JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

The Protonation States of the Active-Site Histidines in (6−4) Photolyase

Karmen Condic-Jurkic,[†,§] Ana-Sunčana Smith,[‡,§] Hendrik Zipse,[∥] and David M. Smith*[,†,⊥]

[†]Department of Organic Chemistry and Biochemistry, Ruđer Bošković Institute, Bijenička 54, 10000 Zagreb, Croatia

[‡]Institute of Theoretical Physics, University Erlangen−Nürnberg, Staudtstrasse 9, 91058 Erlangen, Germany

[§]Excellence Cluster, Engineering of Advanced Materials, University Erlangen−Nürnberg, Nägelsbachstrasse 49b, 91052 Erlangen, Germany

[∥]Department of Chemistry, Ludwig-Maximilians Universität, Butenandtstrasse 13, 82131 München, Germany

[⊥]Computer-Chemie-Centrum, Universität Erlangen−Nürnberg, Nägelsbachstrasse 25, 91052 Erlangen, Germany

Supporting Information

ABSTRACT: The active sites of the (6−4) photolyases contain two conserved histidine residues, which, in the *Drosophila melanogaster* enzyme, correspond to His365 and His369. While there are nine combinations in which the three possible protonation states of the two histidines (with protons on Nδ (HID), Nε (HIE), or both Nδ and Nε (HIP)) can be paired, there is presently no consensus as to which of these states is present, let alone mechanistically relevant. EPR hyperfine couplings for selected protons of the $FADH^{\bullet}$ radical have previously been used to address this issue. Our QM/MM calculations show, however, that the experimental couplings are equally well reproduced by each of the nine combinations. Since the EPR results seemingly cannot be used to unequivocally assign the protonation states, the pK_a values of the two histidines were calculated using the popular PROPKA, H++, and APBS approaches, in various environments and for several lesions. These techniques consistently indicate that, at pH = 7, both His365 and His369 should be neutral, although His369 is found to be more prone to becoming protonated. In a comparative approach, a series of molecular dynamics simulations was performed with all nine combinations, employing various reference crystal structures and different oxidation states of the FAD cofactor. The overall result of this approach is in agreement with our pK_a results. Consequently, although the introduction of the reduced cofactor results in an increased stability for selected protonated states, particularly the His365═HID and His369═HIP combination, the neutral combination His365═HID and His365═HIE stands out as the most relevant state for the activity of the enzyme.

INTRODUCTION

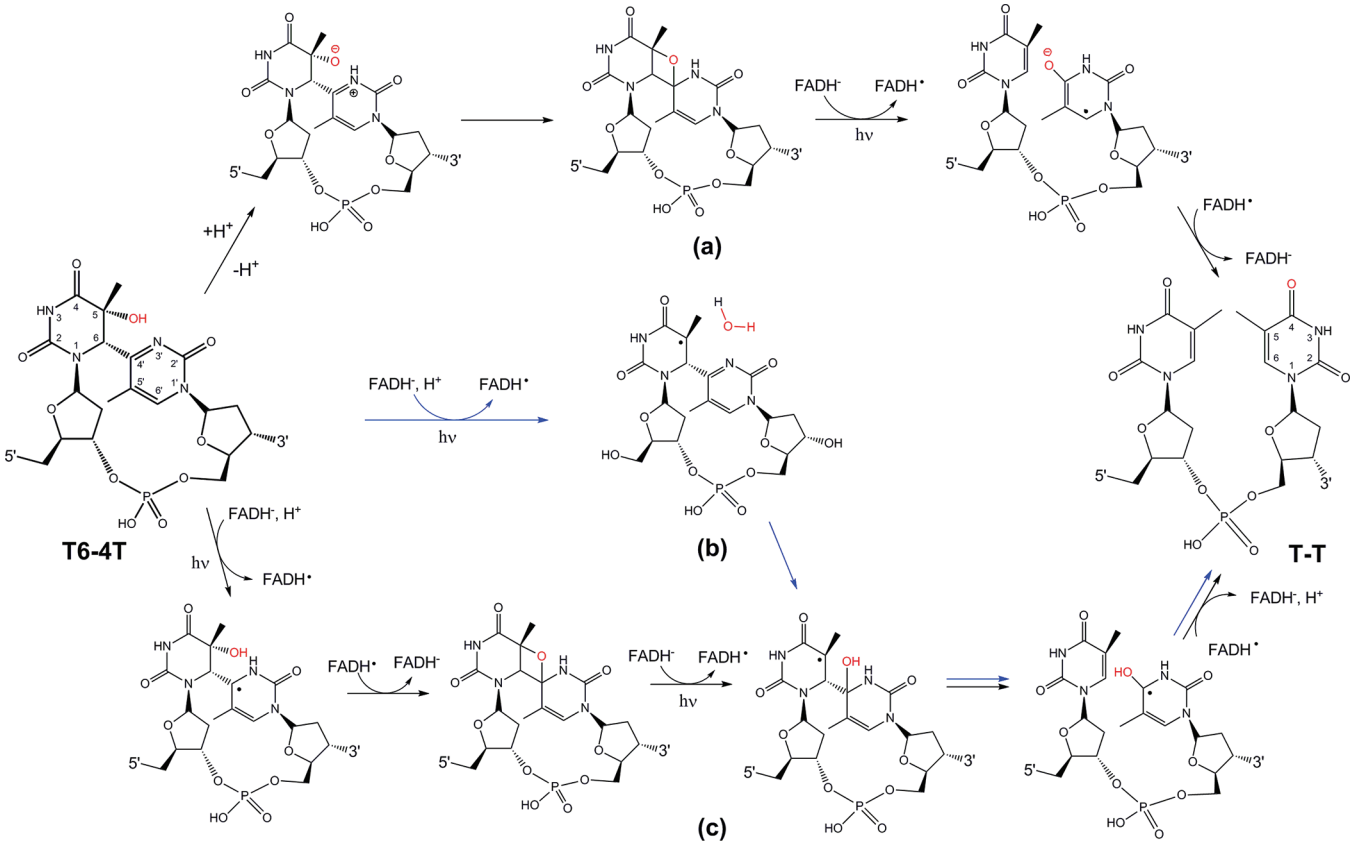
Ultraviolet radiation can cause harmful damage to DNA. For example, the exposure of two adjacent pyrimidine bases, such as the thymine−thymine (T−T) pair shown in Scheme 1, to UV light (in the 200−300 nm range) may result in dimerization. The most common forms of the corresponding lesions are the cyclobutane pyrimidine dimers (CPD) and the pyrimidine(6−4)pyrimidone photoproducts (see, e.g., T6−4T in Scheme 1).[1] Because of their high mutagenic and carcinogenic potential, the repair of these lesions is of utmost importance for cell survival. The group of light-dependent enzymes capable of reforming the initial monomers by photoreactivation is known collectively as the DNA photolyases. The formation of the CPD lesion and its repair by CPD photolyase has been well studied over the years, both experimentally[2] and computationally.[3,4] Much less is known, however, about the mechanism of repair of (6−4) photoproducts by (6−4) photolyase. While both types of enzymes show certain sequence similarities and require reduced $FADH^{-}$ in their active site for catalysis,[5] their repair mechanisms are believed to differ.[6]

It has been widely accepted that the formation of the (6−4) photoproduct is a result of a Paterno−Büchi cycloaddition reaction of the C5═C6 (5′) and C4═O/NH (3′) double bonds of the two neighboring pyrimidines (for T−T or T−C), which proceeds through an oxetane (for T6−4T) or an azetidine (for T6−4C) intermediate.[7,8] The initial proposal of the repair mechanism of the (6−4) photolyase included the thermal formation of this same cyclic intermediate as the first step,[9] catalyzed by two highly conserved histidines in the active site.[10] One histidine was proposed to act as a base, deprotonating the migrating functional group for the nucleophilic attack on C4′, while the other should act as an acid, donating the proton to the acylimine (Scheme 1a).[11] Subsequent electron donation from the excited FAD cofactor should enable the cycloreversion of the radical anion of the oxetane/azetidine,[12] a process which appears feasible on the basis of some experimental[13] and computational[14] studies. The final step involves the return of the electron to FAD. The energy gap between the T6−4T lesion and its oxetane isomer, predicted by theoretical calculations to be between 14.5 and 16.6 kcal mol^{-1},[15] has usually been identified as a major disadvantage of this mechanism. Similarly, the oxetane intermediate has not thus far been experimentally observed.

The determination of the crystal structure of (6−4) photolyase from *Drosophila melanogaster*, complexed with the T6−4T lesion, was a major step forward in the understanding of the repair.[16] On the basis of structural and biochemical data,

Received: August 12, 2011
Published: January 24, 2012

Scheme 1. (6-4) Photolyase Repair Mechanisms Proposed by (a) Hitomi et al,[11] (b) Maul et al. (blue arrows),[16] and (c) Sadeghian et al.[19a]



[a]Protons are donated or accepted by His365 and His369, respectively, but the residues are omitted for clarity. For the same reason, the water molecule participating in mechanism c is not shown.

a modified mechanistic proposal lacking the oxetane intermediate has emerged. Oxetane formation requires the protonation of the acylimine, a step that was reportedly not supported by the positions of the histidine residues. Instead, it was suggested that one of the histidines (His365) should protonate the migrating functional group, rather than deprotonating it, thus making it a better leaving group after the electron injection from FAD (Scheme 1b). The resulting water molecule could attack the acylimine to form a radical intermediate, which would then rapidly fragment into the initial pyrimidine bases. In this scenario, the loss of a proton along with electron transfer back to FAD would complete the catalytic cycle.

A subsequent computational study addressed the repair mechanism of the (6−4) photoproduct from an energetic point of view.[17] Using a model system that omitted the enzyme, the study investigated the possible fates of the radical anion of the T6−4T dimer, which is formed after electron capture from the excited FAD cofactor. It was found that pathways on the electronic ground state of the radical anion, either with or without an oxetane intermediate, were associated with high activation energies. The nonoxetane pathway, which involves a direct hydroxide transfer (analogous to the Scheme 1b without the addition or removal of a proton), was found to have a low-lying excited state whose relaxation could lead to lesion repair. In order to make productive use of this state, however, the reaction would require either a re-excitation (an additional photon) or, as the authors seemingly prefer, a directed nonadiabatic relaxation of the hot radical-anion state toward the product configuration. This study also highlighted the potential importance of the (non)formation of the intramolecular O5H−N3′ hydrogen bond. According to the spectroscopic analysis of the pyrimidine(6−4)pyrimidone photoproduct, the formation of this intramolecular hydrogen bond is responsible for the remarkably low pK_a value for the N3′ atom.[18]

During the preparation of the present article, an additional computational investigation appeared in the literature.[19] In agreement with the previous results,[17] the most recent study failed to identify a low-energy rearrangement mechanism on the ground state surface of the radical anion. Rather, through the extensive use of QM/MM calculations that explicitly included the protein environment, the authors advocate what is essentially a two-photon process. That is, the electron transfer from the FAD to the lesion is proposed to lead to the formation of an oxetane intermediate, which transfers its excess electron back to the cofactor (Scheme 1c). The oxetane formation is catalyzed by a protonated His365, and the additional stabilization is gained from a hydrogen bond formed between a water molecule and the hydroxyl group of the lesion. In the second step, another electron transfer from FAD, which presumably needs to be re-excited, is required to split the oxetane intermediate into the original monomers.

Despite the fact that the existing theoretical studies[17,19] seem to prefer a mechanism involving the excited state of the lesion or a re-excitation of either the lesion or the cofactor, two

dx.doi.org/10.1021/ct2005648 | J. Chem. Theory Comput. 2012, 8, 1078−1091

independent investigations have recently provided evidence that the repair reaction takes place entirely on the ground state of the reduced lesion. In an experimental study,[20] the repair photocycle of (6−4) photolyase was investigated by means of ultrafast spectroscopy, applied to the wild-type enzyme as well as inactive mutants. The authors report that, according to their results, a proton transfer to the anionic lesion is a key step in the repair pathway. They find that, after this step takes place, the ordinarily rapid back electron transfer (50 ps) from the lesion to the cofactor is completely blocked, and the subsequent ground-state repair occurs with 100% efficiency. Indeed, this back electron transfer has been suggested to be the main reason for the low repair efficiency of (6−4) photolyase. In a complementary study,[21] TD-DFT calculations have demonstrated that the initially absorbed photon does not carry sufficient energy to initiate the electron transfer and to simultaneously excite the radical anion of T6−4T. Similarly, using arguments based on the photon flux density of solar radiation at the Earth's surface in the range where $FADH^-$ and the lesion absorb, the authors argue against the involvement of a second photon in the repair mechanism. These arguments are likely to be more relevant to a mechanism of the type proposed in ref 17, where the intermediate is a short-lived radical anion, than a mechanism like that presented in ref 19, where the intermediate is a relatively stable closed-shell system. Namely, the latter species could have a lifetime that is sufficiently long for it to absorb a second, lower-energy photon and thus constitute a viable reaction intermediate.

Despite the success of previous studies, further work is required to reach a consensus as to the repair mechanism of (6−4) photolyase. One of the issues with the computational studies[17,19] may involve inadequate treatment of the protein environment. That is, even though the most recent study was performed with state of the art QM/MM techniques,[19] it is not entirely clear that the protonation states of the key active-site residues were assigned correctly. Given that all of the proposed mechanisms involve acid−base chemistry of one form or another and that the active site contains two conserved histidine residues, this is clearly an important issue. Indeed, one may argue that its resolution is a prerequisite for further mechanistic studies, especially those employing computational methods.

It is commonly thought that, in the (6−4) photolyase from *D. melanogaster* (see Figure 1, prepared with VMD[22]), His365 is protonated while His369 is neutral.[16,20] This assumption is usually based upon results from an EPR/ENDOR study performed on the (6−4) photolyase from *Xenopus laevis*, which exploited the formation of the $FADH^\bullet$ radical.[23] On the basis of changes in the principal components and the intensities of the hyperfine couplings of selected FAD protons (H8 and H1′), induced by changing pH values and introducing point mutations, the authors argued that His354 (corresponding to His365 in *D. melanogaster*) is protonated, while His358 (corresponding to His369 in *D. melanogaster*) is neutral at pH 9.5. It is important to mention that (6−4) photolyase repair activity exhibits strong pH dependence, as shown by the experimental study of Hitomi et al.[11] According to their measurements, *X. laevis* (6−4) photolyase reaches its maximum activity at pH 8.5, which decreases significantly upon approaching pH 6. The optimal activity of *D. melanogaster* (6−4) photolyase has been measured at pH 7.8.[16] Another interesting observation in this context comes from Li et al., who also studied the repair dynamics and the steady-state enzyme activity over a pH range of 7−9 and did not note any associated changes.[20] At the time of the aforementioned EPR study, however, the (6−4) photolyase crystal structure had not been solved. Given the relative positioning of the cofactor and the active site histidines apparent in the structure from *D. melanogaster*,[16] it is not entirely clear that the hyperfine coupling constants on the hydrogens of the cofactor are the most appropriate probes of the protonation states of the histidine residues.

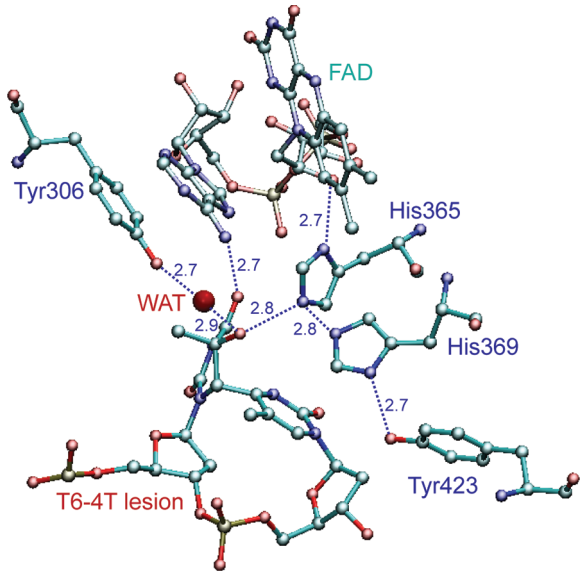


Figure 1. Hydrogen bond network connecting the T6−4T lesion, catalytic residues (His365, His369, Tyr423), and the FAD cofactor in the 3CVU crystal structure[16] of the active site of the (6−4) photolyase. Also shown are relevant interatomic distances in Ångstroms and the positions of Tyr306 and the oxygen atom of a structural water molecule.

It is against this background that we present here a systematic investigation of the protonation states of the conserved histidine residues in (6−4) photolyase by means of various computational techniques. An individual histidine residue may adopt one of two neutral tautomers, with a proton at the Nε (HIE) or Nδ (HID) position, as well as one protonated form, with protons on both Nε and Nδ (HIP). With these three possible states for each histidine, there are a total of nine possible ways in which the protons can be distributed among two histidine residues. In the case of (6−4) photolyase, which has a complex H-bonding network connecting the lesion, catalytic residues and the cofactor, it is difficult to definitively discard any of the nine possibilities, purely on the basis of structural or intuitive criteria. With this in mind, and in the interest of remaining systematic, we have explicitly taken all nine combinations into account.

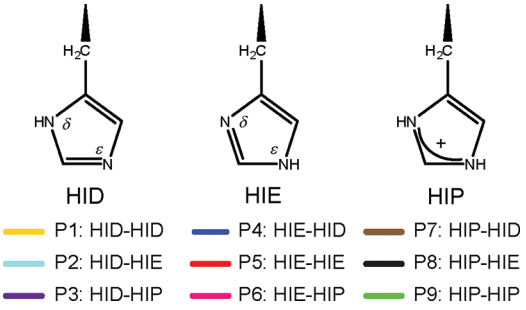
As the first step in our analysis, we have examined the suitability of using the hyperfine coupling constants of selected hydrogens of the $FADH^\bullet$ radical as a quantitative indicator for the histidine protonation states. Specifically, using a combined QM/MM approach, we have calculated the relevant hyperfine coupling constants for all nine protonation states and compared them to the measured values. In the next step, we have applied a range of popular methods for estimating the pK_a values of the histidine residues in an effort to determine if one can arrive at a definitive assignment of the protonation state in this way. For this purpose, we have investigated techniques of varying

sophistication, ranging from basic Web servers to more elaborate methods based on the solution of the Poisson–Boltzmann equation and free-energy cycles. Finally, we have carried out explicit molecular dynamics simulations on all nine possible combinations and monitored their structural evolution with respect to the crystal structure. Indeed, due to the interesting nature of the results of this latter aspect, we have applied an analogous approach to structures of the (6–4) photolyase complexed with variations of the T6–4C lesion[24,25] as well as of the repaired T–T product.[16] With the combination of these varied approaches, we are aiming to provide more definitive information concerning the protonation states of the conserved histidine residues and, in this manner, pave the way for further mechanistic studies.

COMPUTATIONAL DETAILS

Structural Models. Virtually all aspects of our study required the definition of a structural model. The initial configurations were taken from the appropriate crystal structure of (6–4) photolyase in complex with a DNA oligomer and the FAD cofactor. The relevant PDB entries were 3CVU, 3CVY,[16] 2WB2,[24] and 2WQ7,[25] which contained the T6–4T lesion, the repaired T–T dimer, the T6–4C lesion, and the methylated mT6–4C lesion, respectively. Before further processing, the original PDB files were slightly modified by removing duplicate entries and assigning the protonation states of histidines, other than His365 and His369, based on their local environments. For each crystal structure, nine alternative versions were prepared according to the nine possible protonation states of His365 and His369 (Scheme 2).

Scheme 2. Possible Combinations of Protonation States (P*n*) of His365 and His369[a]



[a]Each state is defined as P*n*: HIX365-HIX369 (X = D, E, or P) and associated with a certain color. This notation will be used throughout the text and Supporting Information.

Parameter Assignment. For standard nucleic acid residues, we assigned classical force-field parameters according to the AMBER99 (ff99) force field,[26] supplemented with the refined *parmbsc0* parameters (ff99bsc0).[27,28] For standard amino-acid residues, we employed the AMBER99SB (ff99SB) version of the AMBER protein force field.[29] For the nonstandard residues, such as the DNA lesions and tautomers and the various oxidation states of the FAD cofactor (oxidized FAD, $FADH^{\bullet}$ radical, and fully reduced $FADH^{-}$), missing parameters were derived using the Antechamber[30] suite and the GAFF force field[31] available in the AMBER9 program package.[32] Initial coordinates for the auxiliary FAD and DNA lesions were extracted from the crystal structures, and the geometries, with appropriately added hydrogen atoms, were optimized in the gas phase with Gaussian 03[33] at the HF/6-31G(d) level of theory (UHF for radical species). The electrostatic potential was determined for the optimum structures at the same level of theory. The final atom-centered point charges were fit to the resulting potential using the RESP procedure, and the requisite parameter files are included in the Supporting Information.[34]

Although the composite thymine monomers of the (6–4) lesions appear as separate residues in the PDB files, for ease of charge derivation, they were treated together as a single residue. We adopted a similar approach when treating an alternative tautomer of the repaired TT dimer. Accordingly, the ESP charges for the (di)nucleotides were derived such that the 5′-phosphate of the dimer was capped with a methoxy ($O-CH_3$) group, while the terminal O3′ of the dimer was capped with a proton.[35] During the RESP fitting, the charges on the capping groups were required to sum to zero and were later excluded from the final library to allow the correct connectivity with the rest of the DNA molecule.

While the use of electrostatic potentials calculated using HF/6-31G(d) could be considered somewhat outdated, we have chosen this approach primarily to arrive at a balanced and consistent treatment of the electrostatic parameters for the protein, the lesions, the DNA fragments, and the FAD cofactor. Indeed, the same reasoning lies behind the often employed combination of ff99SB and ff99bsc0 (both based of HF/6-31G(d) RESP charges) for systems that contain both amino and nucleic acids.[36] Force fields based on more sophisticated electrostatic potentials are available for proteins. For example, AMBER03 uses charges derived from B3LYP/cc-pVTZ calculations in a polarizable continuum, which incidentally correlate quite well with those present in ff99.[37] However, analogous charge sets are not yet available for DNA fragments, where HF/6-31G(d) charges are still widely employed.[28,38] Thus to avoid mixing different charge-derivation protocols within our system, we have chosen to use charges based on the HF-RESP combination, throughout our study.

Using the standard (ff99SB, ff99bsc0) and custom-built libraries, a topology file containing all necessary parameters for each system was built using the LEaP module of AMBER9. For those systems that were further subjected to molecular dynamics simulations in explicit solvent, each system was solvated with TIP3P waters in a truncated octahedron box, with an edge length of ~40 Å. All crystal water molecules present in the PDB file were retained. Sodium ions were added to neutralize the charge. The number of added ions (15–20) varied with the protonation states of the histidine residues, the length of DNA strands, and the oxidation state of FAD for each system. Overall, the sizes of the simulated systems ranged from 61862 (3CVY) to 74279 (2WQ7) atoms (see Table S5). Prior to any calculations, hydrogen bonds were oriented so as to produce an optimal (or at least nonclashing) interaction between the residues of interest (His365, His369, Tyr423, FAD, and the two nucleotide residues).

Calculation of EPR Parameters. Given that the hyperfine couplings (hfc's) of selected protons of the $FADH^{\bullet}$ radical were considered as probes for the protonation states of the nearby histidine residues in the EPR/ENDOR study[23] of the *X. laevis* (6–4) photolyase, we set out to calculate these hfc's using, primarily, the Gaussian 03 software package.[33]

Previous calculations of hfc's relevant to the ribonucleotide-reductase-catalyzed reaction[39] had found that the B3LYP/TZVP level of theory provided acceptable accuracy at a

dx.doi.org/10.1021/ct2005648 | J. Chem. Theory Comput. 2012, 8, 1078–1091

reasonable cost. In light of the relatively large systems considered herein, we have chosen this approach as our quantum mechanical benchmark.

We used the isolated $FADH^{\bullet}$ radical as a simple model for the purpose of method validation. The initial geometry was extracted from the crystal structure, and hyperfine couplings were calculated directly with B3LYP/6-31G(d). The structure was subsequently optimized at the same level of theory, and hfc's were again computed with B3LYP/6-31G(d) and B3LYP/TZVP.

The influence of the surrounding protein on the hfc's was investigated with a series of QM/MM calculations, carried out within the ONIOM[40] formalism available in Gaussian 03[33] for the P2 protonation state. Three different sizes of the QM region, which was treated with B3LYP/6-31G(d), were trialed in these calculations. The simplest model (QM1) consisted of only the $FADH^{\bullet}$ radical. Beyond this, we expanded the high-level model system to include the most important catalytic residues (His365, His369, and Tyr423) in QM2 and, additionally, the T6−4T lesion in QM3 (see Figure 1). Only the side chains of the amino acids, which were terminated between the $C_{\alpha}-C_{\beta}$ bonds, were retained in the QM region. The connections between the lesion and the DNA chain were introduced by placing link atoms between C4′ and C5′ ribose carbon atoms on both ends of the lesion.

The MM layer of the QM/MM models, whose parametrization was described above, consisted of the entire protein and the DNA duplex molecules surrounded by the 3000 water molecules closest to FAD. The starting geometry for the P2 state was extracted after the solvated, protonated crystal structure had been allowed to relax (see steps *i* and *ii* in the MD section below). Thereafter, a mobile region, consisting of all residues partially within 15 Å of the Nϵ atom of His365, as well as all water molecules partially within a 12 Å radius of the same atom, was defined. With the remainder of the system frozen, the structure was optimized with mechanical embedding[40] in combination with the appropriate QM region (QM1−3). Subsequent evaluations of the hfc's within the QM/MM framework were carried out within the electrostatic embedding formalism.[41] After settling on QM3 as the most appropriate QM region, the above protocol was applied to calculate the hfc's for all nine protonation states (Scheme 2).

Structure-based pK_a Calculations. The initial estimation of the pK_a values of His365 and His369 was performed using the PROPKA 2.0 server. Apart from being fully automated and very fast, PROPKA is a structure-based empirical method for the pK_a prediction of the ionizable residues in the proteins.[42] pK_a emerges as a sum of two terms ($pK_a = pK_{a,model} + \Delta pK_a$). $pK_{a,model}$ is associated with the unperturbed value for each residue (6.5 for His), whereas the shift (ΔpK_a) is treated as an environmental perturbation. Furthermore, for each ionizable group, ΔpK_a is given as a sum of perturbations due to desolvation, hydrogen bonding, and charge−charge interactions ($\Delta pK_a = \Delta pK_{Desolvation} + \Delta pK_{HB} + \Delta pK_{ChgChg}$).[43] A pdb file is used as input to the server, which provides pK_a values in a fully automated fashion.

For comparison, we also used the H++ server.[44] In addition to pK_a (strictly $pK_{1/2}$) values for ionizable residues, the server provides an output structure containing the missing hydrogens added according to the results of the pK_a evaluations.[45] The automated calculation is based on continuum solvent methodology,[46] within the framework of either the generalized Born (GB) or linearized Poisson−Boltzmann (PB) models, and accounts for each of the 2^N protonation microstates (when N ionizable residues are active).[47] In our calculations, we used the PB model and kept the default values for ionic strength (0.15 M) and the external dielectric constant ($\varepsilon_{ext} = 80$). We used an internal dielectric constant (ε_{in}) of 6, while the pH value was set to 7.

In principle, a ligand bound to the protein and its net charge can be included in H++ calculation, which is then parametrized in an automated procedure, although only one ligand can be processed per run. Because of this limitation, and the fact that we did not obtain satisfactory automatic parametrizations of the lesions or the FAD, we chose to complete our H++ calculations in three sequential steps.

The first step in each calculation involved the application of the recently introduced *reduce* algorithm,[48] which is used to identify the preferred orientation for ambiguously placed heavy atoms, as well as to differentiate and assign the most probable histidine tautomers. The second step involved the calculation of the pK_a values and the associated protonation states for all ionizable residues in the resulting structure, in the presence of an approximate parametrization for FAD (oxidized or reduced) but in the absence of the lesion (except for 3CVY where the T−T pair was present). The final refinement step involved recalculating the pK_a values for the two residues of interest, in the explicit presence of the properly parametrized FAD (oxidized or reduced) and the lesion.

This final step was achieved with the use of appropriate PQR files, which were prepared according to the previously outlined parametrization (ff99SB, ff99bsc0, etc.), without the crystal waters. Using these files, we evaluated the pK_a's of the two histidines together with all other titratable residues as well as in calculations where only a single target residue was allowed to titrate. Importantly, the relevant neutral tautomers (HID and HIE) for use in the PQR-based calculations were assigned on the basis of the results of the second step above.

A related approach for obtaining protein pK_a values is the Adaptive Poisson−Boltzmann Solver (APBS).[49] In cases of a single titration event, the estimation of the pK_a value of the chosen residue is based on a rigorous free energy cycle involving the transfer of the titratable group from solution to the protein. The transfer free energies (ΔpK_a) for protonated and deprotonated species are derived from the numerical solution of the Poisson−Boltzmann equation[50] using the Finite Element ToolKit.[51] The requisite model pK_a values in water (6.5 for His) were taken from the table provided by the developers.[52]

Using APBS, we have evaluated the transfer free energies associated with all possible single titration events from the states listed in Scheme 2, each in the presence the relevant DNA lesion and either oxidized or reduced FAD. We used the PDB2PQR[53] server to generate a template input file for the electrostatic energy calculations with recommended grid parameters for a given system. The PQR files used for APBS were analogous to those used for H++, with the added modifications required to close the free energy cycle. All calculations were carried out at 300 K with a solvent dielectric constant of 80, an ionic strength of 0.15 M, and a protein dielectric constant of 6.

In line with standard practice, all pK_a calculations were performed directly on the crystal structures in the absence of water molecules. According to ref 16, the crystals of T6−4T were grown in a reservoir buffer with a pH of either 7.0 or 8.6, after which they were rinsed in a cryoprotection solution of pH

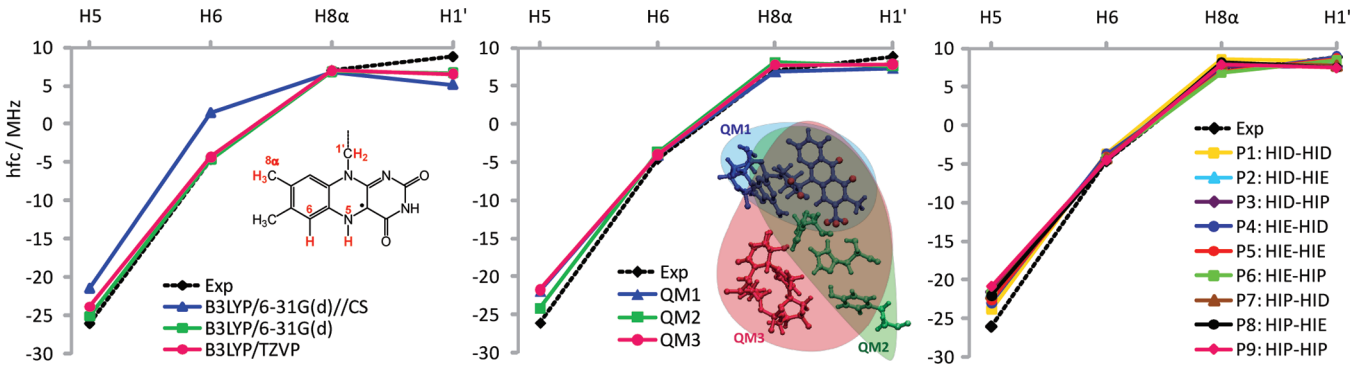


Figure 2. Calculation of hyperfine couplings of selected protons (colored in red) on FADH• radical compared to experimental data:[23] (a) gas-phase method validation (CS stands for crystal structure geometry), (b) QM/MM model validation for the P2 state, (c) calculated values for all nine possible protonation states. Numerical values are given in Table S1 in the Supporting Information.

7.8. In this sense, one could argue that the use of solutions with elevated pH might introduce a bias toward neutral unprotonated states. In our opinion, however, such a bias is likely to be very minor.

Classical MD Simulations. The nine fully built systems (Scheme 2) were treated within periodic boundary conditions. Long range electrostatic interactions were calculated with the Particle-Mesh Ewald (PME) technique with a default nonbonded cutoff of 8.0 Å to limit the direct space sum. The temperature in all simulations was controlled by coupling the system with the Berendsen thermostat.[54] An integration time step of 2 fs was used, and the SHAKE algorithm was employed to constrain bonds involving hydrogen atoms during dynamics.

The equilibration of the system was carried out in several steps: (*i*) Steepest descent minimization was applied to the enzyme−DNA−FAD complex (solute) with harmonic positional restraints on solvent molecules (5 kcal/mol Å^2). (*ii*) Minimization was repeated with restraints on the solute (5 kcal/mol Å^2) and no restraints on the solvent. (*iii*) Heating dynamics was performed with continued solute restraints at a constant volume (NVT). Thereby, the temperature was increased from 0 to 300 K over 60 ps and kept at that value for another 30 ps. (*iv*) Minimization was carried out with reduced solute restraints (2.5 kcal/mol Å^2). (*v*) The system was again heated, as described above, with reduced solute restraints (2.5 kcal/mol Å^2). (*vi*) A total of 150 ps of constant pressure (NPT) dynamics at 300 K, with isotropic position scaling at pressure of 1 bar and a pressure relaxation time of 0.2 ps, was performed. Harmonic restraints in this run were applied only to the DNA duplex (2.5 kcal/mol Å^2) at 300 K. (*vii*) The equilibration was finished with an unrestrained NVT simulation at 300 K (150 ps).

The MD production runs of 2 ns were carried out with a constant volume at 300 K, saving the snapshots every 4 ps. All of the data obtained through molecular dynamics simulations were subsequently processed and analyzed using the *ptraj* module of the AMBER9 program package.

THE EPR HYPERFINE STRUCTURE

Using information obtained from an EPR study of the stable FADH• radical in the (6−4) photolyase from *X. laevis*,[23] the active-site histidine residues were assigned a protonation state that would correspond to either P7 or P8 (Scheme 2) in the *D. melanogaster* enzyme. Specifically, the principal values of the hyperfine couplings of four selected protons (H5, H6, H1′, H8*α*) were extracted by performing simulation and deconvolution of the pulsed ENDOR spectra to fit the experimental data. The spectra of FADH• radical were measured in the wild type enzyme and in the H354A and H358A mutants, in the pH range between 6 and 9.5 and in the absence of substrate. The strongest shifts and intensity changes were observed for the H1′ and H8*α* signals in the H354A mutant, which was interpreted as being indicative of a change in the protonation state of His358 when going from pH 6 to 9.5. The spectra of the H358A mutant resembled those of the wild type more closely at both pH values. On the basis of these observations, and assuming that both histidines were protonated at pH 6, it was suggested that His354 (corresponding to His365) remains protonated over the measured pH range, while His358 (corresponding to His369) becomes neutral at higher pH values. Indeed, it is at a higher pH value (8.5) that the *X. laevis* (6−4) photolyase shows its maximum activity.

To investigate the connection between the hyperfine couplings (hfc's) of the FADH• radical and the protonation states of the nearby histidine residues, we set out to calculate the hfc's from first principles, as described in the methods section. Interestingly, the hfc's calculated for the isolated FADH• radical in the gas phase (Figure 2a) show reasonable agreement with the experimental data. The results obtained with B3LYP/6-31G(d) for the geometry extracted directly from the crystal structure (*D. melanogaster*) show the largest deviations from the experimental values, which were obtained from the *X. laevis* enzyme at pH 8. The hfc's resulting from optimizing the geometry at this same level of theory, however, parallel the trends in the experimental data remarkably well. The use of the larger TZVP basis set in conjunction with B3LYP has been previously shown to be a reliable method for this type of calculation.[39] The results presented in Figure 2a clearly show that, in this instance, the smaller basis set (6-31G(d)) can be safely used without any notable loss of accuracy.

To establish a method that could potentially capture the effect of changing the histidine protonation states on the hfc's, we employed the ONIOM[B3LYP/6-31G(d):AMBER] method and tested it on three models with differing QM regions (QM1−3). The results in Figure 2b, which were obtained for protonation state P2 (Scheme 2), show that the calculated hfc's are not strongly dependent on the size of the QM region in such a treatment. Thus, while the smaller and more tractable QM1 could be expected to perform satisfactorily, we elected to use the more inclusive QM3 representation for reasons of completeness.

Figure 2c shows the hfc's obtained with the QM3/MM model for all nine possible protonation states of His365 and His369. From this figure, it is immediately clear that any of the nine combinations of HID, HIE, and HIP could equally well give rise to hfc's very close to the experimental values. This result indicates that any assignment of the protonation states of the active site histidines in (6–4) photolyase, made on the basis of the hfc's for the $FADH^{\bullet}$ radical, should be treated with some caution.

As outlined in the methods section, the results shown in Figure 2c were obtained for single protein conformations closely resembling the crystal structure. One could consider a more sophisticated comparison on the basis of protein conformations that more accurately represent the true equilibrium structure of each protonation state in solution. Indeed, we find that such an approach slightly improves the already close agreement with the experimental values, for the P2 protonation state (Figure S1 in Supporting Information). A similar analysis across all nine possibilities (Figure S2) confirms that, even when using relaxed protein structures, the hfc's do not constitute an adequate measure with which to distinguish the different protonation states. An extra level of sophistication, involving the calculation of the hfc's for numerous structures of a given conformational ensemble, could also be considered.[55] In the current application, however, we were primarily focused on comparing the different protonation states on an equal footing. In this sense, we wished to remove the static and dynamic structural variations from the comparison and emphasize the situation where the only difference between the nine examples shown in Figure 2c was, indeed, the protonation states of the active-site histidines.

Given the results arising from that analysis, it is rather difficult to assert that one can use the hfc values of the $FADH^{\bullet}$ radical to arrive at any definitive assignment of the protonation states of the histidine residues in the active site of (6–4) photolyase. On this basis, we feel that the investigation of alternative methods for the determination of the likelihood of these states is warranted, and it is to this endeavor that we now turn our attention.

THE pK_A VALUES

The assignment of proton positions for a structure derived from X-ray crystallography is a ubiquitous problem in the modeling community. Standard approaches range from the use of intuition to sophisticated algorithms designed to calculate pK_a values and automate the proton placement. In this section, we investigate the performance of several such approaches. In addition to attempting to determine the implications for the active-site histidines in (6–4) photolyase, we also wish to exemplify the typical results, sensitivities, and limitations of these widely used procedures.

Empirical Approach. One of the fastest means to estimate the pK_a values of protein residues is through the use of the PROPKA server.[43] The empirical procedure estimates the pK_a values for all ionizable sites, on the basis of their environment. Table 1 shows the corresponding pK_a values for His365 and His369, obtained with several different crystal structures of (6–4) photolyase.

In the absence of the FAD and DNA-lesion ligands, the calculated pK_a values for His369 are uniformly low but positive. For His365, the values are even lower (Table 1, columns 2 and 3). Under these circumstances, the PROPKA method clearly predicts neither of the active site histidines to be protonated at a pH of ~7. The large deviations from the model pK_a value of 6.5 are due primarily to the desolvation term, which contributes $\Delta pK_{a,Des} \sim -4.5$ for both residues. The remaining deviation arises from charge–charge interactions and hydrogen bonding, which have a stronger influence on His365 than on His369.

Table 1. Estimated pK_a Values for His365 and His369 in (6-4) Photolyase Using PROPKA 2.0 Server and Different Crystal Structures Available at the PDB

PDB code: lesion	No ligands[a]		Ligands[a]	
	His365	His369	His365	His369
3CVU:T6-4T	-1.85	1.61	-8.09	-0.09
3CVY:T-T	-3.14	0.50	-8.20	-0.10
2WB2:T6-4C	-1.55	1.75	∞[b]	-1.45
2WQ6:T6(dew)4C	-1.77	1.58	-7.88	-0.14
2WQ7:mT6-4C	-1.93	1.53	∞[b]	-0.31
3FY4 (A)	-0.84	2.25	-7.57	-3.37

[a]The ligands are FAD and the DNA lesion, except for 3CVY:T–T, where FAD is the only ligand. [b]Problems with determining the contribution of the hydrogen bond between the His365 and N6A atom of FAD to the pK_a shift.

The effect of the FAD and DNA-lesion may be included in the PROPKA procedure, although one cannot distinguish between the oxidized and reduced forms of FAD. The inclusion of these ligands causes the pK_a values of both monitored residues to shift toward even lower values (Table 1, columns 4 and 5). This effect, which is bordering on nonphysical, is significantly stronger for His365, which participates in additional charge–charge interactions with the ligands. These additional interactions are not apparent in the case of His369, whose pK_a shift is mainly due to further increased desolvation effects. The relative importance of the DNA residues and the FAD cofactor can be gauged by comparing the results for 3CVY:T–T, where only FAD is considered as a ligand, to the remaining results.

Overall, the PROPKA methodology shows His369 to be more basic than His365. Although the absolute pK_a values are much lower than might be intuitively expected, it is clear that neither residue is predicted to be protonated at physiological pH. This preference for histidine neutrality persists irrespective of the absence or the presence of the DNA and the FAD ligand. While aspects of these results may be instructive, their reliability should be considered against the fact that the FAD charge cannot presently be explicitly accounted for. We therefore compare these findings to results obtained from more sophisticated approaches to obtaining pK_a values. The latter involve explicitly accounting for the electrostatic effects of the environment with the Poisson–Boltzmann equation as provided by the H++ server[44] and the APBS approach[49] (see Methods).

Histidine Protonation States in the T6–4T Lesion. With its default settings, H++ provides pK_a values of all potentially charged residues in the protein, on the basis of self-consistent solution of the Poisson–Boltzmann equation for the lesion-related pdb structure of the protein. Since only one ligand is permitted in this way, the calculation is performed without the lesion and with an approximate parametrization of the FAD cofactor (Table 2, columns 2 and 3). The *reduce* algorithm[48] can be used to estimate the protonation states of the system (Table 2, column 4). For the case of the T6–4T/

Table 2. Estimated pK_a Values for His365 and His369 in *D. melanogaster* (6-4) Photolyase Obtained with H++ Server and APBS Software Package (All Calculations Were Carried out at 300 K with an Internal Dielectric Constant of 6 and an Ionic Strength of I = 0.15 M)

Lesion/FAD	Simultaneous titration of all residues					Single residue titration			
	H++, pdb, reduce, no lesion			H++ PQR		H++ PQR		APBS	
	His 365	His 369	State	His 365	His 369	His 365	His 369	His 365	His 369
T6-4T / FAD	-3.5	5.5	P2	-18.6	3.7	2.5	5.9	0.2	3.4
T6-4T / $FADH^-$	0.2	7.6	P3	-14.7	7.0	5.7	7.7	2.7	4.6
T-T / FAD	-6.6[a]	6.2[a]	P2[a]	-13.5	4.3	2.3	6.6	0.5	3.7
T-T / $FADH^-$	0.1[a]	10.0[a]	P3[a]	-4.7	7.7	5.4	8.2	2.7	4.4
T6-4C / FAD	-5.5	2.8	P2	-9.8	2.0	0.5	5.5	-0.4	2.9
T6-4C / $FADH^-$	-1.8	5.9	P2	-6.2	5.9	3.4	7.3	2.0	4.1
T6-4pC / FAD	-5.5	2.8	P2	-16.5	-2.1	-8.0	2.9	-8.4	1.8
T6-4pC / $FADH^-$	-1.8	5.9	P2	-12.3	0.2	-4.1	4.7	-4.6	2.5
mT6-4pC / FAD	/	/	/	-18.0	-2.7	-8.7	2.4	-9.3	1.0
mT6-4pC/$FADH^-$	/	/	/	-13.7	-0.3	-5.5	4.3	-6.8	2.1

[a]With the entire DNA.

FAD (3CVU), this yields pK_a values for His365 and His369 of −3.5 and 5.5, respectively. This result, which was obtained with all possible titratable residues active, is in agreement with the PROPKA predictions in that neither histidine is expected to be protonated at pH = 7. Consistently, at this pH value, the most likely protonation state is found to be the neutral state P2 (HID−HIE).

Despite the fact that all of the crystal structures were resolved with the FAD cofactor in the oxidized state (FAD), it is known that the reduced anionic form ($FADH^-$) is required for the enzyme activity.[5] Following the same procedure discussed above and in the methods section, we have also calculated the pK_a values in the presence of the reduced cofactor. For the T6−4T structure (3CVU) in the absence of the lesion but in the presence of an approximate $FADH^-$ parametrization (Table 2, row 2), H++ yields pK_a values for His365 and His369 of 0.2 and 7.6, respectively. At a pH of 7, this result implies that His369 is likely to experience some degree of protonation, and the most likely such state is found to be P3 (HID−HIP).

Once the initial characterization of the protonation states has been achieved, a more rigorous parametrization of FAD (oxidized and reduced) and the lesion can be included, and pK_a values of *all* ionizable residues can be recalculated. As can be seen from Table 2 (columns 5 and 6), this reduces pK_a values of both histidines in the active site, in a similar manner to that observed in Table 1. For the oxidized FAD, the reduction in the case of His365 (to −18.6) is more dramatic than that for His369 (to 3.7). The large, and presumably nonphysical, negative value obtained for His365 is best interpreted to mean that, under these conditions, this residue is unlikely to be protonated at any accessible pH and that the histidines remain in the P2 state. Inclusion of the more realistic parametrizations of the reduced cofactor $FADH^-$ and the lesion again reduces both pK_a values. While His365 (−14.7) again appears very unlikely to be protonated, His369 (7.0) is expected to be half-protonated at pH 7. Hence both the P2 and P3 states could be potentially observable.

Determination of the Intrinsic pK_a. Another informative measure of acidity is the intrinsic pK_a for a given residue. This value is obtained by calculating the free energy difference between the protonated and nonprotonated forms, under the condition that all other residues remain in their standard protonation states. The intrinsic pK_a values for His365 and His369, given by H++ (Table 2 columns 7 and 8) for the T6−4T lesion (3CVU) in the presence of oxidized FAD and the lesion, are 2.5 and 5.9, respectively. Although these values are higher than those obtained when all ionizable residues are active, they still indicate the (neutral) state P2 to be the most likely at a pH of 7 (or higher). The APBS results for the equivalent system (Table 2, last two columns) are systematically lower than the H++ predictions (0.2 and 3.4 for His365 and His369) but concur that neither histidine is likely to be protonated under these conditions.

Calculations with H++ on the level of a single residue in the presence of $FADH^-$ give intrinsic pK_a values of 5.7 and 7.7 for His365 and His369, respectively. These values are again higher than those which account for changes in the other ionizable residues and place the pK_a of His369 above 7. The equivalent APBS results for the intrinsic pK_a values (2.7 and 4.6 for His365 and His369), however, are again systematically lower than those obtained from H++. In the case of T6−4T with the reduced lesion, the APBS approach does not find the pK_a of His369 to be greater than the pH of 7. On this basis, the APBS methodology favors the neutral protonation state P2 over the charged P3 state that is preferred by H++. Irrespective of this fact, the two methods agree that the presence of the cofactor in its reduced state increases the pK_a values of both residues. The magnitude of this effect is somewhere between 1 and 4 pK_a units but is uniformly smaller with APBS than with H++.

Histidine Protonation States in the T−T Dimer. The results obtained in this case (Table 2, rows 3 and 4) largely reproduce the trends discussed for the T6−4T lesion, although the effect of including the ligands appears to be somewhat reduced. This is largely due to the fact that the two key thymine residues are already present in the direct analysis of the crystal structure. In the presence of the oxidized FAD cofactor, the neutral protonation state P2 (HID−HIE) is preferred for the T−T dimer and the pK_a of His369 is below 7 in all cases.

The introduction of the reduced $FADH^-$ cofactor shows an initial preference for the charged P3 state (HID−HIP), which persists for all results obtained with H++. The intrinsic pK_a of His369 obtained with APBS in the presence of $FADH^-$ (4.4) is,

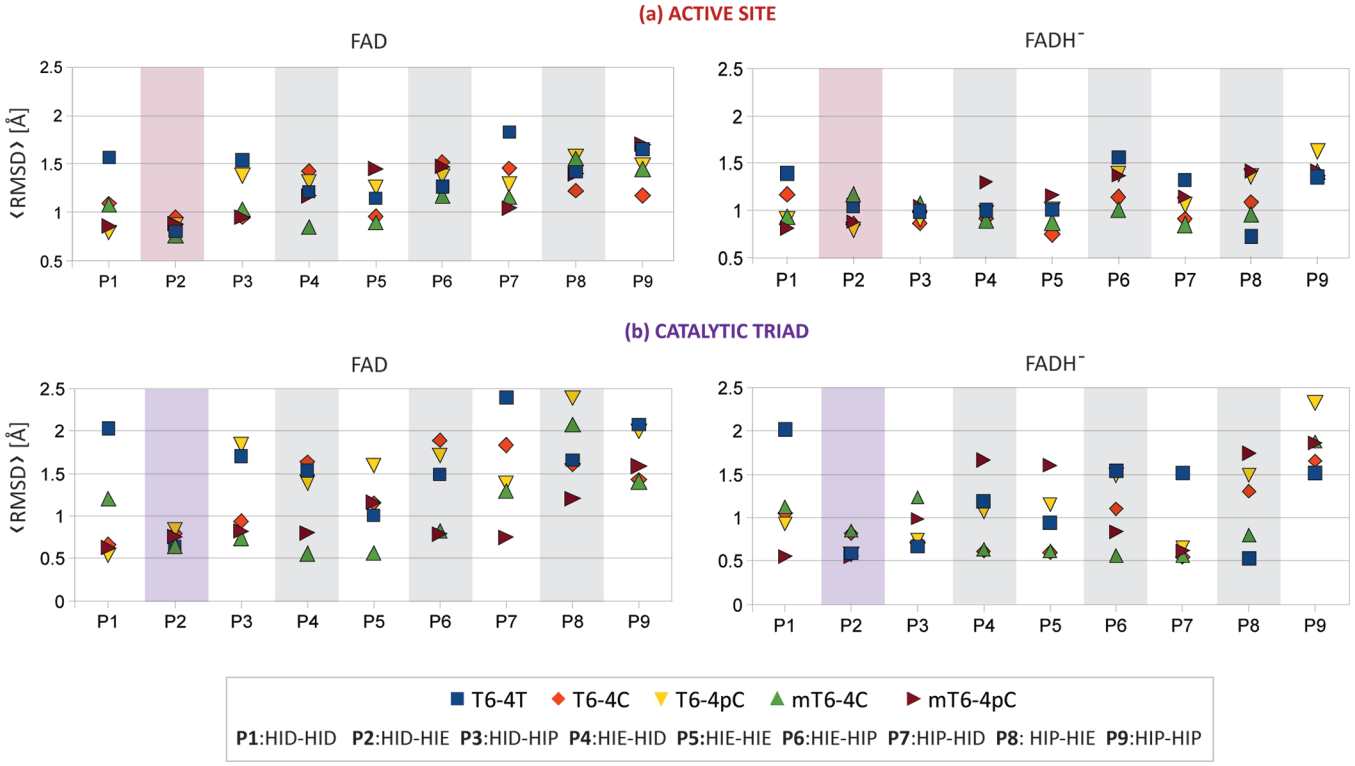


Figure 3. Average RMS deviations of (a) the active site (FAD, lesion, His365, His369, Tyr423) and (b) the catalytic triad (His365, His369, Tyr423) from their positions in the crystal structure during 2 ns MD simulations of systems containing: T6−4T lesion (3CVU), T6−4C lesion (2WB2), N4-methyl T6−4C lesion (2WQ7), repaired T−T dimer (3CVY). T6−4C and N4-methyl T6−4C lesions come in their deprotonated (T6−4C, mT6−4C) and protonated (T6−4pC, mT6−4pC) form. The original RMSD curves can be found in the Supporting Information.

however, below 7, indicating a preference for the P2 state. Interestingly, the APBS pK_a for His369 (4.4) is actually lower for T−T than for T6−4T (4.6).

Protonation of Histidines in the T6−4C Lesions. The structure containing the T6−4C lesion (2WB2) carries with it the added complication that the primary amine group of the lesion may be in the protonated or nonprotonated form. This factor does not influence the initial H++ calculations (Table 2, columns 2−4), where the P2 state is predicted to be the most favorable for both forms of the FAD cofactor. Inclusion of realistic ligand parametrizations continues this preference. The only pK_a value to (marginally) exceed 7 corresponds to the H++ result for the intrinsic pK_a of His369 (7.3), in the presence of the neutral amine and FADH$^-$. Protonation of the amine (T6−4pC) significantly suppresses the pK_a values of both histidines and is associated with a very clear preference for the P2 state. For the methylated T6−4C lesion (2WQ7), where we only considered the protonated form of the lesion (mT6−4pC), we were unable to obtain a result directly from H++ with the pdb file. The direct inclusion of the lesion and the FAD cofactor, however, produced results very similar to those obtained without the methyl group.

Perspective of the pK_a Study. The majority of the results presented in this section point to an overall neutral protonation state of the two histidine residues at a pH of 7. The *reduce* algorithm finds the P2 combination to be the most likely such state. Certain circumstances lead the H++ procedure to predict the potential occurrence of the protonated P3 state at neutral pH. However, it is interesting to reiterate that, in all such cases, the intrinsic pK_a's from the more rigorous APBS procedure are uniformly below 7. Furthermore, the APBS method systematically produces lower pK_a values than H++, even when the two are directly comparable.

Despite the apparent semidefinitive nature of these results, several cautionary statements are in order. It is evident from Table 2 that the results can vary widely depending on the way in which the calculations are performed. Indeed, certain circumstances can result in disturbingly large negative pK_a values for His365, the physical meaning of which is difficult to fathom. There is also a marked sensitivity of the results to input parameters such as the internal dielectric constant and the ionic strength (see Tables S2 to S4 in the Supporting Information). On this basis, one needs to treat the conclusions derived from such results with some apprehension. It is because of this uncertainty that we have also probed an alternative route to assigning the correct protonation state for the photolyase enzymes. Namely, we have systematically investigated the structural stability of various protonation states with classical molecular dynamics simulations. The associated results are presented in the subsequent section.

MOLECULAR DYNAMICS STUDY

In addition to the spectroscopic (EPR) and energetic (pK_a) criteria discussed above, we have employed a structural criterion to address the question of the most likely protonation states in the (6−4) photolyases. Specifically, we have run classical MD simulations of all nine possible protonation states and monitored their respective deviations relative to the relevant crystal structures. The rationale behind choosing such a criterion derives from the observation that the (6−4) photolyase enzymes can repair the DNA lesion in their crystalline form,[16] meaning that the conformation of the active

dx.doi.org/10.1021/ct2005648 | J. Chem. Theory Comput. 2012, 8, 1078−1091

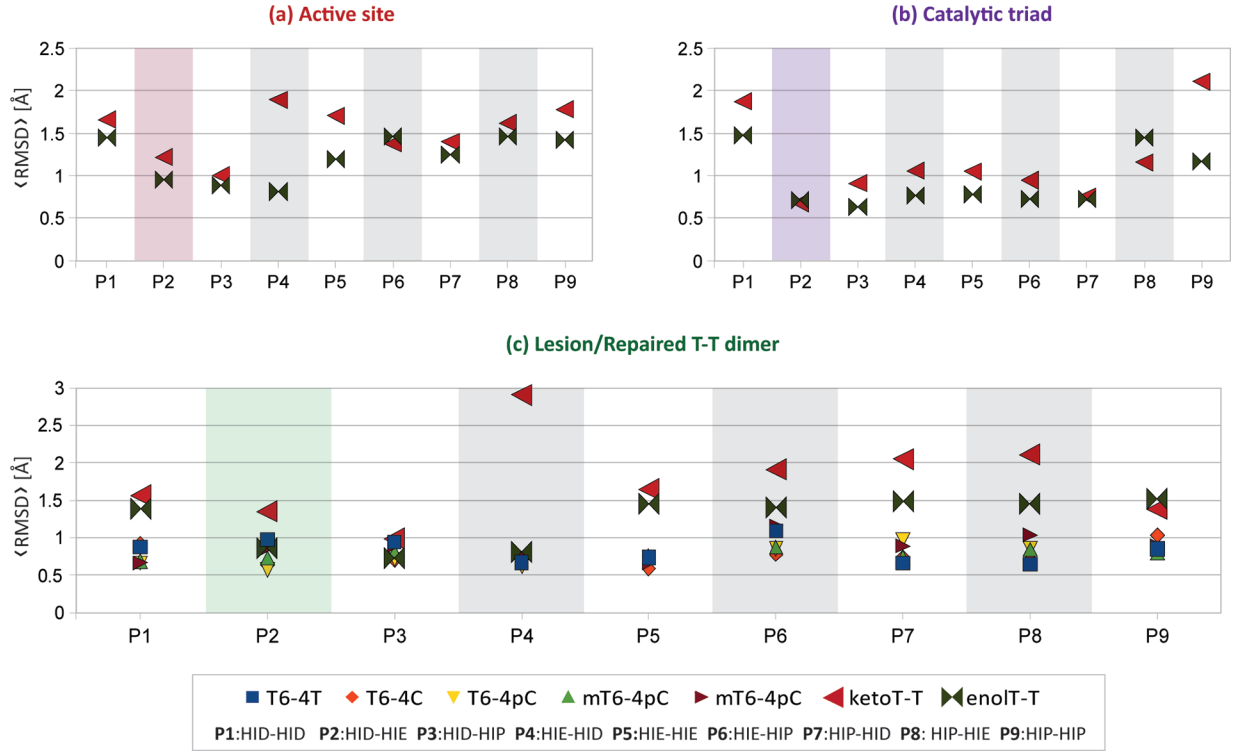


Figure 4. RMS deviations from the crystal structure positions during 2 ns simulations for (a) active site for keto and enol forms of the repaired T−T dimer, (b) catalytic triad (His365, His369, Tyr423), and (c) the lesion or the repaired T−T dimer. The original RMSD curves can be found in the Supporting Information.

site present therein is a productive one. Protonation states that maintain the crystal structure arrangement throughout unrestrained MD simulations are therefore more likely to be relevant for repair than those that deviate strongly from it. Because the crystal structures were resolved with oxidized FAD, while the reduced form ($FADH^-$) is required for activity, we have performed simulations in the presence of both cofactor states. For completeness, we have also simulated the radical form ($FADH^•$), whose associated results are shown in the Supporting Information.

Figure 3 shows a summary of the MD simulations that we carried out for systems containing the T6−4T lesion (3CVU), the T6−4C lesion (2WB2), and the N4-methyl T6−4C lesion (2WQ7). For the T6−4C and N4-methyl T6−4C lesions, we performed simulations in which the primary and secondary amino groups were both protonated (T6−4pC, mT6−4pC) and deprotonated (T6−4C, mT6−4C), respectively. Each depicted point in Figure 3 represents an average of the RMSD deviation, from the relevant crystal structure, over 2 ns of production dynamics. The corresponding RMSD time series can be found in the Supporting Information (Figures S8 and S9). For selected examples, we elected to extend the simulations to a length of up to 5 ns. The corresponding results, which are presented in Figures S11 and S12, show that, in such cases, the behavior over the longer simulation parallels that observed in the shorter one. This is especially true for those protonation states that remain close to the crystal structure and supports the hypothesis that the results are not artificially biased by short simulation times.

For the systems presented in Figure 3, it can be seen that the average RMSD values for the active site (Figure 3a, defined as His365, His369, Tyr423, FAD, and the lesion) parallel the analogous values for the catalytic triad (Figure 3b, His365, His369, and Tyr423) rather well. This implies that the motion of the catalytic residues comprises a major component of the dynamics in the active-site and thus serves to highlight the importance of protonation states of the two histidine residues.

In the presence of the neutral form of the cofactor (FAD), the neutral states (P1, P2, P4, and P5) generally exhibit lower deviations from the crystal structures than states involving a protonated histidine (P3, P6−9). Particularly consistent is the state P2. This is in agreement with the pK_a calculations presented above, which indicated a clear preference for neutral overprotonated states in the presence of FAD (Figure 3a). The introduction of the reduced cofactor ($FADH^-$) results in an increased stability of the protonated states, which is manifested in slightly less dispersed average RMSD values for the triad (Figure 3b). Specifically, the state P3 appears to be particularly stable. This observation is also consistent with the calculated pK_a values, including the observed shift toward higher values when FAD is reduced to $FADH^-$.

Even though the P2 state and, to a lesser extent, the P3 state exhibit the lowest RMSD deviations in general, it is obvious from Figure 3 that several other states show low deviations for specific structures. Despite the low RMSD values, however, not all of these states correspond to the preservation of the initial hydrogen bond network. Namely, even though the residues may stay close to their initial crystal-structure positions, they frequently experienced a modified orientation, with a corresponding rearrangement of the hydrogen bonds. This situation occurs, for example, for P7 in combination with $FADH^-$ (Figure 3b and Figure S5). On the other hand, the largest RMSD deviations generally (but not exclusively) correspond to the complete fragmentation of the H-bond network in the active site (Figure S5). It has been observed that the hydrogen bond between His369 and Tyr423 is disrupted

dx.doi.org/10.1021/ct2005648 | J. Chem. Theory Comput. 2012, 8, 1078−1091

most frequently. This causes an increased mobility of His369, which, in many cases, results in the loss of the H bond to His365 and a consequent divergence from the initial geometry. In this context, it is worth mentioning that the P2 state is the only arrangement capable of restoring the original hydrogen bonding once it has been perturbed. This feature was not observed for any other state.

Simulations with the Repaired DNA. In addition to the results on the lesions presented in Figure 3, we have also carried out a series of simulations on the repaired T−T dimer (3CVY, Figure 4) in the presence of $FADH^-$. Since keto−enol tautomerization of the 3′ base is likely to be the last step of the repair (see, e.g., Scheme 1), we have performed simulations with both the enol and keto forms of the repaired 3′ thymine. The upper panels of Figure 4 show the corresponding average RMSD deviations of the active site and catalytic triad, respectively (see also Figure S10). In contrast to the results for the lesions in Figure 3, it is clear that the variations in the catalytic triad do not correlate exceedingly well with the overall active-site behavior. Instead, the overall active-site deviations are much more closely related to the variations of the T−T dimer (lower panel Figure 4). Indeed, those variations can be seen to be stronger for the repaired T−T bases than for any of the lesions shown in Figure 3.

The deviation of the catalytic triad from its crystal structure position is seen to be relatively insensitive to the protonation state of the histidine residues (upper left panel in Figure 4), with all but three states (P1, P8, and P9) remaining comparatively close to their original positions. The DNA bases show a stronger dependence on the protonation states (lower panel Figure 4). Once again, the P2 and P3 states stand out as being able to maintain the closest resemblance to the crystal structure conformations.

Even though both the keto and enol forms of the repaired thymine dimer show stronger deviations than the lesions, the effect is definitely more pronounced for the keto form (lower panel Figure 4). The variations of the enol form, on the other hand, are of the same order as those of the lesions. The reduced mobility of the enol form, relative to the keto, appears to be due to its ability to engage in hydrogen bonds, through the hydroxyl group, in a similar manner as the lesions. The larger amplitude of the DNA motion, especially in the case of the keto form, leads toward the loss of the hydrogen bonds between the 3′-thymine and the histidine residues, allowing for a greater flexibility of the active site.

The greater flexibility observed for the repaired DNA is of interest from the perspective of the product release. Generally, the binding of the DNA strands to the repair enzyme is achieved by flipping the lesion out of the double helix and into the active site. Thus, there is a need to restore the helical structure after the repair is complete. The results presented in Figure 4 indicate that the final enol−keto tautomerization, and the resulting hydrogen bond losses, could provide the trigger that reduces the strength of the binding to the enzyme and allows this restoration to take place. We see evidence of this in the behavior of the DNA backbone in the simulations. The bend in the backbone, which is caused by the presence of the lesion, tends to flatten out in the repaired DNA. This effect is considerably more pronounced in the keto form, pointing to an increased freedom for the DNA strand to return to its preferred helical form.

A FREE ENERGY CYCLE

To better understand the meaning of the pK_a calculations, and their connection to the molecular dynamics simulations, it is instructive to consider the cyclical connectivity of the protonation states via single titration events. Such a cycle is shown for the four neutral and four singly charged protonation states in Figure 5. Using the APBS results for T6−4T

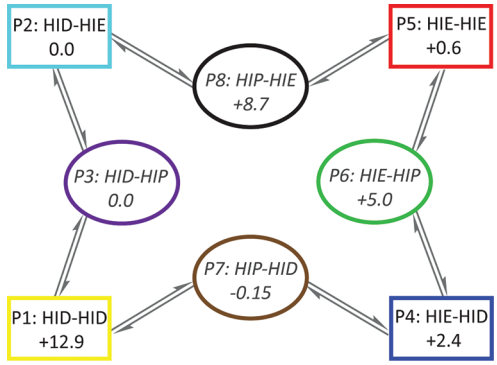


Figure 5. Cyclical representation of the four neutral (rectangles) and four singly charged (ovals) protonation states connected by single titration events. The APBS free energies (in kJ mol^{-1}), which are relative to P2 and P3, respectively, are shown for the case of T6−4T in the presence of $FADH^-$. See also Table S3.

($FADH^-$) shown in Table S3, and the expression $\Delta G^\circ = -RT \ln K_a$, one can assign relative free energies separately to the neutral and the charged states. Setting the respective free energies of P2 and P3 to zero, each four-membered cycle closes to within less than 1 pK_a unit. The actual free energy values are therefore given as an average of the forward and reverse paths to a given state from P2 or P3, as appropriate.

The final calculated pK_a values of the histidine residues are crucially dependent on the selection of the appropriate neutral reference state. In the analysis of Table 2, the neutral reference state was assigned as P2 on the basis of the *reduce* algorithm. This assignment is strongly supported by the molecular dynamics simulations as well as the neutral free energy cycle (Figure 5). Direct protonation of the P2 state can yield only P3 or P8, the former of which is shown to be favored by the pK_a calculations (Table 2), the free energy cycle (Figure 5), and the molecular dynamics simulations (Figure 3). On this basis, the P2 to P3 titration event, which is documented in Table 2, would indeed appear to be the most relevant.

Interestingly, the P7 state is found to be slightly lower in energy than P3 by the free-energy cycle (Figure 5). P7 cannot arise directly from P2, and the two neutral states that could result in its direct formation (P1 and P4) are not likely to be competitive with P2 (Figures 3 and 5). However, P7 could be formed from P3 through a proton transfer between the two Nε atoms of His369 and His 365 (see Figure 1). While our molecular dynamics calculations indicate that P7 would not be structurally stable on the nanosecond time scale (especially for T6−4T, Figures 3 and S5), this result opens the intriguing possibility that the P7 state could be a viable intermediate in the repair mechanism,[19] if protonation indeed plays a role.

DISCUSSION AND CONCLUSIONS

It has been established that the active site of the (6−4) photolyases, capable of repairing UV-induced lesions in DNA strands, contains two histidines (His365 and His 369 in *D.*

melanogaster) and a tyrosine, which are key residues in catalysis. On the basis of the EPR/ENDOR study of the *Xenopus laevis* (6–4) photolyase, several proposed mechanisms assume that the two histidines act as an acid–base pair with His365 being the protonated residue. However, because the two histidines are relatively closely positioned in the active site, it may be expected to be difficult to resolve their protonation states by the use of the EPR-derived hyperfine couplings (hfc's). To investigate this phenomenon, we used the ONIOM QM/MM approach to calculate the relevant EPR hyperfine couplings for all nine possible combinations (P1–P9) that two histidines with various protonation states can adopt. Surprisingly, all combinations resulted in hfc's in good agreement with the experimentally measured values. Under such circumstances, it seems difficult to make use of the EPR data as a definitive criterion for assigning the protonation states of the histidines.

In order to resolve the problem, we proceeded to explore the active site from both the energetic and structural points of view, for various oxidation states of the FAD cofactor and several available lesions. The energetics aspects of the system are reflected in calculations of the pK_a values for the two histidines. The results based on the solution of the Poisson–Boltzmann equation, combined with the *reduce* algorithm, tend to favor an overall neutral state in the presence of the oxidized FAD cofactor, identifying the P2 state (HID–HIE) as the most relevant one. The introduction of the reduced cofactor ($FADH^-$) increases the pK_a values of both histidines and, in selected cases, causes the pK_a of His369 to exceed the reference pH of 7. In these cases, the state P3 (HID–HIP) emerges as the most likely combination. In this context, it is worth noting that the H++ procedure results in pK_a values that are systematically higher than those from APBS, which are uniformly below 7. At the same time, it is pertinent to recall that (6–4) photolyase exhibits strong pH dependence and reaches maximum activity around pH ∼ 8. It is reasonable to expect that, under these conditions, the histidines would prefer a neutral state, even according to the higher pK_a values obtained from H++.

Interestingly, the neutral P2 state (HID–HIE) also stands out from the structural point of view, as reflected in the results from molecular dynamics simulations. Of the neutral states, P2 exhibits the lowest structural deviations from all of the experimentally derived structures (Figure 3). The P2 state also exhibits generally low deviations from the crystal structures in the presence of the reduced cofactor ($FADH^-$). This same HID–HIE combination is found to be unique in its ability to maintain the hydrogen bond network between the catalytic residues for all types of lesions, resulting in the lowest average triad deviations for all simulated systems. This is true irrespective of the oxidation state of the cofactor, including the $FADH^•$ variant, whose results are shown in the Supporting Information.

According to the structural criterion, the best performing of the protonated states is P3 (HID–HIP). Although this state fails to keep the crystal structure arrangement in the presence of oxidized FAD, the inclusion of $FADH^-$ significantly improves its ability to maintain the initial structure. Indeed, under these conditions, the P3 state shows a performance comparable to that of P2. Once again, there is an interesting correlation between the structural and energetic criteria, in the sense that the pK_a/*reduce* results also identified P3 to be the most likely of the protonated combinations. Analysis of a free energy cycle connecting eight of the protonation states further supports the appropriateness of P2 as the neutral reference state and P3 as its relevant protonated counterpart. Interestingly, P7 is found to be associated with a free energy marginally lower than P3 and could thus participate in the repair mechanism, insofar as protonation does take place,

We find that the stability of the active site is intimately related to the stability of the resident hydrogen bond network. Once again, the P2 state is found to be the only combination capable of restoring the network after a disruptive fluctuation. It appears that the stability of the network may also be significantly affected by water molecules present in the active site. This is relevant in the context of a recent suggestion that a water molecule, hydrogen bonded to the OH group of the lesion and to Nε of His369, plays a significant role in the mechanism of the repair.[19] This water has a high crystallographic B-factor.[16] Its position is thus not well resolved, and it might be quite dynamic. In our simulations, a number of water molecules surrounding the active site are found, often making a hydrogen bond with the OH group of the lesion. The "bridging" conformation, however, is rarely present. We have also observed that preservation of the crystal structure conformation tends to prevent the entrance of additional water molecules to the active site, while the disruption of the initial hydrogen bonding network is much more permissive in this respect. This effect can be seen in the radial distribution of water around key atoms involved in the active-site network (Figure S6) and in plots of the water density in the active site (Figure S7). The states that remain closest to the crystal structures (P2 and P3) can also be classified as the "driest", in that they are associated with less water in and around the active site than those states that diverge from the experimental reference.

The success of state P2 has been further confirmed in the set of simulations performed with the repaired DNA bound to the active site of the enzyme. We observe that the repaired DNA undergoes larger fluctuations and drifts away from the active site, especially when the keto tautomer is present. This is consistent with a recent study of product release,[56] which suggested that restoration of the original bases occurs on a picosecond time scale, while the conformational changes leading toward product dissociation take place on the observed time scale of 50 μs.

In conclusion, it is useful to recall that the traditional view of the (6–4) photolyase enzymes has His365 as the protonated residue in the active site. In contrast, our systematic analysis of spectroscopic, structural, and energetic aspects provides strong evidence that the P2 (HID–HIE) combination is the dominant protonation state for the His365 and His369 residues. We find that the presence of the reduced cofactor ($FADH^-$) endows His369 with a higher propensity to become protonated (HIP), although subsequent proton transfer to His365 is possible. However, considering the pH associated with maximum activity, it is quite likely that protonation does not play a role in the mechanism. With the proton distribution in the active site of (6–4) photolyase thus established, we may confidently focus our attention toward the characterization of a consistent mechanism for the lesion repair.

■ ASSOCIATED CONTENT

Supporting Information

Figures S1–S12, Tables S1–S5, and force field parameter tables for nonstandard residues. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: David.Smith@irb.hr.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was financially supported by the Croatian Ministry of Science (Project No. 098-0982933-2937), SFB 749, and the EAM Cluster of Excellence at FA University Erlangen. The computational resources were provided by University Computing Centre (SRCE), University of Zagreb, and Croatian National Grid Infrastructure (CRO-NGI). We thank Thomas Carell for fruitful discussions and for allowing us to present and discuss our work at the DFG-SFB749 symposium. We are grateful to Zlatko Brkljača for proof reading the manuscript. A.-S.S. thanks the Ruđer Bošković Institute for hospitality during the preparation of this manuscript.

REFERENCES

(1) Sancar, A. *Chem. Rev.* **2005**, *103*, 2203−2227.

(2) Chang, C.-W.; Guo, L.; Kao, Y.-T.; Li, J.; Tan, C.; Li, T.; Saxena, C.; Liu, Z.; Wang, L.; Sancar, A.; Zhong, D. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 2914−2919.

(3) Harrison, C. B.; O'Neill, L. L.; Wiest, O. *J. Phys. Chem.* **2005**, *109*, 7001−7002.

(4) Masson, F.; Laino, T.; Röthlisberger, U.; Hutter, J. *ChemPhysChem* **2009**, *10*, 400−410.

(5) Cichon, M. K.; Arnold, S.; Carell, T. *Angew. Chem., Int. Ed.* **2002**, *41*, 767−770.

(6) Sancar, A. *J. Biol. Chem.* **2008**, *283*, 32153−32157.

(7) Rahn, R. O.; Hosszu, J. L. *Photochem. Photobiol.* **1969**, *10*, 131−137.

(8) Clivio, P.; Fourrey, J.-L.; Gasche, J.; Favre, A. *J. Am. Chem. Soc.* **1991**, *113*, 5481−5483.

(9) Kim, S. T.; Malhotra, K.; Smith, C. A.; Taylor, J. S.; Sancar, A. *J. Biol. Chem.* **1994**, *269*, 8535−8540.

(10) Zhao, X.; Liu, J.; Hsu, D. S.; Zhao, S.; Taylor, J. S.; Sancar, A. *J. Biol. Chem.* **1997**, *272*, 32850−32590.

(11) Hitomi, K.; Nakamura, H.; Kim, S.-T.; Mizukoshi, T.; Ishikawa, T.; Iwai, S.; Todo, T. *J. Biol. Chem.* **2001**, *13*, 10103−10109.

(12) Joseph, A.; Prakash, G.; Falvey, D. E. *J. Am. Chem. Soc.* **2000**, *122*, 11219−11225.

(13) Asgatay, S.; Petermann, C.; Harakat, D.; Guillaume, D.; Taylor, J. S.; Clivio, P. J. *J. Am. Chem. Soc.* **2008**, *130*, 12618−12619.

(14) Borg, O. A.; Eriksson, L. A.; Durbeej, B. *J. Phys. Chem.* **2007**, *111*, 2351−2361.

(15) Heelis, P. F.; Liu, S. *J. Am. Chem. Soc.* **1997**, *119*, 2936−2937.

(16) Maul, M. J.; Barends, T. R. M.; Glas, F. A.; Cryle, M. J.; Domratcheva, T.; Schneider, S.; Schlichting, I.; Carell, T. *Angew. Chem., Int. Ed.* **2008**, *47*, 10076−10080.

(17) Domratcheva, T.; Schlichting, I. *J. Am. Chem. Soc.* **2009**, *131*, 17793−17799.

(18) Yamamoto, J.; Tanaka, Y.; Iwai, S. *Org. Biomol. Chem.* **2009**, *7*, 161−166.

(19) Sadeghian, K.; Bocola, M.; Merz, T.; Schütz, M. *J. Am. Chem. Soc.* **2010**, *132*, 16285−16295.

(20) Li, J.; Liu, Z.; Tan, C.; Guo, X.; Wang, L.; Sancar, A.; Zhong, D. *Nature* **2010**, *466*, 887−891.

(21) Harbach, P. H. P.; Borowka, J.; Bohnwagner, M.-V.; Dreuw, A. *J. Phys. Chem. Lett.* **2010**, *1*, 2556−2560.

(22) Humphrey, W.; Dalke, A; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33−38.

(23) Schleicher, E.; Hitomi, K.; Kay, C. W. M.; Getzoff, E. D.; Todo, T.; Weber, S. *J. Biol. Chem.* **2007**, *282*, 4738−4747.

(24) Glas, A. F.; Schneider, S.; Maul, M.; Hennecke, U.; Carell, T. *Chem.—Eur. J.* **2009**, *15*, 10387−10396.

(25) Glas, A. F.; Kaya, E.; Schneider, S.; Heil, K.; Fazio, D.; Maul, M.; Carell, T. *J. Am. Chem. Soc.* **2010**, *132*, 3254−3255.

(26) Wang, J. M.; Cieplak, P; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049−1074.

(27) Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E. III; Laughton, C. A.; Orozco, M. *Biophys. J.* **2007**, *92*, 3817−3829.

(28) Svozil, D.; Šponer, J. E.; Marchan, I.; Pérez, A.; Cheatham, T. E.; Forti, F.; Luque, F. J.; Orozco, M.; Šponer, J. *J. Phys. Chem. B* **2008**, *112*, 8188−8197.

(29) Hornak, V; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins* **2006**, *65*, 712−725.

(30) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. *J. Mol. Graphics Modell.* **2006**, *25*, 247−260.

(31) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157−1174.

(32) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. E.; Wang, J.; Duke, R. J.; Luo, R.; Merz, K. M.; Pearlman, D. A.; Crowley, M.; Walker, R. C.; Zhang, W.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Wong, K. F.; Paesani, F.; Wu, X.; Brozell, S.; Tsui, V.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Mathews, D. H.; Schafmeister, C.; Ross, C. S.; Kollman, P. A. *AMBER 9*; University of California: San Francisco, CA, 2006.

(33) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision D.02; Gaussian, Inc.: Wallingford, CT, 2004.

(34) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269−10280.

(35) Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A. *J. Comput. Chem.* **1995**, *16*, 1357−1377.

(36) See, for example: Beierlein, F. R.; Kneale, G. G.; Clark, T. *Biophys. J.* **2011**, *101*, 1130−1138.

(37) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. *J. Comput. Chem.* **2003**, *24*, 1999−2012.

(38) Zgarbová, M.; Otyepka, M.; Sponer, J.; Mládek, A.; Banáš, P.; Cheatham, T. E.; Jurečka, P. *J. Chem. Theory Comput.* **2011**, *7*, 2886−2902.

(39) Zipse, H.; Artin, E.; Wnuk, S.; Lohman, G. J. S.; Martino, D.; Griffin, R. G.; Kacprzak, S.; Kaupp, M.; Hoffman, B. M.; Bennati, M.; Stubbe, J.; Lees, N. *J. Am. Chem. Soc.* **2009**, *131*, 200−211.

(40) Vreven, T.; Morokuma, K.; Farkas, Ö.; Schlegel, H. B.; Frisch, M. J. *J. Comput. Chem.* **2003**, *24*, 760−769.

(41) Vreven, T.; Byun, K. S.; Komaromi, I.; Dapprich, S.; Montgomery, J. A. Jr.; Morokuma, K.; Frisch, M. J. *J. Chem. Theory Comput.* **2006**, *2*, 815−826.

(42) Li, H.; Robertson, A. D.; Jensen, J. H. *Proteins* **2005**, *61*, 704−721.

(43) Bas, D. C.; Rogers, D. M.; Jensen, J. H. *Proteins* **2008**, *73*, 765−783.

(44) H++: web-based computational prediction of protonation states and pK of ionizable residues in macromolecules. http://biophysics.cs.vt.edu/H++/index.php (accessed Jul 2011).

(45) Gordon, J. C.; Myers, J. B.; Folta, T.; Shoja, V.; Heath, L. S.; Onufriev, A. *Nucleic Acids Res.* **2005**, *33*, W368−W371.
(46) Bashford, D.; Karplus, M. *Biochemistry* **1990**, *29*, 10219−10225.
(47) Anandakrishnan, R.; Onufriev, A. *J. Comp. Biol.* **2008**, *15*, 165−184.
(48) Word, M. J.; Lovell, S. C.; Richardson, J. S.; Richardson, D. C. *J. Mol. Biol.* **1999**, *285*, 1735−1747.
(49) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10037−10041.
(50) Bank, R.; Holst, M. *SIAM Rev.* **2003**, *45*, 291−323.
(51) Holst, M. *Adv. Comput. Math.* **2001**, *15*, 139−191.
(52) Nielsen, J. E.; Vriend, G. *Proteins* **2001**, *43*, 403−412.
(53) Dolinsky, T. J.; Nielsen, J. E.; McCammon, J. A.; Baker, N. A. *Nucleic Acids Res.* **2004**, *32*, W665−W667.
(54) Berendsen, H. J. C.; Postma, J. P. M; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(55) Pauwels, E.; Declerck, R.; Verstraelen, T.; De Sterck, B.; Kay, C. W. M.; Van Speybroeck, V.; Waroquier, M. *J. Phys. Chem. B* **2010**, *114*, 16655−16665.
(56) Kondoh, M.; Hitomi, K.; Yamamoto, J.; Todo, T.; Iwai, S.; Getzoff, E. D.; Terazima, M. *J. Am. Chem. Soc.* **2011**, *133*, 2183−2191.
(57) Domratcheva, T. *J. Am. Chem. Soc.* **2011**, *133*, 18172−18182.

NOTE ADDED IN PROOF

We would like to draw the attention of the reader to a recent and relevant study that was published after the submission of our work.[57] That work considered the effect of protonation on the pertinent photolyase excitation spectra and concluded that a neutral combination of the active-site histidines, coincidentally corresponding to P2 (Scheme 2), was the most promising.