *n* (as exemplified in Table I) in which the corresponding residues appear in the respective form.

II. UV-VIS SPECTRA

A. Comparison ZINDO versus TDDFT

Figure 1 compares the optical absorption spectra calculated with the ZINDO method (panels A-D) to those obtained at the TDDFT (BVP86/6-31G*) QM/MM level (panels E, F). The TDDFT absorption bands of the zwitterion and the anion are too much blue-shifted, and none of the absorption bands is close to the measured absorption at 2.18 eV. As shown in Table III, the use of larger basis sets does not lead to a substantial improvement of the TDDFT results. The absorption band of the neutral species is described more accurately at the TDDFT level as compared to ZINDO. However, the relative order of the absorption bands of the different protonation states predicted by ZINDO and TDDFT is the same (Z < A << N < N+), providing further support for our protonation state assignment. The reason for the failure of TDDFT to quantitatively predict the excitation energies of the zwitterion and the anion is that the $S_0 \rightarrow S_1$ transitions involve intramolecular charge transfer from the phenyl to the imidazolinone ring,[33] which is known to be usually rather poorly described by TDDFT.[34,35] We performed a RASSCF(18,7+4+5)[2,2]/6-31G* calculation to further analyze the failure of TDDFT for the ionic protonation states. The $S_1(\pi,\pi*)$ state is a strongly doubly-excited state, explaining the too much blue-shifted TDDFT excitation energy.

B. Absorption of the GFP chromophore

To assess the accuracy of the ZINDO and TDDFT methods for asFP595, we calculated absorption energies of the anionic and the neutral GFP chromophore, which is very similar to MYG. Because in GFP, the dependence of the chromophore's absorption energy on the surrounding amino acids is weak,[36] we used an isolated (p-hydroxybenzylidene)imidazolone as a model chromophore. ZINDO predicts absorption energies of 2.65 eV and 3.65 eV for the geometry-optimized

(HF/6-31G*) anionic and neutral chromophores, respectively, whereas TD-BVP86/6-31G* predicts 3.12 eV and 3.39 eV. The measured absorption maxima of the anionic and neutral GFP are at 2.59 eV and 3.12 eV, respectively.[37] Thus, the red-shifted absorption of the ionic species is well described at the ZINDO level, whereas the excitation energy of the neutral species is overestimated by about 0.5 eV. In contrast, TDDFT is more accurate for the neutral species, but fails to predict the correct absorption of the anion with an error of about 0.5 eV. Because for the red-shifted absorption regime of asFP595 (around 2.18 eV), ZINDO apparently performs better than TDDFT, we based our protonation state assignment in this regime on the ZINDO spectra only. For the more blue-shifted regime (around 2.78 eV), both ZINDO and TDDFT spectra were taken into account.

C. Influence of the protein matrix

Panels (C) and (D) in Figure 1 show the spectra of the asFP595 chromophore in vacuo, calculated from the same ensemble of structures as (A) and (B). The electrostatic influence of the protein matrix shifts the absorptions of the zwitterion and of the anion to the blue by about 0.1 eV to 0.2 eV, whereas the neutral states are slightly red-shifted. In addition to this electrostatic effect on the chromophore by the surrounding amino acids, the non-planarity of the chromophore imposed by the protein matrix also leads to a blue-shift in absorption, irrespective of the protonation state (Table III). With a blue-shift of about 0.5 eV, this shift is especially pronounced for the neutral chromophore. However, neither the electrostatic nor the steric effects of the protein matrix on the chromophore change the order the states.

The degree of chromophore planarity in asFP595 is quantified in Figure 2, which shows the distributions of the two ring-bridging dihedral angles of the chromophore in the ensembles used for the excitation calculations.The distributions are independent from the protonation state, but depend on the chromophore conformation. In the trans conformation, the protein matrix enforces the phenyl ring to deviate from planarity by as much as 20 degrees (Figure 2A). In contrast, the cis chromophore is considerably more planar (Figure 2B), explaining the red-shift. The imidazolinone moiety slightly deviates from planarity by about 5 to 10 degrees, irrespective of the chromophore conformation (Figure 2C and D). The finding that the cis chromophore is more planar than trans supports the hypothesis that ultra-fast radiationless deactivation of the dark trans state is facilitated by the non-planarity of the chromophore.[33]

III. CONTROLS AND ERROR DISCUSSION

The MD simulation of the doubly anionic state allows to test our assumption that the close contact between the chromophore imidazolinone ring and Glu215 under physiological conditions requires a hydrogen bond, and, hence, the doubly anionic chromophore needs not be considered. Already at the start of the 7.5 ns force field MD simulation of this protonation state, the distance between the imidazolinone moiety and Glu215 quickly increased from the x-ray value of 2.7 Å to about 4 Å, ruling out this protonation state under physiological conditions.

To check whether the π-stacking of the His197 imidazole ring on the chromophore, which persisted throughout the simulations, affects the TDDFT excitation, we performed a QM/MM TDDFT calculation for which the imidazole side chain of His197 was included within the QM subsystem. A structure from the zwitterion ensemble was taken for this purpose. According to these calculations, inclusion of the His197 imidazole ring did not alter the nature of the $\pi \rightarrow \pi*$ absorption in the chromophore, and induced only a small red-shift of about 0.02 eV. This finding supported the ZINDO calculations in that the His197 influences the chromophore excitation mainly via electrostatic interactions, which are captured at the force field level, and that its π-system is not directly involved in the excitation. Furthermore, as enlarging the QM subsystem did not alter the results, this control also demonstrates that the size of the original QM subsystem was sufficient for calculating the TDDFT spectra.

To test how the calculated absorption wavelengths depend on the basis set, we repeated the calculations on representative frames from the MD ensembles using the more complete 6-31+G* and 6-311G* basis sets. The results of these calculations are given in Table III. The absorption energies hardly changed upon increasing the basis set size from 6-31G* to 6-31+G* or 6-311G*.

From the many successful minimization studies on small molecules,[38] one might suggest to directly compare the ground-state energies of the different protonation states. To test the feasibility of this approach, we calculated the QM/MM ground-state energies of wt asFP595 at the B3LYP/6-31G* level for the 100 minimized structures taken from the MD simulations for each protonation state. To this end, the His197 and Glu215 sidechains were included into the QM subsystem in addition to the chromophore. As can be expected from the frustrated and complex energy landscape, the obtained energy distributions were much broader than the expected energy differences between the protonation states (data not shown) and thus allow no assignment. Similarly large energy fluctuations in proteins were found by Warshel *et al.*[39,40] and recently also measured in

single molecule experiments.[41]

Our continuum electrostatics estimates of protonation probabilities for the chromophore in the protein matrix depend critically on the accuracy of the reference pK_A-value in aqueous solution. Although it is possible to calculate these values accurate to within half a pK-unit,[26] such computations require Complete Basis Set (CBS) methods which are prohibitively expensive for molecules as large as the asFP595 chromophore. Instead, we used the less accurate density functional calculations described in the Methods. Assuming similar and systematic errors for both the cis and trans conformation, we consider the observed differences for the two isomers sufficiently accurate.

---

[1] R. W. W. Hooft, C. Sander, G. Vriend, *Proteins — Structure, Function, and Genetics* **1996**, *26*, 363.

[2] G. Vriend, *J. Mol. Graph.* **1990**, *8*, 52.

[3] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, J. R. Haak, *J. Chem. Phys.* **1984**, *81*, 3684.

[4] B. Hess, H. Bekker, H. J. C. Berendsen, J. G. E. M. Fraaije, *J. Comp. Chem.* **1997**, *18*, 1463.

[5] S. Miyamoto, P. A. Kollman, *J. Comp. Chem.* **1992**, *13*, 952.

[6] T. Darden, D. York, L. Pedersen, *J. Chem. Phys.* **1993**, *98*, 10089.

[7] D. van der Spoel, B. Hess, E. Lindahl, G. Groenhof, A. E. Mark, H. J. C. Berendsen, *J. Comput. Chem.* **2005**, *26*, 1701.

[8] W. L. Jorgensen, J. Tirado-Rives, *J. Am. Chem. Soc.* **1988**, *110*, 1657.

[9] M. Andresen, M. C. Wahl, A. C. Stiel, F. Gräter, L. V. Schäfer, S. Trowitzsch, G. Weber, C. Eggeling, H. Grubmüller, S. W. Hell, S. Jakobs, *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 13070.

[10] F. Maseras, K. Morokuma, *J. Comput. Chem.* **1995**, *16*, 1170.

[11] P. Amara, M. J. Field, *Theor. Chem. Acc.* **2003**, *109*, 43.

[12] G. Groenhof, M. Bouxin-Cademartory, B. Hess, S. P. de Visser, H. J. C. Berendsen, M. Olivucci, A. E. Mark, M. A. Robb, *J. Am. Chem. Soc.* **2004**, *126*, 4228.

[13] M. J. Frisch, et al. Gaussian, Inc., Wallingford, CT, 2004.

[14] S. H. Vosko, L. Wilk, M. Nusair, *Can. J. Phys.* **1980**, *58*, 1200.

[15] A. D. Becke, *Phys. Rev. A* **1988**, *38*, 3098.

[16] J. P. Perdew, *Phys. Rev. B* **1986**, *33*, 8822.

[17] S. J. A. van Gisbergen, F. Kootstra, P. R. T. Schipper, O. V. Gritsenko, J. G. Snijders, E. J. Baerends,

Phys. Rev. A **1998**, *57*, 2556.

[18] S. J. A. van Gisbergen, A. Rosa, G. Ricciardi, E. J. Baerends, *J. Chem. Phys.* **1999**, *111*, 2499.

[19] A. V. Nemukhin, I. A. Topol, S. K. Burt, *J. Chem. Theory Comput.* **2006**, *2*, 292.

[20] G. M. Ullmann, E.-W. Knapp, *Eur. Biophys. J.* **1999**, *28*, 533.

[21] D. Bashford, M. Karplus, *Biochemistry* **1990**, *29*, 10219.

[22] G. M. Ullmann, *QMCT: a Monte Carlo titration program* **2005**.

[23] A. D. MacKerell, et al., *J. Phys. Chem. B* **1998**, *102*, 3586.

[24] A. Bondi, *J. Phys. Chem.* **1964**, *68*, 441.

[25] D. Bashford, D. A. Case, C. Dalvit, L. Tennent, P. E. Wright, *Biochemistry* **1993**, *32*, 8045.

[26] M. D. Liptak, G. C. Shields, *J. Am. Chem. Soc.* **2001**, *123*, 7314.

[27] S. Miertus, E. Scrocco, J. Tomasi, *Chem. Phys.* **1981**, *55*, 117.

[28] M. Cossi, V. Barone, R. Cammi, J. Tomasi, *Chem. Phys. Lett.* **1996**, *255*, 327.

[29] E. Cances, B. Mennucci, J. Tomasi, *J. Chem. Phys.* **1997**, *107*, 3032.

[30] V. Barone, M. Cossi, J. Tomasi, *J. Comput. Chem.* **1998**, *19*, 404.

[31] M. Cossi, V. Barone, B. Mennucci, J. Tomasi, *Chem. Phys. Lett.* **1998**, *286*, 253.

[32] A. F. Bell, X. He, R. M. Wachter, P. J. Tonge, *Biochemistry* **2000**, *39*, 4423.

[33] T. A. Schüttrigkeit, T. von Feilitzsch, C. K. Kompa, K. A. Lukyanov, A. P. Savitsky, A. A. Voityuk, M. E. Michel-Beyerle, *Chem. Phys.* **2006**, *323*, 149.

[34] M. Wanko, M. Garavelli, F. Bernardi, T. A. Niehaus, T. Frauenheim, M. Elstner, *J. Chem. Phys.* **2004**, *120*, 1674.

[35] A. Dreuw, M. Head-Gordon, *J. Am. Chem. Soc.* **2004**, *126*, 4007.

[36] T. Laino, R. Nifosi, V. Tozzini, *Chem. Phys.* **2004**, *298*, 17.

[37] P. Leiderman, D. Huppert, N. Agmon, *Biophys. J.* **2006**, *90*, 1009.

[38] H. B. Schlegel, *J. Comput. Chem.* **2003**, *24*, 1514.

[39] A. Warshel, *Proc. Natl. Acad. Sci. USA* **1984**, *81*, 444.

[40] M. Klahn, S. Braun-Sand, E. Rosta, A. Warshel, *J. Phys. Chem. B* **2005**, *109*, 15645.

[41] B. P. English, W. Min, A. M. van Oijen, K. T. Lee, G. B. Luo, H. Y. Sun, B. J. Cherayil, S. C. Kou, X. S. Xie, *Nat. Chem. Biol.* **2006**, *2*, 87.

Figures

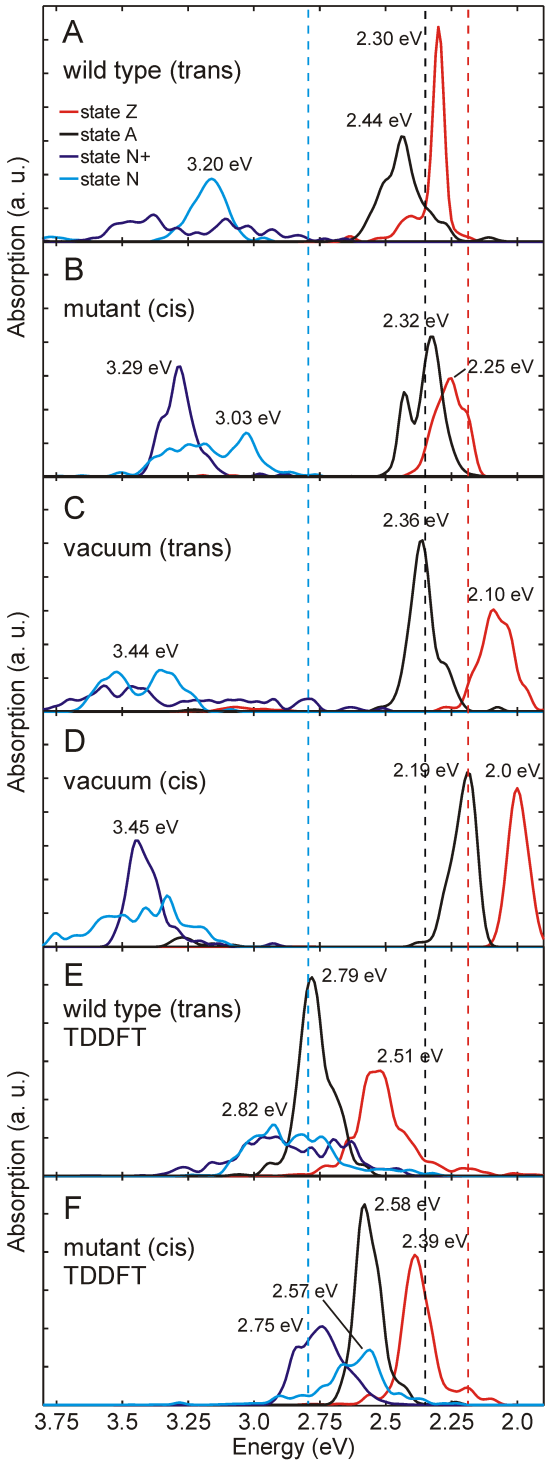
A
wild type (trans)
state Z
state A
state N+
state N
2.30 eV
2.44 eV
3.20 eV
B
mutant (cis)
2.32 eV
2.25 eV
3.29 eV
3.03 eV
C
vacuum (trans)
2.36 eV
2.10 eV
3.44 eV
D
vacuum (cis)
2.19 eV
2.0 eV
3.45 eV
E
wild type (trans)
TDDFT
2.79 eV
2.51 eV
2.82 eV
F
mutant (cis)
TDDFT
2.58 eV
2.39 eV
2.57 eV
2.75 eV
Absorption (a. u.)
Absorption (a. u.)
Absorption (a. u.)
3.75 3.5 3.25 3.0 2.75 2.5 2.25 2.0
Energy (eV)

Figure 1: Calculated UV-vis spectra of the different asFP595 protonation states considered in this study. Spectra (A) to (D) were obtained at the ZINDO level, (E) and (F) were calculated using TD-BVP86/6-31G* within a QM/MM framework. For spectra (A) and (B), the majority of the chromophore binding pocket was included into the quantum calculation. Spectra (C) and (D) are based on the same ensemble of structures as (A) and (B), but the amino acids surrounding the chromophore were omitted in the ZINDO calculation.

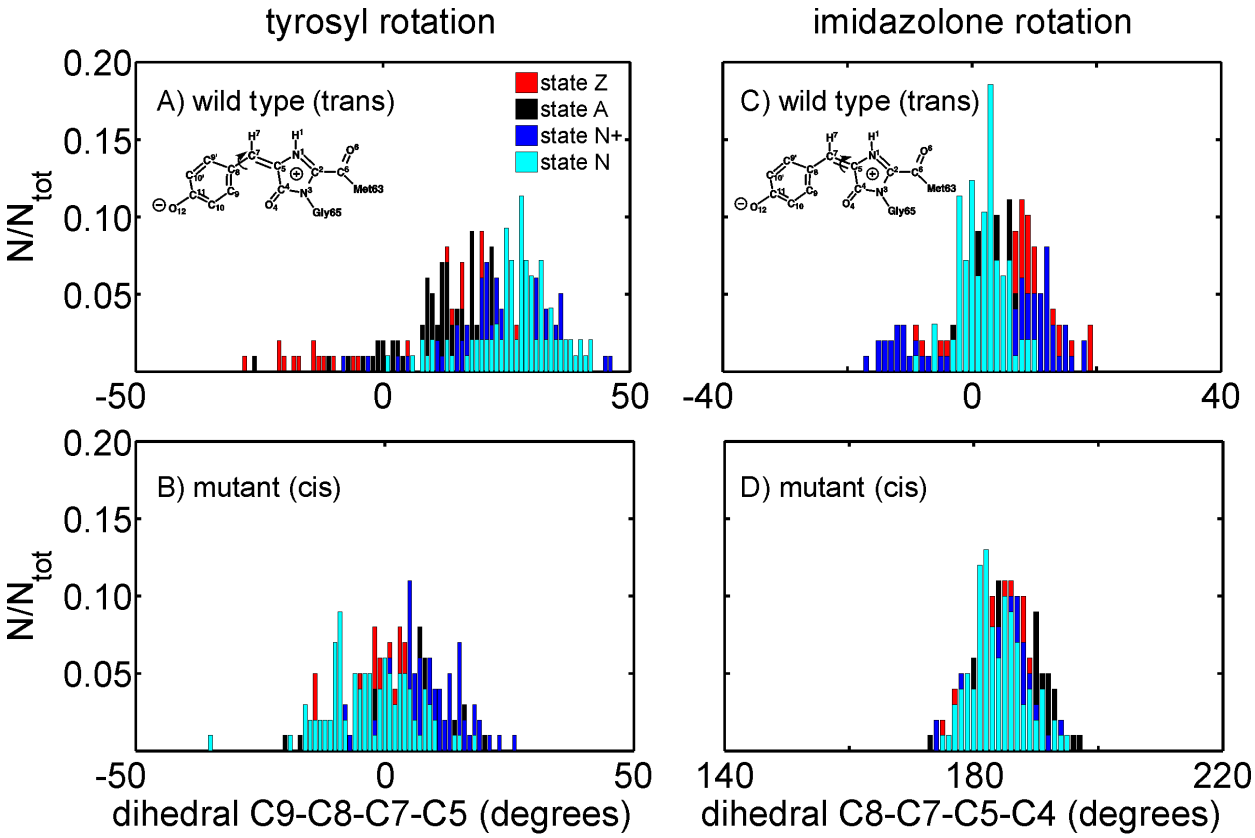

Figure 2: Histogram of the ring-bridging dihedral angles of the asPF595 chromophore in the protein matrix. (A) and (B) show the C9-C8-C7-C5 dihedral angle in the trans wild type and in the cis mutant, respectively. (C) and (D) show the respective C8-C7-C5-C4 dihedral angle distributions. See Figure 1b for atom numbering.

Tables

(A) wildtype protein with trans chromophore

chromophore	protonation state										energy	population
	Glu215	His197	His1	His24	His105	His169	His173	His175	Glu203	His231		
Z	0	e	d	d	e	e	d	d	-	e	0.000	6.848
Z	0	e	d	d	e	d	d	d	-	e	0.265	4.387
Z	0	e	d	e	e	e	d	d	-	e	0.501	2.954
Z	0	e	d	d	e	e	d	e	-	e	0.577	2.599
Z	0	e	d	e	e	d	d	d	-	e	0.762	1.906
Z	-	+	d	d	e	e	d	d	-	e	0.798	1.795
Z	0	e	d	d	e	e	d	d	-	d	0.814	1.746
Z	0	e	d	d	e	d	d	e	-	e	0.832	1.695
Z	-	d	d	d	e	e	d	d	-	e	0.857	1.625
Z	0	e	d	d	e	e	d	d	0	e	1.027	1.221
Z	-	d	d	d	e	d	d	d	-	e	1.048	1.180
Z	-	+	d	d	e	d	d	d	-	e	1.059	1.158
Z	0	e	d	d	e	d	d	d	-	d	1.080	1.118
Z	0	e	d	e	e	e	d	e	-	e	1.088	1.103
Z	0	e	d	d	d	e	d	d	-	e	1.128	1.031
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

(B) mutant protein with cis chromophore

chromophore	protonation state										energy	population
	Glu215	His197	His1	His24	His105	His169	His173	His175	His231			
A	0	e	d	d	e	e	d	d	e		0.000	5.001
N	0	e	d	d	e	e	d	d	e		0.000	4.996
N	0	e	d	d	e	e	d	e	e		0.147	3.905
A	0	e	d	d	e	d	d	d	e		0.210	3.513
N	0	e	d	d	e	d	d	d	e		0.350	2.780
A	0	e	d	d	e	e	d	e	e		0.423	2.458
N	0	e	d	d	e	d	d	e	e		0.488	2.206
A	0	e	d	d	e	e	e	d	e		0.596	1.840
A	0	e	d	d	e	d	d	e	e		0.625	1.753
N	0	e	d	e	e	e	d	d	e		0.723	1.485
N	0	e	d	d	e	e	e	d	e		0.738	1.449
A	0	e	d	e	e	e	d	d	e		0.794	1.319
A	0	e	d	d	e	d	e	d	e		0.808	1.290
A	0	e	d	d	e	e	d	d	d		0.851	1.200
N	0	e	d	d	e	e	d	d	d		0.858	1.185
N	0	e	d	e	e	e	d	e	e		0.878	1.145
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮		⋮	⋮

Table I: List of the lowest-energy protonation states (population > 1 %) calculated for (A) the wild type protein and (B) the cis mutant structure, both at $pH = 7$. The chromophore, Glu215, His197, His1, His24, His105, His169, His173, His175, Glu203 (for the wild type) and His231 were part of the subset for which all state energies were calculated (more than 50,000 in total). Labels indicating the protonation form are Z, N and A (zwitter, neutral, anionic) for the chromophore, 0 and - (neutral/protonated and negatively charged/deprotonated) for glutamate residues, +, d and e (doubly protonated, singly protonated delta-tautomer, singly protonated epsilon-tautomer) for the histidines. Energies are given relative to the lowest-energy protonation state (in kcal/mol), populations are in %.

residue	form	population in				
		trans		cis (mutant)		
		wt	mutant	xray	min. N	min. A
chromophore	Z	96	99	0	1	1
	N	3	1	49	92	61
	A	1	0	51	7	38
Glu215	0	64	8	100	84	100
	-	36	92	0	16	0
His197	+	16	1	0	19	2
	d	15	87	13	0	3
	e	69	12	87	81	95
His1	+	5	7	7	7	9
	d	95	93	93	93	91
	e	0	0	0	0	0
His24	+	7	5	4	1	3
	d	65	74	75	37	65
	e	28	21	21	62	32
His105	+	1	3	3	2	3
	d	13	12	12	0	12
	e	86	85	85	98	85
His169	+	4	3	3	1	1
	d	38	38	37	30	30
	e	58	59	60	69	69
His173	+	1	1	1	2	2
	d	88	77	77	59	63
	e	11	22	22	38	35
His175	+	9	6	6	3	4
	d	66	59	59	45	75
	e	25	35	35	52	21
Glu203	0	16	0	0	0	0
	-	84	100	100	100	100
Lys207	+	100	87	100	100	100
	0	0	13	0	0	0
His231	+	7	8	7	8	10
	d	19	18	18	17	69
	e	74	74	75	75	21

Table II: Population of different protonation forms of individual residues at $pH = 7$. Residues listed were part of the subset of N residues for which all possible protonation state energies were calculated for at least one of the structures considered. Calculations were run for the wild type protein (wt), a mutant structure with the chromophore in the trans conformation, and the for same mutant with the chromophore in the cis conformation. For the cis mutant, the crystal structure (xray) and two energy minimized structures were considered, as outlined in the paper (min. N: minimized with the chromophore in its neutral form, min. A: minimized with the chromophore in its anionic form). Labels for protonation forms are as in Table I.

	TD-BVP86						ZINDO					
state	cis (mutant)			trans (wild type)			cis (mutant)			trans (wild type)		
	6-31G*	6-311G*	6-31+G*	6-31G*	6-311G*	6-31+G*	non-planar protein	non-planar vacuum	planar vacuum	non-planar protein	non-planar vacuum	planar vacuum
Z	2.39	2.38	2.34	2.51	2.51	2.46	2.25	2.00	1.90	2.30	2.10	1.98
N	2.62	2.60	2.58	2.82	2.82	2.79	3.03	3.44	3.00	3.20	3.44	2.96
N+	2.74	2.73	2.70	2.82	2.82	2.80	3.29	3.45	3.00	3.20	3.44	2.96
A	2.58	2.56	2.56	2.79	2.77	2.75	2.32	2.19	2.16	2.44	2.36	2.20

Table III: Calculated absorption energies in eV of different asFP595 protonation states. To elucidate the basis set dependence of the TDDFT absorption energies, representative structures were taken from the maxima of the respective absorption peaks in Fig. 1. The planar species were optimized in vacuo prior to the excitation calculation.