JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Article
pubs.acs.org/JACS

Theory Uncovers an Unusual Mechanism of DNA Repair of a Lesioned Adenine by AlkB Enzymes

Binju Wang,[†] Dandamudi Usharani,[†,⊥] Chunsen Li,[‡,§] and Sason Shaik*[,†]

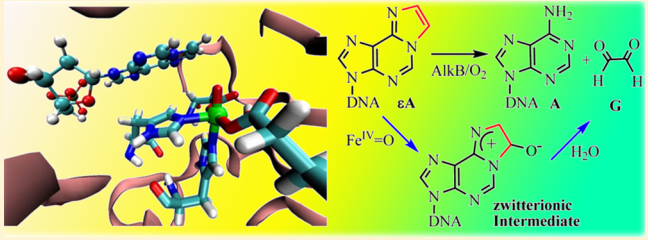
[†]Institute of Chemistry and The Lise Meitner-Minerva Center for Computational Quantum Chemistry, Hebrew University of Jerusalem, 91904 Jerusalem, Israel

[‡]State Key Laboratory of Structural Chemistry, Fujian Institute of Research on the Structure of Matter, Chinese Academy of Sciences, Fuzhou, Fujian 350002, China

[§]Fujian Provincial Key Laboratory of Theoretical and Computational Chemistry, Xiamen, Fujian 361005, China

Supporting Information

ABSTRACT: DNA-base lesions cause cancer and propagate into the genome. We use in-protein QM/MM calculations to study the repair of etheno-bridged adenine (εA) by the iron(IV)-oxo species of AlkB enzymes. Recent experimental investigations, using mass-spectrometry and *in crystallo* isolation, suggested that εA was repaired by formation of an epoxide (εA-ep) that further transforms to a glycol (εA-gl), ending finally in adenine and glyoxal. Theory reproduces the experimentally observed barrier for the rate-determining step and its pH dependence. However, as we show, the *mass-spectrometrically identified species are side-byproducts* unassociated with the repair mechanism. The repair is mediated by a zwitterionic species, of the same molecular mass as the epoxide, which transforms to an intermediate that matches the *in crystallo* trapped species in structure and mass, but is NOT the assumed εA-gl iron-glycol complex. Verifiable/falsifiable predictions, regarding the key protein residues, follow. The paper underscores the indispensable role of theory by providing atomistic descriptions of this vital mechanism, and guiding further experimental investigations.



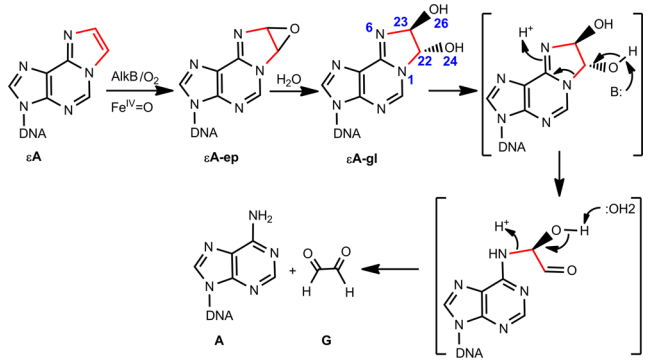
1. INTRODUCTION

DNA bases are subject to a variety of modifications, which, on the one hand, are required to "control" DNA, and, on the other, may cause DNA damage.[1−10] Much like the two-faced Roman god, Janus, these reactions are both "bad and good". For example, methylation of cytosine (C) in the fifth position acts as a vital signal for gene silencing, while the methylation at the third position (the sp^2 N-site) is a lesion that requires repair.

The damaging lesions have dire consequences, leading to diseases like cancer, and to the propagation of the damaged bases into the genome. Therefore, *it is hugely important to elucidate the mechanisms by which enzymes repair these lesions*. One of these harmful lesions is the exocyclic etheno bridge, which occurs in cytosine and adenine, presumably by peroxidized lipids, formed by oxidative stress. The so formed lesion in adenine (A) is shown in Scheme 1, and labeled εA. In a series of beautiful experiments, it was shown by He et al.[11,12] and Delaney et al.[13] that the family of enzymes called generically AlkB are responsible for the repair. AlkB belongs to the superfamily of nonheme iron-α-ketoglutarate-dependent enzymes, which activate O_2 and generate the high-valent iron(IV)-oxo intermediate that performs a variety of oxidative processes.[14−24]

On the basis of detectable intermediates, which were characterized by mass spectrometric techniques and *in crystallo* isolation, He et al.[11,12] and Delaney et al.[13] proposed tentatively the general mechanism shown in Scheme 1. In this mechanism,

Scheme 1. Proposed DNA Repair of the Etheno-Bridged Adenine (εA)[a]



[a]The lesion is marked in red. The active species of the enzyme AlkB is labeled Fe^{IV}=O. The atom numbering on the glycol intermediate (εA-gl) follows ref 12.

an initial epoxidation of the etheno-bridge lesion of εA by the iron(IV)-oxo species creates the epoxide intermediate (εA-ep), which was identified mass-spectrometrically based on its *m*/*e* values. Subsequently, εA-ep opens up by nucleophilic attack of a water molecule to yield the glycol (εA-gl) intermediate, also

Received: August 3, 2014
Published: September 9, 2014

   dx.doi.org/10.1021/ja507934g | J. Am. Chem. Soc. 2014, 136, 13895−13901

identified by m/e values in mass spectrometry. The latter then breaks down by acid/base catalysis to yield the repaired adenine (A) and glyoxal (G). The beautiful *in crystallo* experiment of He et al.,[12] further enabled trapping the glycol intermediate (εA-gl) ligated to the iron center of the AlkB enzyme.

We were drawn to investigate this repair mechanism because of the following reasons: (a) While mass spectrometry provides information on the atomic constitution and stoichiometry of the species, the mass-spectral data do not provide precise structures, and hence both the "epoxide and glycol intermediates" may not be the ones actually detected in the experiment. (b) X-ray crystallography does not assign hydrogen atoms, and the *in crystallo* trapped intermediate may not be εA-gl, but an isomer of it. (c) Epoxides require boiling water to open up to glycols.[25] (d) None of the experiments gave a clear clue as to the identity of the crucial acidic/basic residues that are involved in the catalysis and the breakdown of the epoxide (εA-ep). Thus, the true identity of the intermediates and the mechanistic pathways require direct scrutiny and validation, which is the focus of the present paper.

To elucidate the nature of this vital repair mechanism, we used quantum mechanical/molecular mechanical (QM/MM) calculations,[26] which can provide an *in silico* detailed atomistic description of the mechanism, to test its key conclusions and make new predictions, which can be verified or falsified by further experiment. As is reported below, the iron(IV)-oxo species of AlkB *cannot epoxidize* εA, and instead it leads to *a zwitterionic species* of *the same molecular mass* as the epoxide. The zwitterion then undergoes consecutive protonations and C−N bond cleavages to form the repaired adenine (A) and glyoxal (G). The *in crystallo trapped species is not the glycol complex but rather an isomer of it.*

2. COMPUTATIONAL METHODS

2.1. Setup of the System. The initial structure of the enzyme−substrate complex was prepared on the basis of the recently determined X-ray structure of εA-containing glycol intermediate (PDB code[12] 3O1U). In the double-stranded DNA, only one unit of adenine phosphate (εA) that acts as substrate was kept, while other base pairs were omitted. The oxygen (O24) of glycol originated from the oxo transfer of iron(IV)-oxo species according to the experiments,[12] while O26 of the glycol originated from the neighboring water (Wat2). Thus, we used O24 as the oxo moiety of iron(IV)-oxo species and O26 as Wat2. The iron(IV)-oxo species is a hexa-coordinate complex that includes Wat1, His131, His187, Asp133 and succinate.

To add hydrogens to the protein, we inspected the protonation states of titratable residues (His, Glu, Asp). We assigned protonation states on the basis of pK_a values from the PROPKA software[27] in combination with careful visual inspections. All arginine and lysine residues were protonated, while all glutamic acid and aspartic acid residues were deprotonated. All histidine residues were used as singly protonated according to the local hydrogen bond networks and pK_a values. The histidine residues namely His72, His131, His172, and His187 were protonated at the δ position, while His66, His97, and His197 were protonated at the ε position. The resulting total charge of the protein was zero.

After adding all hydrogens, the positions of the hydrogen atoms were optimized with 200 steps of steepest descent and 200 steps of adapted basis Newton−Raphson method using the CHARMM27 force fields implemented in the CHARMM program.[28] The resulting protein was solvated with a 16 Å layer of TIP3P water. Then, to attain equilibrium of the inner solvent layer, we followed these four steps: (1) optimization of inner solvent layer for 1000 steps of steepest descent and 1000 steps of adapted basis Newton−Raphson method, (2) heating slowly the protein from 0 to 300 K for 15 ps with a 1 fs time step, (3) equilibrating the solvent layer for 15 ps at 300 K with a 1 fs time step, and (4) resolvation of the protein to fill up the interspace of solvent layer. We repeated these steps three times until only 36 extra water molecules were being incorporated into the interspace of the protein during the last resolvation step.

After these procedures, a productive MD run was performed for 2.5 ns. From this run we chose three snapshots in the last 1500 ps for QM/MM calculations. During all these MD simulations, the coordinates of the entire Fe(IV)═O unit and the metal-ligating residues as well as the outer 8 Å of the solvent layer were kept frozen.

The representative structures of the active site from the above MD simulations are shown in Supporting Information Figure S1. For comparison, we also run MD simulation with Asp135 protonated. As can be seen clearly from Supporting Information Figure S1, the active structures in both cases are quite similar to each other. The only main difference is that Lys134 will not form a water chain with N6 when Asp135 is protonated.

2.2. QM/MM Methodology. All QM/MM calculations were performed using ChemShell,[29] combining Turbomole[30] for the QM part and DL_POLY[31] for the MM part. The electronic embedding scheme[32] was used to account for the polarizing effect of the enzyme environment on the QM region. Hydrogen link atoms with the charge-shift model[29] were applied to treat the QM/MM boundary. In QM/MM geometry optimizations, the QM region was treated by the hybrid UB3LYP[33] density functional with two basis sets. For geometry optimization: a double-ζ basis set LACVP for iron and 6-31G(d,p) for all other atoms, labeled B1. The energies are further corrected with a larger basis set of LACV3P++** for all atoms, labeled B2. The CHARMM27 force field was employed throughout this study for the MM region. All the transition states (TS) were located by relaxed potential energy surface (PES) scans followed by full TS optimizations using the P-RFO optimizer implemented in the HDLC code.[34] To further test the TS species, we carried out frequency calculations for **TS1**, using a reduced system of 66 atoms, which is the same as the one used for the QM-only calculations in Supporting Information Figure S10. As indicated in Supporting Information Figure S16, both QM/MM and QM-only calculations lead to very similar structures and a single negative mode with an imaginary frequency for each. This QM/MM frequency calculation confirms the characterization of **TS1** as a *bona fide* QM/MM-transition state and supports the usage of the TS location procedure used here. Since these frequency calculations are very time-consuming, we did not perform them for the other species in the study.

Our QM/MM calculations are based on an amply tested procedure,[35−43] which has proven to be very reliable[38,42,43] for iron-based metalloenzymes. After the initial preparatory steps of the enzyme using protein data bank (PDB) entry 3O1U, we selected from the 2.5 ns molecular dynamics (MD) equilibration trajectory three snapshots, **Sn1−3**, and subjected them to QM/MM optimization, using the following QM-sub systems, QM_I−QM_{III} based on the active site in Figure 1.

2.3. Gas Phase QM-Only Calculations. All the QM calculations were performed in the gas phase using the Gaussian 09 package.[44] The QM atoms include the iron-oxo complex and the εA substrate (see Supporting Information Figure S10). The geometries of all transition states, reactants and intermediates involved in the reaction were fully optimized at the UB3LYP/B1 level, while the energies were further corrected with UB3LYP/B2 level in gas phase. The correlation between the stable structures and the transition states was verified by analyzing the corresponding imaginary frequency mode, as well as by intrinsic reaction coordinate (IRC) calculations.

3. RESULTS AND DISCUSSION

3.1. Formation of an Oxy-Zwitterion. Figure 2a shows the key steps of the activation of the doubly bonded moiety of the lesion of εA, and some follow-up steps. Thus, starting from the reactant complex (**RC**), quite surprisingly, the O-transfer reaction leads via **TS1** to a zwitterion intermediate O^--εA^+ (**IC1**), rather than the "normal" carbon-based radical

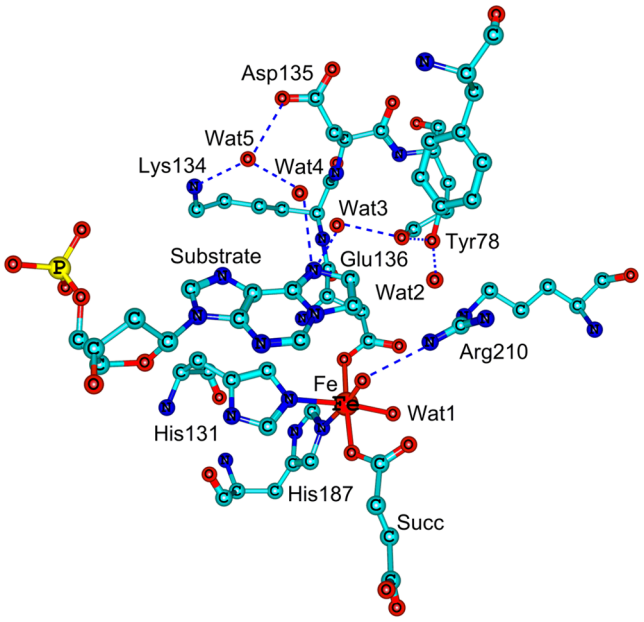


Figure 1. A representative structure of the active site from the equilibrium MD trajectory (**Sn1**). Hydrogen atoms are omitted for clarity. The QM regions considered in QM/MM are QM_I = $Fe^{IV}(O)$ complex + Substrate + Glu136 + Tyr78 + Wat2 + Wat3; QM_{II} = QM_I + Asp135 + Wat4 + Wat5); QM_{III} = QMI + Arg210 (for more details, see Supporting Information Figure S2).

intermediate that undergoes ring closure to the epoxide product (*ε*A-ep), as generally observed during oxygenation of C=C double bonds by iron^{IV}-oxo species.[35,45−47] As shall be seen, **IC1** is the key species that controls this "non-natural" mechanism.

To ascertain the nature of **IC1**, we show in Figure 2b key geometric, spin and charge density features of this intermediate. It is seen that the spin density on the *ε*A moiety (labeled Subs) of **IC1** is virtually zero, and the substrate is in fact, a carbocation with a charge of +0.95, while O24 is an oxy-anion. Figure 2c shows further that the doubly occupied orbital (HOMO-1) in the reactant complex (**RC**), which corresponds to $\pi(C_{22}{=}C_{23})$ of the *ε*A lesion (mixed with the π orbitals of the adenine) is transformed in **IC1** to a vacant orbital, which has a major contribution of the $2p_z$ orbital on C23 of **IC1**. Clearly, by all criteria **IC1** *is a zwitterion* O^--*ε*A^+ *intermediate.*

Turning back to Figure 2a, it is seen that the zwitterion intermediate, **IC1** (O^--*ε*A^+), can, in turn, undergo two alternative subsequent transformations. One (in blue) involves a nucleophilic attack by the crystal water molecule Wat2 that remains bound to Tyr78 (Supporting Information Figure S5) throughout the MD simulation. Thus, Wat2 attacks the carbocation center C23 of O^--*ε*A^+, and coupled with the proton abstraction through the Glu136-Tyr78 bridge, this leads via **TS2** to **IC2**, which is *a deprotonated glycol (a glycoxide) complex* (*ε*A-gl-H).

We tested also a variety of alternative putative intermediates, which may lead to an epoxide.[46,48] The corresponding radical intermediate, the metallaoxetane, and the electron transferred (ET) intermediates were all found to be much less stable than the **IC1** (see the Supporting Information, page S6). Furthermore, the radical intermediate and the metallaoxetane collapsed to **IC1** upon geometry optimization; while the electron transferred (ET) intermediate formation is endothermic by 117.4 kcal/mol (at UB3LYP/B2). The only alternative path we found is the one shown by the green energy profile in Figure 2a. In this alternative pathway (in green), O^--*ε*A^+ (**IC1**) is converted to *the epoxide* (*ε*A-ep). This however does not occur by way of a simple ring closure, *because ring closure has a much higher barrier* (Supporting Information Figure S6) and *cannot compete with the hydrolysis of* O^--*ε*A^+ *via* ***TS2***. Instead, O^--*ε*A^+ (**IC1**) undergoes initially a C22−N1 bond cleavage via **TS3**, leading to the aldehyde species, **IC3**, which then undergoes ring closure to the epoxide (*ε*A-ep) labeled **IC4**. As can be seen from Figure 2a, the epoxide is less stable than either the zwitterionic **IC1** intermediate, O^--*ε*A^+, or its hydrolytic product **IC2**, *ε*A-gl-H. Using two other snapshots (**Sn2** and **Sn3**) leads to the same results as those presented in Figure 2a for **Sn1**. Thus, in all three snapshots, the oxygenation of *ε*A generated the zwitterion O^--*ε*A^+ that underwent hydrolysis to the glycoxide complex (*ε*A-gl-H), which is more stable than the epoxide. All these results for the various snapshots are summarized in Supporting Information Tables S1 and S2 and Figures S7−S9.

In a catalytic cycle, this result means that the epoxide **IC4** (*ε*A-ep) will revert back to **IC2** (*ε*A-gl-H), unless the hydrolysis of epoxide is more facile than that of the zwitterion hydrolysis (*ε*A-gl-H formation). As will be seen, *the findings below rule out the option that* ***IC4*** *has a low hydrolytic barrier!*

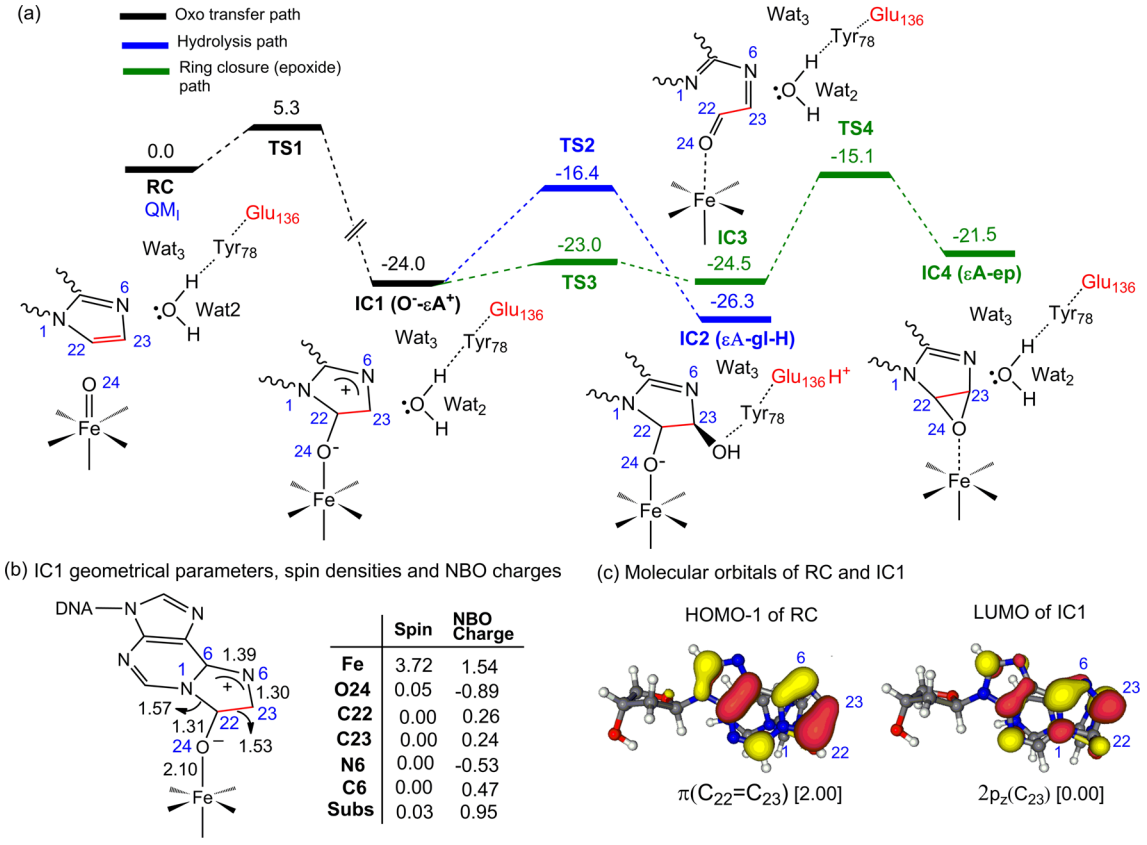
3.2. Epoxide Appears To Be a Dead-End Species. To ascertain whether the instability of *ε*A-ep (**IC4**) vis-à-vis the aldehyde species (**IC3**) is intrinsic or due to the protein environment, we carried out QM-only calculations in the gas phase. These calculations showed that the epoxide could be formed in the gas phase via a similar pathway as in Figure 2a. However, the so formed epoxide (**IC4-QM**) *was found to be unstable toward C22−N1 bond cleavage, and underwent cyclo-reversion* to yield the aldehyde species (**IC3-QM**) which lies 6 kcal/mol lower in energy (see Supporting Information Figure S10) vs 3 kcal/mol in the protein than the epoxide (Figure 2a).

Thus, *the epoxide is intrinsically unstable* in the gas phase. The reason for the intrinsic instability of the epoxide is the strong hyperconjugative interaction, between the filled π-orbitals of the adenine moiety of *ε*A-ep with the lesion's σ^*(C−O) orbitals of the epoxide ring, which in turn results in a facile symmetry-allowed[49] cyclo-reversion of the lesion in the epoxide ring.

To test whether the epoxide is an actual intermediate en-route to the repair pathway, we compared the hydrolysis of both the epoxide and the zwitterion in Figure 3. Our QM/MM calculations show that the nucleophilic attack of the water on the epoxide is energetically unfavorable; the energy keeps rising without giving the hydrolytic products (either *ε*A-gl-H or *ε*A-gl). We note in this regard the experimental evidence, which shows that epoxide hydrolysis is not a facile process and it generally requires boiling water,[25] and even transition metal-catalyzed hydrolysis[50] is not facile as the hydrolysis of **IC1**. In fact, as shown in Figure 3, all attempts to hydrolyze the epoxide led to the zwitterion intermediate O^--*ε*A^+ (**IC1**). Thus, *the epoxide complex εA-ep is a dead-end side product species* and not an intermediate along the repair pathway. This means that, if formed, the epoxide **IC4** (*ε*A-ep) will revert back to **IC1** (O^--*ε*A^+), which will proceed to **IC2** (*ε*A-gl-H) that continues the repair pathway elucidated by the QM/MM results below.

3.3. Completion of the Repair Mechanism via IC2. The remainder of the repair mechanism proceeds from the glycoxide complex *ε*A-gl-H intermediate (**IC2** in Figure 2a).

dx.doi.org/10.1021/ja507934g | J. Am. Chem. Soc. 2014, 136, 13895−13901



	Spin	NBO Charge
Fe	3.72	1.54
O24	0.05	-0.89
C22	0.00	0.26
C23	0.00	0.24
N6	0.00	-0.53
C6	0.00	0.47
Subs	0.03	0.95

Figure 2. (a) UB3LYP/B2 relative QM/MM energies (kcal/mol) for the oxygenation of εA by AlkB-iron^{IV}-oxo and the further transformation of the zwitterion intermediate O$^-$-εA$^+$ (using the QM$_I$ subsystem and **Sn1**). The **IC1** has two pathways, one is hydrolysis and formation of εA-gl-H intermediate (blue), and the other is ring closure for epoxide, εA-ep (green). Key intermediates along the reaction energy profile are schematically drawn. (b) The UB3LYP/B1 optimized geometric details (in Å) of **IC1**(O$^-$-εA$^+$) along with spin and NBO charges at UB3LYP/B2. (c) The lowest unoccupied molecular orbital (LUMO) of **IC1** shown side by side with the HOMO-1 of εA in **RC**. The main orbital characters and their populations (in brackets) are indicated underneath the orbital drawings.

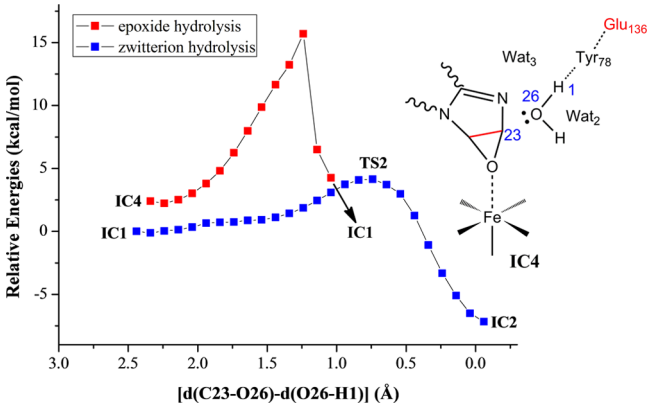


Figure 3. Comparison of the scanned B3LYP/B1 relative QM/MM energy profiles (in kcal/mol) for the hydrolysis of O$^-$-εA$^+$ (blue squares) and the epoxide **IC4** (red squares). The key atom numberings are shown in **IC4** in right-hand structure. Note that the nucleophilic attack of the water on the epoxide ultimately collapses to the zwitterion intermediate O$^-$-εA$^+$.

Note that whereas experiment suggested that the glycol complex (εA-gl) is an intermediate along the repair pathway, the fact is that the glycol complex cannot be formed easily, since the proton abstracted from Wat2 by Glu136 (see **IC2** in Figure 2a) has no way of being transferred to O24 to protonate the glycoxide complex, εA-gl-H (see details later).

Figure 4 shows the transformation of **IC2** to the repaired adenine (A) and the glyoxal (G). Initially, the water environment around **IC2** rearranges and establishes an H-bond network from N6 to the proton in $^+$H-Glu136. The so formed **IC5** species undergoes protonation at N6, thus generating the N6-protonated zwitterionic intermediate εA-gl-zw (**IC6**), which is *an isomer of the glycol complex*. Subsequent cleavage of the C22−N1 bond via **TS6** yields to **IC7** that has a short C22−O24 bond of 1.22 Å, which is a typical value of a C═O double bond. At the same time, the Fe−O24 bond in **IC7** undergoes lengthening to 2.32 Å.

Starting from **IC7**, in the midpoint of Figure 4, we investigated the cleavage of C23−N6 bond, and as expected (when N6 is monoprotonated), this requires a very high barrier (>30 kcal/mol in Supporting Information Figure S11). To break the C23−N6 bond easily, the N6 position of **IC7** has to accept an additional proton and become the NH_2 group of the repaired adenine. Therefore, we tested the feasibility of N6 protonation by incorporating Wat4, Wat5, and Asp135 into the QM region, resulting in QM$_{II}$, and the corresponding **IC7′** species.

Starting from **IC7′** in Figure 4, the Asp135 transfers the proton to N6 through the water chain of Wat4 and Wat5 (via **TS7**), yielding a doubly protonated N6 in **IC8**. *This is the rate-determining step for the mechanism, with a barrier of 18.6 kcal/mol.* The subsequent C23−N6 bond dissociation (via **TS8**) indeed requires a small barrier to generate the final products

dx.doi.org/10.1021/ja507934g | J. Am. Chem. Soc. 2014, 136, 13895−13901

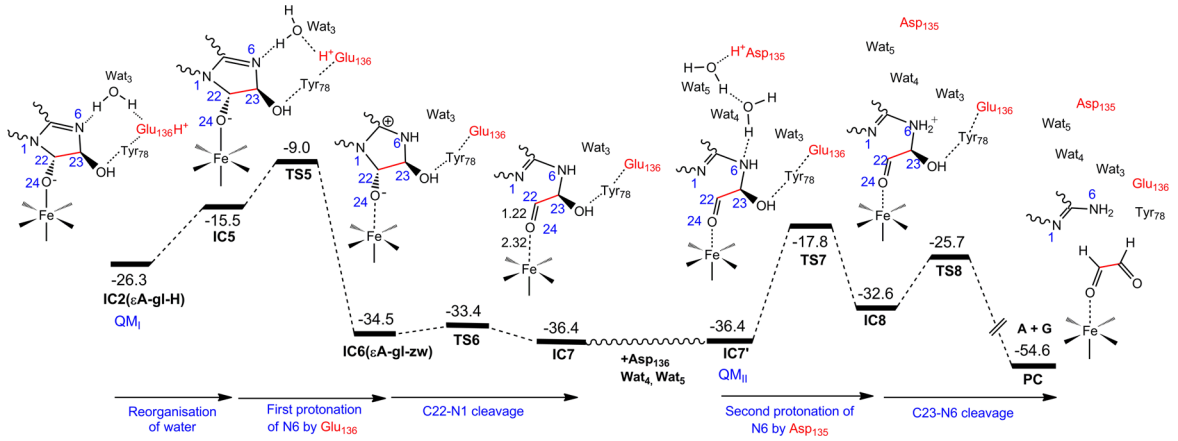


Figure 4. UB3LYP/B2 relative QM/MM energies (in kcal/mol) for the conversion of εA-gl-H intermediate (**IC2**) to the glyoxal complex and the repaired adenine base (**PC**) via a zwitterionic intermediate εA-gl-zw (**IC6**). In **IC7**, we indicate also the Fe−O and C═O bond lengths (2.32 and 1.22 Å, respectively).

(**PC**), the repaired adenine base and the glyoxal complex. In this latter process, the cleavage of C23−N6 bond is catalyzed by Glu136, which can abstract the proton from the O26−H group that is bonded to C23, and thereby facilitate the C−N bond cleavage.

We also considered the proximal Lys134 as the proton donor in the protonation of **IC7**. However, with Lys134, whose pK_a is 13.8 (vis-à-vis 5.5 for Asp135), the second protonation of N6 has a very high barrier (>50 kcal/mol in Supporting Information Figure S12). As such, Lys134 is incapable of causing the second protonation of N6, since it is a much weaker proton donor compared to Asp135. These findings and the crucial role of Asp135 are consistent with experimental data. Thus, as shown by Figure 4, the second protonation that forms **IC8** *depends on Asp135 and constitutes the rate-determining step*. This is in accord with experimental findings that the efficiency of the repair mechanism exhibits a pH dependency within the pH range of pH = 4−7.[11] Furthermore, the calculated QM/MM barrier, 18.6 kcal/mol, is in a reasonably close agreement with the experimental value of 21.5 kcal/mol that can be estimated from the rate constant in ref 11 using the Eyring equation.

3.4. Is Glycol an Intermediate along the Repair Path?

Inspection of the structure of the active site of the enzyme showed that the glycol formation could transpire only via O-protonation of **IC6** by Arg 210. To test the feasibility of this process, we incorporated Arg210 into the QM_I, resulting in QM_{III}, and carried out the protonation pathways of the glycoxide complex, εA-gl-H (**IC5**/**IC5′**). Figure 5 compares the formation of the glycol complex εA-gl (**IC12**) vis-à-vis the formation of the isomeric species εA-gl-zw (**IC6**), starting in both pathways from the glycoxide complex, εA-gl-H.

As is clearly evident from Figure 5, the formation of εA-gl-zw (**IC6**) is more favorable thermodynamically compared to the formation of the glycol εA-gl (**IC12**), by 20 kcal/mol! Furthermore, the optimized structures of εA-gl in Figure 5 reveal that the Fe−O24 distance has increased to 3.68 Å compared with only 2.10 Å for εA-gl-zw. Thus, the protonation of O24 to form the glycol complex dissociates the Fe−O24 bond. By contrast, Figure 5 shows that the Fe−O distance of εA-gl-zw (2.10 Å) is very close to the experimental values (2.02 Å), which was ascribed to the "glycol" complex.[12] The short Fe−O distance of εA-gl-zw is caused by the negative charge on

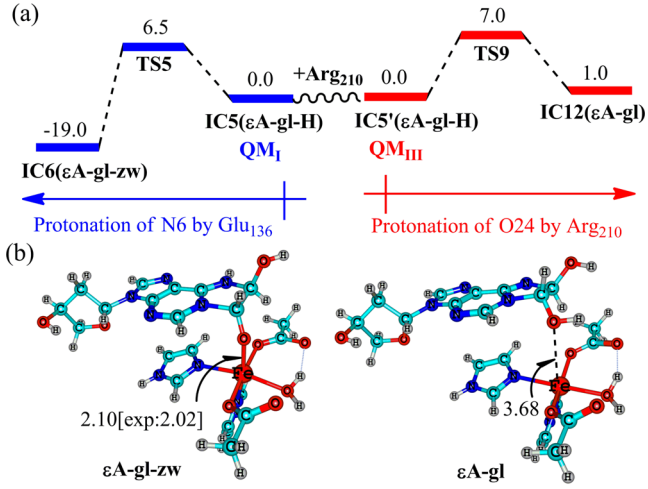


Figure 5. (a) UB3LYP/B2 relative QM/MM energies (in kcal/mol) for the glycol formation and its N-protonated isomer formation. (b) Optimized Fe−O24 distances (in Å) for the glycol and its N-protonated isomer εA-gl-zw. The Fe−O24 distance in the crystal structure is shown in brackets alongside the QM/MM value for εA-gl-zw.

the oxygen. When this oxygen gets protonated, the Fe−O bond becomes very weak and hence very long. As such, the QM/MM glycol-complex (**IC12**) has *a much too long Fe−O24 bond to match the in crystallo species*. On the other hand, the computed εA-gl-zw in Figure 5b, matches the *in crystallo* species very well.

We also investigated the putative transformation of the glycol εA-gl complex into the repaired adenine and glyoxal complex, and found a barrier of 33.2 kcal/mol (Supporting Information Figure S15). All these calculations lead to the conclusion that *the glycol is not an intermediate along the repair pathway*.

As we stated at the onset of this present paper, X-ray crystallography does not assign the location of H atoms, and hence, it cannot distinguish εA-gl from εA-gl-zw. However, with the advent of the QM/MM results in Figure 5, we can use the computed Fe−O distance to match the *in crystallo* isolated intermediate with the isomer of the glycol having N-protonated moiety, εA-gl-zw, which is 20 kcal/mol more stable than the glycol intermediate (Figure 5). The role of theory in providing precise atomistic information here is indispensable.

dx.doi.org/10.1021/ja507934g | J. Am. Chem. Soc. 2014, 136, 13895−13901

4. CONCLUSIONS

Our QM/MM investigation of the DNA repair of adenine with an etheno-bridge lesion (εA, in Scheme 1) by the iron^{IV}-oxo active species of the AlkB enzyme reveals a novel mechanism. This mechanism starts with the formation of an oxy zwitterion O^{-}-εA^{+} (**IC1**), formed by attack of iron^{IV}-oxo on the C=C double bond of the lesion (Figure 2a). This is followed by **IC1** hydrolysis and then protonation of the N6 position of **IC2** (εA-gl-H) to form **IC6**, εA-gl-zw, which is a nitrogen-protonated isomer of glycol. Subsequently, **IC6** undergoes C22−N1 bond cleavage and a second protonation on the N6 via a proton-shuttle path formed by the Asp135 residue. *The second protonation is the rate-determining step of the mechanism*, which is followed by C23−N6 cleavage and a facile formation of glyoxal and the repaired adenine. Our rate-determining step and its respective barrier fit well with the experimental findings of pH dependence of the repair and its rate.

The QM/MM results rule out the repair of the C=C lesion via an epoxide formation. It is quite probable that the "epoxide" identified tentatively by mass spectrometry[13] is an aldehyde (**Ald-1** in Supporting Information Figure S17), which is formed by rearrangement of the liberated zwitterionic intermediate **IC1**. **Ald-1** has the same *m*/*e* value as the epoxide and, as shown by our calculations (Supporting Information Figure S17), it is much more stable than the epoxide. Similarly, the mass spectrometrically identified "glycol" intermediate appears more likely to be another aldehyde (**Ald-2** in Supporting Information Figure S17) also formed by rearrangement of the liberated zwitterionic intermediate εA-gl-zw (**IC6**). Furthermore, the results show, with certitude, that the *in crystallo* entrapped species[12] is not a ferrous-glycol complex but rather its isomer εA-gl-zw (**IC6**), which involves Fe−O(−) bond and a nitrogen protonated glycol εA-gl-H, with an Fe−O bond length that fits well with the *in crystallo* bond length. These results underscore the role of theory as a partner to experiment by providing atomistic descriptions of the key intermediates vis-à-vis observed ones in this vital mechanism.

This mechanism may have general implication on repairs of such lesion, i.e., 1,N^6-ethenoadenine (εA), 3,N^4-ethenocytosine (εC), and 1,N^2-ethenoguanine (εG), and perhaps also other lesions. Furthermore, we have identified two crucial residues that control the mechanism; one lines up with the experimental finding of a pH dependence of the rate, and the other is new. Thus, in accord with experiment, Asp 135 is responsible for the rate-determining protonation. However, our calculations show additionally that Glu136 is an extremely important residue for the overall repair reactions. This residue is involved in two key steps: one is the hydrolysis of O^{-}-εA^{+} (Figure 2a), where Glu136 acts as base catalyst, and the second is the C23−N6 cleavage (Figure 4), where it also acts as base catalysis. On the other hand, the proposition in the experimental work that the Arg210 acts as base catalysis in C22−N1 cleavage is not substantiated by our QM/MM calculations, which rule out the participation of Arg210. These predictions of crucial residues can be easily tested to verify or falsify the proposed novel mechanism, and guide the further experimental mechanistic investigations.

■ ASSOCIATED CONTENT

Supporting Information

Natural orbitals, spin densities and NBO charges for the key species. Optimized geometries of all the structures. Total energies at B2 and Cartesian coordinates of all the structures. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

sason@yfaat.ch.huji.ac.il

Present Address

⊥D.U.: Lipidomic Centre, Central Food Technological Research Institute, Council of Scientific and Industrial Research, Mysore, 570 020, India.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The manuscript is dedicated to Helmut Schwarz, a Mensch and a fellow scientist, on occasion of his 70th birthday. S.S. acknowledges the support of the ISF (Grant 1183/13). B.W. thanks J. C. Delaney, J. M. Essigmann, Q. H. Weng, and C.-H. Lu for helpful comments on the mass spectrometric experiments.

■ REFERENCES

(1) Ito, S.; Shen, L.; Dai, Q.; Wu, S. C.; Collins, L. B.; Swenberg, J. A.; He, C.; Zhang, Y. *Science* **2011**, *333*, 1303−1307.
(2) Cedar, H.; Bergman, Y. *Nat. Rev. Genet.* **2009**, *10*, 295−304.
(3) Bergman, Y.; Cedar, H. *Nat. Struct. Mol. Biol.* **2013**, *20*, 274−281.
(4) Shrivastav, N.; Li, D.; Essigmann, J. M. *Carcinogenesis* **2010**, *31*, 59−70.
(5) Law, J. A.; Jaccobsen, S. E. *Nat. Rev. Genet.* **2010**, *11*, 204−220.
(6) Spruijt, C. G.; Gnerlich, F.; Smits, A. H.; Pfaffeneder, T.; Jansen, P. W. T. C.; Bauer, C.; Münzel, M.; Wagner, M.; Müller, M.; Khan, F.; Eberl, H. C.; Mensinga, A.; Brinkman, A. B.; Lephikov, K.; Müller, U.; Walter, J.; Boelens, R.; van Ingen, H.; Leonhardt, H.; Carell, T.; Vermeulen, M. *Cell* **2013**, *152*, 1146−1159.
(7) Pfaffeneder, T.; Spada, F.; Wagner, M.; Brandmayr, C.; Laube, S. K.; Eisen, D.; Truss, M.; Steinbacher, J.; Hackner, B.; Kotljarova, O.; Schuermann, D.; Michalakis, S.; Kosmatchev, O.; Schiesser, S.; Steigenberger, B.; Raddaoui, N.; Kashiwazaki, G.; Müller, U.; Spruijt, C. G.; Vermeulen, M.; Leonhardt, H.; Schär, P.; Müller, M.; Carell, T. *Nat. Chem. Biol.* **2014**, *10*, 574−581.
(8) Shen, L.; Song, C. X.; He, C.; Zhang, Y. *Annu. Rev. Biochem.* **2014**, *83*, 585−614.
(9) Zheng, G.; Fu, Y.; He, C. *Chem. Rev.* **2014**, *114*, 4602−4620.
(10) Wu, H.; Zhang, Y. *Cell* **2014**, *156*, 45−68.
(11) Mishina, Y.; Yang, C. G.; He, C. *J. Am. Chem. Soc.* **2005**, *127*, 14594−14595.
(12) Yi, C.; Jia, G.; Hou, G.; Dai, Q.; Zhang, W.; Zheng, G.; Jian, X.; Yang, C. G.; Cui, Q.; He, C. *Nature* **2010**, *468*, 330−333.
(13) Delaney, J. C.; Smeester, L.; Wong, C.; Frick, L. E.; Taghizadeh, K.; Wishnok, J. S.; Drennan, C. L.; Samson, L. D.; Essigmann, J. M. *Nat. Struct. Mol. Biol.* **2005**, *12*, 855−860.
(14) Falnes, P. O.; Johansen, R. F.; Seeberg, E. *Nature* **2002**, *419*, 178−182.
(15) Trewick, S. C.; Henshaw, T. F.; Hausinger, R. P.; Lindahl, T.; Sedgwick, B. *Nature* **2002**, *419*, 174−178.
(16) Que, L., Jr. *Acc. Chem. Res.* **2007**, *40*, 493−500.
(17) Nam, W. *Acc. Chem. Res.* **2007**, *40*, 522−531.
(18) Costas, M. *Coord. Chem. Rev.* **2011**, *255*, 2912−2932.
(19) Borovik, A. S. *Chem. Soc. Rev.* **2011**, *40*, 1870−1874.
(20) Krebs, C.; Fujimori, D. G.; Walsh, C. T.; Bollinger, J. M., Jr. *Acc. Chem. Res.* **2007**, *40*, 484−492.
(21) Wong, S. D.; Srnec, M.; Matthews, M. L.; Liu, L. V.; Kwak, Y.; Park, K.; Bell, C. B., III; Alp, E. E.; Zhao, J. Y.; Yoda, Y.; Kitao, S.; Seto, M.; Krebs, C.; Bollinger, J. M.; Solomon, E. I. *Nature* **2013**, *499*, 320−323.

(22) Liu, H.; Llano, J.; Gauld, J. W. *J. Phys. Chem. B* **2009**, *113*, 4887−4898.
(23) Fang, D.; Lord, R. L.; Cisneros, G. A. *J. Phys. Chem. B* **2013**, *117*, 6410−6420.
(24) Quesne, M. G.; Latifi, R.; Gonzalez-Ovalle, L. E.; Kumar, D.; de Visser, S. P. *Chem.—Eur. J.* **2014**, *20*, 435−446.
(25) Wang, Z.; Cui, Y. T.; Xu, Z. B.; Qu, J. *J. Org. Chem.* **2008**, *73*, 2270−2274.
(26) The QM/MM was originally conceived in: Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227−249.
(27) Olsson, M. H.; Søndergard, C. R.; Rostkowski, M.; Jensen, J. H. *J. Chem. Theory Comput.* **2011**, *7*, 525−537.
(28) Brooks, B. R.; Brooks, C. L., III; MacKerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. *J. Comput. Chem.* **2009**, *30*, 1545−1614.
(29) Sherwood, P.; de Vries, A. H.; Guest, M. F.; Schreckenbach, G.; Catlow, C. R. A.; French, S. A.; Sokol, A. A.; Bromley, S. T.; Thiel, W.; Turner, A. J.; Billeter, S.; Terstegen, F.; Thiel, S.; Kendrick, J.; Rogers, S. C.; Casci, J.; Watson, M.; King, F.; Karlsen, E.; Sjovoll, M.; Fahmi, A.; Schäfer, A.; Lennartz, C. *J. Mol. Struct. (THEOCHEM)* **2003**, *632*, 1−28.
(30) Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. *Chem. Phys. Lett.* **1989**, *162*, 165−169.
(31) Smith, W.; Forester, T. R. *J. Mol. Graphics* **1996**, *4*, 136−141.
(32) Bakowies, D.; Thiel, W. *J. Phys. Chem.* **1996**, *100*, 10580−10594.
(33) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5652.
(34) Billeter, S. R.; Turner, A. J.; Thiel, W. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2177−2186.
(35) Shaik, S.; Cohen, S.; Wang, Y.; Chen, H.; Kumar, D.; Thiel, W. *Chem. Rev.* **2010**, *110*, 949−1017.
(36) Senn, H. M.; Thiel, W. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198−1229.
(37) For a related method, see: Lin, H.; Truhlar, D. G. *Theor. Chem. Acc.* **2007**, *117*, 185−199.
(38) Altun, A.; Kumar, D.; Neese, F.; Thiel, W. *J. Phys. Chem. A* **2008**, *112*, 12904−12910.
(39) Usharani, D.; Zazza, C.; Lai, W.; Chourasia, M.; Waskell, L.; Shaik, S. *J. Am. Chem. Soc.* **2012**, *134*, 4053−4056.
(40) Kumar, D.; Thiel, W.; de Visser, S. P. *J. Am. Chem. Soc.* **2011**, *133*, 3869−3882.
(41) Dong, G.; Shaik, S.; Lai, W. Z. *Chem. Sci.* **2013**, *4*, 3624−3635.
(42) Unno, M.; Chen, H.; Kusama, S.; Shaik, S.; Ikeda-Saito, M. *J. Am. Chem. Soc.* **2007**, *129*, 13394−13395.
(43) Lai, W.; Chen, H.; Matsui, T.; Omori, K.; Unno, M.; Ikeda-Saito, M.; Shaik, S. *J. Am. Chem. Soc.* **2010**, *132*, 12960−12970.
(44) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; 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.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, Revision B. 01; Gaussian, Inc.: Wallingford, CT, 2009.
(45) (a) de Visser, S. P.; Ogliaro, F.; Harris, N.; Shaik, S. *J. Am. Chem. Soc.* **2001**, *123*, 3037−3047. (b) Kumar, D.; de Visser, S. P.; Shaik, S. *Chem.—Eur. J.* **2005**, *11*, 2825−2835.
(46) Groves, J. T. *J. Chem. Educ.* **1985**, *62*, 928−931.
(47) Hirao, H.; Kumar, D.; Thiel, W.; Shaik, S. *J. Am. Chem. Soc.* **2005**, *127*, 13007−13018.
(48) Arasasingham, R. D.; He, G.-X.; Bruice, T. C. *J. Am. Chem. Soc.* **1993**, *115*, 7985−7991.
(49) Woodward, R. B.; Hoffmann, R. *The Conservation of Orbital Symmetry*; Verlag Chemie: Weinheim, 1970.
(50) Tokunaga, M.; Larrow, J. F.; Kakiuchi, F.; Jacobsen, E. N. *Science* **1997**, *277*, 936−938.