Spin-projected QM/MM Free Energy Simulations for Oxidation Reaction of Guanine in B–DNA by Singlet Oxygen

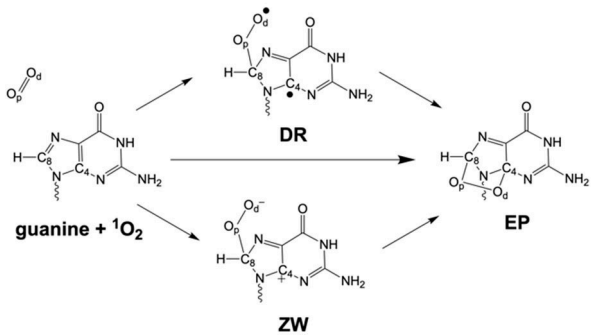
Toru Saito*[a] and Yu Takano[a]

Guanine is the most susceptible base to oxidation damage induced by reactive oxygen species including singlet oxygen ($^1O_2$, $^1\Delta_g$). We clarify whether the first step of guanine oxidation in B–DNA proceeds via either a zwitterionic or a diradical intermediate. The free energy profiles are calculated by means of a combined quantum mechanical and molecular mechanical (QM/MM) method coupled with the adaptive biasing force (ABF) method. To describe the open-shell electronic structure of $^1O_2$ correctly, the broken-symmetry spin-unrestricted density functional theory (BS–UDFT) with an approximate spin projection (AP) correction is applied to the QM region. We find that the effect of spin contamination on the activation and reaction free energies is up to ~8 kcal mol^{-1}, which is too large to be neglected. The QM(AP–ULC–BLYP)/MM-based free energy calculations also reveal that the reaction proceeds through a diradical transition state, followed by a conversion to a zwitterionic intermediate. Our computed activation energy of 5.2 kcal mol^{-1} matches experimentally observed range (0 ~ 6 kcal mol^{-1}).

1. Introduction

Reactive oxygen species such as singlet oxygen ($^1O_2$, $^1\Delta_g$), hydroxyl radical (OH$^\bullet$) and superoxide anion ($O_2^{-\bullet}$) contribute to oxidation reactions in biological systems.[1–2] Singlet oxygen, commonly abbreviated as $^1O_2$, is the first excited electronic state of oxygen molecule, lying 22.4 kcal mol^{-1} in energy above the triplet ground state ($^3\Sigma_g^+$). $^1O_2$ can be generated by photosensitized reactions or be obtained from chemical sources such as hydrogen peroxide and endoperoxides. It exhibits a large reactivity toward electron-rich olefins and aromatic compounds.[3] The target biomolecules including nucleic acids, unsaturated lipids, and amino acids mostly undergo pericyclic reactions. While these oxidation reactions give rise to several diseases such as porphyria and skin cancer, they are useful in the natural product synthesis.[4] In the present work, we focus on $^1O_2$-induced oxidative DNA damage implicated in various biological processes..[5,6] In DNA, guanine is the most susceptible base to oxidation damage, and is mainly oxidized to 8-oxo-7,8-dihydroguanine (8-oxoG). According to extensive experimental and computational studies, it is verified that the formation of 8-oxoG is followed by a transient guanine 4,8-endoperoxide intermediate (**EP**) generated by the [4 + 2] cycloaddition of $^1O_2$ to the imidazole ring of guanine.[7–19] It is still unclear whether the reaction proceeds through either a concerted mechanism or a stepwise mechanism through a zwitterionic (**ZW**) or a diradical (**DR**) intermediate (Scheme 1).

Marchetti and Karsili studied the reaction mechanism of the formation of **EP** using an isolated system consisting of $^1O_2$ and guanine only (hereafter called QM-only model).[14] They performed multi-configurational CASPT2/cc-pVDZ calculations in the gas phase along an approximate reaction coordinate based on a linear interpolation in internal coordinates obtained at the spin-restricted second-order Møller-Plesset (RMP2)/cc-pVDZ level.[20–22] Their results suggest that the [4 + 2] cycloaddition is virtually barrierless and proceeds via a concerted mechanism. Dumont et al. performed quantum mechanical and molecular mechanical molecular dynamics (QM/MM MD) simulations to investigate the reaction mechanism in aqueous solution and in



Scheme 1. Proposed two reaction mechanisms of an endoperoxide (**EP**) formation. Singlet oxygen ($^1O_2$) reacts with guanine to form either a concerted mechanism or a stepwise mechanism through a zwitterionic (**ZW**) or a diradical intermediate (**DR**).

[a] Dr. T. Saito, Prof. Y. Takano
Department of Biomedical Information Science
Graduate School of Information Science
Hiroshima City University
3-4-1 Ozuka-Higashi, Asa-Minami-Ku, 731-3194 Hiroshima (Japan)
E-mail: tsaito@hiroshima-cu.ac.jp

Supporting information for this article is available on the WWW under https://doi.org/10.1002/cphc.202000978

© 2021 The Authors. ChemPhysChem published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution Non-Commercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

B–DNA.[15,16] The spin-restricted density functional theory (RDFT) method, namely the RLC–BLYP/6-31+G* method, suggested a stepwise mechanism, in which the reaction starts with an attack of the proximal oxygen atom (O_p) of $^1O_2$ on the C8 atom of guanine to form **ZW**, and the following nucleophilic attack of the distal oxygen atom (O_d) of the peroxo group on the C4 atom leads to **EP**. They concluded that the B–DNA environment stabilizes **ZW** and significantly enhances the formation of **EP** because the calculated barrier for the rate-limiting process is small (6.2 kcal mol^{-1}) and the overall reaction is strongly exergonic (~ −60 kcal mol^{-1}).

However, it has been pointed out that energy profiles based on spin-restricted calculations are in general not accurate for the open-shell electronic structure of $^1O_2$, due to the inherent inability to describe static correlation effects arising from the double degeneracy of π* orbitals.[23–28] The broken-symmetry spin-unrestricted DFT (BS–UDFT) calculation can handle such static correlation, but its broken-symmetry singlet state significantly suffers from the spin contamination from triplet ground state.[29–31] Several methods have been proposed to remove spin contamination from electronic energies for broken-symmetry singlet states.[32–34] Thapa et al. investigated both the zwitterionic and diradical pathways using the QM-only model.[19] Yamaguchi's approximate spin projection (AP) method described by eq. (1) (see Computational Method) was applied to a range of DFT functionals including LC–BLYP. Solvent effects were implicitly included by means of SMD solvation model.[35] The single-point SMD/NEVPT2/6-31+G** calculations on the SMD/BS-UB3LYP-optimized stationary points were also performed for comparison.[36] They found that inclusion of implicit solvent effects can stabilize **ZW**, and the stepwise zwitterionic pathway was found to be more preferable. It was also indicated that the effect of spin contamination is present in transition states and intermediates. The activation enthalpy relative to isolated reactants predicted by AP–ULC–BLYP (8.7 kcal mol^{-1}) was comparable to that by NEVPT2 (7.1 kcal mol^{-1}), assuring the applicability of the AP method to the investigation of guanine oxidation as well as other singlet oxygen reactions.

Motivated by these computational results, we wish to provide a comprehensive description of guanine oxidation in B–DNA by applying the AP method to QM/MM MD simulation. The adaptive biasing force (ABF) method was employed to determine free energy landscape.[37] In light of the previous studies mentioned above and computational efficiency, the LC–BLYP/6-31G* model was chosen to treat the QM region (Figure 1a).[38–43]

The present approach allows to make a more accurate and unbiased analysis, retaining the benefits of accounting for both the B–DNA environment and static correlation. With regard to singlet oxygen reactions, the use of AP–UDFT is advantageous over more rigorous multi-configurational calculations and much faster semi-empirical calculations.[44,45] Multi-configurational calculations are prohibitively expensive for QM/MM free energy calculations, since several tens of thousands of electronic structure calculations are needed to persist several tens of picoseconds. We notice that semi-empirical methods cannot describe the $^1O_2$ reduction process appropriately, even if a

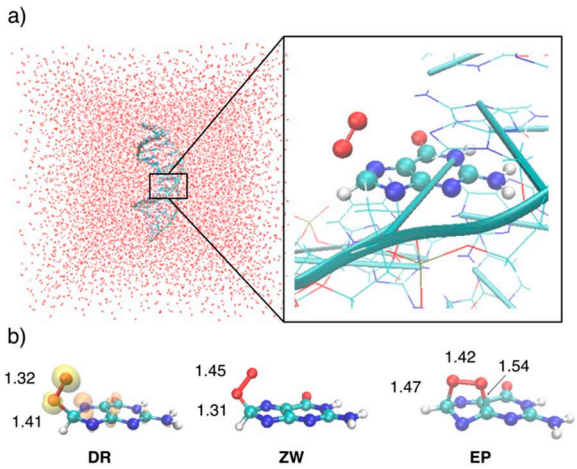


Figure 1. a) QM/MM model with bulk water molecules (left) and QM region (right) used in this study. b) Final snapshots (**DR**, **ZW**, **EP**) from the QM/MM MD equilibration exploited as the initial configurations for ABF simulations, with isovalue surfaces of positive (orange) and negative (yellow) spin density (isovalue = 0.01).The RLC–BLYP/6-31G* method was used for **ZW** and **EP**, and the BS–ULC–BLYP/6-31G* method was used for **DR**.

parameter set tuned to open-shell systems is utilized.[46] To illustrate the applicability of the spin projection method to QM/MM MD simulations, the performance of AP–ULC–BLYP is compared with that of RLC–BLYP and BS–ULC–BLYP. We demonstrate how inclusion of static correlation and removal of spin contamination improve the accuracy of free energies. Emphasis is also placed on whether the reaction proceeds via either **ZW** or **DR**. Although several QM/MM MD studies have been reported for open-shell systems such as metalloproteins,[47–49] to the best of our knowledge, this is the first study that deals with a reaction on the open-shell singlet surface using spin projection.

2. Results and Discussion

2.1. Singlet-Triplet Energy Gap for O_2

The adiabatic singlet-triplet energy gap (ΔE_{ST}) for $^1O_2$ has often been examined to check the accuracy of theoretical methods for describing of its electronic structure.[50–54] We show how the electronic energies of the singlet state vary between RLC–BLYP, BS–ULC–BLYP, and AP–ULC–BLYP. The 6-31G* basis set was used. The energy of the reference triplet ground state was calculated at the BS–ULC–BLYP/6-31G* level, with an assumption that the state is free from spin contamination. The ΔE_{ST} values obtained using RLC–BLYP, BS–ULC–BLYP, and AP–ULC–BLYP are 39.5, 11.0, and 22.0 kcal mol^{-1}, respectively. It is seen that the RLC–BLYP calculation is unsatisfactory because of static correlation error. The resulting ΔE_{ST} value of 39.5 kcal mol^{-1} significantly overestimates the experimental value of 22.4 kcal mol^{-1}. The singlet state calculated by BS–ULC–BLYP has an $\langle S^2\rangle^{BS}$ value of 1.0029, which is typical for diradical species containing an equal mixture of singlet and

© 2021 The Authors. ChemPhysChem published by Wiley-VCH GmbH

triplet. The severe spin contamination by the triplet ground state causes an underestimation of the ΔE_{ST} value (11.0 vs. 22.4 kcal mol^{-1}). The AP–ULC–BLYP calculation successfully corrects the energy of the singlet state, leading to a good agreement with the experimental data (22.0 vs. 22.4 kcal mol^{-1}). On the basis of the results, we wish to propose a hypothesis that the three different calculations will give rise to different free energy profiles for singlet oxygen reactions. The RLC–BLYP calculation is likely to underestimate activation barriers and overestimate exergonicity of a target reaction, whereas the opposite is true for the BS–ULC–BLYP calculation. The AP–ULC–BLYP is expected to lie between the two calculations and to provide a much better description.

2.2. Diradical Pathway

The QM/MM ABF simulations using BS–ULC–BLYP and AP–ULC–BLYP for the QM region were started with the **DR** intermediate shown in Figure 1. We will hereafter refer to, for example, QM(BS–ULC–BLYP)/MM ABF as BS–ULC–BLYP for convenience. The electronic structure of **DR** can be viewed as a perfect diradical with an $\langle S^2\rangle^{BS}$ value of 1.0215 calculated at the BS–ULC–BLYP/6-31G* level. The reaction coordinate was described by the C8-O_p distance. The one-dimensional potentials of mean force (1D-PMFs) for the diradical pathway calculated at the BS–ULC–BLYP and AP–ULC–BLYP levels are shown in Figure 2. Also displayed in Figure 2 are representative snapshots of the reactant state ($\mathbf{RS_{AP}}$), transition state ($\mathbf{TS1_{AP}}$), and **DR** intermediate state ($\mathbf{DR_{AP}}$) and corresponding spin density populations obtained with AP–ULC–BLYP. Analogous results are found in their BS–ULC–BLYP counterparts ($\mathbf{RS_{BS}}$, $\mathbf{TS1_{BS}}$, $\mathbf{DR_{BS}}$). The obtained $\mathbf{RS_{AP}}$ has a C8···Op separation of ~3.0 Å and

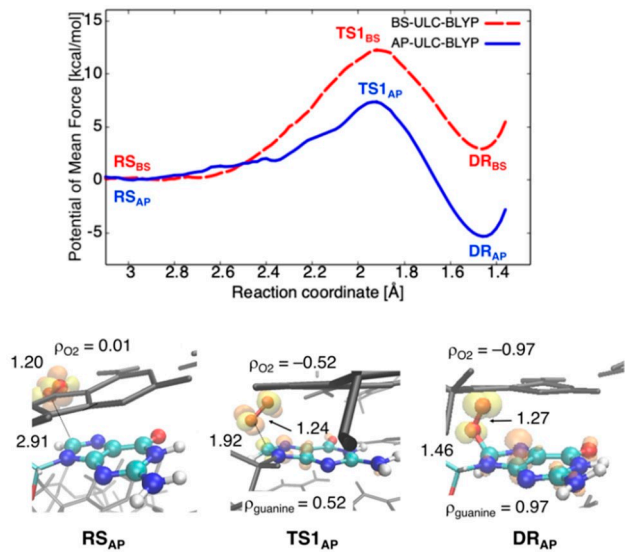


Figure 2. Top: 1D-PMFs for the diradical pathway obtained with the BS–ULC–BLYP/MM ABF and AP–ULC–BLYP/MM ABF simulations. Bottom: Representative snapshots of $\mathbf{RS_{AP}}$, $\mathbf{TS1_{AP}}$, and $\mathbf{DR_{AP}}$ calculated by the AP–ULC–BLYP/MM ABF simulations, with isovalue surfaces of positive (orange) and negative (yellow) spin density are plotted (isovalue = 0.01).

the O_p-O_d bond length of 1.19 ± 0.03 Å. In $\mathbf{TS1_{AP}}$, the C8-O_p bond-forming distance is shortened to 1.95 ± 0.03 Å, while the O_p-O_d bond is slightly elongated to 1.23 ± 0.03 Å. Compared with these values in $\mathbf{TS1_{BS}}$ (1.90 ± 0.02 and 1.26 ± 0.03 Å), $\mathbf{TS1_{AP}}$ can be indicated as a more reactant-like transition state. Upon the *syn*-addition of $^1O_2$, the $\mathbf{DR_{AP}}$ intermediate is generated, with C8-O_p and O_p-O_d bond distances of 1.43 ± 0.03 Å and 1.30 ± 0.03 Å, respectively. These critical structural parameters are comparable to those in **DR** used as the initial structure. All computed $\langle S^2\rangle^{BS}$ values are ~1.0 throughout the simulation, confirming that the reaction follows the diradical pathway. Spin densities on the $^1O_2$ and guanine moieties also characterize the diradical nature of the three states.

The superior performance of AP–ULC–BLYP over BS–ULC–BLYP is found in the estimation of relative free energies. The activation and reaction free energies computed with BS–ULC–BLYP are 12.3 and 2.5 kcal mol^{-1}, whereas those with AP–ULC–BLYP are decreased to 7.1 and −5.6 kcal mol^{-1}. Clearly, the C8-O_p bond formation is found to be more facile and the exergonicity of the reaction is reproduced by eliminating the large effects of spin contamination.

2.3. Zwitterionic Pathway

The QM/MM-based ABF simulations for the zwitterionic pathway were started with the **ZW** intermediate shown in Figure 1, which has a closed-shell electronic structure ($\langle S^2\rangle^{BS} = 0$). The C8-$O_p$ distance was used for the reaction coordinate in line with the diradical pathway. Figure 3 depicts the 1D-PMFs corresponding to the zwitterionic pathway obtained with the QM/MM ABF simulations (QM=RLC–BLYP, BS–ULC–BLYP, and AP–ULC–BLYP).

The RLC–BLYP method provided the reaction composed of $\mathbf{RS_R}$, $\mathbf{TS2_R}$, and $\mathbf{ZW_R}$. In $\mathbf{RS_R}$, the C8···O_p separation is 2.58 ± 0.08 Å, which is comparable to the length of 2.66 ± 0.03 Å reported in a previous QM/MM MD study.[16] These values are shorter than those computed with BS–ULC–BLYP and AP–ULC–BLYP for the diradical pathway (~3.0 Å), but the O_p-O_d bond length of 1.20 ± 0.02 Å seems to support that the interaction between $^1O_2$ and guanine is negligibly small, The C8···O_p distance of 2.06 ± 0.08 Å in $\mathbf{TS2_R}$ is slightly longer than that for the *anti*-addition of $^1O_2$ in the B–DNA environment (~1.95 Å) reported by Dumont and co-workers.[16] The C8-O_p and O_p-O_d bond distances of 1.31 ± 0.02 Å and 1.46 ± 0.03 Å in $\mathbf{ZW_R}$ are compatible with the initial **ZW** structure (Figure 1). The activation and reaction free energies are calculated to be 2.7 and −26.7 kcal mol^{-1}, while those for the *anti*-addition are 6.2 and −25.0 kcal mol^{-1}.[16] It indicates that *syn*-addition is more favorable over *anti*-addition in the B–DNA environment. The result agrees well with a previous QM-only model calculation that employed a range of theoretical methods and showed that the *syn*-addition is favored by 4~6 kcal mol^{-1} over the *anti*-addition judging from the activation barrier.[19]

The BS–ULC–BLYP method using the *Guess=Mix* keyword (see Computational Method) gave only closed-shell solutions before the onset of the restricted/unrestricted instability, due to

© 2021 The Authors. ChemPhysChem published by Wiley-VCH GmbH

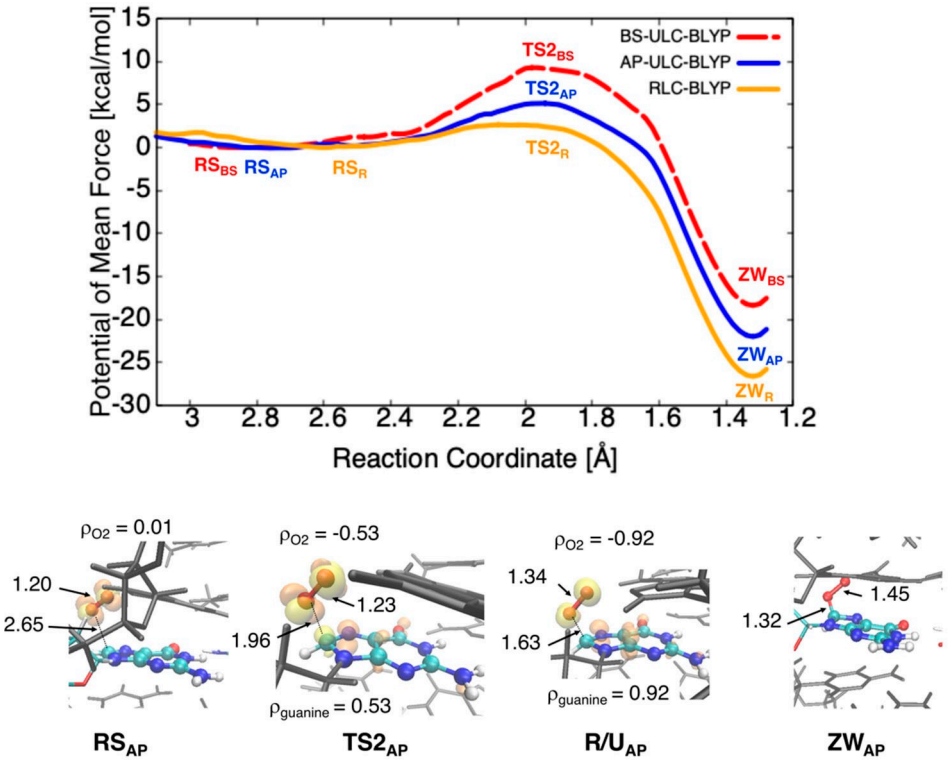


Figure 3. Top: 1D-PMFs for the zwitterionic pathway obtained with the RLC–BLYP/MM ABF, BS–ULC–BLYP/MM ABF, and AP–ULC–BLYP/MM ABF simulations. Bottom: Representative snapshots of $\mathbf{RS_{AP}}$, $\mathbf{TS2_{AP}}$, $\mathbf{R/U_{AP}}$, and $\mathbf{ZW_{AP}}$ calculated by the AP–ULC–BLYP/MM ABF simulations, with isovalue surfaces of positive (orange) and negative (yellow) spin density (isovalue = 0.01).

the closed-shell nature of **ZW** served as the starting point. As the sampling progressed, however, open-shell electronic structures having non-zero $\langle S^2\rangle^{BS}$ values gradually appeared. Such symmetry breaking led to a lowering of the potential energy due to spin contamination from the triplet ground state. With regard to the simulations using AP–ULC–BLYP, calculations for the triplet state were also conducted to compute the spin-projected energies in regions beyond the onset of the restricted/unrestricted instability. Then, every BS–ULC–BLYP/MM calculation yielded a broken-symmetry solution with complete diradical character ($\langle S^2\rangle^{BS} \sim 1.0$) in the vicinity of a region with a C8···O_p distance of ~1.6 Å (denoted as $\mathbf{R/U_{BS}}$ and $\mathbf{R/U_{AP}}$, Figure 3). It may be because the MM region also underwent gradual geometric changes so that it can fit into the open-shell electronic structure of the QM region. This behavior caused the resampling of the already sampled points, but no minimum corresponding to **DR** started to appear. It means that the **ZW** intermediate is much more stable than the **DR** intermediate, being in good agreement with previous studies using QM calculations.

Consequently, 1D-PMFs obtained using BS–ULC–BLYP and AP–ULC–BLYP methods resulted in having only one minimum corresponding to $\mathbf{ZW_{BS}}$ and $\mathbf{ZW_{AP}}$. On the basis of their QM-only model calculations, Thapa et al. indicated that a zwitterionic intermediate (virtually identical to **ZW**) can be formed directly via a transition state having a weak diradical character with $\langle S^2\rangle^{BS} = 0.34$ and the C8···O_p distance of 1.87 Å.[19,55]

Such behavior is not observed in our computed results. We find that the nature of electronic structure changed from open-shell to closed-shell in $\mathbf{R/U_{BS}}$ and $\mathbf{R/U_{AP}}$ after the diradical transition states ($\mathbf{TS2_{BS}}$, $\mathbf{TS2_{AP}}$). The C8···O_p distances in $\mathbf{TS2_{BP}}$ and $\mathbf{TS2_{AP}}$ (1.98 ± 0.05 and 1.98 ± 0.04 Å) are longer than that for the QM-only model. From this comparison, it can be argued that the explicit B–DNA environment is likely to stabilize the open-shell electronic structure and to lead to an earlier transition state. The simulations using BS–ULC–BLYP can reproduce the free energy profile qualitatively. Both the relative energies in $\mathbf{TS2_{BS}}$ (9.3 kcal mol^{-1}) and $\mathbf{ZW_{BS}}$ (−18.4 kcal mol^{-1}) are lower than those observed in $\mathbf{TS1_{BS}}$ (12.3 kcal mol^{-1}) and $\mathbf{DR_{BS}}$ (2.5 kcal mol^{-1}) for the diradical pathway. However, one must keep in mind that these values contain spin contamination. The computed activation barrier of 9.3 kcal mol^{-1} exceeds the experimentally observed range of 0~6 kcal mol^{-1} for reactions of singlet oxygen.[56]

The AP–ULC–BLYP-based simulations can remove the spin contamination effect, thereby decreasing the activation and reaction free energies to 5.2 and −22.0 kcal mol^{-1}. Unlike BS–ULC–BLYP, the activation free energy of 5.2 kcal mol^{-1} reproduces well the experimentally observed range (0~6 kcal mol^{-1}). Compared with the results for the diradical pathway, the zwitterionic pathway is found to be more exergonic, with a smaller activation barrier. As such, the zwitterionic pathway is expected to occur.

© 2021 The Authors. ChemPhysChem published by Wiley-VCH GmbH

2.4. Evaluation of Spin Contamination and Static Correlation

Comparison is made between the free energy profiles computed with RLC–BLYP, BS–ULC–BLYP, and AP–ULC–BLYP methods, to quantitatively evaluate the static correlation and spin contamination errors. The differences between the three QM methods in the ΔE_{ST} value for $^1O_2$ are reflected in the calculated activation and reaction free energies, supporting the aforementioned hypothesis. For the diradical pathway, spin contamination errors (BS–ULC–BLYP vs. AP–ULC–BLYP) on the activation and reaction free energies are calculated to be 5.2 and 8.1 kcal mol^{-1}, respectively. To characterize the spin contamination effect along the reaction coordinate, vertical ΔE_{ST} values in the three states were additionally evaluated on the AP–ULC–BLYP geometries taken from the snapshots depicted in Figure 2. The values calculated by BS–ULC–BLYP (AP–ULC–BLYP) are 10.9 (21.9), 5.8 (11.7), and 0.4 (0.8) kcal mol^{-1} for **RS$_{AP}$**, **TS1$_{AP}$**, and **DR$_{AP}$**, respectively (see also Table S1 in the Supporting Information). It underlines that the magnitude of spin contamination decreases in the order **RS$_{AP}$** > **TS1$_{AP}$** > **DR$_{AP}$**. The AP–ULC–BLYP method is needed to be used for sampling configurations close to **RS** and **TS1**, where the potential energies differ significantly with or without AP correction. Otherwise, the reaction is predicted to be endergonic, which is qualitatively inconsistent with previous experimental and computational findings.

Concerning the zwitterionic pathway, the spin contamination errors (BS–ULC–BLYP vs. AP–ULC–BLYP) on the activation and reaction free energies turn out to be 4.1 and 3.6 kcal mol^{-1}. These values are large, but less prominent than those for the diradical pathway. This is presumably due to the closed-shell nature and remarkable stability of **ZW**. As most of the sampled configurations were in the vicinity of **ZW**, the spin contamination effects in **RS** and **TS2** may not significantly affect the free energy landscape. A similar trend is observed in the static correlation errors (RLC–BLYP vs. AP–ULC–BLYP) on the activation and reaction free energies (i.e., 2.5 and 4.7 kcal mol^{-1}). The error on the activation barrier is smaller as compared to that arising from spin contamination. The present analysis apparently provides quantitative evidence that the 1D-PMF computed with RLC–BLYP can show better agreement with the experimental data, as suggested in previous computational studies.[13,16] Although one might argue that RLC–BLYP is computationally efficient for studying guanine oxidation, the small activation barrier stems from the fact that **RS$_R$** does not correspond to the correct energetic reference point (details in Supporting Information). As such, AP–ULC–BLYP is more reliable in that it is capable of characterizing the energetically correct reference state (**RS$_{AP}$**) and diradical **TS2$_{AP}$**, and of providing the smooth conversion from the open-shell surface to the closed-shell one.

2.5. Ring-closure Reaction

Then, we explored the subsequent intramolecular ring-closure reaction independently starting from **EP** presented in Figure 1.

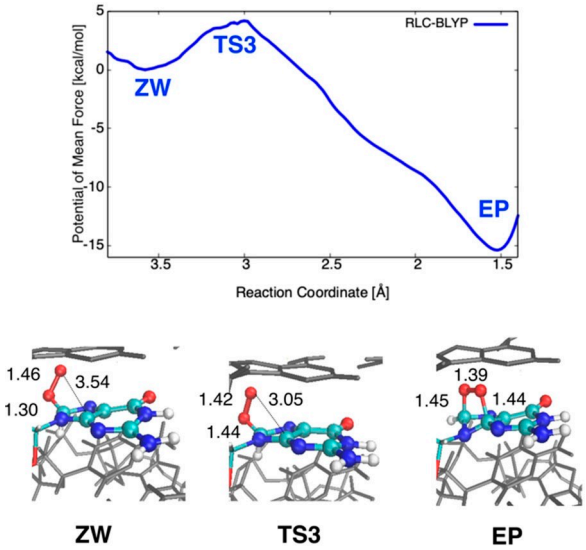


Figure 4. Top: 1D-PMFs for the ring closure reaction obtained with the RLC–BLYP/MM ABF simulations. Bottom: Representative snapshots of **ZW**, **TS3** and **EP**.

Since no open-shell species is considered to be involved in this process, the QM subsystem was treated by the RLC–BLYP/6-31G* method. The RLC–BLYP/MM-based ABF simulations were carried out using the reaction coordinate described by the C4-O_d distance. Figure 4 shows the obtained 1D-PMF. The obtained reactant state is essentially identical to **ZW$_R$**. We call the state **ZW** hereafter. Representative snapshots of the reactant state, transition state, and intermediate state (denoted as **ZW**, **TS3**, and **EP**) are also depicted.

The C4-O_d bond formation requires an activation barrier of 4.2 kcal mol^{-1} in **TS3**, where the C4···O_d is shortened to 3.03 ± 0.07 Å from 3.53 ± 0.04 Å in **ZW**. Similarly, the O_p-O_d distance is decreased to 1.43 ± 0.03 Å from 1.45 ± 0.04 Å, whereas the O8-O_d is stretched to 1.36 ± 0.03 Å from 1.31 ± 0.02 Å. A previous QM/MM MD study that used two C8-O_p and C4-O_d bonds successively as reaction coordinates predicted that the C4-O_d bond formation step is almost barrier-free with a large reaction free energy of −37.0 kcal mol^{-1}.[16] Meanwhile, our results suggest that formation of **ER** from **ZW** is slightly facile than the C8-O_p bond formation in terms of the smaller activation free energy (4.2 vs. 5.2 kcal mol^{-1}) and the resulting **EP** lies −15.4 kcal mol^{-1} below **ZW**. Nevertheless, our findings are consistent with the notion that the reaction proceeds via a stepwise mechanism, the first C8-O_p bond formation is the rate-limiting step, and the overall reaction is strongly exergonic (~ −37.4 kcal mol^{-1}).

3. Conclusions

In this study, cycloaddition reaction of $^1O_2$ with guanine in B–DNA is explored by means of QM/MM ABF calculations, to assess whether the reaction proceeds via either a zwitterionic or a diradical intermediate. We demonstrate how inclusion of

© 2021 The Authors. ChemPhysChem published by Wiley-VCH GmbH

static correlation via the broken-symmetry spin-unrestricted approach and removal of spin contamination by Yamaguchi's AP method improve the accuracy of free energies. Our simulations highlight that the free energy profile vary depending on correlation treatment in the QM region. The AP–ULC–BLYP calculation gives a realistic free energy profile by accounting for the effects of static correlation and spin contamination. It turns out that the static correlation errors on the activation and reaction free energies are 2.5 and 4.7 kcal mol^{-1} for the zwitterionic pathway, while spin contamination errors on these quantities range from 3.6 to 8.1 kcal mol^{-1}. These errors are sufficiently large not to be neglected. We emphasize that the AP method is, in particular, essential for investigating the diradical pathway, where the spin contamination effects are pronounced. According to the free energy profiles obtained with the accurate and robust QM-(AP–ULC–BLYP)/MM ABF simulations, the zwitterionic pathway is expected to occur. Our computed results match the experimental findings, with the notion that the reaction proceeds via a stepwise mechanism, the first C8-O_p bond formation is the rate-limiting step, and the overall reaction is strongly exergonic. The computed activation free energy of 5.2 kcal mol^{-1} shows agreement with the experimentally observed range (0 ~ 6 kcal mol^{-1}).

4. Computational Method

The X-ray crystallographic structure of the B–DNA dodecamer d(CGCGAATTCGCG)$_2$ was taken from the Protein Data Bank (PDB code: 1BNA, resolution 1.9 Å).[57] We constructed three QM/MM models (**DR**, **ZW**, and **EP**), each of which has the same QM region consisting of the second guanine from the 5′-end and $^1O_2$ (17 atoms). Hydrogen atoms were added by using the CHARMM-GUI input generator.[58] The N and C atoms in the N-glycosidic bond were set to the boundary between the QM and MM subsystems for both models as presented in Figure 1. The initial structures of the QM regions for **ZW** and **EP** were optimized at the RLC–BLYP/6-31G* level of theory, while the BS–ULC–BLYP/6-31G* model was used to optimize that for **DR**. Geometry optimizations for the QM subsystem were carried out by using the Gaussian 09 program package.[59] The systems were solvated within a 70 Å cubic box of TIP3P water molecules and neutralized by 22 Na^+ ions using the Visual Molecular Dynamics (VMD) program.[60] Then, classical 1 ns MD simulations were performed for each model in the NPT ensemble at 300 K with a time step of 2.0 fs by using NAMD.[61] A cut-off of 14 Å was employed with a switching for nonbonded interactions, and electrostatic interactions were treated by the particle mesh Ewald method.[62] During the MD simulations all the atoms in the QM region represented by the CHARMM General Force Field were kept fixed.[63] The CHARMM36 force field and TIP3P water models were exploited to describe the rest of the system.[64]

All QM/MM MD simulations were conducted with NAMD. The QM subsystem was described by the Gaussian 09 program package, whereas the MM subsystem was treated by the same parameters as in the classical MD simulations mentioned above. The energies of the open-shell singlet state were corrected using Yamaguchi's approximate spin projection (AP) method given by:[34]

$$E^{AP} = \alpha E^{BS} - (\alpha - 1)E^{T} \quad (1)$$

with

$$\alpha = \frac{\langle S^2 \rangle^{T}}{\langle S^2 \rangle^{T} - \langle S^2 \rangle^{BS}} \quad (2)$$

where E^{BS} and E^{T} represent the total energies for the broken-symmetry singlet and triplet states obtained with BS–ULC–BLYP. The $\langle S^2 \rangle^{BS}$ and $\langle S^2 \rangle^{T}$ are the corresponding expectation values of total spin angular momentum. We made a wrapper script that supports the interface to the Gaussian 09 program and adapts to the standardized NAMD QM/MM interface.[65] The forces on the MM point charges ($\mathbf{F}_{MM}$) derived from the interaction with the QM atoms were calculated by means of the electric field ($\mathbf{E}_{MM}$) at point charges (q_{MM}) as:

$$\mathbf{F}_{MM} = q_{MM}\mathbf{E}_{MM} \quad (3)$$

The *Charge*, *Density*, and *Prop = (Read, Field)* keywords were applied to calculate $\mathbf{E}_{MM}$ as with AMBER-Gaussian and DL_POLY-Gaussian interfaces.[66,67] The present NAMD-Gaussian interface gives the same results as the native NAMD-ORCA interface, provided that QM calculations are performed at the same level of theory (see Supporting Information including Table S2 for details).[68] All QM/MM MD simulations were conducted in the NVT ensemble with a time step of 0.5 fs. The link atom method was used with the Z3 scheme for manipulating MM point charges.[69]

In this study, we explored three reactions; (i) formation of **DR**, (ii) formation of **ZW**, and (iii) formation of **EP** via **ZW**. First, the three systems (**DR**, **ZW**, and **EP**) were equilibrated without any constraints for 2.0 ps, respectively. The RLC–BLYP/6-31G* method was employed for the closed-shell **ZW** and **EP**. The BS–ULC–BLYP/6-31G* method was used to compute **DR** with an open-shell character. Starting with the final snapshots obtained from the equilibration step (Figure 1), 1D-PMFs for the reactions (i)–(iii) were computed, coupled with the ABF method implemented as a part of the collective variables module of NAMD. The RLC–BLYP, BS–ULC–BLYP, and AP–ULC–BLYP methods were exploited as the QM regions for the reaction (ii). The QM region for the reaction (i) was treated by BS–ULC-LYP and AP–ULC–BLYP because the reaction takes place on the open-shell surface, whereas RLC–BLYP was sufficient for the reaction (iii) occurring on the closed-shell surface. The C8-O_p bond distance was chosen as the reaction coordinate for the reactions (i) and (ii), in a range from 1.36 to 3.10 Å, and from 1.28 to 3.10 Å, respectively. The reaction coordinate for the reactions (iii) was set to the C4-O_d bond distance ranging from 1.40 to 3.80 Å. For all cases, the bin width was set to 0.02 Å, and a threshold of 100 samples prior to the application of the bias was used. Concerning the reaction (ii), the restricted/unrestricted instability necessarily takes place, since the reactant $^1O_2$

© 2021 The Authors. ChemPhysChem published by Wiley-VCH GmbH

is a diradical and the **ZW** intermediate is closed-shell. Thus, during the ABF simulation, the initial guess for every QM step was generated by using the *Guess=Mix* keyword (see also Supporting Information). ABF simulations were run for 20~40 ps in total (i.e., 40,000~80,000 QM calculations).

Acknowledgements

T.S. acknowledges the fund for the Promotion of Joint International Research (Fostering Joint International Research (B)) from the Japan Society for the Promotion of Science (JSPS) (No. 18KK0194). Y.T acknowledges the fund for the Grants-in-Aid for Scientific Research (C) from the JSPS (No. 19K06589). Y.T. is also grateful to the Ministry of Education, Culture, Sports, Science and Technology (MEXT) for a Grant-in-Aid for Scientific Research on Innovative Areas "3D active-site science" (No. 26105012) and Transformative Research Areas (A) "Progressive condensed matter physics inspired by hyper-ordered structures" (No. 20H05883). Some computations were performed using Research Center for Computational Science, Okazaki, Japan.

Conflict of Interest

The authors declare no conflict of interest.

Keywords: oxidation reaction · QM/MM MD · singlet oxygen · DNA · spin contamination

[1] P. Di Mascio, G. R. Martinez, S. Miyamoto, G. E. Ronsein, M. H. G. Medeiros, J. Cadet, *Chem. Rev.* **2019**, *119*, 2043–2086.
[2] Y. Nosaka, A. Nosaka, *Chem. Rev.* **2017**, *117*, 11032–11336.
[3] A. A. Frimer, *Chem. Rev.* **1979**, *79*, 359–387.
[4] A. A. Ghogare, A. Greer, *Chem. Rev.* **2016**, *116*, 9994–10034.
[5] S. Tada-Oikawa, S. Oikawa, J. Hirayama, K. Hirakawa, S. Kawanishi, *Photochem. Photobiol.* **2009**, *85*, 1391–1399.
[6] W. P. Roos, A. D. Thomas, B. Kaina, *Nat. Rev. Cancer* **2016**, *16*, 20–33.
[7] T. P. Devasagayam, S. Steenken, M. S. Obendorf, W. A. Schulz, H. Sies, *Biochemistry* **1991**, *30*, 6283–6289.
[8] C. Sheu, C. S. Foote, *J. Am. Chem. Soc.* **1993**, *115*, 10446–10447.
[9] I. Girault, D. Molko, J. Cadet, *Free Radical Res.* **1994**, *20*, 315–325.
[10] Y. Ye, J. G. Muller, W. Luo, C. L. Mayne, A. J. Shallop, R. A. Jones, C. J. Burrows, *J. Am. Chem. Soc.* **2003**, *125*, 13926–13927.
[11] J. Cadet, J.-L. Ravanat, G. R. Martinez, M. H. G. Medeiros, P. Di Mascio, *Photochem. Photobiol.* **2006**, *82*, 1219–1225.
[12] J. Cadet, T. Douk, J.-L. Ravanat, *Acc. Chem. Res.* **2008**, *41*, 1075–1083.
[13] W. Lu, J. Liu, *Chem. Eur. J.* **2016**, *22*, 3127–3138.
[14] R. Grüber, A. Monari, E. Dumont, *J. Phys. Chem. A* **2014**, *118*, 11612–11619.
[15] B. Marchetti, T. N. V. Karsili, *Chem. Commun.* **2016**, *52*, 10996–10999.
[16] E. Dumont, R. Greber, E. Bignon, C. Morell, Y. Moreau, A. Monari, J.-L. Ravanat, *Nucleic Acids Res.* **2016**, *44*, 56–62.
[17] E. Dumont, R. Greber, E. Bignon, C. Morell, J. Aranda, J. L. Ravanat, I. Tuñón, *Chem. Eur. J.* **2016**, *22*, 12358–12362.
[18] Y. Sun, W. Lu, J. Liu, *J. Phys. Chem. B* **2017**, *121*, 956–966.
[19] B. Thapa, B. H. Munk, C. J. Burrows, H. B. Schlegel, *Chem. Res. Toxicol.* **2016**, *29*, 1396–1409.
[20] B. Thapa, B. H. Munk, C. J. Burrows, H. B. Schlegel, *Chem. Eur. J.* **2017**, *22*, 12358–12362.
[21] B. O. Roos, K. Andersson, *Chem. Phys. Lett.* **1995**, *245*, 215–223.
[22] C. Møller, M. S. Plesset, *Phys. Rev.* **1934**, *46*, 618–622.
[23] T. H. Dunning Jr. *J. Chem. Phys.* **1989**, *90*, 1007–1023.
[24] A. Maranzana, G. Ghigo, G. Tonachini, *J. Am. Chem. Soc.* **2000**, *122*, 1414–1423.
[25] F. Sevin, M. L. McKee, *J. Am. Chem. Soc.* **2001**, *123*, 4591–4600.
[26] A. Leach, K. N. Houk, *Chem. Commun.* **2002**, 1243–1255.
[27] T. Darden, D. York, L. Pedersen, *J. Chem. Phys.* **1993**, *98*, 10089–10092.
[28] T. Saito, S. Nishihara, Y. Kataoka, Y. Nakanishi, T. Matsui, Y. Kitagawa, T. Kawakami, M. Okumura, K. Yamaguchi, *Chem. Phys. Lett.* **2009**, *483*, 168–171.
[29] T. Saito, S. Nishihara, Y. Kataoka, Y. Nakanishi, Y. Kitagawa, T. Kawakami, M. Okumura, K. Yamaguchi, *J. Phys. Chem. A* **2010**, *114*, 7967–7974.
[30] R. J. Bartlett, *Annu. Rev. Phys. Chem.* **1981**, *32*, 359–401.
[31] J. F. Stanton, *J. Chem. Phys.* **1994**, *101*, 4371–4374.
[32] A. I. Krylov, *J. Chem. Phys.* **2000**, *113*, 6052–6062.
[33] C. Sosa, H. B. Schlegel, *Int. J. Quantum Chem.* **1986**, *29*, 1001–1015.
[34] K. Yamaguchi, F. Jensen, A. Dorigo, K. N. Houk, *Chem. Phys. Lett.* **1988**, *149*, 537–542.
[35] C. A. Jiménez-Hoyos, T. M. Henderson, T. Tsuchimochi, G. E. Scuseria, *J. Chem. Phys.* **2012**, *136*, 164109.
[36] A. V. Marenich, J. C. Cramer, D. G. Truhlar, *J. Phys. Chem. B* **2009**, *113*, 6378–6396.
[37] C. Angeli, R. Cimiraglia, S. Evangelisti, T. Leininger, J.-P. Malrieu, *J. Chem. Phys.* **2001**, *114*, 10252–10264.
[38] E. Darve, D. Rodoríguez-Gómez, A. Pohorille, *J. Chem. Phys.* **2008**, *128*, 144120.
[39] A. D. Becke, *Phys. Rev. A* **1988**, *38*, 3098–3100.
[40] C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B* **1988**, *37*, 785–789.
[41] A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648–5652.
[42] H. Iikura, T. Tsuneda, T. Yanai, K. Hirao, *J. Chem. Phys.* **2001**, *115*, 3540–3544.
[43] W. J. Hehre, R. Ditchfield, J. A. Pople, *J. Chem. Phys.* **1972**, *56*, 2257–2261.
[44] P. C. Hariharan, J. A. Pople, *Theor. Chim. Acta* **1973**, *28*, 213–222.
[45] D. Roca-Sanjuan, F. A. Aquilante, R. Lindh, *WIREs Comput. Mol. Sci.* **2012**, *2*, 585–603.
[46] W. Thiel, *WIREs Comput. Mol. Sci.* **2014**, *4*, 145–157.
[47] T. Saito, Y. Kitagawa, Y. Takano, *J. Phys. Chem. A* **2016**, *120*, 8750–8760.
[48] M. Alfonso-Prieto, X. Biarnés, P. Vidossich, C. Rovira, *J. Am. Chem. Soc.* **2009**, *131*, 11751–11761.
[49] H. Zhou, B. Wang, F. Wang, X. Yu, L. Ma, A. Li, M. T. Reetz, *Angew. Chem.* **2019**, 131, 774–778; *Angew. Chem. Int. Ed.* 2019, 58, 764–768.
[50] B. Wang, Z. Cao, C. Rovira, J. Song, S. Shaik, *J. Am. Chem. Soc.* **2019**, *141*, 9284–9291.
[51] L. V. Slipchenko, A. I. Krylov, *J. Chem. Phys.* **2002**, *117*, 4694–4708.
[52] J. Lee, M. Head-Gordon, *J. Chem. Phys.* **2019**, *150*, 244106.
[53] T. Saito, Y. Takano, *Chem. Lett.* **2019**, *48*, 1441–1444.
[54] D. H. Ess, E. R. Johnson, X. Hu, W. Yang, *J. Phys. Chem. A* **2011**, *115*, 76–83.
[55] D. H. Ess, T. C. Cook, *J. Phys. Chem. A* **2012**, *116*, 4922–4929.
[56] The BS-UB3LYP/6-31+G(d,p) method coupled with the SMD model was exploited.
[57] W. T. Yang, D. B. Min, *Lipids in Food Flavors* **2009**, *2*, 15–29.
[58] H. R. Drew, R. M. Wing, T. Takano, C. Broka, S. Tanaka, K. Itakura, R. E. Dickerson, *Proc. Natl. Acad. Sci. USA* **1981**, *78*, 2179–2183.
[59] J. Lee, X. Cheng, J. M. Swails, M. S. Yeom, P. K. Eastman, J. A. Lemkul, S. Wei, J. Buckner, J. C. Jeong, Y. Qi, S. Jo, V. S. Pande, D. A. Case, C. L. Brooks, III., A. D. MacKerell, Jr., J. B. Klauda, W. Im, *J. Chem. Theory Comput.* **2016**, *12*, 405–413.
[60] *Gaussian 09 (Revision D.01)*, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, B. G. Janesko, F. Lipparini, G. Zheng, J. L. Sonnenberg, W. Liang, M. Hada, M. Ehara, R. F. K. Toyota, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. J. A. Montgomery, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, P. V. Parandekar, N. J. Mayhall, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, D. J. Fox, Gaussian, Inc., Wallingford CT, **2009**.
[61] K. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graphics* **1996**, *14*, 33–38.

© 2021 The Authors. ChemPhysChem published by Wiley-VCH GmbH

[62] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, K. Schulten, *J. Comput. Chem.* **2005**, *26*, 1781–1802.
[63] T. Darden, D. York, L. Pedersen, *J. Chem. Phys.* **1993**, *98*, 10089–10092.
[64] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, A. D. Mackerell Jr., *J. Comput. Chem.* **2010**, *31*, 671–690.
[65] J. Huang, A. D. MacKerell, Jr., *J. Comput. Chem.* **2013**, *34*, 2135–2145.
[66] M. C. R. Melo, R. C. Bernardi, T. Rudack, M. Scheurer, C. Riplinger, J. C. Phillips, J. D. C. Maia, G. B. Rocha, J. V. Ribeio, J. E. Stone, F. Neese, K. Schulten, Z. Luthey-Schulten, *Nat. Methods* **2018**, *15*, 351–354.
[67] T. Okamoto, K. Yamada, Y. Koyano, T. Asada, N. Koga, M. Nagaoka, *J. Comput. Chem.* **2011**, *32*, 932–942.
[68] T. Saito, W. Thiel, *J. Phys. Chem. B* **2014**, *118*, 5034–5043.
[69] F. Neese, *WIREs Comput. Mol. Sci.* **2018**, *8*, e1327.
[70] H. Lin, D. G. Truhlar, *J. Phys. Chem. A* **2005**, *109*, 3991–4004.

Manuscript received: December 1, 2020
Revised manuscript received: January 17, 2021
Accepted manuscript online: January 18, 2021
Version of record online: February 12, 2021

© 2021 The Authors. ChemPhysChem published by Wiley-VCH GmbH