DNA Repair

DOI: 10.1002/anie.201403334

Ribose-Protonated DNA Base Excision Repair: A Combined Theoretical and Experimental Study**

Keyarash Sadeghian, Denis Flaig, Iris D. Blank, Sabine Schneider, Ralf Strasser, Dimitrios Stathis, Malte Winnacker, Thomas Carell, and Christian Ochsenfeld**

Abstract: *Living organisms protect the genome against external influences by recognizing and repairing damaged DNA. A common source of gene mutation is the oxidized guanine, which undergoes base excision repair through cleavage of the glycosidic bond between the ribose and the nucleobase of the lesion. We unravel the repair mechanism utilized by bacterial glycosylase, MutM, using quantum-chemical calculations involving more than 1000 atoms of the catalytic site. In contrast to the base-protonated pathway currently favored in the literature, we show that the initial protonation of the lesion's ribose paves the way for an almost barrier-free glycosidic cleavage. The combination of theoretical and experimental data provides further insight into the selectivity and discrimination of MutM's binding site toward various substrates.*

While DNA repair is of central importance for the stability of the genome of organisms,[1,2] the underlying processes such as lesion recognition and repair are still poorly understood. Repair enzymes have to recognize single lesions within a vast majority of undamaged bases. The most prominent endogenous DNA damage is the oxidized guanine, 7,8-dihydro-8-oxoguanine (8OG). 8OG has the ability to form a Hoogsteen base pair with adenine, which is the basis for the G:C→A:T transversion mutation (Figure 1).[3]

Despite the existence of crystal structures showing 8OG-containing DNA in complex with the catalytically incompetent repair enzymes, one of the major unsolved questions is how 8OG is specifically recognized and excised in the genome.[4–12] It is still not clear how the enzymes are able to exquisitely select 8OG and avoid harmful excision of canonical guanidine (G). In bacteria, 8OG is repaired by MutM, also known as formamidopyrimidine glycosylase (Fpg).[13] MutM first recognizes 8OG and then guides it into the active site (lesion recognition complex, LRC).[14] The glycosidic bond is subsequently cleaved, creating an abasic intermediate. A crystal structure trapped after sodium borohydride treatment reveals that the enzyme, at some point, forms a covalent bond (Schiff base) between the anomeric center of 8OG and a proline residue in the active site.[15] However, the chemistry leading to this intermediate and contributing to the ability of the enzyme to discriminate G from 8OG is still unknown.

The currently favored hypothesis is that MutM protonates the 8OG base, thereby converting it into a good leaving group, which facilitates the nucleophilic attack of the anomeric C1′ carbon by the proline nucleophile.[16] The heterocycle of 8OG is more electron-rich than that of G, hence making the former a better leaving group. Using newly developed linear-scaling quantum-chemical methods we report a thorough computational investigation of the repair mechanism of MutM. We address MutM's exquisite 8OG selectivity as well as the specific catalytic role of central amino acids in the active site. The results of our theoretical study are in full agreement with the presented experimental repair studies with different G derivatives and a new crystal structure showing 8OG embedded in a DNA duplex in complex with a wild-type and hence catalytically competent MutM (for details see the Supporting Information, SI). The theoretical treatment of the complex DNA-repair mechanism is a severe challenge for quantum chemical calculations, forcing us to utilize newly developed linear-scaling quantum-chemical methods.[17,18] These new methods enable us to calculate interaction energies with more than 1000 atoms of the active site and to study the full repair pathway with more than 500 atoms in the QM region of the QM/MM (quantum mechanics/molecular mechanics) approach. Such large QM regions have recently been shown to be crucial for reliable QM/MM results in other molecular systems.[19–23] The linear-scaling QM/MM calculations make it possible for the first time to clarify the repair mechanism and provide insight into the interaction energies between MutM and various substrates. We have used the dispersion-corrected BLYP-D3 density funtional,[24–26] the split-valence polarized (SVP) basis set,[27] and the AMBER force field[28] to study the reaction pathways. The dispersion-corrected Hartree–Fock approximation was used for the QM part of the QM/MM interaction energy calculations (for full details see SI-2).

[*] Dr. K. Sadeghian, Dr. D. Flaig, I. D. Blank, Prof. Dr. C. Ochsenfeld
Chair of Theoretical Chemistry, Department of Chemistry
University of Munich (LMU)
Butenandtstrasse 7, 81377 Munich (Germany)
Dr. K. Sadeghian, Dr. D. Flaig, I. D. Blank, Dr. S. Schneider, Dr. R. Strasser, Dr. D. Stathis, Dr. M. Winnacker, Prof. Dr. T. Carell, Prof. Dr. C. Ochsenfeld
Center for Integrated Protein Science (CIPSM) at the Department of Chemistry, University of Munich (LMU)
Butenandtstrasse 5–13, 81377 Munich (Germany)
E-mail: thomas.carell@uni-muenchen.de
christian.ochsenfeld@uni-muenchen.de
Dr. S. Schneider
Department of Chemistry, Chair of Biochemistry
TU München, Lichtenbergstrasse 4, 85747 Garching (Germany)

[**] We acknowledge financial support by the DFG funding initiatives SFB749 (TP A4 and C7) and the Excellence Cluster EXC114 (CIPSM).

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201403334.

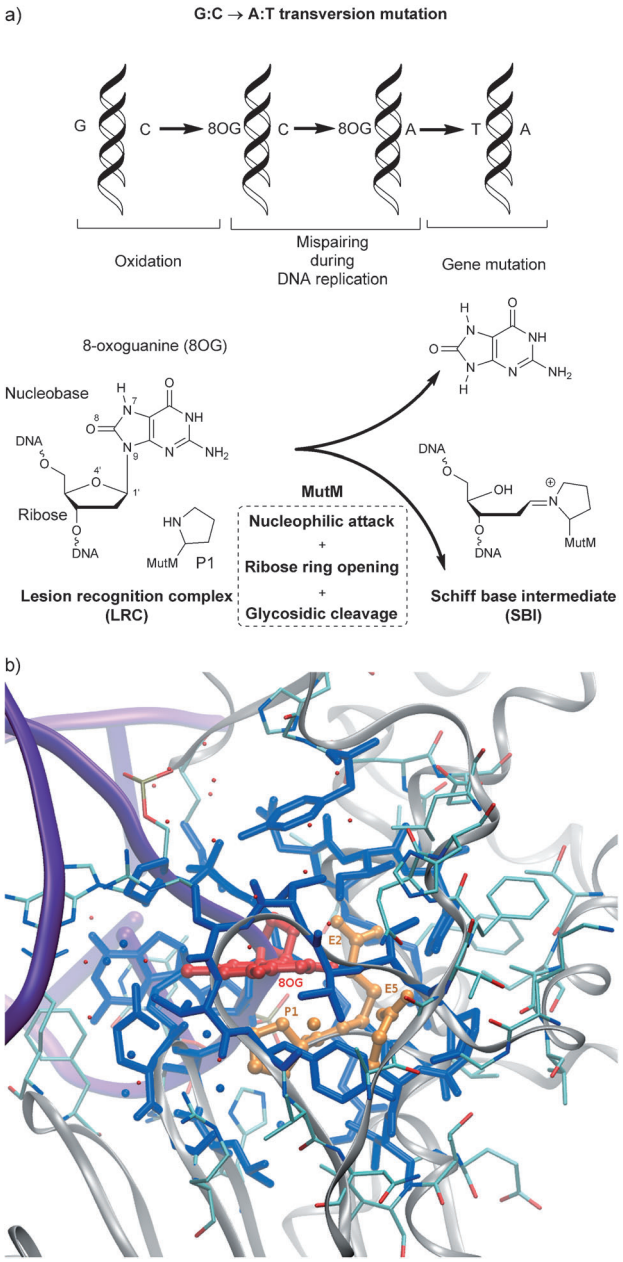

Figure 1. a) Top: 8OG can form a Hoogsteen base pair with A during DNA replication leading to G:C→A:T transversion mutations. Bottom: schematic representation of the glycosylase activity of MutM connecting the LRC and SBI states. b) A close-up view of the MutM–DNA complex showing the QM regions used in our QM/MM calculations. The QM region used to investigate the repair pathway includes not only 8OG (red) and the P1, E2, and E5 (orange) of MutM, but also the catalytic water (W) and all residues (blue) directly interacting with these fragments (→ QM region = 509 atoms). For the calculations of the QM/MM interaction energies, we have extended the QM region to 1115 atoms, which are shown in green. Hydrogens are omitted for clarity. DNA and protein backbone are shown in pink and silver ribbon representations, respectively.

The currently preferred nucleobase-protonated pathway involves the initial cleavage of the glycosidic bond after protonation/H-bonding of the 8OG base (Figure 2a). This protonation increases the leaving group activity of the nucleobase. The subsequently formed oxycarbenium ion is thought to be stabilized by the E2 residue. The essential proline nucleophile, P1, then reacts with the oxycarbenium ion as depicted in Figure 2a to form an imine that was trapped using sodium borohydride. We denote the intermediates of this mechanism as BA1 and BA2. The final product is the Schiff base intermediate (SBI), which includes the O8-protonated tautomer (SBI-O8p). Although this mechanism is in principle chemically feasible, our QM/MM data, shown in Figure 2b, mark this mechanism as surprisingly unlikely: The highest energy point found along this pathway is 71 kcal mol^{-1} above the LRC. We also confirmed the BLYP-D3/MM energetics by comparison to single-point RIMP2/MM and B2PLYP/MM results (see SI-3). In addition, we find that the critical intermediates BA1 and BA2, featuring a bond between P1 and the ribose, lie energetically so high (45 and 34 kcal mol^{-1} above the LRC, respectively) that both of them are thermodynamically not accessible. The 8OG base may assist the proton transfer required to open up the ribose ring by temporarily accepting the proton of P1 (→ barrierless formation of the BA2 state). However, also the SBI-O8p state is energetically so unfavorable (+25 kcal mol^{-1}) that we have to exclude a glycosylase-driven repair reaction that is initiated by conversion of 8OG into a better leaving unit. For a brief comparison between the currently favored 8OG repair mechanism and the repair of uracil by the DNA uracil glycosylase[29] see SI-3.

In light of the high energies observed for the intermediates of the base-protonated pathway, we investigated other reaction pathways. We considered a mechanism, in part proposed earlier,[30] in which the nucleobase does not function as an electron-accepting unit (after protonation) but as an electron-donating substructure (Figure 2a). This ribose-protonated pathway involves the opening of the ribose as the first step. The latter is facilitated by protonated E2, a known critical residue in the active site. The ribose ring-opening is accompanied by the nucleophilic attack of P1 leading to the first intermediate, RA1 (Figure 2a). P1 may undergo deprotonation, forming the second intermediate RA2, followed by the reprotonation of E2 (→RA3). The final step in this pathway is the cleavage of the glycosidic bond to finally give the iminium ion (SBI-N9).

In contrast to the first mechanism, in the ribose-centered scenario the intended glycosidic cleavage takes place at the very last stage. The QM/MM data show clearly that this pathway is energetically more favorable throughout, with an indeed almost barrierless glycosidic bond cleavage reaction. The highest energy point on this path is only 13 kcal mol^{-1} above the LRC and involves a proton transfer from P1 to the O8 position of 8OG. Also the final SBI state (SBI-N9p) is 17 kcal mol^{-1} below the LRC, implying that the overall reaction along the ribose-protonated pathway is thermodynamically significantly more favored than the nucleobase-protonated one. The mechanism also explains the eminent function of E2 in the active site. Site-directed mutagenesis showed that in the E2Q mutant of MutM the repair function is fully abolished.[14] In the ribose-protonated mechanism the protonated form of E2 is essential for enabling ribose ring opening, hence allowing the nucleophilic attack of P1.

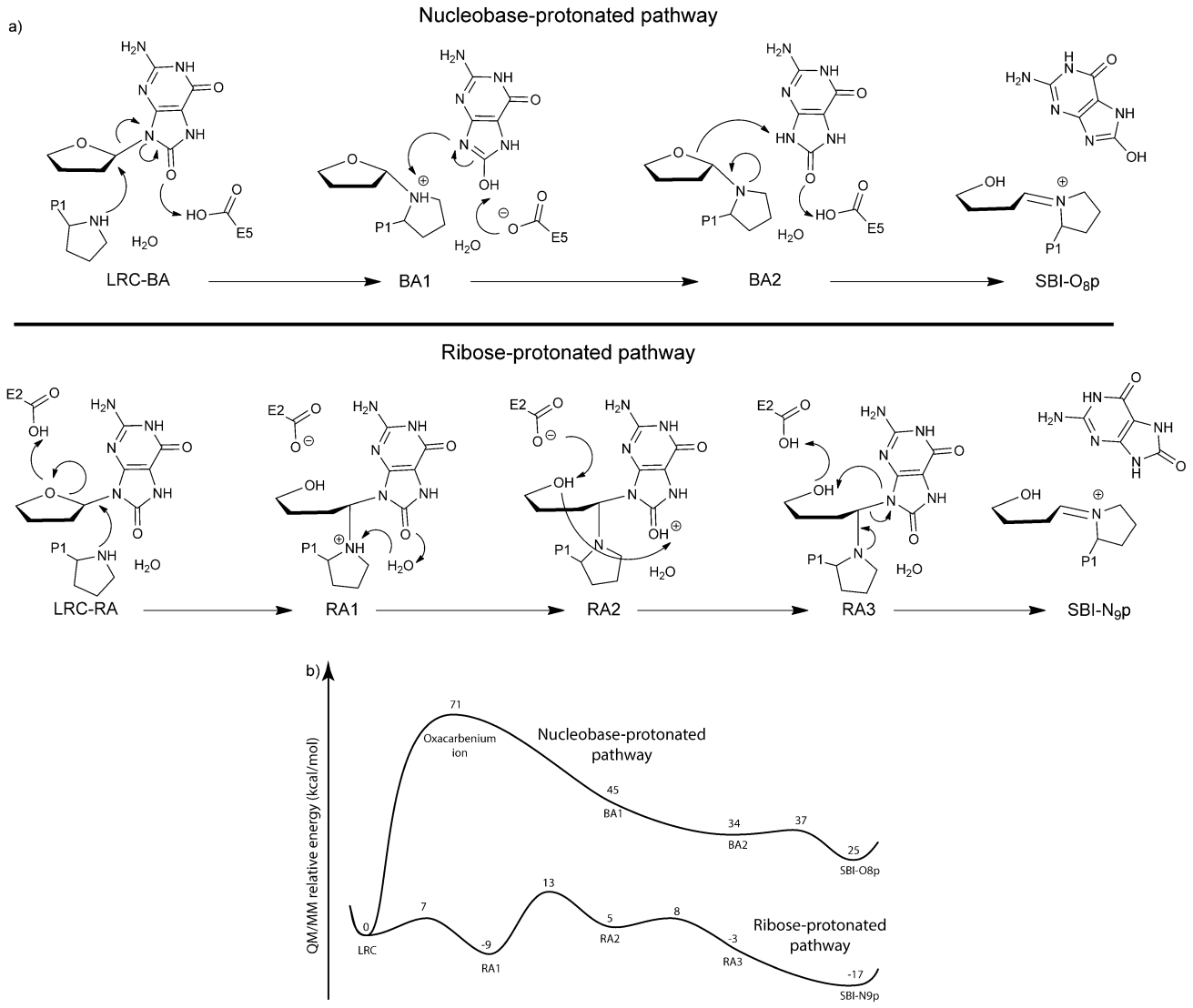

Figure 2. Overview of the nucleobase- and ribose-protonated pathways. a) Schematic representation of the studied mechanisms. b) Our QM/MM data obtained for both pathways. The QM/MM (BLYP-D3/AMBER) structure optimizations performed to obtain these QM/MM relative energies (with respect to the LRC) involve 509 atoms in the QM region. On the base-protonated pathway, we locate a potential oxacarbenium ion, which also has been discussed as a reaction intermediate in the context of the base excision repair by uracil DNA glycosylase.[29] For 8OG repair, however, we find this ion to be a very high-energy transition state. Furthermore, we observe that a single water molecule, directly interacting with P1, can play a crucial catalytic role in the ribose-protonated pathway. It stabilizes the RA1 intermediate (Figure SI-16) and mediates the proton transfer to make the RA2 formation more feasible. The presence of such a water molecule replacing the R79 residue is motivated in Section SI-1. The influence of zero-point energy correction is found to be minor (see SI-4).

The rate-determining step of the ribose-protonated mechanism is the proton transfer step connecting the RA1 (-9 kcal mol^{-1}) and RA2 (5 kcal mol^{-1}) structures. As evident from Figure 2a, the energetically highest point on this path is 13 kcal mol^{-1} above the LRC (i.e. 22 kcal mol^{-1} above RA1) and connects the RA1 and RA2 intermediates. At the first glance, these values may seem high compared with metabolic enzymes; however, as pointed out by O'Brien,[31] DNA repair operates under different kinetics: it is a rare event compared with the synthesis of relatively large amounts of substrates processed by metabolic pathways.

The ribose-protonated pathway presented here in full detail revives an idea proposed earlier,[30] which seems to receive no attention in recent literature. A possible reason for neglecting this idea is that a base-invariant mechanism implies that any nucleobase (even G) that can find its way into the active site can, in principle, be cleaved from DNA. For this to happen, G must first successfully pass through both the base-encounter and the base-extrusion steps before reaching the base-stabilization stage (LRC in the case of 8OG). However, experimental and theoretical studies[7,8] suggest that MutM discriminates G already at the early encounter and extrusion stages. Our results, presented further below, strongly support the idea that MutM can also differentiate 8OG from G, even at the final LRC stage.

We have chosen a combination of experimental and computational approaches to probe the extent of MutM's active site discrimination towards various substrates. On the

experimental side, we synthesized a series of G derivatives with different substituents at C8. The chemical groups introduced at this position are expected to influence the electronic density in the vicinity of the glycosidic bond. The structures of these G derivatives and the repair data obtained with short single-stranded oligonucleotides, in which the 8OG analogues were incorporated, are depicted in Figure 3c. Our data show fast repair of the 8OG-containing oligonucleotide (Figure 3a, lane 1). It is clearly visible that the 8-Br-G and canonical G are not repaired (lanes 2 and 4). Interestingly, we observe a fast repair of the NH_2-substituted 8OG analogue (lane 3). The electron-donating amino group increases the electron density of the G derivative thereby converting the base into a better electron donor as needed for the ribose-protonated mechanism. Apparantly the 8-Br-G, which is even less electron-rich than G, is not repaired because it is an inadequate electron donor to support the ribose-protonated process. Besides these electronic properties, the binding of the compounds into the active site also needs to be considered.

To address the issue of MutM discrimination computationally, we performed QM/MM calculations, which allow us to compare the interaction of MutM with the experimentally investigated substrates. Our HF-D3/MM calculations show that G is discriminated at the LRC stage by 20 kcal mol^{-1}. The mean absolute error of the HF-D3 method has been benchmarked to below 1 kcal mol^{-1}.[32] To check the convergence of the obtained value with respect to QM size (Figure 3b), we included up to 1115 atoms into the QM region (see SI-2). To converge the G discrimination within 2 kcal mol^{-1} with respect to the QM size, one must include at least 600 atoms into the QM region. QM-only calculations exhibit a much slower convergence as depicted in Figure 3b, so that even more atoms need to be included in the QM sphere for converged results. Our calculations indicate that the S219 residue, in H-bonding interaction with the substrate, is mainly responsible for the discrimination.

Furthermore, we have repeated these QM/MM calculations by replacing 8OG with the 8-NH_2-G and 8-Br-G substrates, thereby assuming the *syn* binding mode as already observed for 8OG (see SI-5 for further details). The QM/MM calculations yield very similar binding affinities for 8-NH_2-G and 8OG (Figure 3). The 8-Br-G substrate on the other hand is 15 kcal mol^{-1} less stable than 8OG and therefore less likely to be bound to MutM in the same orientation as 8OG. Overall, MutM seems to be selective when it comes to stabilizing substrates in its binding site. To gain deeper insight into the exact arrangements of the important residues E2 and E5 relative to the lesion, we solved a crystal structure of MutM in complex with carbocyclic 8OG-containing DNA (PDB: 4CIS; see SI-7 for further details).

In conclusion, our linear-scaling QM/MM calculations with large QM regions (more than 500 atoms for the pathway and more than 1000 atoms in computing interaction energies) unravel the base excision repair mechanism for the glycosylase activity of the bacterial DNA repair enzyme, MutM. We show that through the opening of the ribose ring of 8OG, the final cleavage of the glycosidic bond becomes an almost barrier-free process. Our proposed mechanism is in full agreement with the experimental observations, in which the

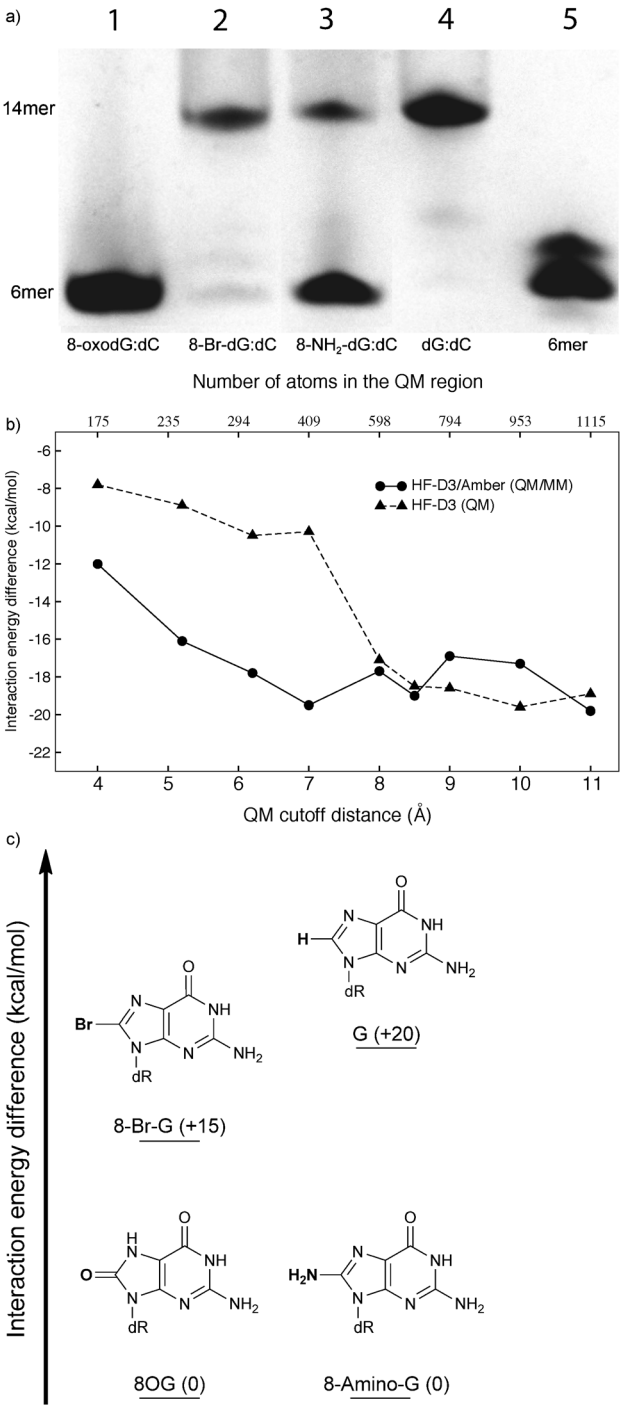

Figure 3. Repair activity (experimental data) and binding discrimination (theoretical data) of MutM toward different substrates. a) Activity assay of MutM with C8-modified nucleobases: 1) 8OG:C, 2) 8-Br-G:C, 3) 8-NH_2-G:C, 4) G:C, and 5) single-stranded 6-mer DNA (see SI-6). b) Convergence of the interaction energy difference (kcal mol^{-1}) between 8OG versus G with the protein environment (negative values indicate the preference of binding of MutM toward 8OG). The fluctuations (in the order of 2 kcal mol^{-1}) observed in the QM/MM values for $R > 8$ Å are due to the inclusion of charged fragments in the QM region and their interaction with the remaining point-charge field (see SI-5 for the list of fragments). Such a long-range effect on the calculated QM/MM energies emphasizes the choice of sufficiently large QM regions in QM/MM calculations. c) QM/MM MutM–DNA interaction energy difference for different G derivatives, with 8OG–MutM taken as reference.

modifications of the ribose ring or mutations of the amino acids directly interacting with the ribose prohibit the glycosylase activity of MutM. We have computed the extent of discrimination by MutM towards different substrates by calculating the total interaction energies. We have shown that the catalytic site of MutM imposes considerable discrimination towards G, hence providing further support for the proposed ribose-protonated mechanism. This pathway is in agreement with the experimental activity data demonstrating MutM's selectivity in binding and repair affinity.

Received: March 14, 2014
Revised: May 12, 2014
Published online: July 25, 2014

Keywords: DNA repair · molecular modeling · oxidative damage · quantum chemistry · ribose

[1] E. C. Friedberg, *Nature* **2003**, *421*, 436.
[2] T. Lindahl, R. D. Wood, *Science* **1999**, *286*, 1897.
[3] T. Lindahl, *Nature* **1993**, *362*, 709.
[4] J. C. Fromme, G. L. Verdine, *Nat. Struct. Biol.* **2002**, *9*, 544.
[5] P. C. Blainey, A. M. van Oijen, A. Banerjee, G. L. Verdine, X. S. Xie, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 5752.
[6] A. Banerjee, W. L. Santos, G. L. Verdine, *Science* **2006**, *311*, 1153.
[7] Y. Qi, M. C. Spong, K. Nam, A. Banerjee, S. Jiralerspong, M. Karplus, G. L. Verdine, *Nature* **2009**, *462*, 762.
[8] Y. Qi, M. C. Spong, K. Nam, M. Karplus, G. L. Verdine, *J. Biol. Chem.* **2010**, *285*, 1468.
[9] K. Nam, G. L. Verdine, M. Karplus, *J. Am. Chem. Soc.* **2009**, *131*, 18208.
[10] A. Banerjee, W. Yang, M. Karplus, G. L. Verdine, *Nature* **2005**, *434*, 612.
[11] S. D. Bruner, D. P. Norman, G. L. Verdine, *Nature* **2000**, *403*, 859.
[12] J. C. Fromme, S. D. Bruner, W. Yang, M. Karplus, G. L. Verdine, *Nat. Struct. Biol.* **2003**, *10*, 204.
[13] S. S. David, V. L. O'Shea, S. Kundu, *Nature* **2007**, *447*, 941.
[14] J. C. Fromme, G. L. Verdine, *J. Biol. Chem.* **2003**, *278*, 51543.
[15] R. Gilboa, D. O. Zharkov, G. Golan, A. S. Fernandes, S. E. Gerchman, E. Matz, J. H. Kycia, A. P. Grollman, G. Shoham, *J. Biol. Chem.* **2002**, *277*, 19811.
[16] P. J. Berti, J. A. B. McCann, *Chem. Rev.* **2006**, *106*, 506.
[17] J. Kussmann, M. Beer, C. Ochsenfeld, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2013**, *3*, 614 – 636.
[18] J. Kussmann, C. Ochsenfeld, *J. Chem. Phys.* **2013**, *138*, 134114.
[19] C. V. Sumowski, C. Ochsenfeld, *J. Phys. Chem. A* **2009**, *113*, 11734.
[20] L. Hu, P. Söderhjelm, U. Ryde, *J. Chem. Theory Comput.* **2011**, *7*, 761.
[21] R.-Z. Liao, W. Thiel, *J. Chem. Theory Comput.* **2012**, *8*, 3793.
[22] C. V. Sumowski, B. B. T. Schmitt, S. Schweizer, C. Ochsenfeld, *Angew. Chem.* **2010**, *122*, 10147; *Angew. Chem. Int. Ed.* **2010**, *49*, 9951.
[23] D. Flaig, M. Beer, C. Ochsenfeld, *J. Chem. Theory Comput.* **2012**, *8*, 2260.
[24] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* **2010**, *132*, 154104.
[25] A. D. Becke, *Phys. Rev. A* **1988**, *38*, 3098.
[26] C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B* **1988**, *37*, 785.
[27] A. Schäfer, H. Horn, R. Ahlrichs, *J. Chem. Phys.* **1992**, *97*, 2571.
[28] D. Case, et al. "AMBER 10", Technical Report, University of California, San Francisco, **2008**.
[29] A. R. Dinner, G. M. Blackburn, M. Karplus, *Nature* **2001**, *413*, 752.
[30] A. K. McCullough, M. L. Dodson, R. S. Lloyd, *Annu. Rev. Biochem.* **1999**, *68*, 255.
[31] P. J. O'Brien, *Chem. Rev.* **2006**, *106*, 720 – 752.
[32] S. Grimme, S. Ehrlich, L. Goerigk, *J. Comput. Chem.* **2011**, *32*, 1456 – 1465.

