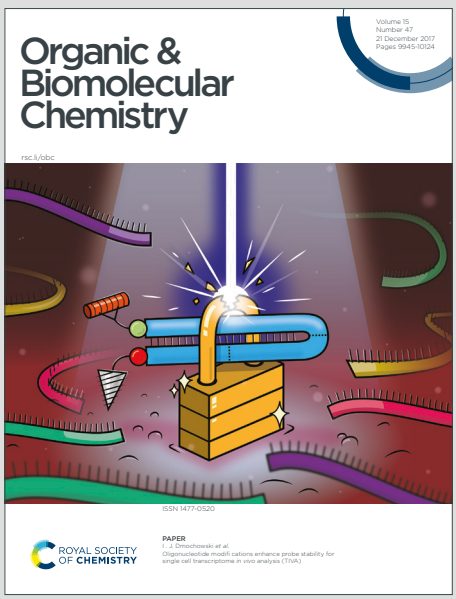
Organic & Biomolecular Chemistry

Accepted Manuscript

This article can be cited before page numbers have been issued, to do this please use: B. L. Grigorenko, M. Khrenova, D. D. Jones and A. Nemukhin, *Org. Biomol. Chem.*, 2023, DOI: 10.1039/D3OB01450A.



ROYAL SOCIETY OF CHEMISTRY

rsc.li/obc

View Article Online
DOI: 10.1039/D3OB01450A

Histidine-assisted reduction of arylnitrenes upon photo-activation of phenyl azide chromophores in the GFP-like fluorescent proteins

Bella L. Grigorenko[a,b], Maria G. Khrenova[a,c*], D. Dafydd Jones[d], Alexander V. Nemukhin[a,b]

[a] Chemistry Department, Lomonosov Moscow State University, Moscow, Russian Federation

[d] Emanuel Institute of Biochemical Physics, Russian Academy of Sciences, Moscow, Russian Federation

[c] Bach Institute of Biochemistry, Moscow, Russian Federation

[d] School of Biosciences, Molecular Biosciences Division, Cardiff University, Cardiff, UK

Abstract

The photochemically active sites of the proteins sfGFP^{66azF} and Venus^{66azF} from the green fluorescent protein (GFP) family contain a non-canonical amino acid residue p-azidophenylalanine (azF) instead of Tyr66 in GFP. Light-induced decomposition of azF in these sites leads to the reactive arylnitrene (nF) intermediates followed by formation of phenylamine-containing chromophores. We report the first study of a reaction mechanism of reduction of arylnitrene intermediates in sfGFP^{66nF} and Venus^{66nF} using molecular modeling methods. The Gibbs energy profiles for the elementary steps of chemical reaction in sfGFP^{66nF} are computed using molecular dynamics simulations with the quantum mechanics/molecular mechanics (QM/MM) potentials. Structures and energies along the reaction pathway in Venus^{66nF} are evaluated using the QM/MM approach. According to the results of simulations, arylnitrene reduction is coupled with oxidation of the histidine side chain His148 located near the chromophore.

View Article Online
DOI: 10.1039/D3OB01450A

Introduction

Engineered variants of fluorescent proteins considerably expand the area of applications of these important biomarkers as compared to their ancestor, the green fluorescent protein (GFP)[1,2]. In particular, incorporation of an artificial amino acid, p-azido-L-phenylalanine (azF), at a specific position in the polypeptide chain enables photo-activation that is required for optogenetics. In this work, we consider the variants based on the superfolder GFP [3] and Venus [4] proteins, in which the Tyr66 amino acid responsible for formation of the phenolic ring (P-ring) of the matured chromophore in GFP-like proteins, is replaced by azF. In the corresponding proteins termed sfGFP^{66azF} [5,6] and Venus^{66azF} [7], photolysis of the phenyl azide releases N_2 leading to a nitrene intermediate. Arylnitrene intermediates, in turn, initiate various chemical reactions [8], which, in particular, can lead to stable fluorescent species. As tentatively suggested previously [5–7], a possible reaction route is reduction of a nitrene to an amine; however, a detailed reaction mechanisms has not been revealed and other pathways typical for the arylnitrene chemistry [8] cannot not be rejected. Indeed, alternative mechanisms have been reported when phenyl azide chemistry is placed in other regions of sfGFP outside of but close to the chromophore [5] including ring expansion to an azepinone on reaction of the nitrene with water when His148 is replaced by azF [9].

The goal of the present work is to model chemical reactions initiated by the arylnitrene intermediates after photochemical decomposition of the azide group in sfGFP^{66azF} and Venus^{66azF}. The major conclusion of these simulations is that reduction of the nitrene intermediate to the phenylamine chromophore is coupled with oxidation of the nearby histidine side chain, His148, in both proteins. To illustrate the scheme, we show in Fig. 1 the composition of the chromophore-containing pocket in the light-induced state of the sfGFP^{66azF} protein taking atomic coordinates from the crystal structure PDB ID 4J89 [5]. This structure was obtained after the loss of N_2 from the phenyl azide group in sfGFP^{66azF}. We show in Fig. 1 the molecular groups essential for the chromophore maturation in the GFP-like proteins [2,10–14], i.e., the Glu222, Arg96, Ser205 side chains, as well as the His148 side chain. The crystal structure (PDB 4J89) reports two alternative conformations of the His148 side chain; one of those is shown in colored balls and sticks, another one in yellow balls and sticks in Fig. 1. We also emphasize the presence of multiple water molecules observed in the crystal structure shown as red dots in Fig. 1. The side chain of His148 is solvent exposed on the protein's surface in contrast to fluorescent proteins with a tyrosine residue at 66th position[9,15–17]. Thus, conformational flexibility and presence of nearby water molecules in sfGFP^{66azF} and Venus^{66azF} are clear. Crystal structures of the sfGFP^{66azF} and Venus^{66azF} proteins in the dark state, i.e. with the p-azido-L-phenylalanine moiety at position 66, are available as well (PDB ID 4J88 [5], PDB ID 6SM0 [7], respectively). These structures also show

View Article Online
DOI: 10.1039/D3OB01450A

the location of the solvent accessible His148 side chain near the chromophore and the presence of water molecules near the histidine side chain. As we see below, water molecules actively participate in reactions of the chromophore maturation in these proteins.

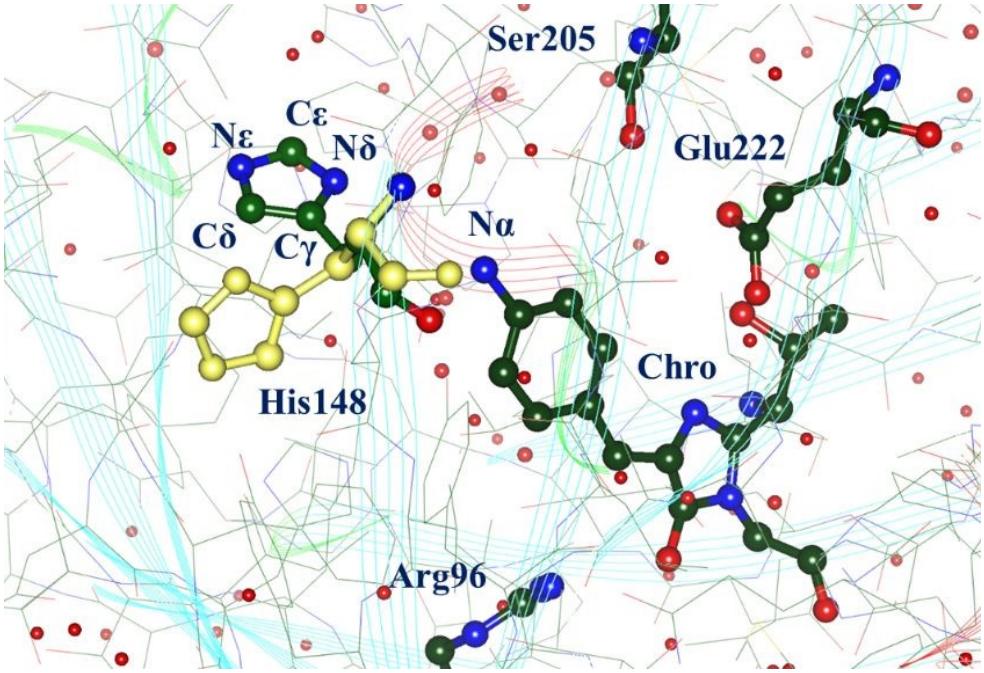


Figure 1. The chromophore-containing pocket in the light-induced state of the sfGFP^{66azF} protein using coordinates of the crystal structure PDB ID 4J89 [5]. One of the alternative conformations of the His148 side chain is shown in yellow balls and sticks. In this and other figures, carbon atoms are colored green, oxygen – red, nitrogen – blue. Red dots denote positions of water molecules.

The conclusion on involvement of the imidazole side chain of His148 in chemical reactions upon chromophore maturation in photoactive proteins is not unexpected. Occurrence of the oxo-forms of the histidine side chain is documented in the experimental studies of the miniSOG protein[18]. The important experimental work of Schöneich [19] describes mechanisms of metal-catalyzed oxidation of histidine to the oxo-histidine in peptides and proteins. A computational study by Safaei et al. [20] considers mechanisms of oxidation reactions of the imidazole initiated by hydroxyl radicals, similarly to our case, as shown below. Finally, the work by Morris et al. [6] reports a tentative formation of the anilino radical near the His148 side chain. As revealed in the present work, this is the first step in the phenylamine-containing chromophore maturation.

In this work, we study reactions of the nitrene intermediates (denoted here as sfGFP^{66nF} and Venus^{66nF}) in the photochemistry of the phenyl azide variants of sfGFP and Venus using modern computational methods of molecular modeling. Specifically, we use quantum mechanics/molecular mechanics (QM/MM)[21] and molecular dynamics (MD) with QM/MM

View Article Online
DOI: 10.1039/D3OB01450A

potentials (QM/MM MD)[22] to estimate reaction energy profiles and to reveal mechanisms of formation of the phenylamine chromophores.

Methods

Chemical reactions regarding the formation of phenylamine-containing chromophores in the sfGFP and Venus variants were modeled starting from the phenylnitrene intermediates. The available crystal structures of protein macromolecules served as a source of atomic coordinates of heavy atoms. The molecular model for sfGFP^{66nF} was created using the structure PDB ID 2B3P [3] of the original variant of sfGFP; then the hydroxyl group of Tyr66 was replaced by nitrogen. The molecular model for simulations with Venus^{66nF} was initially based on the crystal structure PDB ID 6SM0 [7] of pre-matured Venus^{66azF} containing the phenyl azide group. A model of Venus^{66nF} was constructed by manual deletion of the N_2 group from the computationally derived structure of the hydroperoxyl intermediate [7] of the chromophore. As discussed in Ref. [7], this structure appeared after cyclization, dehydration and partial oxidation of the initial polypeptide in Venus^{66azF}, but final oxidation of the hydroperoxyl intermediate to the matured chromophore should be a subsequent reaction step after photolysis of phenyl azide. Conversion of the X-ray structures to full-atom molecular models was performed as follows: hydrogen atoms were added manually using molecular mechanics tools; the side chains of Arg and Lys were assumed to be positively charged, and the side chains of Glu and Asp as negatively charged. The model protein molecules were fully surrounded by explicit water molecules.

Simulations were carried out using the QM/MM-based approaches. In the case of sfGFP^{66nF}, the QM subsystem included the chromophore, the side chains of His148, Ser205, Thr203, Arg96, Gln94, Gln69, backbone of Asn146, Ser147, and three water molecules. The energies and energy gradients in QM were computed at the Kohn-Sham density functional theory (DFT) level with the PBE0 functional [23] with the D3 dispersion correction [24]. The 6-31G** basis set was employed. Electronic embedding scheme was utilized, partial atomic charges from the MM atoms contributed to the one-electron part of the Hamiltonian of the QM subsystem polarizing it. Hydrogen atoms were added to the quantum subsystem as capping atoms on the QM-MM border. The MM subsystem described by the CHARMM36 [25] and TIP3P [26] force field parameters comprised the rest of the protein and solvent shells, respectively. To compute the Gibbs energy profiles, we carried out MD simulations with these QM/MM potentials. Molecular dynamic runs were performed in the NPT ensemble with the periodic boundary conditions at $p = 1$ atm and $T = 300$ K. Particle Mesh Ewald (PME) was utilized for electrostatic calculations. The cutoff

View Article Online
DOI: 10.1039/D3OB01450A

distance for all non-bonded interactions was 12 Å with switching to smoothing potential at 10 Å. Sodium ions were added to neutralize the system. The trajectories were computed using NAMD [27] and TeraChem [28] software packages and specialized interface [29]. The unrestricted DFT version was applied for the triplet and singlet radical state calculations, whereas the restricted version was used for the singlet closed-shell states. Collective variables (CV) were selected for each chemical of an entire reaction (see Results section). The selected CVs were divided into windows covering the range of corresponding reaction coordinate at each step. The harmonic potential was added to obtain CV distribution at each window. Force constants of additional harmonic potentials were 40 kcal/mol/Å^2 for regions closer to minima and 120 kcal/mol/Å^2 in the transition state regions. Each trajectory was about 10 ps calculated with the 1 fs integration time step. The distribution overlaps were checked prior to the calculation of the Gibbs energy profile of the reaction step. The umbrella integration [30] and weighted histogram analysis (WHAM)[31] were utilized to construct the Gibbs energy profiles. In the case of Venus^{66nF}, the QM subsystem included the hydroperoxyl intermediate of the pre-matured chromophore, the side chains of His148, Arg96, Ser205, Glu222, Tyr203, fragments of the main chain and 28 water molecules (192 atoms in total). These water molecules were selected to solvate the reaction region and allow proton transfers via different hydrogen bond networks during the reaction if required. The energies and energy gradients in QM were computed using the PBE0-D3/6-31G* approximation, whereas the AMBER99 force field parameters [32] were used in MM. The NWChem program[33] was employed in QM/MM calculations. Electronic embedding scheme was utilized and hydrogen atoms were added to the quantum subsystem as capping atoms on the QM-MM border. The unrestricted DFT version was applied for the triplet state calculations and the restricted version was used for the singlet states. This particular approach was used in the preceding simulations of chromophore maturation in Venus^{66azF} [7]; therefore, we did not modify the computational protocol in the present work.

We utilized different force field parameters in QM/MM and QM/MM MD simulations as AMBER is better implemented in NWChem and CHARMM in NAMD. It has already been demonstrated that for enzymatic reactions utilization of these two different force fields leads to the same results. [34] In QM/MM MD simulations we extended the basis set as we were able to reduce the size of the QM subsystem from 192 to 129 atoms.

We use the following convention to designate species along reaction pathways in the figures and in the manuscript. The molecular group, which finally forms the chromophore are denoted as Chro irrespective to a specific form of this group at every reaction step. Thus, the name Chro refers to the initial phenylnitrene-containing molecule, the final aminophenyl-containing chromophore, or any other pre-matured species. Molecular groups at the position 148 are always

View Article Online
DOI: 10.1039/D3OB01450A

termed His with possible extensions oxo-His or hydrated His. Below we pay attention to the transformations in the areas of the active site containing P-ring of Chro, His and water molecules in both proteins. Other important molecular groups (Arg96, Ser205, Glu222) are included to the model systems, but they will not be shown in the figures below.

Mass spectrometry was performed as outlined previously.[7]

The files with atomic coordinates of all key structures occurred at the reaction pathways are available in the general-purpose open-access repository Zenodo and can be accessed via https://doi.org/10.5281/zenodo.8327920.

Results

Figures 2, 3 illustrate structural features of the chromophore and closely associated residues (termed active site from hereon in) in the sfGFP^{66nF} and Venus^{66nF} models in the triplet spin state as revealed in the present work using the QM/MM and QM/MM MD simulations. Fig. 2 shows a typical frame along QM/MM MD trajectories in the vicinity of the starting point in sfGFP^{66nF}. Fig. 3 shows a minimum energy point on the potential energy surface in Venus^{66nF} optimized by the QM/MM method.

We intentionally apply related but different QM/MM-based approaches to model the two similar variants of original GFP to show that the major conclusion of the present work on the coupled reactions of the nitrene intermediate reduction and the histidine oxidation holds in both cases irrespective on details of simulations. Both model systems show a typical motif of the Chro-Ser205-Glu222 fragment of the chromophore-containing pocket in the GFP-like proteins [2,35]. The His148 side chain and a water molecule denoted Wat1 are located close enough to the nitrogen atom Nα of Chro despite some differences in hydrogen-bonding patterns in both systems.

View Article Online
DOI: 10.1039/D3OB01450A

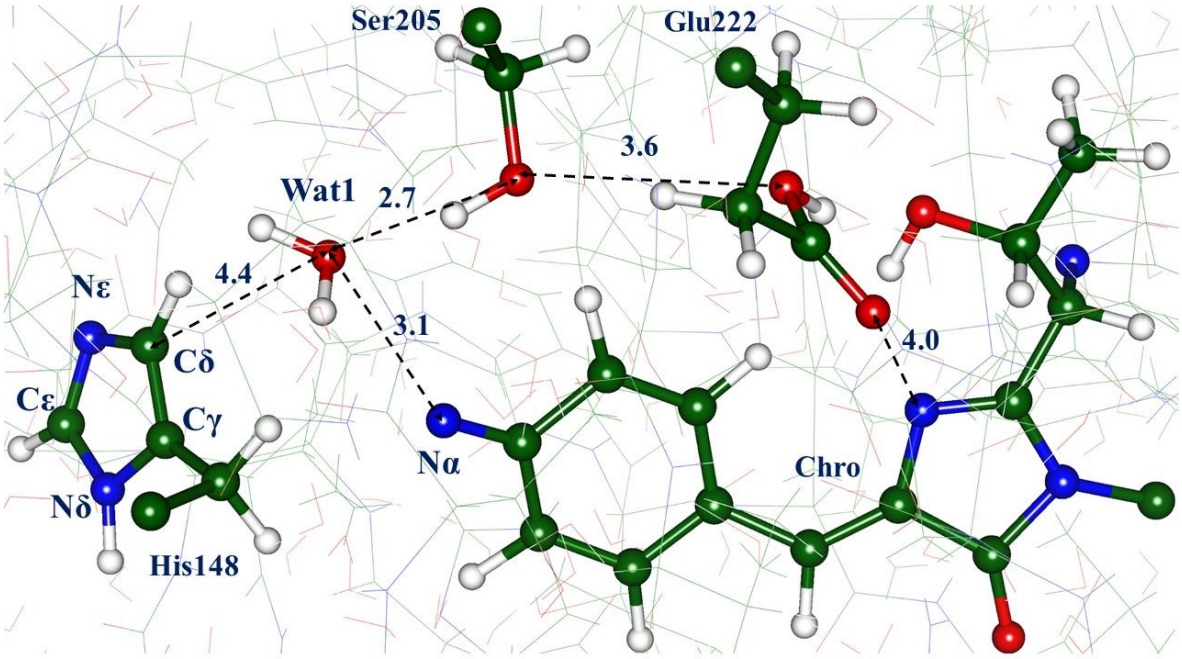


Figure 2. The active site in the sfGFP^{66nF} model system.

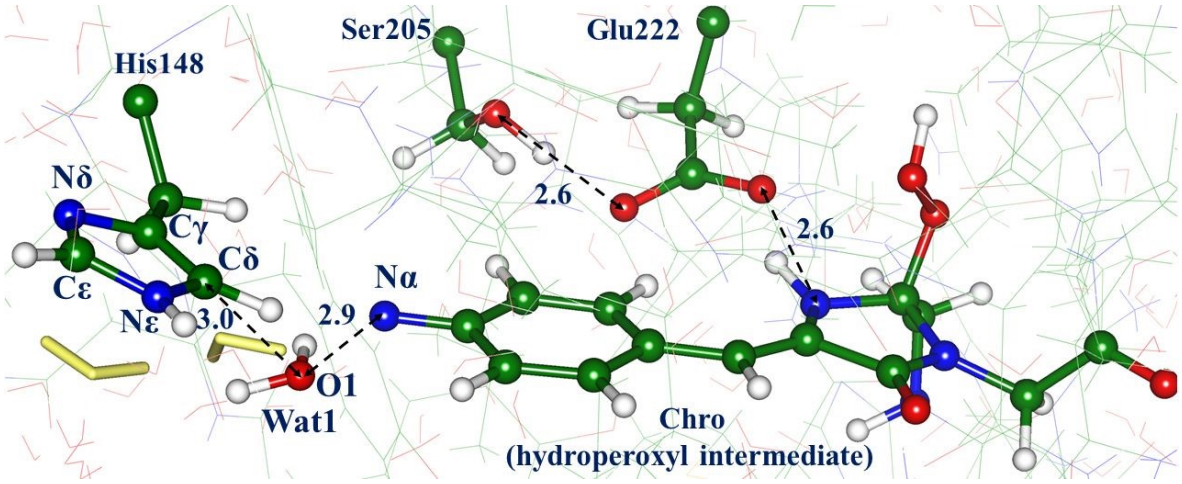


Figure 3. The active site in the Venus^{66nF} model system in the configuration of the hydroperoxyl intermediate on the chromophore maturation pathway.

Reduction of phenylnitrene in sfGFP^{66nF}

QM/MM MD trajectories were initiated from the structure of reactants (Fig. 2) termed Reac. We calculated trajectories in both singlet and triplet states to determine the multiplicity of the lowest energy level in the reactant region. For both trajectories, we found that the triplet state was 5-8 kcal/mol lower in energy than the singlet state for all calculated geometry configurations. Therefore, the subsequent calculations were performed in the triplet state. We understand the

View Article Online
DOI: 10.1039/D3OB01450A

limitations of the unrestricted DFT method for the singlet state[36], but we can at least semi-quantitatively evaluate the relative energies of the triplet and singlet states with the same geometry configurations.

An increased spin density on the Nα atom of Chro in the triplet state (Fig. 4b) promotes abstraction of a hydrogen atom from a nearby molecular group. A nearby water molecule Wat1 (Fig. 2) is an ideal candidate for this reaction.

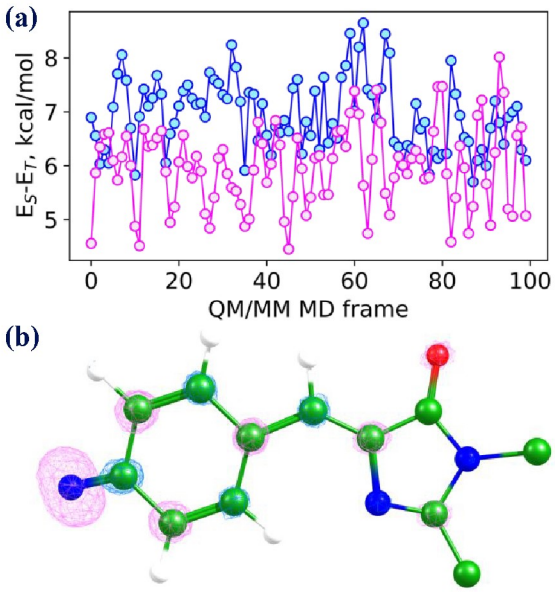


Figure 4. (a) Energy differences of the singlet and triplet states (E_S-E_T) computed along QM(PBE0-D3/6-31G**)/MM(CHARMM) MD trajectories in the reactants region in sfGFP^{66nF}. Energy frames from the trajectories are extracted every 20 fs. Here and in the following figures blue and magenta colors correspond to geometry configurations on the triplet and singlet states, respectively. (b) Spin density on the phenylnitrene-containing Chro evaluated at the 0.013 a.u. isovalue. Pink surfaces correspond to the excess of the α-spin electron density, blue – to the β-spin density.

We explored QM(UDFT)/MM MD calculations to construct the Gibbs energy profile for the reaction step from Reac to the first intermediate Int1 using a difference of two distances: d(O-H) in Wat1 and d(N(Chro)-H(Wat1)) as a collective variable in the umbrella sampling technique. The scheme of the reaction is illustrated in Fig. 5a; the computed energy profile is shown in Fig. 5b.

View Article Online
DOI: 10.1039/D3OB01450A

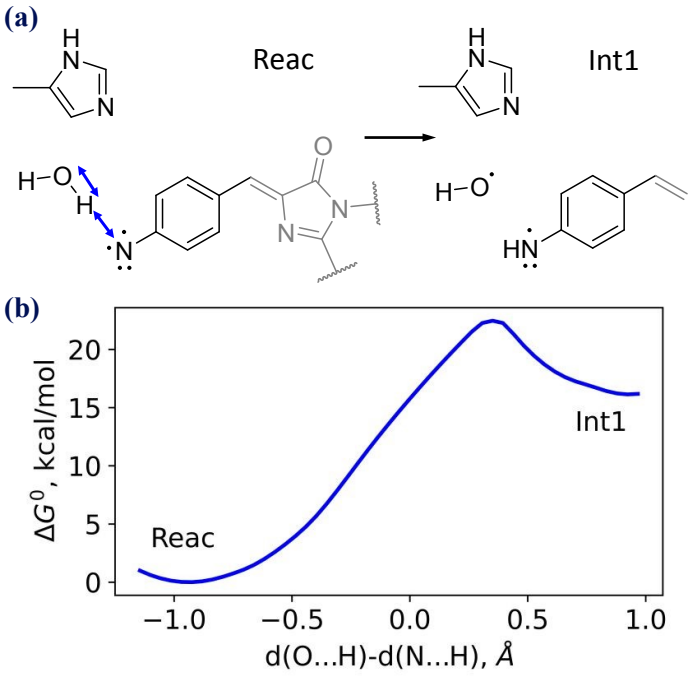


Figure 5. (a) The first step of the phenylnitrene reduction in sfGFP^{66nF} calculated at the QM(PBE0-D3/6-31G**)/MM(CHARMM) MD level in the triplet state. (b) The computed Gibbs energy profile for the step Reac → Int1 in the triplet spin state. The reaction coordinates are clarified by blue arrows in panel (a).

The minimum energy structure Int1 formally comprises the hydroxyl radical, which can react with the imidazole ring of His148. Computational results by Safaei et al. [20] show that the hydroxyl radical can attack different atoms in the imidazole ring; in our case the Cδ atom is a target (see Fig. 6a). The computed Gibbs energy profile for the Int1 → Int2 reaction step along the coordinate d(Cδ-O(Wat1)) is shown in Fig. 6b.

View Article Online
DOI: 10.1039/D3OB01450A

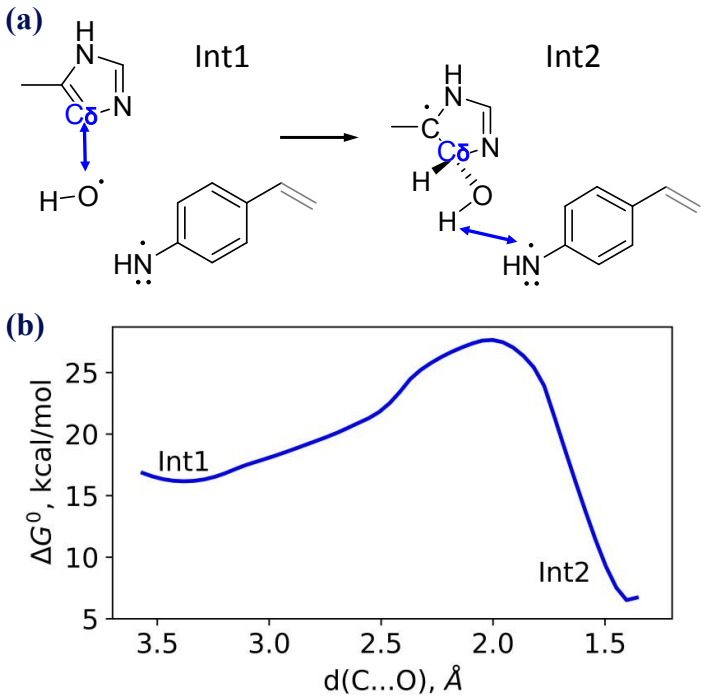


Figure 6. (a) The Int1 → Int2 elementary step in the phenylnitrene reduction in sfGFP^{66nF} calculated at the QM(PBE0-D3/6-31G**)/MM(CHARMM) level in the triplet state. (b) The computed Gibbs energy profile in the triplet spin state. The reaction coordinate is clarified by a blue arrow in panel (a).

According to the present calculations, the reaction intermediates Int1 and Int2 include the anilino radical of Chro tentatively proposed in the experimental work by Morris et al. [6].

The electronic structure of the model system in the Int2 reaction intermediate region (Fig. 7a) changes significantly. We calculated energy difference between the singlet and triplet state in the Int2 region in QM/MM MD simulations on the triplet state (Fig. 7b). The singlet state turned out to be 1–5 kcal/mol lower in energy in contrast to the preceding minima regions. This small difference indicates that this is a region of triplet and singlet surfaces crossing. At some frame of the MD simulation in the triplet state, we switched to the singlet state and continued the trajectory run. After the triplet to singlet cross over the energy difference between triplet and singlet states gradually increased along the QM/MM MD trajectory. It resulted in a considerable energy stabilization (more than 15 kcal/mol) of the singlet state relative to the triplet state. The following computations of the reaction steps were performed in the singlet state. Panels (c) and (d) in Fig. 7 show the spin density isovalues on the imidazole ring of the hydrated His and the Chro in both spin states. Both states are biradical with the spatially separated radicals; one is mostly localized on the Cγ of the hydrated His and the other one on the chromophore nitrogen P ring region. It

View Article Online
DOI: 10.1039/D3OB01450A

seems that significant spatial separation of radicals leads to the stabilization of the singlet biradical state.

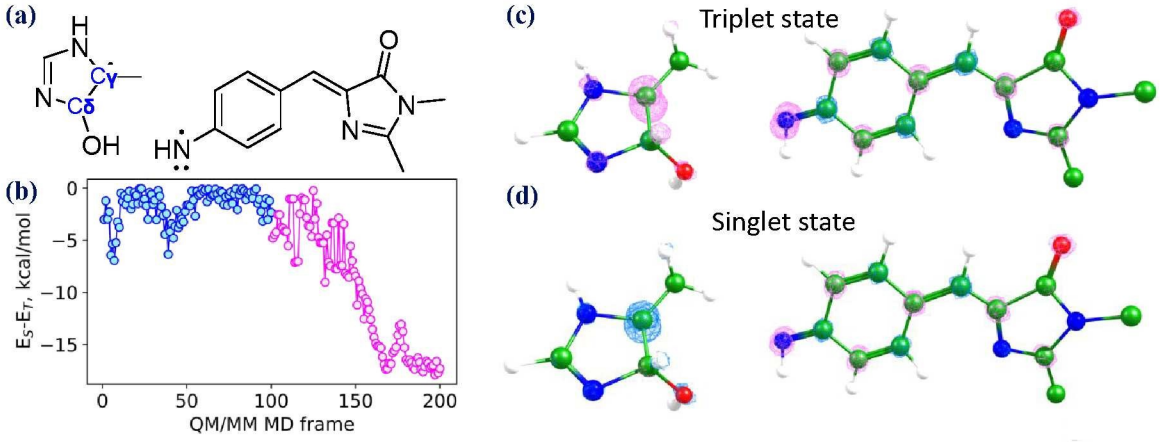


Figure 7. Features in the Int2 reaction intermediate. (a) Chemical structure of the Int2. (b) Energy differences of the singlet and triplet states (E_S-E_T) calculated at a set of QM(PBE0-D3/6-31G**)/MM(CHARMM) frames. Blue points correspond to the MD trajectory in the triplet state and magenta is its progress after switch to the singlet state. Energy frames from the trajectories are extracted every 10 fs. (c) Spin density in the T_0 state. (d) Spin density in the S_0 state. In panels (c) and (d), pink surfaces correspond to the excess of the α-spin electron density, blue – to the β-spin density. The contours are shown at the 0.013 a.u. isovalue.

The rest of the reaction pathway Int2 → Int3 → Products shown in Fig. 8 was modeled at the singlet state free energy surface using QM(DFT)/MM MD simulations. The reaction products comprise the phenylamine-containing chromophore and the oxidized histidine side chain Cδ-oxo-His. We shifted the energy of the Int2 in the singlet state by 17 kcal/mol (as it was shown on Figure 7(b)) relative to the Int2 in the triplet state. We are not confident about the exact value, but we stress that the stabilization upon triplet to singlet transition is considerable.

View Article Online
DOI: 10.1039/D3OB01450A

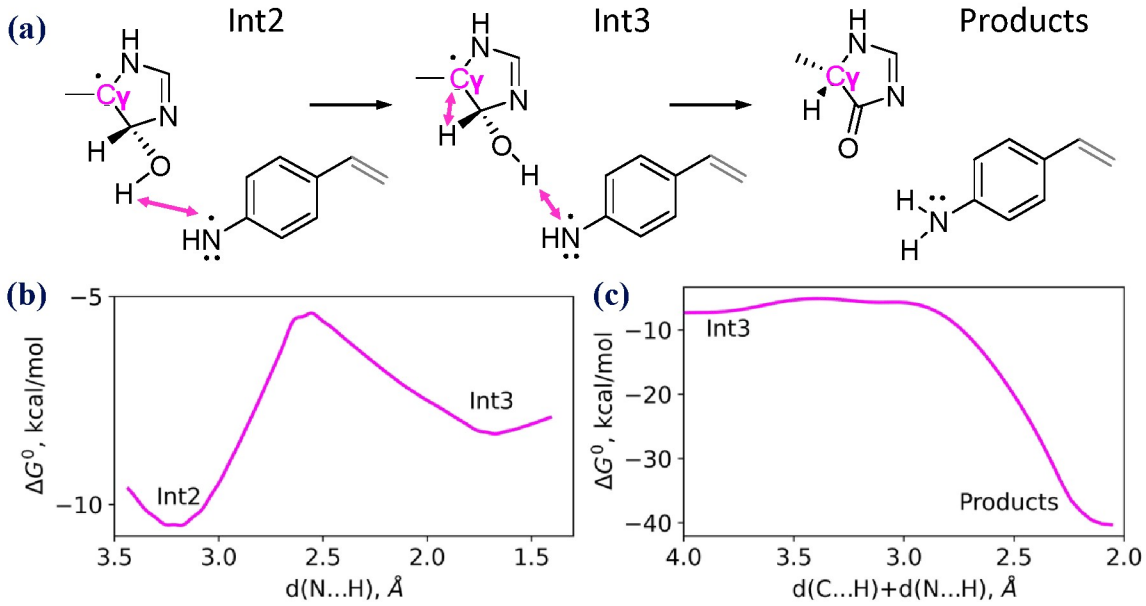


Figure 8. Reaction steps of the phenylnitrene reduction in sfGFP^{66nF} on the singlet state free energy surface calculated at the QM(PBE0-D3/6-31G**)/MM(CHARMM) level. (a) Chemical diagrams of the formation of the oxo-His – amino-Chro products. (b) The computed Gibbs energy profile for the step Int2 → Int3. (c) The computed Gibbs energy profile for the step Int3 → Products. The reaction coordinates are clarified by blue arrows in panel (a).

The analysis of the Gibbs energy profiles (Fig. 5a, Fig. 6b, Fig. 8b,c) shows that the highest energy barrier is about 27 kcal/mol and corresponds to the first elementary step Reac → Int1, whereas the energy of the products is about 20 kcal/mol lower than the level of the reactants.

Reduction of phenylnitrene in Venus^{66nF}

The reaction route from the reactants with the phenyl nitrene-containing Chro in Venus^{66nF} to the products containing the Cδ-oxo-His species and the phenyl amine-containing Chro qualitatively resembles the mechanism described in the preceding section for sfGFP^{66nF}. The use of QM/MM simulations for Venus^{66nF} allows us to analyze structures corresponding to the minimum potential energy points on the pathway from the reactants (REAC) to the products (PROD) via the reaction intermediates INT1 and INT2 (see Fig. 9). We denote these structures using solely capital characters to differentiate them from the sfGFP^{66nF} reaction. For instance, the INT1 structure on the Venus^{66nF} route (Fig. 9b) corresponds to the Int2 structure on the sfGFP^{66nF} route (Fig. 6a), because no minimum on the potential energy surface (PES) between REAC and INT1 resembling Int1 in sfGFP^{66nF} was located.

View Article Online
DOI: 10.1039/D3OB01450A

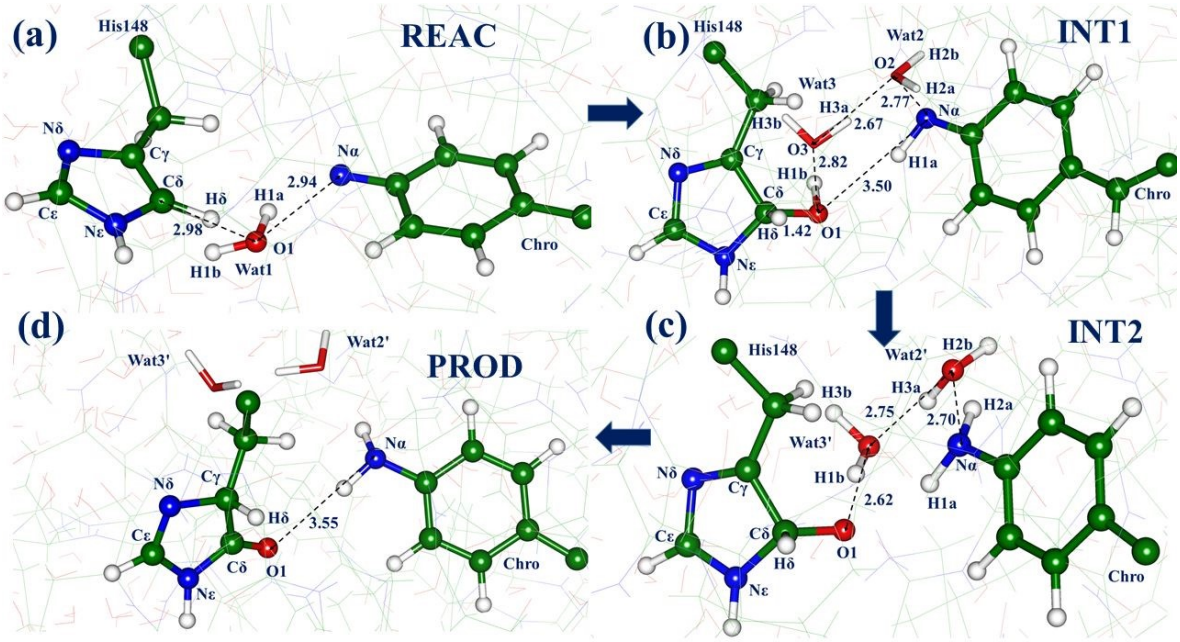


Figure 9. Reaction pathway from the system containing His and phenylnitrene to the system containing oxo-His and phenylamine in Venus^{66nF} calculated at the QM(PBE0-D3/6-31G*)/MM(AMBER) level.

As with sfGFP^{66nF}, the triplet state energy of reactants is lower than the energy of the singlet state, with the energy difference being 12 kcal/mol (Figure 10a). Accordingly, the reaction segment REAC → INT1 proceeds via the triplet state PES. Here, the hydrogen atom H1a from Wat1 is abstracted by the Nα atom of Chro with a simultaneous binding of O1-H1b radical to the Cδ atom of His. The energy of INT1 is 6 kcal/mol lower than the energy of REAC, whereas the energy barrier is about 20 kcal/mol (Figure 10a).

A coupled proton and electron transfer necessary to form the phenylamine-containing Chro at the INT1 → INT2 step is mediated by two additional water molecules Wat2 and Wat3 forming a suitable hydrogen-bond network (see Fig. 9b). The energy of the INT2 structure is 15 kcal/mol lower than the level of INT1. The energy of the singlet state of the reaction intermediate INT1 is higher in energy compared with the corresponding triplet state, whereas in the INT2 it is more stable. Therefore, subsequent reaction steps were considered on the singlet state PES. The final step INT2 → PROD describes migration of the hydrogen atom Hδ from the initial position at Cδ to the Cγ atom of His148, similarly to the reaction in sfGFP^{66nF} (see Fig. 8a, 10b). The energy of PROD is 27 kcal/mol lower than the energy of the preceding intermediate INT2, which is close to the estimates for the step Int3 → Products (Fig. 8c) in the transformation of sfGFP^{66nF}. As with

View Article Online
DOI: 10.1039/D3OB01450A

the latter case, the reaction products for Venus^{66nF} comprise the phenyl amine-containing chromophore and the oxidized histidine side chain Cδ-oxo-His.

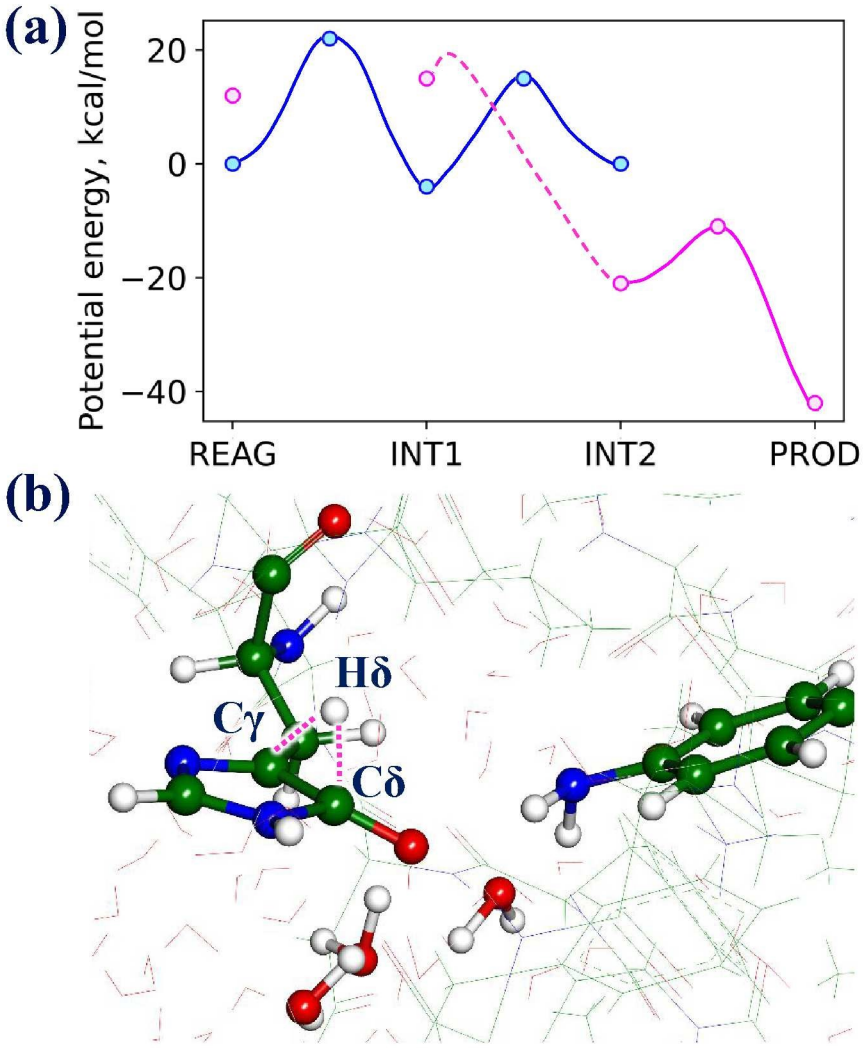


Figure 10. (a) Potential energy profile of chemical transformations in Venus^{66nF} corresponding to chemical structures in Figure 9. Blue correspond to the triplet energy surface. Energy profile in magenta corresponds to the proton transfer from Cδ to Cγ with the 9 kcal/mol energy barrier, corresponding transition state structure is shown on panel (b). All calculations are performed at the QM(PBE0-D3/6-31G*)/MM(AMBER) level.

One aspect to consider here is mass spectrometry studies of Venus^{66azF} described in [7]. According to this analysis, the electrophoresis (SDS-PAGE) showed that after 1 min of UV irradiation a band at ~20 kDa appeared and was constant for between 5-30 mins. This is an indication of fragmentation of the initial protein of ~30 kDa, which is in line with the known fragmentation of GFP-like proteins [37]. Shown here for clarity is the mass spectrum of Venus^{66azF},

View Article Online
DOI: 10.1039/D3OB01450A

which has the expected mass of full length, mature protein with the phenyl azide intact (Figure 11). The electrospray ionization time-of-flight mass spectrum revealed the major product at 19958 Da, which was assigned [7] as a result of cleavage of the C_{65}-$C_{\alpha 65}$ bond in the reaction product and tentative hydroxylation of Chro (see Fig. 11a). However, the results of present simulations show that the same peak observed in the mass spectrum can be assigned to the non-hydroxylated Chro and oxidized His148 (Fig. 11b). Reanalysis of the mass spectra revealed a secondary peak with a mass that could represent the full-length Venus with 66nF converted to the amine and oxygenated product, putatively the Cδ-oxo-His148 species.

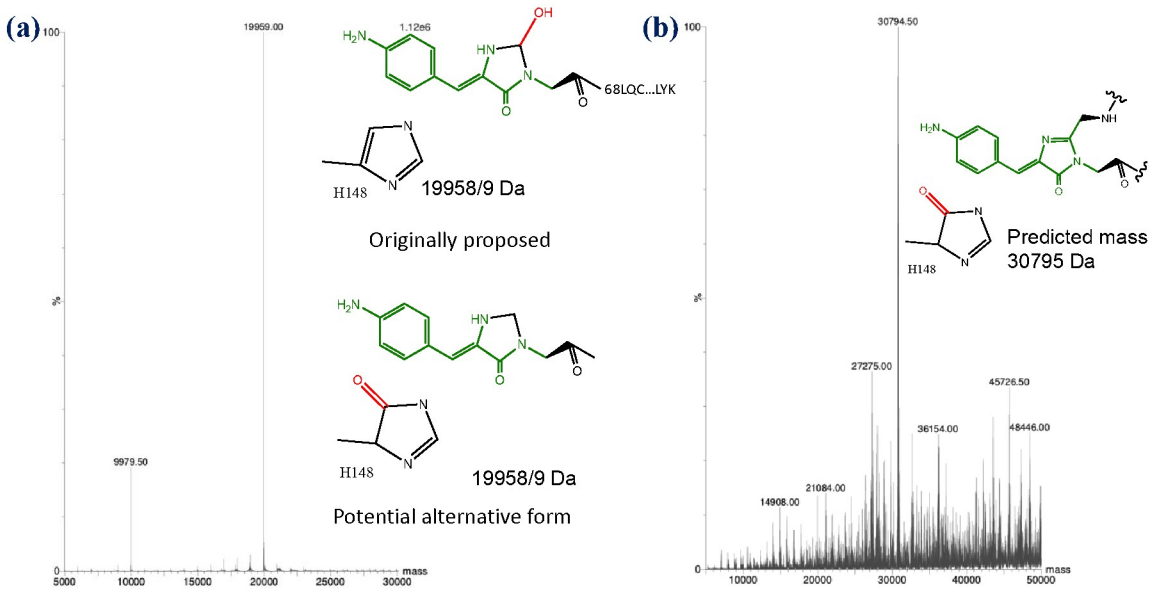


Figure 11. Mass spectra of Venus^{66azF} UV irradiated for 30 min. Mass spectroscopy was carried out as outlined previously [7]. (a) Mass spectrum equivalent to Venus^{66} C-terminal fragment cleaved at the chromophore upon denaturation. (b) Mass spectrum of full length Venus^{66azF} with a major peak corresponding to a mass that putatively represents Cδ-oxo-His species. Full length Venus^{66azF} is 30806 Da with conversion to the amine resulting in the loss of 26 Da (loss of N_2 but protonation of the phenyl nitrene). This leaves a mass deficit of 15 Da compared to the observed peak which can be attributed to the loss of a hydrogen and subsequent addition of an oxygen to the Cδ -His148.

The structure with the Cδ-oxo-His species identified as PROD in Fig. 9d seems to be a local minimum energy point on the PES. Even lower energy structures can be obtained assuming the interaction of reactive His with the surrounding water molecules, which fill the protein cavity. Fig. 12 shows a chain of minimum energy points leading consequently to lower energy points on the singlet state PES initiated by the structure INT2 in Fig. 9c. The result might be complete defragmentation of the imidazole ring of the initial His residue as illustrated in Fig. 12d.

View Article Online
DOI: 10.1039/D3OB01450A

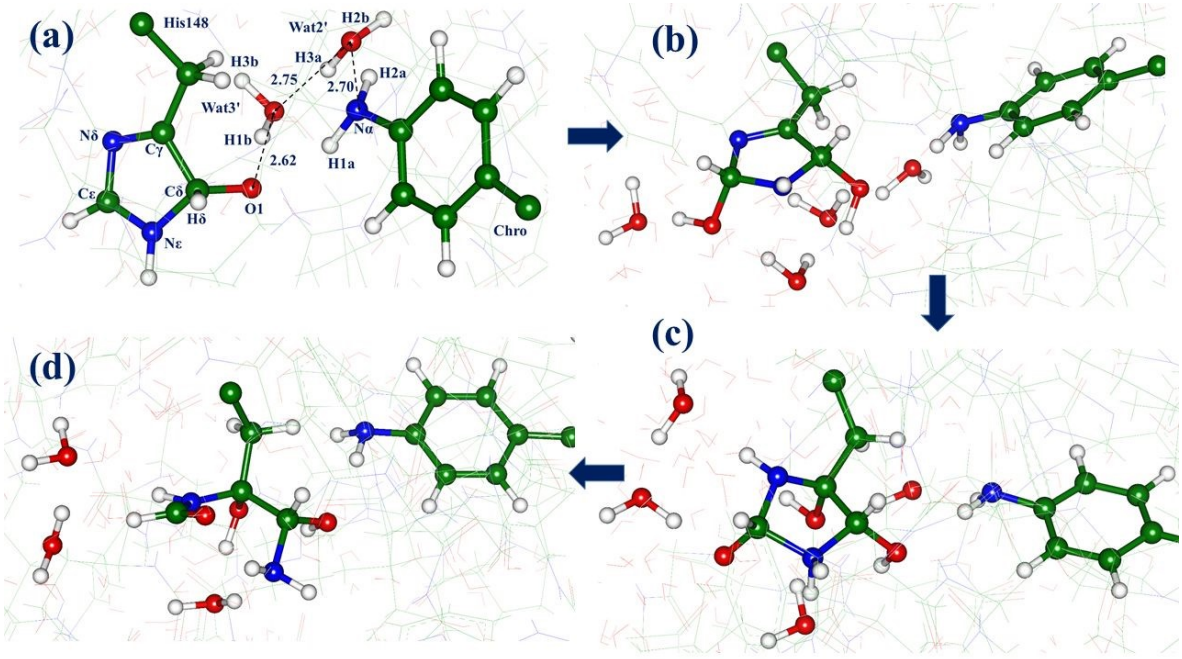


Figure 12. Possible defragmentation of the His species due to its interaction with the surrounding water molecules in Venus^{66nF}.

We note that similar defragmentation of the imidazole ring of histidine after its oxidation in proteins has been suggested previously [19,38–41]. We also note that other scenarios of histidine defragmentation in water starting from the Cδ-oxo-His species can be realized. The one shown in Fig. 12 illustrates a complexity of molecular processes occurring after reduction of the arylnitrene-containing chromophore in Venus^{66azF}.

Discussion

In this work, we consider related, but different, engineered versions of the original GFP protein sfGFP^{66azF} and Venus^{66azF} that contain a non-canonical amino acid residue p-azidophenylalanine instead of Tyr66. An important common feature of these macromolecules is that they include the His148 side chain near the P-ring of the chromophore and this area is solvent accessible. The light-induced decomposition of the azide group leads to arylnitrene intermediates followed by a previously unknown mechanism of amine formation. Simulations carried out in this work show that despite somewhat different conformations of two proteins and distinct computational protocols based on QM(DFT)/MM interaction potentials, the reduction of the arylnitrene intermediates is coupled with the oxidation reaction of His148 leading to the Cδ-oxo-His species. One of the intermediates on these reaction pathways comprises the anilino radical of Chro observed in the experimental studies of sfGFP^{az66F} [6]. Our conclusion on the formation of the

View Article Online
DOI: 10.1039/D3OB01450A

oxo-His species in these macromolecules is in line with the previous experimental results described for miniSOG[18] as well as for the metal-catalyzed oxidation of histidine to the oxo-histidine in peptides and proteins[19]. Our simulations for Venus show that even additional fragmentation of oxo-His cannot be excluded as mentioned in [19]. In both models, we started with the same orientation of the His relative to the chromophore. In both of models, the OH radical is located closer to the Cδ than the Cε atom of His148. Both of these carbon atoms can participate in formation of a covalent bond with the OH radical and the reaction path is determined by the distance between the carbon and oxygen atoms. The side chain of His148 can rotate by 180° around Cβ–Cγ bond compared with our models. If rotation does occur, the Cε of histidine residue is located closer to the OH radical than Cδ. Thus, the Cε –oxo derivative will form.

Besides reporting novel aspects in the chemistry of arylnitrenes, this work also contributes to the fascinating photochemistry of chromophore maturation in fluorescent proteins. Despite multiple experimental and theoretical efforts to dissect the chromophore maturation process into chains of elementary chemical reactions inside protein matrices [2,11–13], many details remain uncovered. Our previous work[7] mainly focuses on the oxidation step in the chromophore synthesis in the Venus^{66azF} protein and presents arguments favoring the amine formation before the final oxidation step is completed. Specifically, the azide-containing chromophore should impede fluorescence until irradiated; the azide group is thought to act as an excited-state quencher until converted to the amine [5,42]. Thus, the structure of the hydroperoxyl intermediate derived computationally previously [7] is considered as a starting point in the present modeling. Considering the entire maturation process, this is an intermediate on the full pathway appeared after cyclization, dehydration and partial oxidation of the initial polypeptide with an artificial amino acid. The present paper completes description of the mechanism. After creation of the amine-containing Chro, final oxidation of the chromophore in Venus^{66azF} should take place.

Conclusion

We have characterized *in silico* the reaction mechanism of reduction of arylnitrene intermediates formed in the light-induced decomposition of chromophores in the engineered variants of the GFP-like proteins sfGFP^{66azF} and Venus^{66azF} containing the non-canonical amino acid residue p-azidophenylalanine instead of tyrosine (residue 66) within the chromophore. Application of the QM/MM-based approaches allowed us to evaluate the reaction energy profiles from the nitrene intermediates to amines and to show that the arylnitrene-containing chromophore

View Article Online
DOI: 10.1039/D3OB01450A

reduction is coupled with the oxidation and possible further decomposition of the histidine side chain His148 located near the P-ring of the chromophore.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This work was supported by the Russian Science Foundation (project #22-13-00012). We acknowledge the use of supercomputer resources of the Joint Supercomputer Center of the Russian Academy of Sciences and the equipment of the shared research facilities of HPC computing resources at Lomonosov Moscow State University.

References

1 E. A. Rodriguez, R. E. Campbell, J. Y. Lin, M. Z. Lin, A. Miyawaki, A. E. Palmer, X. Shu, J. Zhang and R. Y. Tsien, *Trends Biochem. Sci.*, 2017, **42**, 111–129.

2 M. Zimmer, *Chem. Rev.*, 2002, **102**, 759–782.

3 J.-D. Pédelacq, S. Cabantous, T. Tran, T. C. Terwilliger and G. S. Waldo, *Nat. Biotechnol.*, 2006, **24**, 79–88.

4 T. Nagai, K. Ibata, E. S. Park, M. Kubota, K. Mikoshiba and A. Miyawaki, *Nat. Biotechnol.*, 2002, **20**, 87–90.

5 S. C. Reddington, P. J. Rizkallah, P. D. Watson, R. Pearson, E. M. Tippmann and D. D. Jones, *Angew. Chemie Int. Ed.*, 2013, **52**, 5974–5977.

6 J. L. Morris, S. C. Reddington, D. M. Murphy, D. D. Jones, J. A. Platts and E. M. Tippmann, *Org. Lett.*, 2013, **15**, 728–731.

7 H. S. Auhim, B. L. Grigorenko, T. K. Harris, O. E. Aksakal, I. V. Polyakov, C. Berry, G. dos P. Gomes, I. V. Alabugin, P. J. Rizkallah, A. V. Nemukhin and D. D. Jones, *Chem. Sci.*, 2021, **12**, 7735–7745.

8 N. P. Gritsan and M. S. Platz, *Chem. Rev.*, 2006, **106**, 3844–67.

9 A. M. Hartley, H. L. Worthy, S. C. Reddington, P. J. Rizkallah and D. D. Jones, *Chem. Sci.*, 2016, **7**, 6484–6491.

View Article Online
DOI: 10.1039/D3OB01450A

10 T. D. Craggs, *Chem. Soc. Rev.*, 2009, **38**, 2865.

11 D. P. Barondeau, C. D. Putnam, C. J. Kassmann, J. A. Tainer and E. D. Getzoff, *Proc. Natl. Acad. Sci.*, 2003, **100**, 12111–12116.

12 R. M. Wachter, *Acc. Chem. Res.*, 2007, **40**, 120–127.

13 N. V. Pletneva, V. Z. Pletnev, K. A. Lukyanov, N. G. Gurskaya, E. A. Goryacheva, V. I. Martynov, A. Wlodawer, Z. Dauter and S. Pletnev, *J. Biol. Chem.*, 2010, **285**, 15978–15984.

14 B. L. Grigorenko, A. I. A. I. Krylov and A. V. A. V. Nemukhin, *J. Am. Chem. Soc.*, 2017, **139**, 10239–10249.

15 S.-T. D. Hsu, G. Blaser and S. E. Jackson, *Chem. Soc. Rev.*, 2009, **38**, 2951.

16 A. M. Hartley, A. J. Zaki, A. R. McGarrity, C. Robert-Ansart, A. V. Moskalenko, G. F. Jones, M. F. Craciun, S. Russo, M. Elliott, J. E. Macdonald and D. D. Jones, *Chem. Sci.*, 2015, **6**, 3712–3717.

17 H. L. Worthy, H. S. Auhim, W. D. Jamieson, J. R. Pope, A. Wall, R. Batchelor, R. L. Johnson, D. W. Watkins, P. Rizkallah, O. K. Castell and D. D. Jones, *Commun. Chem.*, 2019, **2**, 83.

18 J. Torra, C. Lafaye, L. Signor, S. Aumonier, C. Flors, X. Shu, S. Nonell, G. Gotthard and A. Royant, *Sci. Rep.*, 2019, **9**, 2428.

19 C. Schöneich, *J. Pharm. Biomed. Anal.*, 2000, **21**, 1093–1097.

20 Z. Safaei, A. Shiroudi, E. Zahedi and M. Sillanpää, *Phys. Chem. Chem. Phys.*, 2019, **21**, 8445–8456.

21 M. Mroginski, S. Adam, G. S. Amoyal, A. Barnoy, A. Bondar, V. A. Borin, J. R. Church, T. Domratcheva, B. Ensing, F. Fanelli, N. Ferré, O. Filiba, L. Pedraza-González, R. González, C. E. González-Espinoza, R. K. Kar, L. Kemmler, S. S. Kim, J. Kongsted, A. I. Krylov, Y. Lahav, M. Lazaratos, Q. NasserEddin, I. Navizet, A. Nemukhin, M. Olivucci, J. M. H. Olsen, A. Pérez de Alba Ortíz, E. Pieri, A. G. Rao, Y. M. Rhee, N. Ricardi, S. Sen, I. A. Solov'yov, L. De Vico, T. A. Wesolowski, C. Wiebeler, X. Yang and I. Schapiro, *Photochem. Photobiol.*, 2021, **97**, 243–269.

22 M. G. Khrenova and A. P. Savitsky, in *Theoretical and Computational Photochemistry*, Elsevier, 2023, pp. 337–349.

View Article Online
DOI: 10.1039/D3OB01450A

23 C. Adamo and V. Barone, *Chem. Phys. Lett.*, 1997, **274**, 242–250.

24 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.

25 R. B. Best, X. Zhu, J. Shim, P. E. M. Lopes, J. Mittal, M. Feig and A. D. MacKerell, *J. Chem. Theory Comput.*, 2012, **8**, 3257–3273.

26 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.

27 J. C. Phillips, D. J. Hardy, J. D. C. Maia, J. E. Stone, J. V. Ribeiro, R. C. Bernardi, R. Buch, G. Fiorin, J. Hénin, W. Jiang, R. McGreevy, M. C. R. Melo, B. K. Radak, R. D. Skeel, A. Singharoy, Y. Wang, B. Roux, A. Aksimentiev, Z. Luthey-Schulten, L. V. Kalé, K. Schulten, C. Chipot and E. Tajkhorshid, *J. Chem. Phys.*, 2020, **153**, 044130.

28 S. Seritan, C. Bannwarth, B. S. Fales, E. G. Hohenstein, C. M. Isborn, S. I. L. Kokkila-Schumacher, X. Li, F. Liu, N. Luehr, J. W. Snyder, C. Song, A. V. Titov, I. S. Ufimtsev, L. Wang and T. J. Martínez, *WIREs Comput. Mol. Sci.*, 2021, **11**, e1494.

29 M. C. R. Melo, R. C. Bernardi, T. Rudack, M. Scheurer, C. Riplinger, J. C. Phillips, J. D. C. Maia, G. B. Rocha, J. V Ribeiro, J. E. Stone, F. Neese, K. Schulten and Z. Luthey-Schulten, *Nat. Methods*, 2018, **15**, 351–354.

30 J. Kästner and W. Thiel, *J. Chem. Phys.*, 2005, **123**, 144104.

31 A. Grossfield, 'WHAM: the weighted histogram analysis method', version 2.0.9, http://membrane.urmc.rochester.edu/content/wham.

32 W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, *J. Am. Chem. Soc.*, 1995, **117**, 5179–5197.

33 M. Valiev, E. J. Bylaska, N. Govind, K. Kowalski, T. P. Straatsma, H. J. J. Van Dam, D. Wang, J. Nieplocha, E. Apra, T. L. Windus and W. A. de Jong, *Comput. Phys. Commun.*, 2010, **181**, 1477–1489.

34 T. Vasilevskaya, M. G. Khrenova, A. V. Nemukhin and W. Thiel, *J. Comput. Chem.*, 2015, **36**, 1621–1630.

35 A. Acharya, A. M. Bogdanov, B. L. Grigorenko, K. B. Bravaya, A. V. Nemukhin, K. A. Lukyanov and A. I. Krylov, *Chem. Rev.*, 2017, **117**, 758–795.

36 M. Shoji, H. Isobe, T. Saito, H. Yabushita, K. Koizumi, Y. Kitagawa, S. Yamanaka, T.

View Article Online
DOI: 10.1039/D3OB01450A

Kawakami, M. Okumura, M. Hagiwara and K. Yamaguchi, *Int. J. Quantum Chem.*, 2008, **108**, 631–650.

37 J. Wei, J. S. Gibbs, H. D. Hickman, S. S. Cush, J. R. Bennink and J. W. Yewdell, *J. Biol. Chem.*, 2015, **290**, 16431–16439.

38 D. A. K. Traoré, A. El Ghazouani, L. Jacquamet, F. Borel, J.-L. Ferrer, D. Lascoux, J.-L. Ravanat, M. Jaquinod, G. Blondin, C. Caux-Thang, V. Duarte and J.-M. Latour, *Nat. Chem. Biol.*, 2009, **5**, 53–59.

39 K. Uchida and S. Kawakishi, *FEBS Lett.*, 1993, **332**, 208–210.

40 H. Ihara, Y. Kakihana, A. Yamakage, K. Kai, T. Shibata, M. Nishida, K. Yamada and K. Uchida, *J. Biol. Chem.*, 2019, **294**, 1279–1289.

41 Y. Miyahara, K. Shintani, K. Hayashihara-Kakuhou, T. Zukawa, Y. Morita, T. Nakazawa, T. Yoshida, T. Ohkubo and S. Uchiyama, *Sci. Rep.*, 2020, **10**, 6333.

42 S. C. Reddington, S. Driezis, A. M. Hartley, P. D. Watson, P. J. Rizkallah and D. D. Jones, *RSC Adv.*, 2015, **5**, 77734–77738.