JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

Structural Role of Uracil DNA Glycosylase for the Recognition of Uracil in DNA Duplexes. Clues from Atomistic Simulations

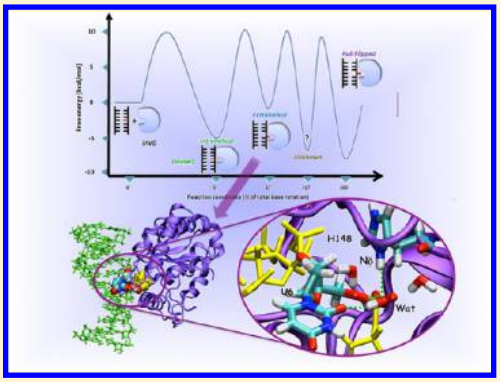
Duvan Franco,[†] Jacopo Sgrignani,[‡] Giovanni Bussi,[†] and Alessandra Magistrato*[,‡]

[†]International School for Advances Studies (SISSA/ISAS), via Bonomea 265, Trieste, Italy

[‡]CNR-IOM-DEMOCRITOS National Simulation Center C/o SISSA, via Bonomea 265, Trieste, Italy

Supporting Information

ABSTRACT: In the first stage of the base excision repair pathway the enzyme uracil DNA glycosylase (UNG) recognizes and excises uracil (U) from DNA filaments. U repair is believed to occur via a multistep base-flipping process, through which the damaged U base is initially detected and then engulfed into the enzyme active site, where it is cleaved. The subtle recognition mechanism by which UNG discriminates between U and the other similar pyrimidine nucleobases is still a matter of active debate. Detailed structural information on the different steps of the base-flipping pathway may provide insights on it. However, to date only two intermediates have been trapped crystallographically thanks to chemical modifications of the target and/or of its complementary base. Here, we performed force-field based molecular dynamics (MD) simulations to explore the structural and dynamical properties of distinct UNG/dsDNA adducts, containing A:U, A:T, G:U, or G:C base pairs, at different stages of the base-flipping pathway. Our simulations reveal that if U is present in the DNA sequence a short-lived extra-helical (EH) intermediate exists. This is stabilized by a water-mediated H-bond network, which connects U with His148, a residue pointed out by mutational studies to play a key role for U recognition and catalysis. Moreover, in this EH intermediate, UNG induces a remarkable overall axis bend to DNA. We believe this aspect may facilitate the flipping of U, with respect to other similar nucleobases, in the latter part of the base-extrusion process. In fact, a large DNA bend has been demonstrated to be associated with a lowering of the free energy barrier for base-flipping. A detailed comparison of our results with partially flipped intermediates identified crystallographically or computationally for other base-flipping enzymes allows us to validate our results and to formulate hypothesis on the recognition mechanism of UNG. Our study provides a first ground for a detailed understanding of the UNG repair pathway, which is necessary to devise new pharmaceutical strategies for targeting DNA-related pathologies.



1. INTRODUCTION

DNA damages provoked by endogenous and exogenous sources (i.e., replication errors, alkylating agents, oxygen radicals, UV light, and X-rays) are omnipresent.[1,2] Cells possess sophisticated repair mechanisms to recognize and correct different types of DNA lesions, in order to preserve unaltered the genetic information and to prevent the onset of several diseases (i.e., cancer, mental retardation, developmental disorders, neurodegenerative and aging related diseases).[3] Multiple pathways exist to repair these several types of DNA lesions. In the base excision repair (BER) path, DNA damages, resulting from deamination, oxidation, and alkylation, are detected and repaired by the coordinate effort of a set of specialized enzymes.[1] One of the most conserved enzymes taking part to this process is uracil DNA glycosylase (UNG), a monofunctional glycosylase, which removes uracil (U), normally occurring in RNA filaments, from single and double strand (ds) DNA.[4] U in DNA arises as a result of incorporation of deoxy−uracil monophosphate instead of deoxy−thymine monophosphate during DNA replication or as result of cytosine deamination. If U is left unrepaired, the first process determines the appearance of the A:U base pair (bp) in the genome,[4] while the latter leads to the appearance of G:U and to G:C → A:T transitions mutation during the first round of replication.[5] UNG can recognize both G:U and A:U with a remarkable preference for the latter bp.[6] The occurrence of these bps in DNA regulatory sequences can hinder recognition by the cognate DNA-binding protein, altering normal cell functions.

One intriguing aspect of UNG, as well as of other BER enzymes,[7] is an extra-helical (EH) recognition mechanism whereby the U exits from the DNA base stack and then it is engulfed into the active site pocket of the enzyme, where, subsequently, U cleavage is promoted by UNG.[8,9] A key question of the BER process is how repair enzymes can efficiently scan large portions of DNA in search of the target base. Once the enzymes have found the damaged site, the base-flipping process takes place via a multistep mechanism (Figure 1).[4,10] However, a detailed understanding of the recognition mechanism of UNG toward U, as well as that of other BER

Received: March 19, 2013
Published: May 25, 2013

dx.doi.org/10.1021/ci4001647 | J. Chem. Inf. Model. 2013, 53, 1371−1387

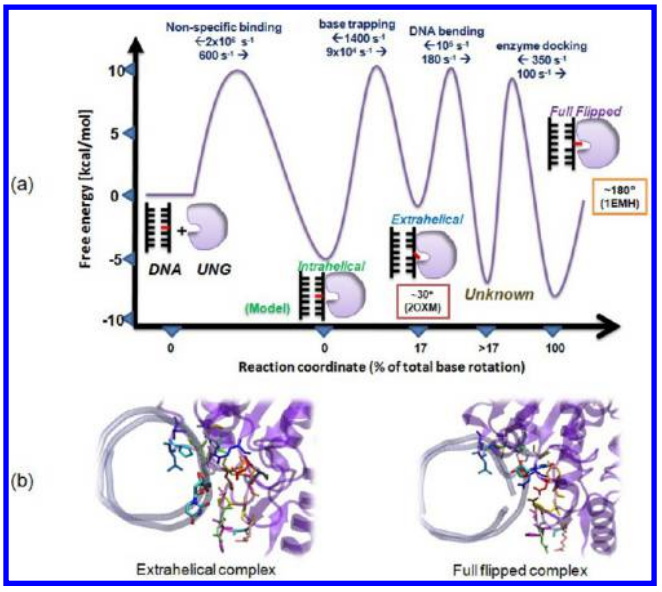


Figure 1. (a). Proposed free energy landscape for U flipping reported with respect to the percentage of base rotation toward the major groove (adapted from ref 7). (i) The initial encounter complex (intrahelical complex, IC); (ii) base trapping step forming the EH complex (PDB entry: 2OXM); (iii) formation of lesion recognition complex, detected only kinetically; (iv) fully flipped (FF) complex (PDB entry: 1EMH). A simplified sketch of the DNA and the protein is provided. (b). View of the **E/DNA_U$_{eh}$A** (left panel) and **E/DNA_U$_{ff}$A** (right panel) adduct selected from a cluster analysis along the classical MD trajectories. In this figure the DNA is depicted as gray licorice, and the U base as is depicted in licorice and colored by atom name, while the protein is in violet ribbons. Protein residues important for the catalytic activity are colored by residue number.

enzymes in search for different types of DNA lesions, remains a challenge.[11]

A general picture of the UNG base-flipping mechanism (Figure 1) was suggested based on structural[4,7,8,12,13] and kinetic information derived from spectroscopic data.[14−17] One of the most debated aspects of base-flipping is whether UNG actively perturbs the DNA structure and facilitates the bp opening[a] or whether the DNA bp "thermal breathing" determines the opening of the damaged bp and UNG captures U when it is already in EH conformation, starting in this manner the base-flipping process. These two mechanisms have been historically labeled "active" and "passive".[17,18] Structural observations, along with imino−proton exchange experiments,[17] suggested a passive role of UNG.[4,19] However, along the flipping path the DNA structure undergoes large structural deformations, which, therefore, certainly require also an active role of the enzyme.

The hypothesized base-flipping mechanism of UNG involves a first (not characterized) step in which UNG binds to the DNA backbone forming an intrahelical complex (IHC), also named encounter complex, in which the base is in its canonical *closed* Watson−Crick (WC) configuration (Figure 1).[b,16] Then, the U is trapped by UNG with formation of an EH complex (EHC), in which the base has broken its H-bond with the complementary base (*open* state) and it undergoes a rotation around the glycosidic bond so that its H-bonding moieties are directed toward the UNG (*syn* conformation, Figure 2).[4] Subsequently, the flipping path proceeds by bending the DNA to form a lesion recognition complex (LRC),[7] which has been, so far, detected only kinetically.[7,14] The DNA bend is supposed to facilitate the formation of a fully flipped complex (FFC) in which U is completely inserted into the active site pocket of

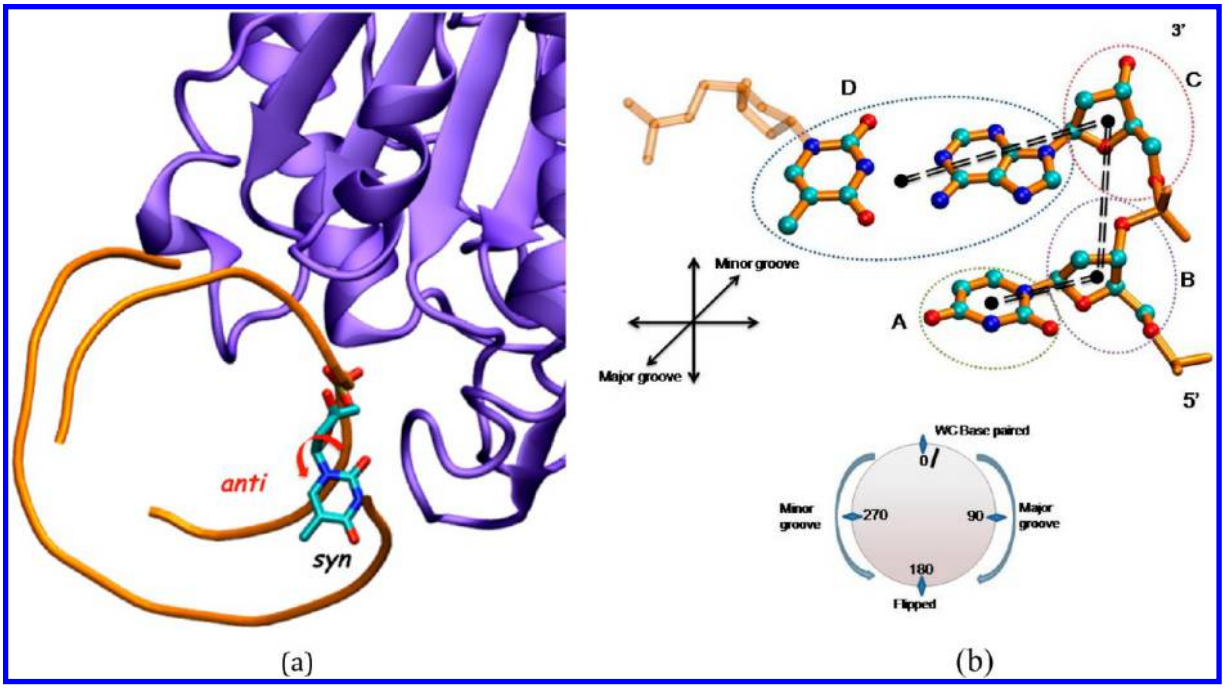


Figure 2. (a) Definition of *syn* and *anti* conformation of the target base. The *syn* conformation is shown and the *anti* is indicated with an arrow, which suggests the rotation around the χ glycosidic bond; the protein is shown in violet ribbons and the DNA in orange tube. The U base is depicted in licorice and colored by the atom name. (b)PDIH pseudodihedral angle formed by centers of mass of the heavy atoms of four groups: (A) pyrimidine ring of target base X (X = U or T), (B) sugar moiety of target base X, (C) sugar moiety of adjacent base to X in 3′ direction and (D) pyrimidine/purine rings of the adjacent base to X in 3′ direction. Black points, representing center of masses of the four groups, are only depicted for clarity. In this example the target base corresponds to U.

dx.doi.org/10.1021/ci4001647 | J. Chem. Inf. Model. 2013, 53, 1371−1387

UNG (Figure 1).[5] Structural information concerning the base-flipping intermediates is available only for the EHC and FFC for which the crystal structures (pdb entries 2OXM[4] and 1EMH,[8] respectively) were solved (Figure 1b).[16,17,20] However, these crystals were obtained using base analogues (vide infra), which may stabilize and lead to the crystallization of off-pathway intermediates. Thus, detailed structural information at atomistic level on the different base-flipping intermediates with the correct nucleobases is not available, and it is inaccessible experimentally at physiologic conditions.

In this respect computer simulations may provide direct structural insights on the different extrusion intermediates. Here our aim was to investigate the structural role of UNG along the base-flipping process and to elucidate whether and at which stage of this process the differences in the structural and dynamic properties of damaged and regular bps may give rise to specificity.[11] To this aim, we carried out extensive force-field based molecular dynamics (MD) simulations to characterize different enzyme/DNA (E/DNA) complexes at different stages of the base-flipping path (IH, EH, and FF, Figure 1) considering regular (T and C) and the damaged (U) nucleobases. Then, we compared the structural and dynamical properties of the E/DNA adducts with those of the corresponding free DNA sequences. Finally, a detailed comparison of our results with the experimental and computational studies carried out on other base-flipping enzymes allowed us to validate and generalize our findings.

Understanding the mechanism of BER enzymes at atomistic level represents, nowadays, an essential step in the development of improved anticancer therapies to counteract multidrug cancer resistance mechanisms.[21]

2. COMPUTATIONAL METHODS

2.1. Nomenclature. We introduce here the nomenclature describing the conformation of the target base along the base-flipping path. In this respect, *closed* refers to a target base, which has a canonical WC configuration and establishes H-bonds with its complementary base, while *open* refers to a base, which has broken the H-bonds with the complementary base.

Syn and *anti* refer to a rotation of the χ dihedral angle around the glycosidic bond, the rotation that takes place between the sugar moiety and the pyrimidine base: in the *anti* conformation the H-bonds of the target base are oriented toward its complementary DNA base (these rotamers are characterized by χ ranging from +90 to 270°); in *syn* conformation the H-bonds of the target base are oriented toward the enzyme (these rotamers are characterized by χ ranging from −90 to 90°) (Figure 2).

Finally, we introduce a pseudo dihedral angle (PDIH, Figure 2), formed by centers of mass of the heavy atoms of four groups: (A) pyrimidine ring of target base X (X = U or T), (B) sugar moiety of target base X, (C) sugar moiety of adjacent base to X in the 3′ direction, and (D) pyrimidine/purine rings of the adjacent base to X in the 3′ direction, and we define the IH conformation as a rotation of 0° of PDIH, namely the base is in canonical WC configuration; the EH conformation as a rotation of 30° of PDIH, and the FF conformation as a rotation of PDIH by 180° with respect to the IH state.

2.2. Model Systems. The structures of the IHCs, for which no crystal structure was available, were built with the HADDOCK program.[22] The remaining E/UNG systems were built starting from the X-ray structures of the EHC and FFC (PDB codes: 2OXM[4] and 1EMH,[8] respectively) of human UNG. These were crystallized using different target and/or complementary bases. Namely, in the first a thymine was paired with 4-methylindole, an adenine analogue unable to form hydrogen bonds. In the second U was replaced with a 2′-deoxypseudouridine, a U analogue in which C5 and N1 were interchanged to produce a ring with an additional N and a modified C−C glycosidic bond.[4,8] All models were built leaving unmodified the protein and adding two G:C bps at the 5′-terminal and one G:C bp at 3′- terminal to the DNA sequence, using 3D-DART server, in order to obtain DNA dodecamers.[23] These latter contained the following nucleotide sequences d[*GGTGTXATCTTG*] (where the target base **X** = U, T, or C), and we considered different complementary (Y = G, A) bases with respect to X. Thus, four different bps (XY = U:G, U:A, T:A, and C:G) were studied.

We also prepared a model for X-ray complex of the EH state (referred as **E/DNA_T$_{eh}$M**), leaving intact both the protein and DNA sequence, which contained a 4-methylindole as complementary base for T (Figure 1b). However, as for the other adducts, the ds DNA was transformed to a dodecamer. This system was used as a reference to check the reliability of our computational protocol. The free DNA models with the target base in the IH and EH conformations were built using the tools of AMBER9.[24]

In summary, in this study we considered four different free DNA oligomers with the bases in IH conformation (**DNA_X$_{ih}$Y**) and four DNA sequences with the bases in EH conformation (**DNA_X$_{eh}$Y**). Moreover, we considered two E/DNA complexes with the bases in IH conformation (**E/DNA_X$_{ih}$Y**), four complexes with the bases in EH conformation (**E/DNA_X$_{eh}$Y**), and three complexes with the bases in FF conformation (**E/DNA_X$_{ff}$Y**). Finally, we also simulated two **E$_{H148G}$/DNA_U$_{eh}$A** and **E$_{H148G}$/DNA_U$_{ff}$A** adducts in which His148 was mutated to Gly, in order to verify the role of this residue, which has been pointed out as important for damage recognition and catalysis.[4,14] A detailed list of the systems studied is shown in Table 1 and Table S1 of the Supporting Information.

Table 1. Nomenclature Used for the Systems Studied in This Work[a]

system type	studied X:Y base pairs
DNA_X$_{ih}$Y	A:U, G:U, A:T, G:C
DNA_X$_{eh}$Y	A:U, G:U, A:T, G:C
E/DNA_X$_{ih}$Y	A:U, A:T
E/DNA_X$_{eh}$Y	A:U, G:U, A:T, G:C, M:T
E/DNA_X$_{ff}$Y	A:U, G:U, A:T

[a]X refers to U, T, and C target bases, while Y refers to A and G complementary bases. M refers to 4-methylindole. The ih, eh, and ff subscripts refer to the intra-helical, extra-helical, and fully flipped conformations of the base which undergoes the flipping.

2.3. Classical Molecular Dynamics Simulations. All MD simulations were performed with the NAMD 2.6 program,[25] using the AMBER99SB force-field[26] modified with parmbsc0 corrections for nucleic acids.[27] The parameters and the atomic charges of deoxy-U were those of the RNA uracil nucleoside contained in the AMBER99SB force field, while for the sugar deoxyribose moiety the force-field parameters and the charges were taken from the T nucleobase.[c] All histidines were protonated on Nε, as in ref 28, except His148, for which all possible protomers were considered. However, the simulations

dx.doi.org/10.1021/ci4001647 | J. Chem. Inf. Model. 2013, 53, 1371−1387

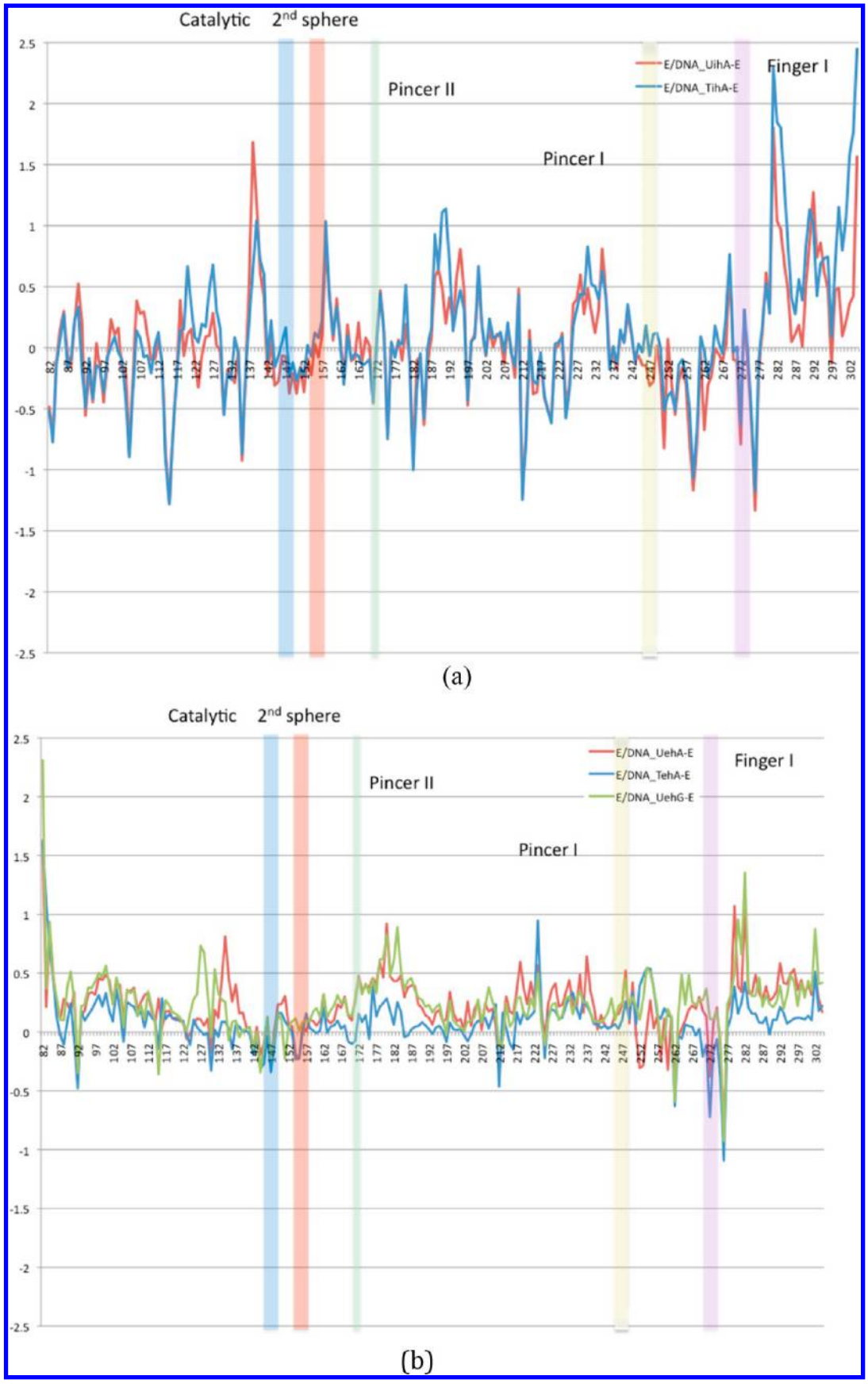


Figure 3. continued

dx.doi.org/10.1021/ci4001647 | J. Chem. Inf. Model. 2013, 53, 1371−1387

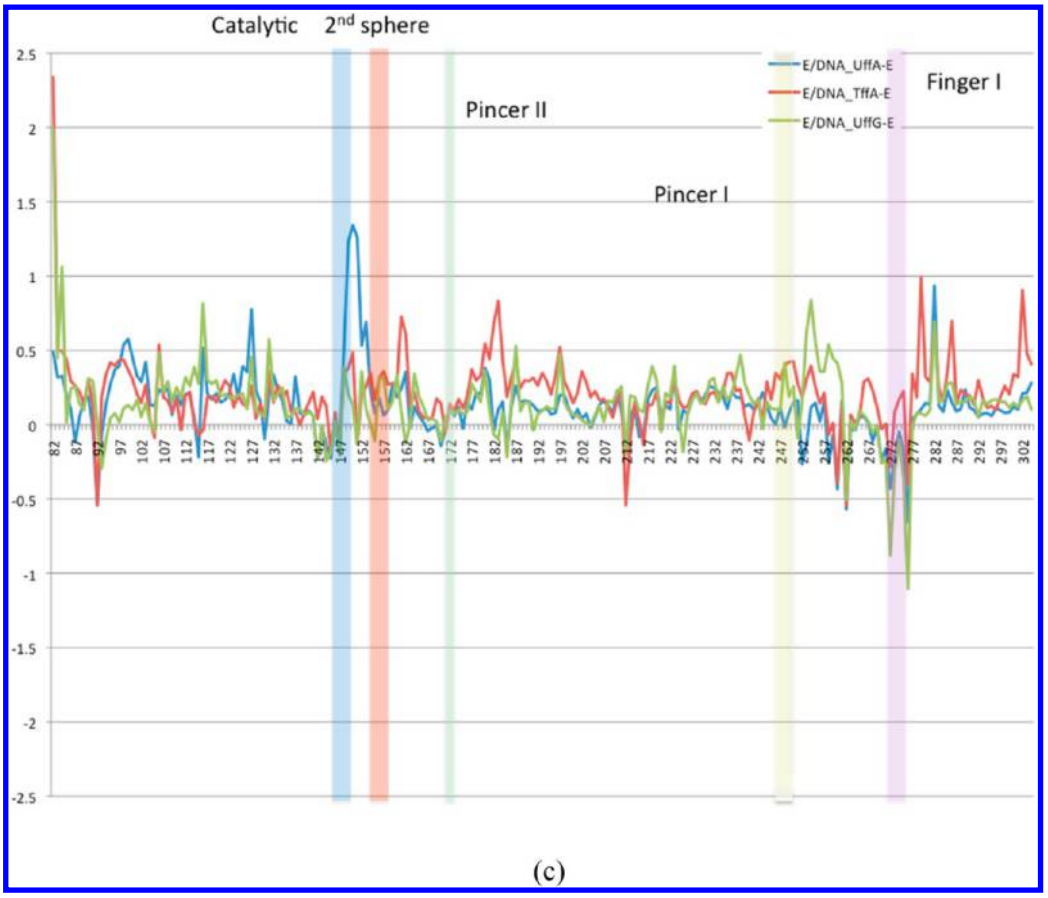


Figure 3. ΔRMSF per UNG residue calculated as the difference between the RMSF of **E/DNA_X$_{ih}$Y**, **E/DNA_X$_{eh}$Y**, **E/DNA_X$_{ff}$Y**, and free UNG in (a), (b), and (c), respectively. The UNG catalytic, second sphere, pincer II, pincer I, and finger regions are indicated with blue, red, green, yellow, and pink rectangles, respectively.

presented here are with a doubly protonated His148, as this protonation state was reproducing at best the crystal structure and was consistent with a previous computational study.[28]

Crystallographic waters were kept in the models, and each system was solvated with explicit TIP3P waters[29] considering a minimal distance of 12 Å between solute and the edge of a cubic box. Sodium cations were added to obtain electroneutrality, requiring 22 Na^+ ions for free DNA systems and 16 for E/DNA adducts.

Geometry optimization and system equilibration were carried out through the following methodology: (i) energy minimization for 5000 steps with harmonic restraints (k = 500 kcal/mol/A^2) applied to the solute to relax the solvent; (ii) energy minimization for 5000 steps to relax the whole system with no restraints; (iii) linear heating with a 150 ps long MD simulation up to 300K° in a NVT ensemble (i.e, temperature was increased 15K° every 7.5 ps), while the solute was restrained with a harmonic restraint of 10 kcal/mol/A^2; (iv) volume equilibration in a NPT ensemble to reach the correct water density of the system; and, finally, (v) MD production runs in a NPT ensemble for 10−26 ns (Table S1). Periodic boundary conditions were used, and the electrostatic interactions were calculated with the Particle-Mesh Ewald (PME) method,[30] using a 12 Å cutoff for the real part, as for the van der Waals (vdW) interactions. NPT simulations at 300 K and 1 atm were performed using the Langevin scheme for temperature control and Nose-Hoover barostat for pressure-coupling.[31,32] A time step of 1.5 fs was set for all of the simulations. H-bond lengths were constrained using the SHAKE algorithm implemented in NAMD.[33]

2.4. Analysis. The Curves+ program[34] was employed to calculate the structural properties of DNA.[35] Root mean square fluctuations (RMSF) and cluster analysis were performed with GROMACS tools in GROMACS 4.0.7.[36] Interaction energies were calculated with NAMD Energy plug-in version 1.4 in VMD 1.8.7,[37] H-bonds, radial distribution function, and water density were calculated along the MD trajectories using *ptraj* module of the AMBER analysis tools.[24] The number of interfacial waters present between UNG and DNA was calculated with PLUMED 1.2.2,[38] integrated to GROMACS 4.0.7[36] (see Supporting Information). The residence time was calculated as the time that a water molecule resides within a solvation shell of 3.5 Å from Nδ,Nε@His148.

MM-GBSA calculations were performed using the MMPBSA.py script[39] available in Amber12. The protein, nucleic acid, and complex conformations were extracted from single MD simulations of the solvated complex. GBSA calculations were performed using the Onuffriev−Bashford−Case model,[40] and the salt concentration was set to 0.1 M. The binding enthalpies of the T and U nucleobases with the catalytic pocket were evaluated setting the idecomp value to 2 (i.e., the 1−4 electrostatic interaction energies and the 1−4 vdw interaction energies are added to the electrostatic and vdW potential terms, respectively).

dx.doi.org/10.1021/ci4001647 | J. Chem. Inf. Model. 2013, 53, 1371−1387

3. RESULTS AND DISCUSSION

3.1. Structural Features of E/DNA Complexes.

3.1.1. Intra-Helical Complexes. Since no structural information of the initial encounter complex exists we have initially constructed this E/DNA adduct in the presence of U and T target bases by docking simulations, and we relaxed the resulting IHC by MD simulations.[22] The ΔRMSF, calculated between the E/DNA adducts and the free UNG, reveals the differences in UNG flexibility due to distinct target bases present in the DNA sequence (Figure 3 and Figure S1, Supporting Information). NMR data suggested that free UNG is rather rigid and that its binding to a base different from U reshapes its conformational free energy landscape, allowing the enzyme to sample the DNA.[18] However, our data show an overall stabilization upon UNG binding to DNA, which is similar for both the **E/DNA_U$_{ih}$A** and **E/DNA_T$_{ih}$A** (Figure S2, Supporting Information). In particular, the formation of the IHC determines a decreased flexibility in selected regions involved either in the base flipping or in the catalytic activity (namely, the catalytic region (residues 144−148), the pincer regions (residues 246−249 and 169), which contact the DNA backbone and contain the U target base both in the EH and in the FF conformations; the finger region (residues 270−274), which fills the hole left empty by the rotation of the target base; and, finally, the second sphere region (residues 153−159), which belongs to a loop that contacts the catalytic region). Moreover, we notice the additional stabilization of residues 251−258, a region which was postulated to be involved in the binding of one DNA strand.[41]

The H-bonds between U, T, and the complementary nucleobase are identical in the simulations of free DNA and are characterized by a A19@N1..N3@U6 = 2.95 ± 0.1 Å (occupancy: 99%) and U (T)6@O4...N6@A19 = 2.98 ± 0.1 Å (99%). However, in the presence of UNG the occupancy of these H-bonds is reduced. Namely, these are A19@N1..N3@U6 = 2.98 ± 0.1 (85%) and U6@O4...N6@A19 = 2.95 ± 0.15(84%) for **E/DNA_U$_{ih}$A** and A19@N1..N3@U6 = 2.96 ± 0.1 (73%) and U6@O4...N6@A19 = 3.0 ± 0.1 (67%) for **E/DNA_T$_{ih}$A**, suggesting that the enzyme perturbs the strength of the H-bonds at the target level already in the encounter complex.

3.1.2. Extra-Helical Complexes. We, then, focused our attention on the structural stability of the **E/DNA_X$_{eh}$Y** systems. An analysis of the ΔRMSF reveals that in **E/DNA_T$_{eh}$A** the key regions involved in the binding of DNA are stabilized upon formation of the EHC.[18] In contrast, for **E/DNA_U$_{eh}$A** there is an overall larger flexibility of the enzyme, especially in the pincer II and in the catalytic regions (Figure 3 and S1,3, Supporting Information). A higher flexibility with respect to the free UNG is also observed for **E/DNA_U$_{eh}$G** and **E/DNA_C$_{eh}$G** (Figure S3, Supporting Information) so that their RMSF difference is very small (Figure S2, Supporting Information). This is consistent with the fact that the EHC is metastable and that in the presence of the target U base UNG may prepare to evolve toward a later lesion recognition complex.[18] In addition, the flexibility of **E/DNA_U$_{eh}$A** is slightly enhanced with respect to the IH state, while the opposite occurs in **E/DNA_T$_{eh}$A** (Figure S1, Supporting Information).

A detailed structural analysis at the target level shows that in all **E/DNA_X$_{eh}$Y** systems the H-bonds between O4′, O2, and N3@T6 and Nε@His148, N@Ala214, and O@His212, respectively (Figure S4, Supporting Information), present in the EHC crystal structure, were lost at the beginning of the simulation (in less than 1 ns) (Figure 4). This aspect is not fully surprising as partially flipped intermediates are highly unstable,[9,10] and other computational studies showed a disagreement between calculations and the crystallographic structures of this kind of intermediates.[10,42] This may be due to the fact that the crystallographic capture of such metastable states requires a variety of trapping techniques, which may stabilize off-pathway intermediates.[10,42,43]

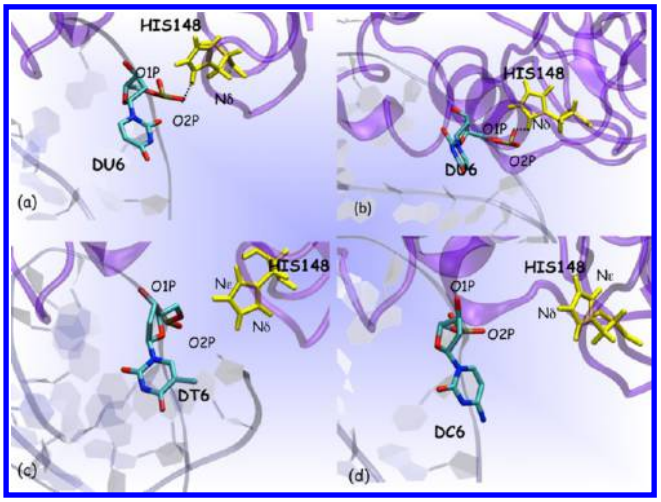


Figure 4. Orientation of the target base after equilibration of **E/DNA_X$_{eh}$Y** systems. **E/DNA_U$_{eh}$A**, **E/DNA_U$_{eh}$G**, **E/DNA_T$_{eh}$A**, and **E/DNA_C$_{eh}$G** are depicted in (a), (b), (c), and (d), respectively. Protein backbone is shown in violet. The target base is colored by atom name, while His148 is colored in yellow.

In order to check that the low stabilities of target bases in the EHCs were not due to artifacts introduced in building our models, we simulated also the crystallographic structure of the EHC (**E/DNA_T$_{eh}$M**). Also in this case the target base was rapidly solvated at the beginning of the MD run, losing its direct H-bond interactions with the enzyme (Figure S4, Supporting Information).

To monitor similarities and differences among the **E/DNA_X$_{eh}$Y** systems we analyzed the evolution of the distances between the rings of the X:Y bp and the evolution of the χ dihedral angle along the simulation (Figure 5). These two parameters allowed us to check the opening of the bp and to differentiate if the target base is in the *syn* or in the *anti* conformation, respectively.

In all systems, the X base returns toward the *closed* conformations. In fact, the average distances between N9@Y and C4@X are 8.5 ± 0.5, 5.5 ± 0.1, 10.1 ± 0.7, 10.8 ± 0.9, and 10.2 ± 0.8 Å for **E/DNA_U$_{eh}$A**, **E/DNA_U$_{eh}$G**, **E/DNA_T$_{eh}$A**, **E/DNA_C$_{eh}$G**, and **E/DNA_T$_{eh}$M**, respectively. This distance was of 15.5 Å in the X-ray structure.

Both T and C assumed an *anti* conformation. However, while T reestablished the H-bond with the complementary DNA base (*closed* state), C is stabilized by a H-bond between C@N4H...O2P@T5 (2.85 ± 0.2 Å; occupancy 66%), which leaves it partially *open*, although the base is in an *anti* conformation (Figure 4). On the contrary, no rotation around the glycosidic bond was observed for U, which remains in the *syn* state (Figures 4 and 5), keeping its H-bonding moieties exposed toward the protein/solvent interface. In **E/DNA_U$_{eh}$A**

dx.doi.org/10.1021/ci4001647 | J. Chem. Inf. Model. 2013, 53, 1371−1387

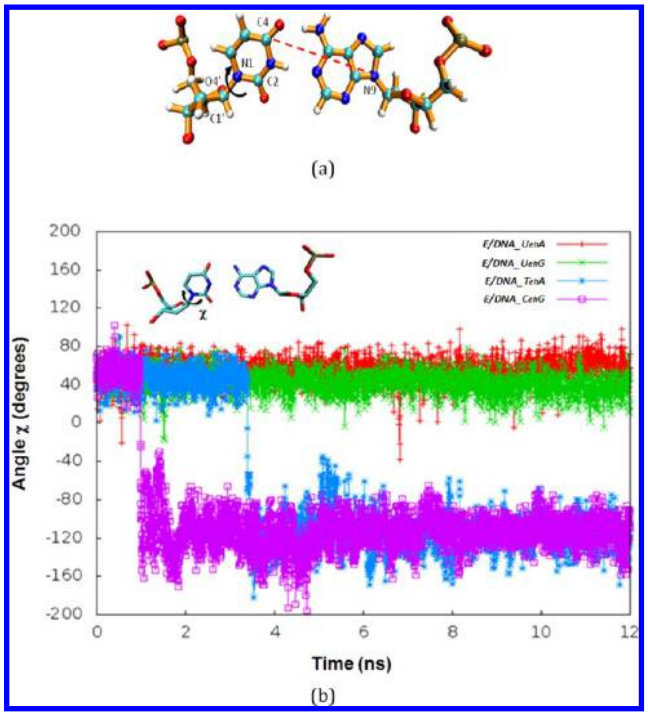


Figure 5. (a) Definition of the distance C4@U and N9@A and of the χ angle (formed by O4′, C1′, N1 and C2) of U for the A:U bp. (b). Plot of the dihedral angles χ (deg) in all **E/DNA_X$_{eh}$Y** complexes vs simulation time (ns).

and **E/DNA_U$_{eh}$G** two H-bonds are formed between His148@Nδ and O2P/O1P@U6 = 2.83 ± 0.18 and 2.88 ± 0.22 Å, respectively (occupancy: 64 and 52%), which contribute to the structural stability of the EHC. The T base assumes a *syn-open* orientation also in the simulation **E/DNA_T$_{eh}$M**, consistently with experimental data.[4] In contrast, in **E$_{H148G}$/DNA_U$_{eh}$A** U returns in the IH *anti-closed* position. Therefore, two aspects appear to be crucial for the existence of short-lived[d] EH *syn-open* intermediate: the presence of the U nucleobase in DNA and of His148 in UNG.

Interestingly, in the simulation of **DNA_X$_{eh}$Y** systems all the target T, C, and U bases immediately return from the EH to the IH canonical WC conformation. This suggests that UNG enhances the lifetime of U in the *open* state in agreement with imino-proton exchange experiments.[4,8]

In the presence of UNG, the formation of a metastable EH intermediate, with a *syn* orientation of the nucleobase, may be congruent with a solvent-exposed capture of the DNA damage proposed on the basis of experimental and computational studies[44] for many classes of proteins[45] (i.e., glycosylases,[9] tryptophan repressor,[46] interferon regulatory factors (e.g., IRF-1 and IRF-2),[47] type II restriction endonucleases,[48] and repair enzymes[44,49]). To verify this hypothesis we first carried out an analysis of the number of interfacial waters (IW) and of the water density in the region of the target base and the residues at the DNA/protein interface. No significant differences were observed in the number of IW among the different E/DNA adducts (Table S2, Supporting Information). Instead, the water density was slightly higher for the systems containing U (Table S3, Supporting Information).

Interestingly, an analysis of water present between U6 and His148 reveals that a water molecule is present (with an occupancy of 60%), forming an H-bond bridge between the U6 and His148 (Figure 6). This molecule has a residence time of

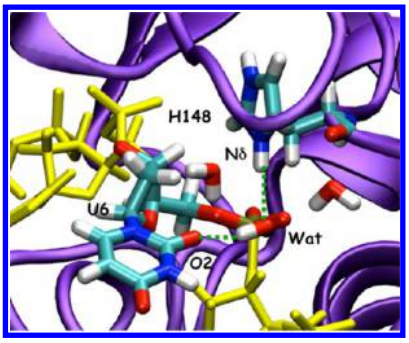


Figure 6. Water network connecting U6 and His148 in **E/DNA_U$_{eh}$A**. The protein backbone is depicted in violet ribbons, the DNA is depicted in yellow licorice, U6, His148, and the water molecule are depicted in licorice and colored by atom names. The H-bond network is represented in green dotted lines.

130 ps. Therefore, we conclude that the EH *syn-open* conformation of U in **E/DNA_X$_{eh}$Y** systems represents a metastable state.[4,17,d] UNG increases the lifetime of this short-lived intermediate bearing an EH *syn* base via a water-mediated H-bond network. In order to complete the analysis of hydration properties of the target base in the different EHCs we have also calculated the RDF of the target base and of His148 (Figure S5, Supporting Information). This analysis shows that **E/DNA_U$_{eh}$A** has a larger number of coordinated water molecules when compared with **E/DNA_U$_{eh}$G** and **E/DNA_T$_{eh}$A**, which have similar hydration properties. **E/DNA_C$_{eh}$G** has, instead, a qualitatively different hydration behavior. This is due to the fact that C has a different conformation with respect to the other target bases due to the formation of an intramolecular H-bond of the amino moiety with the phosphate of the flanking T base.

3.1.3. Fully-Flipped Complexes. We also inspected the structural stability of the FFC considering U and T as a target base to rationalize why U is the preferential substrate of UNG. In fact, it is not yet clear if T undergoes a catalytic noncompetent binding in the UNG active site or if it binds at an alternate location.[6]

An analysis of the ΔRMSF shows that an overall stabilization of UNG occurs in the FFC in presence of U with respect to the EHC (Figure 3 and Figures S1, Supporting Information). In contrast, an opposite trend is observed for T, so that the ΔRMSF between **E/DNA_U$_{ff}$A** and **E/DNA_T$_{ff}$A** shows that in the presence of U a marked stabilization of the UNG flexibility occurs (Figure S2, Supporting Information).[18]

In case of **E/DNA_T$_{ff}$A** we observed T6@O6...Nδ@Asn204 = 3.06 ± 0.19 (occupancy: 79%) and T6@PO2 ...Nε@His148 = 2.78 ± 0.12 (occupancy: 60%), which contribute to stabilizing the distorted DNA backbone. In case of **E/DNA_U$_{ff}$A** we observed U6@O6...Nδ@Asn204 = 3.04 ± 0.18 (occupancy: 85%), U6@PO2...Nε@His268 = 2.83 ± 0.21 (occupancy: 25%), and His148@Nδ...O1P@T5 = 2.78 ± 0.14 (occupancy: 11%) (Figure 7). In both cases a water molecule is present in the vicinity of U/T6@C1′, the carbon that undergoes the nucleophilic attack during the cleavage of the C1′@U6−N1 glycosidic bond catalyzed by UNG. An analysis of the radial distribution function shows that a larger number of waters lies at shorter distances from C1′@U6 with respect to C1′@T6 (Figure S6, Supporting Information). The shorter O@Wat...C1′ distance may be at the basis of lower activation free energy barriers for the nucleophilic attack performed to

dx.doi.org/10.1021/ci4001647 | J. Chem. Inf. Model. 2013, 53, 1371−1387

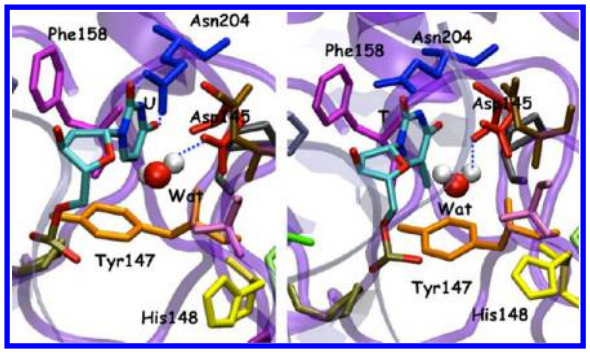


Figure 7. Image of the pocket site of the FFCs containing U and T base on the left and right panel, respectively. The protein backbone is depicted in violet ribbons, and the U and T bases and the putative nucleophilic waters are depicted in licorice and in ball and sticks, respectively, and colored by atom name. Active site residues are depicted in licorice and colored by residue type.

cleave the glycosidic bond of the U base.[50,51] In both cases the putative nucleophilic water H-bonds with Asp145, which may act as a generalized base, consistently with other computational studies.[28,52] Interestingly, a different orientation of the key residue His148, which, in the presence of T, H-bonds to the DNA phosphate backbone, determines a distinct hydration network in the active site between T and U. This different hydration network may be related to the catalytic efficiency of the enzyme.

3.2. DNA Structural Parameters. *3.2.1. Intra-bp Parameters.* We also analyzed several DNA structural parameters to assess whether and how UNG affects the structural properties of DNA in the proximity of the target base. In Table 2, we report three intra-bp parameters (shear, stretch, and opening angle) that describe the internal conformational changes of DNA occurring upon base-flipping.[19] Other intra-bp parameters are reported in Table S4, Supporting Information.[53] To characterize these changes we inspected the target level (TL) (the bp containing the target base, X6Y19) and the two flanking levels in 3′ (T5-A20, FLK1) and in 5′ (A7-T18, FLK2) directions.[53]

Considering the **DNA_X$_{ih}$Y** complexes, no large differences are present in the intra-bp parameters when T or U are in the DNA sequence. This is consistent with the fact that U is virtually indistinguishable from T when the DNA is in canonical B form.[19] G:U at TL is the only exception to this as the opening angle is large, positive and characterized by a large standard deviation,[54] while the shear has a value typical of a non WC bp[19](Table 2). Since the opening angle was demonstrated to play a key role in DNA damage recognition,[54] we also reported the distribution of this angle along the simulations of the different systems investigated (Figure 8). Consistently with the data of Table 2, the opening is slightly enhanced for G:U, while it is smaller for all other investigated bps. As it is expected, the intra-bp parameters show extremely high values when the target base is in the EH conformation (Table 2). However, a comparison between DNA and E/DNA is relevant. The analysis of the intra-bp parameters for the **DNA_X$_{eh}$Y** systems shows that at TL the opening is smaller for A:U than for the other bps. A:T is, instead, characterized by a large negative value, indicating an opening toward the minor groove. Figure 8b shows that the distribution of the opening is narrow for A:T, while it is broader for all other bps.

A comparison of **E/DNA_X$_{ih}$Y** with their corresponding free DNA sequences reveals only that the opening angle assumes a large positive value at TL of A:T. This angle is larger and broader for A:T (Figure 8c) than for A:U, even in the presence of UNG.

Instead, a comparison of **E/DNA_X$_{eh}$Y** systems with their free DNA counterparts reveals that UNG enhances the opening of the target base. In all cases the opening angle has a positive value, since the base is forced to open toward the major groove due to the presence of the enzyme. Even in the EHC the opening angle is larger for A:T than for A:U, assuming the largest value in G:U (Figure 8d).

3.2.2. Interbp and Global DNA Parameters. Selected interbp parameters are reported in Tables S5 and S6, Supporting Information, and we discuss here only the changes of the DNA grooves. Since UNG protrudes inside the DNA minor groove with the finger region, it is expected that the minor groove width will be larger in the E/DNA adducts than in the free DNA oligomers (Table 3). This is, indeed, the case for all IHCs and EHCs at FLK1 and TL, while FLK2 is unaffected by UNG binding. The opening of the minor groove width is reflected in a contraction of the major groove width in all E/DNA complexes.

Interestingly, at TL we observe no differences in the parameters of the IHCs, while in the EH state the major groove width is larger for **DNA_U$_{eh}$A** than for **DNA_T$_{eh}$A**. The same behavior is also observed for the **E/DNA_X$_{eh}$Y** adducts where the enzyme enhances the differences in the major groove width.

Since the UNG finger protrudes from the minor groove side, it may be possible that the recognition of the target base depends on the size of the minor groove in free DNAs, as observed in other DNA binding proteins.[55−58] However, in this case no differences are observed in the minor groove width of the free and complexed IH systems.

We, finally, report the overall axis bend of the DNA sequences in the presence and in the absence of UNG and we monitor its changes along the base-flipping path (Table 4). As expected the free DNA sequences display rather similar values of axis bend when the target base is in either the IH or EH conformations. This bend gradually increases in the presence of UNG, and in **E/DNA_U$_{eh}$A**, it is comparable to the axis bend of the FFC (Figure 9). This is in contrast with what was observed experimentally in the 2OXM crystal structure where the binding of T to the UNG *exosite* is not accompanied by a remarkable bending of DNA.[4,7,8] Figure 9 shows that the overall axis bend of the free DNA sequences is larger and broader for the A:T containing DNA. An opposite trend occurs in the presence of UNG, as the overall bend largely increases when the DNA contains the damaged bp. Moreover, the simulation of **E$_{H148G}$/DNA_U$_{eh}$A** shows that the depletion of the H-bonding moieties of His148 restores the bend to the value it has in **E/DNA_T$_{eh}$A** (Table 4). This confirms the key role of His148 for U recognition.[4] Since a relationship between bend flexibility and the free energy of base-flipping was demonstrated,[54,59−61] our results suggest that one of the reasons why the UNG can distinguish between A:T and A:U bp is that when UNG binds to **DNA_U$_{eh}$A**, a marked DNA bend immediately takes place. Thus, UNG stabilizes the bend conformation of the A:U containing DNA sequences (Figure 9). This may facilitate the target U base to undergo the final part of the base-flipping process with respect to the other bases

dx.doi.org/10.1021/ci4001647 | J. Chem. Inf. Model. 2013, 53, 1371−1387

Table 2. DNA Intra-bp Parameters Relative to T5-A20 bp (FLK1), X6-A(G)19 bp (target level, TL), and A7-T18 bp (FLK2)[a]

	FLK1			TL			FLK2		
system	shear	stretch	opening	shear	stretch	opening	shear	stretch	opening
DNA_U_{ih}A	−0.12 (0.27)	0.03 (0.12)	3.4 (5.2)	−0.10 (0.29)	0.02 (0.13)	3.2 (5.7)	0.07 (0.28)	0.03 (0.13)	4.40 (5.7)
DNA_U_{ih}G	−0.0.9 (0.28)	0.06 (0.14)	6.5 (5.8)	2.17 (0.41)	0.05 (0.11)	6.5 (11.6)	0.08 (0.27)	0.06 (0.12)	5.10 (5.3)
DNA_T_{ih}A	−0.10 (0.29)	0.03 (0.13)	2.7 (5.4)	−0.14 (2.27)	0.02 (0.12)	3.8 (5.7)	0.03 (0.28)	0.03 (0.14)	4.00 (5.6)
DNA_C_{ih}G	−0.12 (0.29)	0.08 (0.14)	7.3 (6.3)	−0.04 (0.30)	0.05 (0.11)	2.3 (3.3)	0.12 (0.29)	0.07 (0.12)	3.90 (5.4)
DNA_U_{eh}A	−0.13 (0.27)	0.03 (0.12)	4.0 (5.2)	−0.15 (0.28)	0.03 (0.13)	1.6 (5.2)	0.04 (0.27)	0.02 (0.12)	5.40 (5.1)
DNA_U_{eh}G	−0.18 (0.30)	0.09 (0.14)	4.7 (5.8)	−3.18 (6.69)	−5.59 (4.87)	−108.1 (59.6)	0.08 (0.28)	0.06 (0.12)	4.80 (4.9)
DNA_T_{eh}A	−0.08 (0.31)	0.05 (0.14)	1.7 (5.9)	−0.41 (2.25)	0.28 (1.45)	−7.6 (19.3)	0.10 (0.34)	0.05 (0.14)	1.10 (5.9)
DNA_C_{eh}G	−0.03 (0.35)	0.11 (0.37)	7.2 (6.4)	3.11 (3.28)	1.25 (3.24)	42.4 (38.0)	0.14 (0.29)	0.06 (0.13)	4.60 (5.5)
E/DNA_U_{ih}A	−0.03 (0.28)	0.02 (0.12)	3.3 (5.6)	−0.18 (0.28)	0.02 (0.14)	3.2 (5.2)	0.07 (0.28)	0.04 (0.13)	2.80 (5.5)
E/DNA_T_{ih}A	0.00 (0.32)	0.00 (0.14)	2.3 (6.0)	−1.63 (1.80)	0.21 (0.45)	14.7 (14.2)	0.07 (0.28)	0.03 (0.12)	4.60 (4.9)
E/DNA_U_{eh}A	−0.18 (0.34)	0.07 (0.20)	0.7 (7.7)	0.19 (3.05)	0.22 (1.76)	6.1 (18.2)	0.08 (0.29)	0.02 (0.12)	1.50 (4.9)
E $_{H148G}$ /DNA_U_{eh}A	−0.04 (0.29)	0.09 (0.14)	3.3 (6.1)	2.84 (1.56)	5.39 (1.54)	66.3 (13.6)	0.15 (0.30)	0.01 (0.12)	3.10 (5.30)
E/DNA_U_{eh}G	−0.06 (0.31)	0.00 (0.18)	7.3 (7.3)	−1.34 (3.71)	1.03 (2.19)	18.1 (29.7)	0.06 (0.28)	0.03 (0.13)	2.20 (5.0)
E/DNA_T_{eh}A	0.10 (0.28)	0.19 (0.27)	10.3 (10.8)	4.28 (1.82)	2.66 (1.28)	34.7 (16.8)	0.07 (0.29)	0.04 (0.13)	4.50 (5.3)
E/DNA_C_{eh}G	−0.03 (0.35)	0.11 (0.37)	13.9 (12.9)	2.60 (1.91)	7.57 (1.76)	80.7 (13.3)	0.24 (0.29)	0.10 (0.15)	3.10 (6.0)
E/DNA_U_{ff}A	−0.05 (0.32)	0.03 (0.14)	6.0 (6.0)	−10.54 (4.28)	−10.01 (3.45)	−89.5 (29.9)	0.05 (0.29)	0.03 (0.15)	3.60 (5.3)
E $_{H148G}$ /DNA_U_{ff}A	0.03 (0.29)	0.06 (0.22)	4.40 (7.40)	11.56 (6.41)	4.35 (4.90)	119.5 (18.3)	0.14 (0.34)	0.02 (0.13)	0.10 (5.70)
E/DNA_U_{ff}G	−3.41 (2.24)	0.0 (0.4)	7.5 (9.2)	−10.9 (6.6)	−16.0 (8.0)	20.6 (34.2)	0.3 (0.4)	0.06 (0.1)	−0.40 (5.0)
E/DNA_T_{ff}A	4.21 (2.21)	−0.53 (0.64)	16.9 (11.8)	−10.82 (4.53)	−11.35 (5.60)	−92.5 (38.2)	0.01 (0.28)	0.05 (0.14)	3.10 (5.6)

[a]Shear and stretch and opening angles (deg). Standard deviations are shown in parentheses.

dx.doi.org/10.1021/ci400164 7 | J. Chem. Inf. Model. 2013, 53, 1371−1387

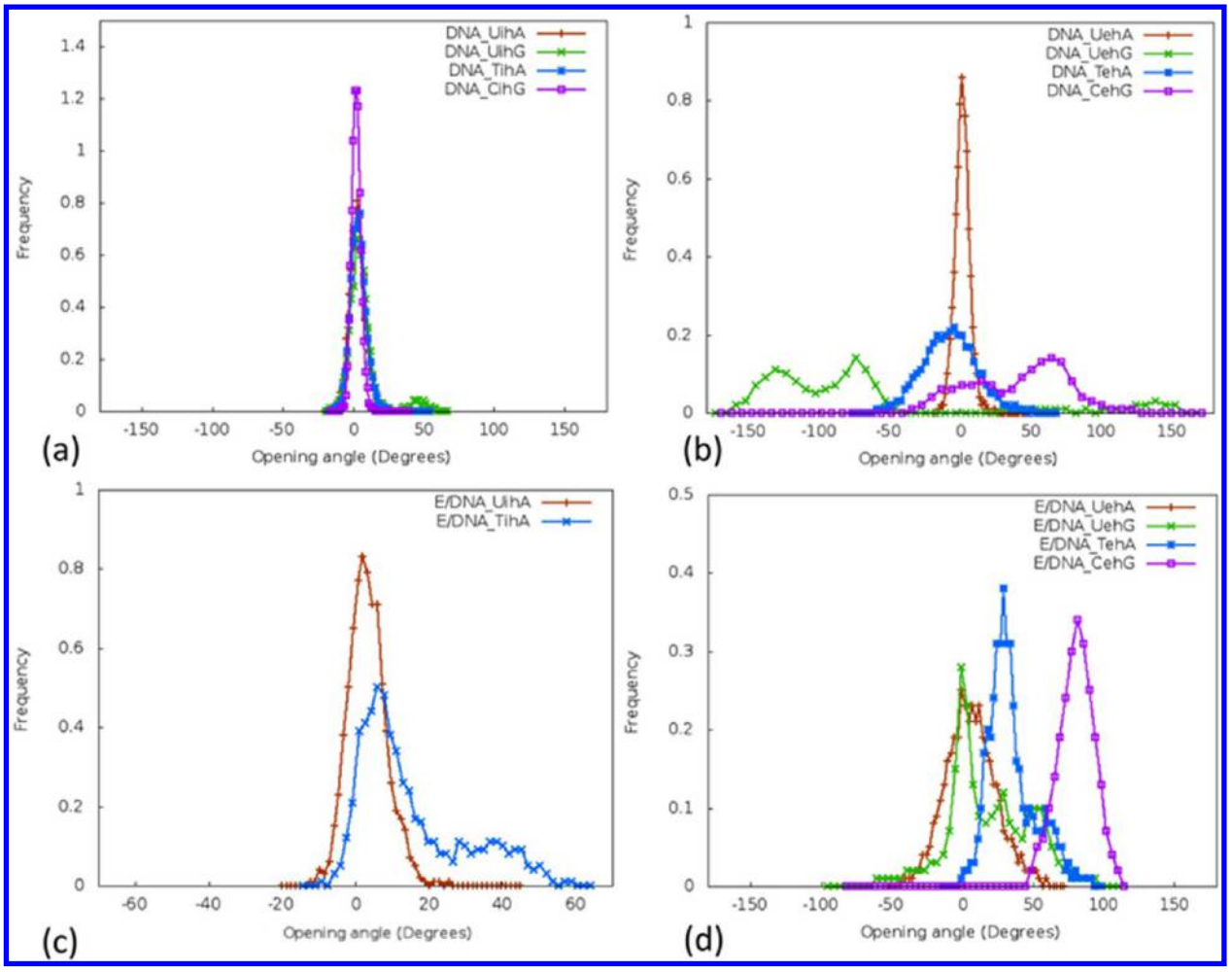


Figure 8. Distribution of opening angles (deg) of the TL during the MD simulation. The **DNA_X$_{ih}$Y**, **DNA_X$_{eh}$Y**, **E/DNA_X$_{ih}$Y**, and **E/DNA_X$_{eh}$Y** are reported in (a), (b), (c), and (d), respectively.

for which a remarkable DNA bending does not occur at the interrogation complex.[61]

More complex is the situation in the G:U DNA containing sequences. In fact, in the G:U bp both bases frequently flip outside the DNA during the simulation, affecting also the DNA overall axis bend. The fact that the axis bend is smaller in the G:U containing DNA sequence with respect to that containing A:U is consistent with other literature data, showing that the G:U bp is recognized in repair processes mainly because of its enhanced opening properties (Table 2).[54] The axis bend among the FF complexes is more similar (Table 4), and in this case, the His148Gly mutation only little affects it.

The different orientation of the target bp in the EH states of both free DNA and of the E/DNA adducts is also visible from the dihedral angles of the nucleobases and of the DNA backbone (Table S7, S8, and S9, Supporting Information). In fact, the distortion of the backbone is remarkably different for a DNA sequence containing U:A with respect to those containing T:A and C:G.

3.3. Energetic Analysis. We report here also relative interaction energies between the target base and UNG, or selected UNG residues, to elucidate similarities and differences among the E/DNA complexes investigated. These energies are approximate due to force-field nature and, therefore, they are reported only for a qualitative comparison. In general all E/DNA adducts are mainly stabilized by electrostatic interactions (Table 5). If we consider ΔE_{int} as the difference in the interaction energy of the target base and UNG for the EHCs and FFCs with respect to the IHCs ($\Delta E_{int}EH = E_{int}EH - E_{int}IH$ and $\Delta E_{int}FF = E_{int}FF - E_{int}IH$, respectively), we see that U is stabilized by ∼46 kcal/mol, while T by ∼25 kcal/mol in the EHCs. We calculated also the binding enthalpies of the base in the EH conformation to UNG by MM_GBSA calculations. In these calculations U has a positive binding enthalpy since in the EH conformation it is mainly stabilized by the solvent (Table 5).

When the target base is in the active site pocket, the $\Delta E_{int}FF$ favors the T in T:A with respect to U in A:U by ∼18 kcal/mol. This is confirmed also by the calculations of the binding enthalpy in which T results to be favored with respect to U by ∼7 kcal/mol (Table 5).

The high stability of T in the active site may result in an increase of the free energy barrier for the cleavage of the glycosidic bond of T.[14] Consistently with this hypothesis the interaction energy of U with UNG in G:U, which is cleaved less efficiently than U in A:U, is similar to that of T in A:T.

In addition, other factors may explain the different catalytic efficiency of the UNG toward the T and U nucleobases. In fact, recent computational studies demonstrated that T is a less effective leaving group with respect to U, when the cleavage of the glycosidic bond has occurred.[52] Finally, the hydration of the

dx.doi.org/10.1021/ci4001647 | J. Chem. Inf. Model. 2013, 53, 1371−1387

Table 3. Grooves Parameters (Å) for T5-A20 bp (FLK1), X6-A(G)19 bp (target level, TL), and T7-A18 bp (FLK2)[a]

	FLK1				TL				FLK2			
system	mgw	mgd	Mgw	Mgd	mgw	mgd	Mgw	Mgd	Mgw	mgd	Mgw	Mgd
DNA_U_{ih}A	6.3 (1.5)	4.5 (0.7)	13.1 (1.6)	7.0 (1.8)	7.1 (1.3)	4.5 (0.9)	13.0 (1.8)	7.3 (2.1)	7.5 (1.2)	4.0 (0.7)	12.8 (2.1)	7.6 (1.7)
DNA_U_{ih}G	7.5 (1.2)	3.8 (0.6)	13.3 (2.0)	7.4 (2.0)	8.2 (1.2)	3.8 (0.7)	13.2 (2.1)	8.7 (1.9)	8.3 (1.2)	3.7 (0.6)	13.4 (2.2)	8.8 (1.6)
DNA_T_{ih}A	6.8 (1.5)	4.3 (0.7)	12.9 (1.6)	7.1 (1.9)	7.2 (1.4)	4.4 (0.9)	12.9 (1.7)	7.2 (2.1)	7.1 (1.4)	4.2 (0.8)	13.0 (1.9)	7.1 (1.7)
DNA_C_{ih}G	7.8 (1.2)	4.2 (0. 7)	12.9 (1.7)	6.6 (2.0)	8.0 (1.0)	4.1 (0.7)	13.3 (1.8)	7.7 (1.8)	7.6 (1.1)	4.2 (0.7)	13.4 (2.0)	7.5 (2.0)
DNA_U_{eh}A	6.9 (1.6)	4.1 (0.8)	13.8 (1.9)	7.4 (1.8)	6.2 (1.5)	4.4 (0.7)	14.2 (2.0)	7.6 (1.7)	5.8 (1.5)	4.4 (0.6)	14.1 (2.1)	7.3 (2.0)
DNA_U_{eh}G	8.2 (1.2)	3.2 (0.9)	13.3 (1.8)	7.6 (1.6)	7.1 (1.5)	4.6 (2.1)	12.3 (2.2)	8.4 (3.8)	7.6 (1.5)	3.6 (0.7)	11.9 (1.7)	7.7 (2.8)
DNA_T_{eh}A	9.0 (1.3)	4.2 (1.0)	12.8 (1.9)	6.8 (2.0)	8.6 (1.5)	3.9 (1.0)	13.2 (2.0)	7.1 (2.4)	7.5 (1.6)	4.8 (0.9)	13.6 (1.8)	6.1 (1.5)
DNA_C_{eh}G	9.5 (1.6)	3.2 (1.3)	13.0 (1.6)	6.4 (2.2)	9.1 (1.6)	4.2 (1.5)	12.8 (1.9)	7.4 (2.4)	6.7 (2.0)	4.1 (0.9)	12.9 (2.4)	7.0 (2.2)
E/DNA_U_{ih}A	9.1 (0.4)	3.5 (0.3)	11.8 (1.7)	8.0 (1.2)	7.8 (0.4)	4.5 (0.4)	11.8 (1.7)	8.4 (1.4)	7.4 (0. 8)	4.1 (0.6)	11.7 (1.6)	8.0 (1.4)
E/DNA_T_{ih}A	9.0 (0.8)	3.4 (0.5)	11.4 (2.0)	8.3 (1.1)	7.9 (0.5)	4.6 (0.9)	11.4 (1.9)	7.7 (1.1)	7.0 (1.4)	3.9 (0.4)	11.1 (1.6)	8.2 (1.3)
E/DNA_U_{eh}A	9.9 (1.4)	2.6 (0.8)	12.9 (3.3)	8.1 (1.9)	9.0 (0.9)	4.5 (0.7)	13.5 (3.0)	7.4 (1.7)	7.6 (0.9)	4.1 (0.6)	13.6 (2.3)	7.7 (1.6)
E/DNA_U_{eh}G	9.5 (0.9)	2.7 (0.5)	14.0 (2.8)	7.2 (1.4)	8.6 (0.8)	4.8 (0.6)	13.9 (2.7)	6.7 (1.6)	7.8 (0.9)	3.7 (0.4)	13.1 (2.1)	7.9 (1.3)
E/DNA_T_{eh}A	10.6 (0.8)	1.9 (0.6)	11.2 (2.1)	8.2 (1.7)	10.1 (0.6)	5.2 (0.5)	10.8 (2.0)	7.3 (0.9)	7.6 (1.0)	4.2 (0.6)	11.4 (2.3)	7.6 (1.2)
E/DNA_C_{eh}G	10.3 (0.7)	2.1 (0.4)	13.9 (2.2)	7.4 (1.8)	10.2 (0.6)	4.6 (0.5)	13.4 (2.1)	7.4 (1.8)	7.8 (1.1)	4.1 (0.6)	13.5 (1.9)	7.8 (1.9)
E/DNA_U_{ff}A	9.7 (0.9)	2.4 (0.6)	11.2 (1.7)	8.9 (1.4)	10.4 (1.4)	5.9 (4.7)	10.2 (1.6)	6.8 (5.0)	7.7 (1.4)	3.9 (0.5)	12.3 (2.0)	7.5 (1.2)
E/DNA_U_{ff}G	10.0 (0.4)	1.5 (0.3)	16.5 (1.8)	8.2 (1.4)	15.8 (2.4)	11.6 (3.6)	9.0 (1.0)	−0.3 (1.5)	3.6 (0.8)	5.1 (0.5)	15.9 (2.3)	5.1 (1.5)
E/DNA_T_{ff}A	9.1 (0.6)	2.5 (0.2)	10.8 (2.9)	8.6 (1.2)	11.6 (2.8)	8.1 (4.3)	9.5 (1.9)	4.9 (4.2)	8.4 (1.0)	3.2 (0.5)	13.0 (2.7)	9.5 (1.4)

[a]minor groove width (mgw), minor groove depth (mgd), major groove width (Mgw), major groove depth (Mgd) are reported. Standard deviations are shown in parentheses.

dx.doi.org/10.1021/ci400164f | J. Chem. Inf. Model. 2013, 53, 1371−1387

Table 4. DNA Overall Axis Bend (deg)[a]

system	axis bend
DNA_U_{ih}A	16 (8)
DNA_U_{ih}G	18 (8)
DNA_T_{ih}A	18 (9)
DNA_C_{ih}G	17 (8)
X-ray EHC (PDB: 2OXM)	11
DNA_U_{eh}A	16 (9)
DNA_U_{eh}G	16 (9)
DNA_T_{eh}A	18 (9)
DNA_C_{eh}G	15 (8)
E/DNA_U_{ih}A	20 (8)
E/DNA_T_{ih}A	13 (7)
E/DNA_U_{eh}A	24 (10)
E_{H148G}/DNA_U_{eh}A	16 (9)
E/DNA_U_{eh}G	16 (11)
E/DNA_T_{eh}A	16 (8)
E/DNA_C_{eh}G	15 (8)
E/DNA_U_{ff}A	26 (11)
E/DNA_U_{ff}G	26 (12)
E/DNA_T_{ff}A	22 (10)
E_{H148G}/DNA_U_{ff}A	24 (10)
X-ray FFC (PDB: 1EHM)	26

[a]Standard deviations are shown in parentheses.

active site suggests that the reaction may be favored in the presence of U.[62]

We also calculated the interaction energies of the target base with selected key residues of UNG, which, if mutated, affect its catalytic activity (Table 6). Concerning the IHCs, we observed that His148 contributes to stabilize the **E/DNA_T_{ih}A** adduct more than **E/DNA_U_{ih}A**. An opposite situation instead takes place in the EHCs. This is due to the formation of a U6@O2P...Nδ@His148 H-bond, which occurs only for **E/DNA_U_{eh}A**.

In contrast, Ser270 and Asp145 exert a higher stabilization in the EHCs when T and C are the target bases, although the energetic differences are rather small. Tyr147 has, instead, a remarkable effect in the stabilization of the FFC when U is present. In fact, the target base in the active site pocket interacts via T shape π-stacking interaction with Tyr147 and, consistently with DFT based calculations, the stabilization of U is larger than T[63,64] (Table 6). Remarkably, in **E/DNA_T_{ff}A** a H-bond is present between His148@Nδ and O2P@T6, determining an additional stabilization of this adduct.

3.4. Damage Detection in UNG and in Other Base-Flipping Enzymes. The recognition and repair of damaged DNA bases are the tasks of the BER pathway. This latter is initiated by a variety of DNA glycosylase enzymes, each displaying a different specificity for a particular DNA lesion. A common mechanistic aspect of the BER enzymes is the base-

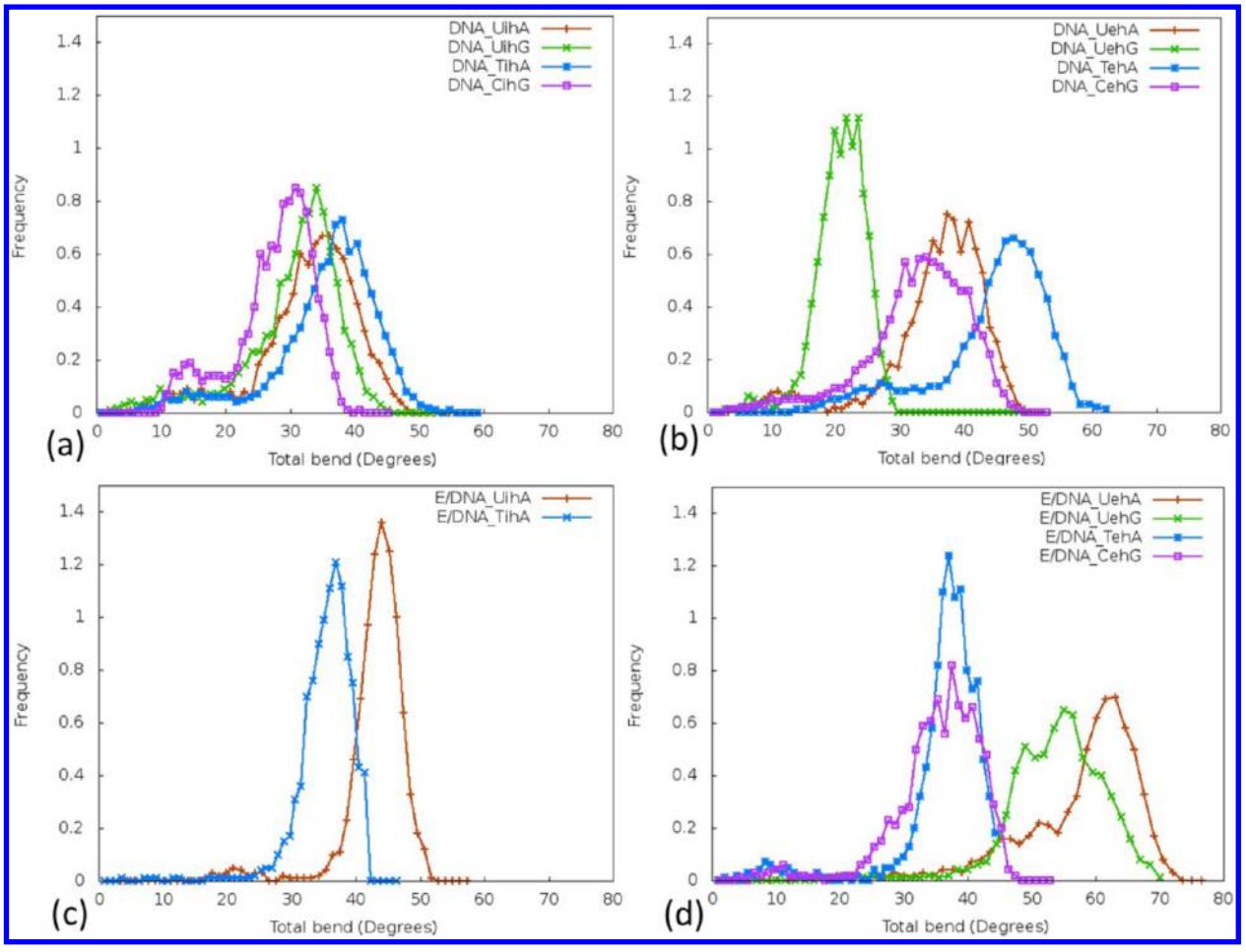


Figure 9. Distribution of overall axis bend (deg) of the DNA during the MD simulation. The **DNA_X_{ih}Y**, **DNA_X_{eh}Y**, **E/DNA_X_{ih}Y**, and **E/DNA_X_{eh}Y** are reported in (a), (b), (c), and (d), respectively.

dx.doi.org/10.1021/ci4001647 | J. Chem. Inf. Model. 2013, 53, 1371−1387

Table 5. Interaction Energies (kcal/mol) between Target Base X (U, T, or C) and UNG and Binding Enthalpies (kcal/mol) of the U and T Base in the EH and FF Conformations Calculated with MM_GBSA[a]

system	electrostatic	van der Waals	total	ΔH
E/DNA_U $_{ih}$A	−29 (9)	−11 (1)	−41 (9)	
E/DNA_T$_{ih}$A	−45 (9)	−7 (2)	−52 (9)	
E/DNA_U$_{eh}$A	−79 (8)	−8 (2)	−87 (8)	1.55 (1.79)
E/DNA_U$_{eh}$G	−64 (8)	−6 (2)	−71 (8)	
E/DNA_T$_{eh}$A	−69 (11)	−8 (2)	−77 (12)	−7.89 (2.68)
E/DNA_C$_{eh}$G	−54 (13)	−6 (2)	−60 (14)	
E/DNA_U$_{ff}$A	−71 (8)	−21 (3)	−92 (8)	−7.84 (3.17)
E/DNA_U$_{ff}$G	−99 (8)	−27 (3)	−126 (8)	
E/DNA_T$_{ff}$A	−97 (16)	−25 (3)	−121 (17)	−14.19 (3.01)

[a]Standard deviations are given in parentheses.

Table 6. Interaction Energies (kcal/mol) between Target Base X (U, T, or C) and Key Residues Involved in U Recognition As Reported in ref [4a]

residue \system	E/DNA_U$_{ih}$A	E/DNA_T$_{ih}$A	E/DNA_U$_{eh}$A	E/DNA_U$_{eh}$G	E/DNA_T$_{eh}$A	E/DNA_C$_{eh}$G	E/DNA_U$_{ff}$A	E/DNA_U$_{ff}$G	E/DNA_T$_{ff}$A
Asp145	22 (6)	23 (8)	**22 (3)**	**20 (4)**	**12 (3)**	**12 (4)**	11 (3)	12 (3)	14 (4)
Tyr147	−14 (3)	−12 (1)	−9 (2)	−8 (3)	−6 (1)	−8 (3)	*−16 (4)*	*−16 (2)*	*−6 (1)*
His148	−25 (8)	−38 (9)	*−69 (11)*	*−66 (7)*	*−34 (11)*	*−17 (12)*	−16 (6)	−23 (6)	−44 (13)
Ser270	−4 (3)	−5 (3)	**−6 (5)**	**−3 (2)**	**−15** (7)	**−14** (7)	−6 (4)	−8 (2)	−5 (3)
His148Gly			−4 (1)				0.2 (0.5)		

[a]Standard deviations are reported in parentheses. Bold residues exert a large influence on the EHCs, while italic residues exert an influence on the FFCs. Only those residues for which differences in the interaction energies between the different states were larger than 5 kcal/mol are listed.

flipping of the damaged base. Thus, in principle, BER enzymes must share a common mechanism for extruding damaged bases from ds DNA and for docking them in their active sites, where the cleavage of the lesion takes place.

Several efforts have been done to characterize the mechanism of the base-flipping process of DNA glycosylases identifying stable or metastable intermediates along the base-flipping pathway. In this work we have structurally characterized different steps of the base-flipping process promoted by UNG in the presence of several target bases with the aim of elucidating how the interrogation process begins and at which stage of the base-flipping path UNG discriminates regular from damaged bases from the structural point of view.[9,54,65−67]

Our study provides evidence that UNG increases the lifetime of U in a EH *syn* state. This may be associated with a large DNA bend imposed by UNG. Since this marked DNA bend occurs only when U is in the EH state, it is likely that the enzyme captures the damaged base when it is already in EH *open* state, induced by thermal fluctuations, consistently with experimental suggestions.[11] Therefore, our results suggest that the enzyme structurally discriminates U with respect to the other bases already at the formation of the interrogation complex.

The metastable EH intermediate identified by our simulations is similar to other transient EH intermediates found in computational studies,[9,10] but it is different from the crystallographic one (PDB: 2OXM), as it is stabilized by a water mediated H-bond network to His148. The fact that crystal structures of crystallographically trapped metastable states do not correspond to intermediates identified in MD simulations is rather common and it is most probably due to the experimental strategies used to capture such transient states.[9,10] Therefore, the crystallographic intermediate displaying a T binding to the UNG *exosite* may be one of the several mestastable intermediates, occurring along the base-flipping path, which should lay at higher energy than the intermediate identified in our simulations.

In order to generalize our findings we compared the structure of the EH *syn-open* intermediate identified in our simulations, and the crystallographic one, with those identified experimentally and computationally for other base-flipping enzymes (Table 7). In particular, we focused on human

Table 7. Crystal Structures of the Interrogation and Lesion Recognition Complexes Reported in the Literature[a]

enzyme	PDB code	substrate	dihedral (deg)
UNG	2OXM[4]	T	87
hOGG1	2I5W[69]	G	−55
hOGG1	1YQK[70]	oxoG	147
MutM	2F5O[69]	G (mg)	18
MutM	3GO8[9]	oxoG	39
MutM	3JR4[42]	G	72
AGT	1YFH[43]	C′	−169

[a]The dihedral angles are measured as the N1 atom of the glycosidic bond, C3′ of the base in the C3′ direction with respect to the target base, C3′ atom and the N1 of the target base.

hOGG1 and bacteria MutM, which are at the basis of the recognition of 8-oxoGuanine (oxoG), on the O6-alkylguanine-DNA alkyltransferase (AGT), which recognizes alkylated G and T[10] and cytosine 5-methyl transferase (CMT), which recognizes C and catalyzes the transfer of a methyl group to it to modulate gene expression.[68] These two latter enzymes, although not belonging to the BER family, share with it the same base-flipping mechanism. For all these enzymes several base-flipping intermediates were identified by X-ray studies,[69,70] and their mechanisms were elucidated by computer simulations.[9,10,42,68]

The crystal structures of the interrogation complexes of MutM and hOGG1 display only minimal structural modi-

dx.doi.org/10.1021/ci4001647 | J. Chem. Inf. Model. 2013, 53, 1371−1387

fications of the target base. For these enzymes no crystal structure exists with a partially EH target base, consistently with the instability of partially flipped intermediates suggested by us and by others.[9,10] In general in the IHCs the recognition of the target base usually comes along with minimal perturbations of the DNA backbone (Table 7 and Figure S6, Supporting Information).[e]

Despite having a fully IH target base, the encounter complexes of MutM and hOGG1 are characterized by a remarkable DNA bend. This suggests that the DNA bend by itself cannot lead the target bases into the EH state. Therefore, in principle, the spontaneous bp DNA breathing may be at the basis of the initial capture of the damage also for other base-flipping enzymes, and, only after the binding of the repair enzyme, the different dynamic behavior of the damaged base with respect to regular bases may lead to its recognition, similarly to what was observed here for UNG.[11]

The situation is different when the LRC is formed. In fact, a very large rotation of the dihedral is observed for hOGG1[42] and for AGT (Figure 10).[43] In both cases, the rotation of the target

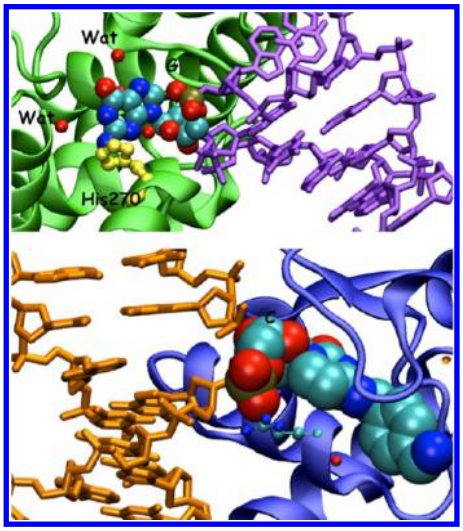


Figure 10. Crystal structure of the LRC of hOOG1 with oxoG completely in a EH position and rotated by 150° toward the enzyme (PDB: 1YQK).[70] His270, acting as a gate during base-flipping, and water molecules in the active site are represented as yellow and red balls, respectively (top). Crystal structure of AGT with a chemically modified and partially flipped C (PDB: 1YFH).[43]

base is significantly larger than that of the partially flipped intermediate trapped crystallographically for UNG (Table 7). Therefore, the presence of the EH *syn-open* intermediate, characterized here for UNG, does not exclude that a LRC may form later in the base-flipping path, similarly to what was observed for hOGG1[70] and AGT.[10] The correlation between DNA bend and the lowering of the free energy barriers for base-flipping is well-known in literature for free DNA sequences[54] and for E/DNA adducts.[9,69] The different DNA bend imposed by UNG for diverse DNA sequences may be associated also to distinct free energy barriers to move different bases along the base-flipping path, although this aspect was not investigated in the present study. In fact, several theoretical studies demonstrate that base-flipping enzymes have a catalytic role during the base-flipping process.[9,10,68] For example, in AGT the kinetic differentiation between regular G and methyl-G base (mG) occurs at the entrance of the active site.[10] Instead, MutM has a remarkable role in reducing the free energy barrier of the damaged base already at the early stage of the base-flipping thanks to a remarkable DNA bending. The DNA bend induced by the enzyme reduces the free energy barrier for the extrusion of the target base by ∼10 kcal/mol.[9,69] The same holds true for CMT for which computational studies demonstrated that the enzyme lowers the free-energy barrier to flip the target C base.[68] All these data point out that a complex and unique picture may exist for BER and other base-flipping enzymes in which an active catalytic role must be played by the enzyme in bending and flipping the damaged base, although for different enzymes this may occur at different steps of the base-flipping path. A detailed investigation of this aspect for UNG requires other computational studies.

4. SUMMARY AND CONCLUSIONS

In this work, we focused on structural and dynamic features that determine the detection of U by UNG. By performing classical MD simulations of regular and damaged DNA sequences we observed that the damaged U base is indistinguishable with respect to regular bases in free DNA sequences.

The difference arises when the UNG binds the DNA oligomer: (i) first, UNG enhances the lifetime of the U in the EH state. The U base, in fact, remains in the *syn-open* conformation, and it is stabilized by a water mediated H-bond network to His148; (ii) second, UNG imposes a remarkable bend to the A:U containing DNA sequence when U is in EH conformation. Since DNA bend is correlated with the free energy cost of base-flipping, the observed marked bend of the DNA in **DNA_U$_{eh}$A** may lower the free energy barrier to push the U base forward along the base-flipping path, as it happens for other BER enzymes.[59,60] However, this aspect requires further computational studies. The observed structural features may be at the basis of the recognition of the A:U with respect to the A:T bp. The simulation performed on **E$_{H148G}$/DNA_U$_{eh}$A** mutant reveals that His148 is mandatory to stabilize the DNA bend, clarifying the key role of this residue as pointed out by mutational studies.[4] In contrast, the G:U bp is characterized by an enhanced opening with respect to a regular WC bp. In fact, G:U is a wobble bp due to the reduced number of H-bonds with respect to G:C .

A comparison of our results with the structural studies of other base-flipping enzymes suggests that the crystal structure of the early recognition complex of UNG bearing a T in the *exosite* is one of the highly unstable base-flipping intermediates. A later LRC may also be present along the base-flipping path of UNG, which may be able to kinetically discriminate the U lesion from regular bases.[11]

Finally, our results indicate that both T and U can bind in the active site being stabilized by a H-bond network to Asn204 and by π-stacking interactions with Tyr147. T is more stable that U inside the catalytic pocket, and probably this may lead to an increase of the free energy barrier for the catalytic reaction.[7] Also in the FFCs His148 plays an important role in tuning the hydration of the catalytic pocket, which may affect the glycosylase efficiency of the enzyme.

We would like finally to remark that in this case molecular simulations have provided insights on an extrusion intermediate, slightly different from the crystallographic one, which would be, otherwise, inaccessible experimentally at physiologic conditions.[f]

dx.doi.org/10.1021/ci4001647 | J. Chem. Inf. Model. 2013, 53, 1371−1387

A detailed understanding of the repair pathways and regulation is important to understand the mechanism of a given number of diseases. In fact, on one hand the selective enhancement of repair pathways may reduce DNA damages, while, on the other hand, the inhibition of DNA repair may help in counteracting the resistance mechanism of several anticancer drugs.[6]

■ ASSOCIATED CONTENT

Supporting Information

Additional supporting Figures S1−S7 and Tables S1−S9 validate and complement the results presented in the article. Tables present water density, number of interfacial waters, and additional DNA parameters. Figures refer to RMSF of the different systems, structural properties of the target base, radial distribution functions, and crystal structures of base flipping enzymes. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: alessandra.magistrato@sissa.it.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We acknowledge the CINECA Grant N. HP10BNQ840, 2011, for the availability of high performance computing resources and support.

■ ADDITIONAL NOTES

[a]This is called a "pinch−push−pull" mechanism in which the protein backbone is pinched together with a consequent remarkable bending of DNA and the U base is kicked out by Leu222.[6]

[b]In different computational and experimental studies the LRC complex is also referred as query complex (QC).

[c]The difference with respect to a unitary charge was distributed on the sugar moiety atoms.

[d]The EH intermediate exists for the entire simulation time, but it is likely that for longer simulation time the U base will return in the IH conformation.

[e]The only exception to this is given by the IC of MutM (PDB: 3JR4) in which the rotation of the G base has occurred toward the mG.

[f]In order to validate our hypothesis it would be useful to perform experimental crystallographic experiments by mutating the base complementary to U (4-methylindole, an adenine analogue unable to form hydrogen bonds) and by mutating in the active site Asp145, which would impair the catalytic activity Tyr147 which stabilizes the substrate in the active site. This should contribute to populate the transient intermediates of the flipping path. Moreover, single-molecule fluorescent energy-transfer experiments on UNG may allow the number of metastable intermediates along the base-flipping path and the transition rates between them to be determined.

■ REFERENCES

(1) Scharer, O. D. Chemistry and biology of DNA repair. *Angew. Chem., Int. Ed.* **2003**, *42*, 2946−2974.

(2) Jackson, S. P.; Bartek, J. The DNA-damage response in human biology and disease. *Nature* **2009**, *461*, 1071−1078.

(3) Wilson, D. M.; Bohr, V. A. The mechanics of base excision repair, and its relationship to aging and disease. *DNA Repair* **2007**, *6*, 544−559.

(4) Parker, J. B.; Bianchet, M. A.; Krosky, D. J.; Friedman, J. I.; Amzel, L. M.; Stivers, J. T. Enzymatic capture of an extrahelical thymine in the search for uracil in DNA. *Nature* **2007**, *449*, 433−437.

(5) Sartori, A. A.; Fitz-Gibbon, S.; Yang, H.; Miller, J. H.; Jiricny, J. A novel uracil-DNA glycosylase with broad substrate specificity and an unusual active site. *EMBO J.* **2002**, *21*, 3182−3191.

(6) Germann, M. W.; Johnson, C. N.; Spring, A. M. Recognition of damaged DNA: structure and dynamic markers. *Med. Res. Rev.* **2012**, *32*, 659−683.

(7) Stivers, J. T. Extrahelical damaged base recognition by DNA glycosylase enzymes. *Chem.—Eur. J.* **2008**, *14*, 786−793.

(8) Parikh, S. S.; Walcher, G.; Jones, G. D.; Slupphaug, G.; Krokan, H. E.; Blackburn, G. M.; Tainer, J. A. Uracil-DNA glycosylase,ÂìDNA substrate and product structures: Conformational strain promotes catalytic efficiency by coupled stereoelectronic effects. *Proc. Natl. Acad. Sci. U.S.A* **2000**, *97*, 5083−5088.

(9) Qi, Y.; Spong, M. C.; Nam, K.; Banerjee, A.; Jiralerspong, S.; Karplus, M.; Verdine, G. L. Encounter and extrusion of an intrahelical lesion by a DNA repair enzyme. *Nature* **2009**, *462*, 762−766.

(10) Hu, J.; Ma, A.; Dinner, A. R. A two-step nucleotide-flipping mechanism enables kinetic discrimination of DNA lesions by AGT. *Proc. Natl. Acad. Sci. U.S.A* **2008**, *105*, 4615−4620.

(11) Friedman, J. I.; Stivers, J. T. Detection of Damaged DNA Bases by DNA Glycosylase Enzymes. *Biochemistry* **2010**, *49*, 4957−4967.

(12) Parikh, S. S.; Mol, C. D.; Slupphaug, G.; Bharati, S.; Krokan, H. E.; Tainer, J. A. Base excision repair initiation revealed by crystal structures and binding kinetics of human uracil-DNA glycosylase with DNA. *EMBO J.* **1998**, *17*, 5214−5226.

(13) Mol, C. D.; Arvai, A. S.; Sanderson, R. J.; Slupphaug, G.; Kavli, B.; Krokan, H. E.; Mosbaugh, D. W.; Tainer, J. A. Crystal structure of human uracil-DNA glycosylase in complex with a protein inhibitor: Protein mimicry of DNA. *Cell* **1995**, *82*, 701−708.

(14) Parker, J. B.; Stivers, J. T. Uracil DNA Glycosylase: Revisiting Substrate-Assisted Catalysis by DNA Phosphate Anions. *Biochemistry* **2008**, *47*, 8614−8622.

(15) Bellamy, S. R. W.; Krusong, K.; Baldwin, G. S. A rapid reaction analysis of uracil DNA glycosylase indicates an active mechanism of base flipping. *Nucleic Acids Res.* **2007**, *35*, 1478−1487.

(16) Krosky, D. J.; Song, F.; Stivers, J. T. The Origins of High-Affinity Enzyme Binding to an Extrahelical DNA Base. *Biochemistry* **2005**, *44*, 5949−5959.

(17) Cao, C.; Jiang, Y. L.; Stivers, J. T.; Song, F. Dynamic opening of DNA during the enzymatic search for a damaged base. *Nat. Struct. Mol. Biol.* **2004**, *11*, 1230−1236.

(18) Friedman, J. I.; Majumdar, A.; Stivers, J. T. Nontarget DNA binding shapes the dynamic landscape for enzymatic recognition of DNA damage. *Nucleic Acids Res.* **2009**, *37*, 3493−3500.

(19) Fadda, E.; Pomes, R. On the molecular basis of uracil recognition in DNA: comparative study of T-A versus U-A structure, dynamics and open base pair kinetics. *Nucleic Acids. Res.* **2011**, *39*, 767−780.

(20) Cao, C.; Jiang, Y. L.; Krosky, D. J.; Stivers, J. T. The Catalytic Power of Uracil DNA Glycosylase in the Opening of Thymine Base Pairs. *J. Am. Chem. Soc.* **2006**, *128*, 13034−13035.

(21) Illuzzi, J. L.; Wilson, D. M. Base excision repair: contribution to tumorigenesis and target in anticancer treatment paradigms. *Curr. Med. Chem.* **2012**, *19*, 3922−3936.

(22) Dominguez, C.; Boelens, R.; Bonvin, A. M. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* **2003**, *125*, 1731−1737.

(23) van Dijk, M.; Bonvin, A. M. J. J. 3D-DART: a DNA structure modelling server. *Nucleic Acids Res.* **2009**, *37*, 235−239.

(24) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Pearlman, D. A.; Crowley, M.; Walker, R. C.; Zhang, W.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Wong, K. F.; Paesani, F.; Wu, X.; Brozell, S.; Tsui, V.;

Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Mathews, D. H.; Schafmeister, C.; Ross, W. S.; Kollman, P. A. *Amber 9*; University of California: San Francisco, 2006.
(25) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781−1802.
(26) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* **2006**, *65*, 712−725.
(27) Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M. Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys. J.* **2007**, *92*, 3817−3829.
(28) Dinner, A. R.; Blackburn, G. M.; Karplus, M. Uracil-DNA glycosylase acts by substrate autocatalysis. *Nature* **2001**, *413*, 752−755.
(29) Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Impey, R.; Klein, M. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(30) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(31) Nosè, S. A unified formulation of the constant temperature molecular-dynamics method. *J. Chem. Phys.* **1984**, *81*, 511−519.
(32) Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **1985**, *31*, 1695.
(33) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the Cartesian equations of motion of a system with constraints; molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327−341.
(34) Lavery, R.; Moakher, M.; Maddocks, J. H.; Petkeviciute, D.; Zakrzewska, K. Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Res.* **2009**, *37*, 5917−5929.
(35) Gervasio, F. L.; Laio, A.; Parrinello, M. Flexible docking in solution using metadynamics. *J. Am. Chem. Soc.* **2005**, *127*, 2600−2607.
(36) Lindahl, E.; Hess, B.; van der Spoel, D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Mod.* **2001**, *7*, 306−317.
(37) Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33−38.
(38) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A.; Parrinello, M. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Comput. Phys. Commun.* **2009**, *180*, 1961−1972.
(39) Miller, B. R. M., T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E MMPBSA.py: An efficient program for end-state free energy calculations. *J. Chem. Theor. Comput.* **2012**, *8*, 3314−3321.
(40) Onufriev, A.; Bashford, D.; Case, D. A. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins* **2004**, *55*, 383−394.
(41) Roberts, V. A.; Pique, M. E.; Hsu, S.; Li, S.; Slupphaug, G.; Rambo, R. P.; Jamison, J. W.; Liu, T.; Lee, J. H.; Tainer, J. A.; Ten Eyck, L. F.; Woods, V. L. Combining H/D exchange mass spectroscopy and computational docking reveals extended DNA-binding surface on uracil-DNA glycosylase. *Nucleic Acids Res.* **2012**, *40*, 6070−6081.
(42) Qi, Y.; Spong, M. C.; Nam, K.; Karplus, M.; Verdine, G. L. Entrapment and structure of an extrahelical guanine attempting to enter the active site of a bacterial DNA glycosylase, MutM. *J. Biol. Chem.* **2010**, *285*, 1468−1478.
(43) Duguid, E. M.; Rice, P. A.; He, C. The structure of the human AGT protein bound to DNA and its implications for damage detection. *J. Mol. Biol.* **2005**, *350*, 657−666.
(44) Brunk, E.; Arey, J. S.; Rothlisberger, U. Role of environment for catalysis of the DNA repair enzyme MutY. *J. Am. Chem. Soc.* **2012**, *134*, 8608−8616.
(45) Reddy, C. K.; Das, A.; Jayaram, B. Do water molecules mediate protein-DNA recognition? *J. Mol. Biol.* **2001**, *314*, 619−632.
(46) Morton, C. J.; Ladbury, J. E. Water mediated protein-DNA interactions: The relationship of thermodynamics to structural detail. *Protein Sci.* **1996**, *5*, 2115−2118.
(47) Fujii, Y.; Shimizu, T.; Kusumoto, M.; Kyogoku, Y.; Taniguchi, T.; Hakoshima, T. Crystal structure of an IRF‚ÄìDNA complex reveals novel DNA recognition and cooperative binding to a tandem repeat of core sequences. *EMBO J.* **1999**, *18*, 5028−5041.
(48) Pingoud, A.; Fuxreiter, M.; Pingoud, V.; Wende, W. Type II restriction endonucleases: structure and mechanism. *Cell. Mol. Life Sci.* **2005**, *62*, 685−707.
(49) Rubinson, E. H.; Gowda, A. S. P.; Spratt, T. E.; Gold, B.; Eichman, B. F. An unprecedented nucleic acid capture mechanism for excision of DNA damage. *Nature* **2010**, *468*, 406−411.
(50) Simona, F.; Magistrato, A.; Dal Peraro, M.; Cavalli, A.; Vila, A. J.; Carloni, P. Common mechanistic features among metallo-beta-lactamases: a computational study of Aeromonas hydrophila CphA enzyme. *J. Biol. Chem.* **2009**, *284*, 28164−28171.
(51) Hong, R.; Magistrato, A.; Carloni, P. Anthrax Lethal Factor Investigated by Molecular Simulations. *J. Chem. Theor. Comput.* **2008**, *4*, 1745−1756.
(52) Przybylski, J. L.; Wetmore, S. D. A QM/QM investigation of the hUNG2 reaction surface: the untold tale of a catalytic residue. *Biochemistry* **2011**, *50*, 4218−4227.
(53) Vargiu, A. V.; Magistrato, A. Detecting DNA mismatches with metallo-insertors: a molecular simulation study. *Inorg. Chem.* **2012**, *51*, 2046−2057.
(54) Fuxreiter, M.; Luo, N.; Jedlovszky, P.; Simon, I.; Osman, R. Role of base flipping in specific recognition of damaged DNA by repair enzymes. *J. Mol. Biol.* **2002**, *323*, 823−834.
(55) Zacharias, M. Minor groove deformability of DNA: a molecular dynamics free energy simulation study. *Biophys. J* **2006**, *91*, 882−891.
(56) Spiegel, K.; Magistrato, A. Modeling anticancer drug-DNA interactions via mixed QM/MM molecular dynamics simulations. *Org. Biomol. Chem.* **2006**, *4*, 2507−2517.
(57) Spiegel, K.; Magistrato, A.; Carloni, P.; Reedijk, J.; Klein, M. L. Azole-bridged diplatinum anticancer compounds. Modulating DNA flexibility to escape repair mechanism and avoid cross resistance. *J. Phys. Chem. B* **2007**, *111*, 11873−11876.
(58) Sgrignani, J.; Magistrato, A. First-Principles Modeling of Biological Systems and Structure-Based Drug-Design. *Curr. Computer-Aided Drug. Des.* **2013**, *9*.
(59) Ramstein, J.; Lavery, R. Energetic coupling between DNA bending and base pair opening. *Proc. Natl. Acad. Sci. U.S.A* **1988**, *85*, 7231−7235.
(60) Ramstein, J.; Lavery, R. Base pair opening pathways in B-DNA. *J. Biomol. Struct. Dyn.* **1990**, *7*, 915−933.
(61) Krosky, D. J.; Song, F.; Stivers, J. T. The origins of high-affinity enzyme binding to an extrahelical DNA base. *Biochemistry* **2005**, *44*, 5949−5959.
(62) Zhachkina, A.; Lee, J. K. Uracil and thymine reactivity in the gas phase: the S(N)2 reaction and implications for electron delocalization in leaving groups. *J. Am. Chem. Soc.* **2009**, *131*, 18376−18385.
(63) Rutledge, L. R.; Campbell-Verduyn, L. S.; Hunter, K. C.; Wetmore, S. D. Characterization of nucleobase-amino acid stacking interactions utilized by a DNA repair enzyme. *J. Phys. Chem. B* **2006**, *110*, 19652−19663.
(64) Cauet, E.; Wintjens, M. R.; Liévin, J.; Biot, C. Histidine-Aromatic Interactions in Proteins and Protein-Ligand Complexes: Quantum Chemical Study of X-ray and Model Structures. *J. Chem. Theor. Comput.* **2005**, *1*, 472−483.
(65) Bergonzo, C.; Campbell, A. J.; de los Santos, C.; Grollman, A. P.; Simmerling, C. Energetic preference of 8-oxoG eversion pathways in a DNA glycosylase. *J. Am. Chem. Soc.* **2011**, *133*, 14504−14506.
(66) Crenshaw, C. M.; Nam, K.; Oo, K.; Kutchukian, P. S.; Bowman, B. R.; Karplus, M.; Verdine, G. L. Enforced presentation of an extrahelical guanine to the lesion recognition pocket of human 8-

oxoguanine glycosylase, hOGG1. *J. Biol. Chem.* **2012**, *287*, 24916−24928.
(67) Hu, J.; Ma, A.; Dinner, A. R. A two-step nucleotide-flipping mechanism enables kinetic discrimination of DNA lesions by AGT. *Proc. Natl. Acad. Sci. U.S.A* **2008**, *105*, 4615−4620.
(68) Huang, N.; Banavali, N. K.; MacKerell, A. D. Protein-facilitated base flipping in DNA by cytosine-5-methyltransferase. *Proc. Natl. Acad. Sci. U.S.A* **2003**, *100*, 68−73.
(69) Banerjee, A.; Santos, W. L.; Verdine, G. L. Structure of a DNA glycosylase searching for lesions. *Science* **2006**, *311*, 1153−1157.
(70) Banerjee, A.; Yang, W.; Karplus, M.; Verdine, G. L. Structure of a repair enzyme interrogating undamaged DNA elucidates recognition of damaged DNA. *Nature* **2005**, *434*, 612−618.