Biophysical Chemistry 152 (2010) 118–127

Contents lists available at ScienceDirect

Biophysical Chemistry

journal homepage: http://www.elsevier.com/locate/biophyschem

Promiscuous DNA alkyladenine glycosylase dramatically favors a bound lesion over undamaged adenine

Anastassia N. Alexandrova *

Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, California, 90095-1569, United States

ARTICLE INFO

Article history:
Received 8 July 2010
Received in revised form 14 August 2010
Accepted 17 August 2010
Available online 26 August 2010

Keywords:
Alkyladenine DNA Glycosylase
Specificity
Free energy perturbation
Thermodynamics

ABSTRACT

The human DNA alkyladenine glycosylase (AAG) is a DNA repair enzyme catalyzing the initial step of DNA repair via lesion excision. Its binding site has an incredible plasticity, and recognizes a variety of DNA lesions resulting from deamination or alkylation of adenine. Based on this plasticity, it is natural to wonder how, and if AAG can discriminate against undamaged adenine. It has even been proposed that it cannot. If, however, AAG is specific, the specificity can be expressed at the stage of the base binding, or base excision. There is also a possibility that the propensity of the base to flip out of the DNA double helix governs its selective subsequent removal. Here, we show that binding to AAG is, in fact, dramatically more thermodynamically favorable for hypoxanthine, at least, a specific lesion, produced by oxidative deamination of adenine, than for undamaged adenine. This preference originates from the more constructive interactions of the lesion with the binding site. Of these, a shorter hydrogen bond between the lesion and the backbone of His136 that does not cause a structural distortion of the base in a major player, in agreement with an earlier kinetic study (P.J. O'Brien and T. J. Ellenberger, J. Biol. Chem. 279 (2004), 9750–9757). Mutating His136 to Pro almost completely eliminates the selectivity. Otherwise, bound adenine and the lesion are positioned very similarly at the binding site, suggesting no predisposition to selective removal. The base flip out of the double helix that precedes the binding has approximately equal thermodynamic facility for the lesion and undamaged adenine. This study employs the Monte Carlo simulations in conjunction with Free Energy Perturbation, and Density Functional Theory (DFT) calculations.

1. Introduction

Deamination and alkylation of DNA bases are inevitable processes that can occur in DNA under its exposure to aggressive chemical agents in the cellular environment. Damaged DNA cannot properly carry the genetic information and must be repaired. DNA repair glycosylases, in particular AAG (EC 3.2.2.21), recognize such lesions, and bind the damaged bases of DNA for subsequent excision [1]. The crystal structure of the AAG protein complexed to damaged DNA was reported in 2000 (Protein Data Bank code 1EWN) [2]. This crystal structure shows that DNA is bound to AAG in a form where the damaged adenine base is rotated out of the DNA double helix. Such base flipping is a phenomenon used by many DNA modifying enzymes, such as methyltransferases, endo- and exonucleases, and glycosylases [3,4]. Base flipping was even conceived to be a more ancient process than carrying the genetic information by DNA [5].

In the DNA-AAG complex [2], the flipped base forms π-stacking interactions with Tyr127 and Tyr159 in the AAG binding site. These are the primary interactions that define the binding affinity. At the same time, the Tyr162 residue of AAG slides into the DNA double helix where it replaces the lesion and π-stacks against the neighboring bases in the strand. The N-glycosidic bond is intact in this DNA-AAG complex. The fact that it is captured in the crystal structure indicates the existence of a minimum on the potential energy surface, whose formation precedes the enzymatic base excision [2]. There remains a possibility that this complex is inactive, and lies far from the reactive one on the potential energy surface. However, Lau et al. [2,6] proposed a reasonable catalytic mechanism based on this crystal structure: Glu125 deprotonates a water molecule, and the forming hydroxyl performs the subsequent nucleophilic attack to cleave the glycosidic bond. Furthermore, it was shown that mutating Glu125 to Ala or Gln makes AAG catalytically inactive, thus supporting the proposed mechanism [2]. Hence, the complex captured in the crystal structure seems to be an adequate starting point for the catalytic bond cleavage. The present study focuses on the structural and energetic aspects of the formation of this complex. Specifically, we interrogate the binding event between AAG and damaged and undamaged DNA in search of

* Tel.: +1 310 825 3769; fax: +1 310 206 4038.
E-mail address: ana@chem.ucla.edu.

0301-4622/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.bpc.2010.08.007

the reasons for specificity toward binding lesions versus undamaged adenine. The specificity of AAG is an open question, and it is important, since AAG participates in the protection of cells from cancer. The concern of whether AAG is specific or not rises, because several base excision enzymes, including AAG, were reported to have highly promiscuous binding sites [7–10]. It was shown, for example, that 3-methyladenine glycosylase (AlkA), and nucleases, can remove undamaged bases at a significant rate [8–10], and the subsequent frequent resynthesis of the bases may in fact cause spontaneous mutations [10].

Lau et al. [2] examined the structure of the DNA-AAG complex and pointed out that the selective recognition of alkylation-damaged bases by AAG could occur at several points along the pathway to base excision: (1) facile base flipping from the DNA double helix, (2) selective binding of damaged bases by AAG, (3) specific positioning of damaged bases that facilitates further excision, and (4) specific electronic effects at the binding site that selectively prepare the glycosydic bond in damaged DNA for cleavage. For each of these points, certain support can be found in literature.

In several studies, it was suggested that the smaller barrier to the flipping of a damaged base out of the double helix with artificially disrupted base pairs maximizes the efficiency with which damaged DNA is recognized by and bound to enzymes [11–14]. For example, a uracil glycosylase exhibited a linear relationship between binding affinity and enthalpic destabilization of the base pair, accessed through manipulation of the chemical nature of the base opposing the lesion [11,12]. A kinetic study by Lyons and O'Brien [10] also gave strong evidence that base pairing in the double helix provides a barrier to base excision by AAG. This study compared glycosylase's activity toward DNA containing a bulge, and DNA containing an oxydatively deaminated adenine paired to different bases. The bulge appeared to be the best context for AAG catalytic efficiency [10]. However, these studies could not directly address excision of undamaged adenine from its usual Watson–Crick (WC) pair context. It is not understood whether this process is even thermodynamically allowed for a normal WC pair, and how its energetics compare to that for WC pairs containing lesions. On the other hand, O'Brien and Ellenberger in their kinetic study [15] proposed that AAG most likely excludes undamaged bases from the active site, due to steric clashes of the amino group on adenine with the binding site of AAG. That constitutes a mechanism for specificity. Additionally, it was suggested that both AAG [16], and AlkA [9] enzymes may discriminate against undamaged purines at the step of the glycosidic bond cleavage, rather than during the binding event. The variety of reported opinions raises a controversy, and renders the question of AAG binding selectivity still open.

In the present work, we compare the binding process for damaged and undamaged DNA to AAG. The considered damaged DNA contains a particular, highly miscoding lesion, hypoxanthine. It might be inappropriate to generalize the results of the present study to other lesions, however. Importantly, the binding process for undamaged DNA is virtually impossible to study experimentally, whereby theory appears to be the only suitable means. The questions that we address are the following: Is base flipping considerably more thermodynamically favorable for the lesion than for undamaged adenine? If adenine is flipped out of the double helix, would its binding to AAG be comparably favorable to lesion's binding? If both the discriminative binding to AAG and unfavorable flipping of adenine contribute to the selectivity of AAG, then which of the two factors contributes to recognition more, in the thermodynamic sense? Also, if AAG can in fact bind adenine without thermodynamic discrimination, do its positioning or electronic effects in the binding site seem to prevent its excision? The present study employs the Monte Carlo simulations in conjunction with Free Energy Perturbation (FEP) methodology, and DFT calculations.

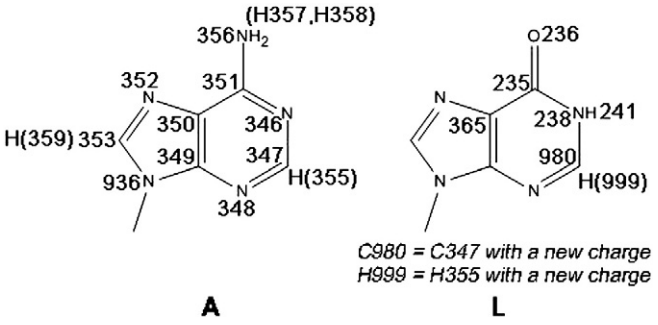
2. System setup and methodology

We considered DNA containing undamaged adenine (**A**), and the lesion (**L**), hypoxanthine, the product of oxidative deamination of adenine (Scheme 1). **L** was reported to be a highly miscoding lesion, relevant to carcinogenesis [17], which makes it an especially interesting target for the present investigation.

The complex of DNA containing a different lesion (the product of alkylation of adenine, εA) and AAG was taken from the Protein Data Bank, PDB code 1EWN (2.1 Å resolution). The DNA was chopped down to five base pairs. The lesion was replaced with **L**. The phosphate groups of the DNA backbone were solvated with Na^+ cations to compensate the negative charge. The protein residues located within 15 Å from the center of the damaged adenine were retained, and the rest of the protein was removed. This way, 134 protein residues were included in the simulations. The total charge on the system was zeroed by means of (de)protonation of residues remote from the binding site and DNA molecule. The reduced size of the system is not anticipated to make a significant impact on our results, since the catalytic action happens far from the system boundaries. A 22 Å water cap centered at the P atom of the damaged nucleotide was placed around the system. The solvent was represented by the TIP4P explicit water model [18]. Water molecules that collided with the protein were removed, and the final cap used in the simulations consisted of 804 water molecules. Similarly, the crystal structure was also used to setup the protein-DNA complexes for DNA containing undamaged **A**. A typical system prepared in this way is shown in Fig. 1A.

For simulations of DNA containing a flipped base in a solution, the initial structures were taken from the complex of DNA with AAG, and AAG was removed. The canonical DNA double helix consisting of five base pairs was prepared based on the crystal structure of a longer double helix (1N4E) [19]. Both types of DNA structures were solvated in a cap containing ca. 1350 TIP4P water molecules (Fig. 1B).

Each system thus prepared was extensively equilibrated using the metropolis Monte Carlo simulations in the NPT ensemble, at 25 °C and 1 atm. The OPLS-AA [20] force field was used to describe the system. The backbone was fixed in the simulations, for both the protein and the DNA molecule, except for the scrutinized adenine or the lesion and the T nucleotide opposing it in the DNA double helix. The angles and torsions were sampled for the entire DNA fragment, and for the protein residues located within 11 Å from the center of the lesion. The bond lengths were sampled only for **A** or **L**. The sodium cations could translate, and the water molecules could translate and rotate. The equilibration consisted of 5×10^6 configurations of sampling where only solvent moves were allowed, and of subsequent 25×10^6 and 15×10^6 configurations, for the DNA-AAG complex and DNA alone, respectively, when both the solvent and the DNA-protein complex were allowed to move. The equilibrated



Scheme 1. Bases considered in this work, with assigned OPLS-AA atom types.

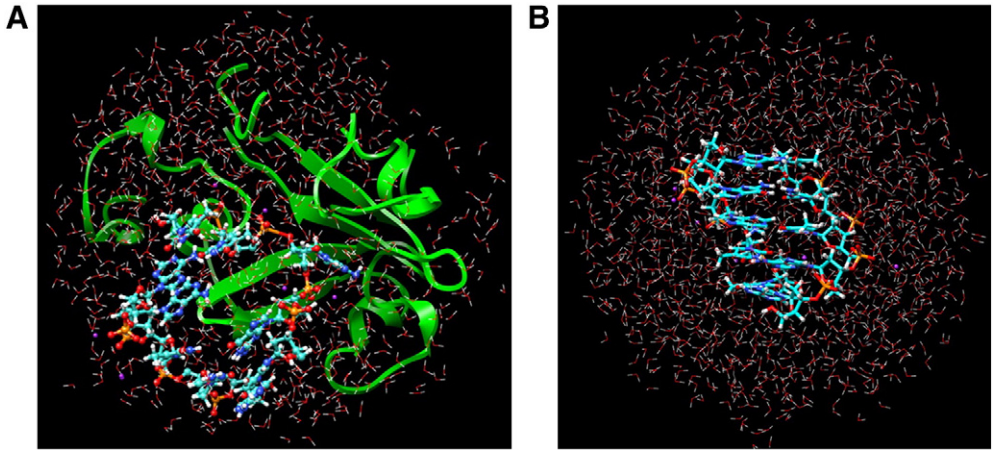


Fig. 1. Computational setup: (A) a part of the AAG protein (its backbone is shown in green) and the damaged DNA fragment bound to it, in a water cap containing 804 water molecules; and (B) a fragment of DNA in a water cap containing 1350 water molecules. The atoms of the DNA molecule are colored as follows: C — cyan, N — blue, O — red, H — white, P — orange, and Na — purple.

systems were then subject to FEP/MC simulations, organized as described in the next paragraphs. During FEP, the system underwent 5×10^6 configurations of solvent equilibration, followed by 10×10^6 configurations of full equilibration, and 25×10^6 configurations of data collection when the perturbation was introduced. All MC simulations were done with *MCPRO 2.0* packages [21].

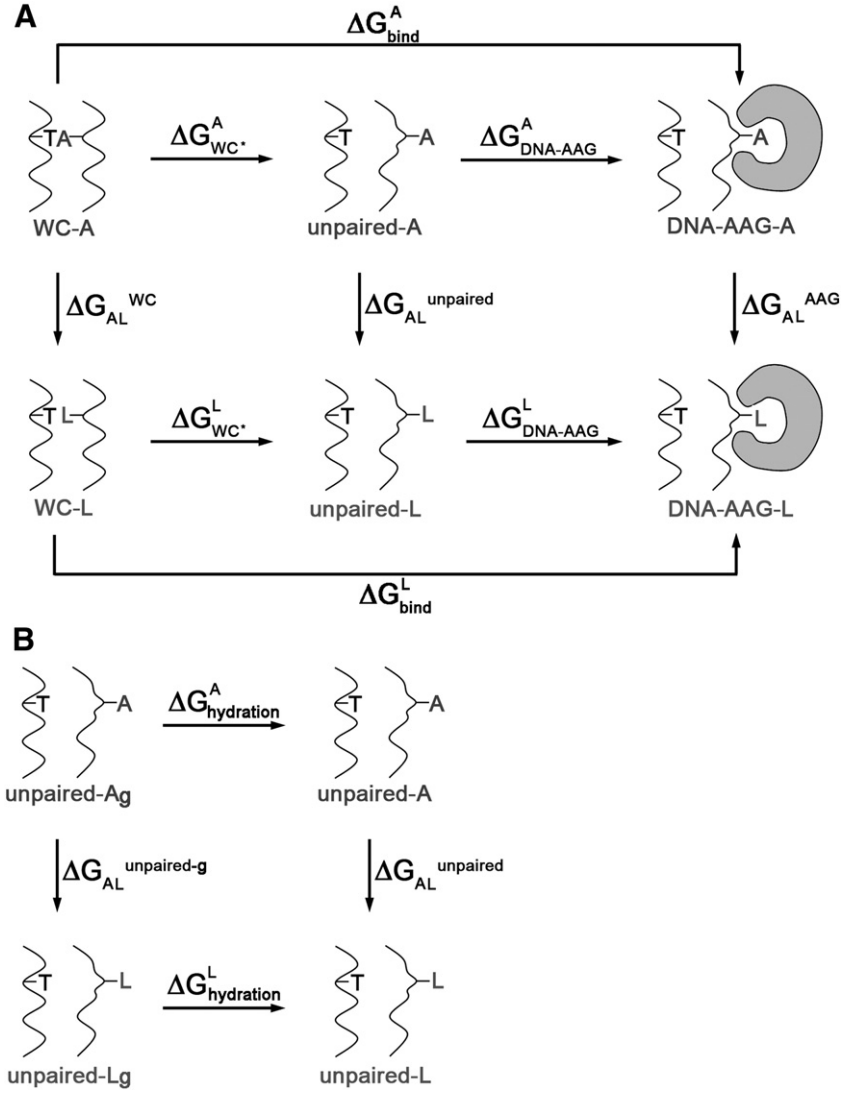


Fig. 2. The thermodynamic cycles used in this work: the top row represents the unpairing of the WC pair, and subsequent binding to AAG for undamaged adenine (**A**); the bottom row represents the same process for the DNA containing a lesion (**L**). The cycle for hydration of DNA containing **A** and **L** is also shown, in the leftmost part of the Figure.

The full process of binding DNA to AAG can be hypothetically split into two events: (1) unpairing of the Watson–Crick (WC) pair accompanied by the rotation of the **A/L** base out of the double helix, and (2) complexation of the flipped base to the AAG active site. The two processes may have complex mechanisms: involve crossing free energy barriers, be multi-step, or even happen simultaneously, in a concerted fashion. However, such a hypothetical split of the binding process is valid, when we are concerned with the associated free energy changes that each step contributes, because free energy is a state function independent of the path. Simulating the entire reaction path is computationally prohibitive. However, it is possible to calculate the relative free energy changes associated with each step, for **A** and **L**. This will allow us to elucidate which step (if any) provides the relatively greater thermodynamic drive for lesion recognition and selective binding by AAG. If the AAG active site is highly discriminating against unmodified adenine as compared to lesions, undamaged DNA would not bind to AAG, even if the WC pair would break. At the same time, if AAG is not that specific, whereas the nucleotide pair is substantially more stable for **A** than for **L**, then WC unpairing would govern the recognition.

The relative free energy changes for **A** and **L** can be computed via FEP simulations, based on the thermodynamic cycles shown in Fig. 2. The following notations are introduced: **WC-A** and **WC-L** are DNA double helixes containing either adenine, or the lesion in a WC pair; **unpaired-A/L(g)** are DNA double helixes containing an unpaired WC pair and flipped adenine or the lesion, where (**g**) denotes the systems in the gas phase; **DNA-AAG-A/L** are complexes between **unpaired-A/L** and AAG. The top row in Fig. 2A represents the binding process for the DNA containing no lesion. In this hypothetical event, first, **A** and T unpair, and the free energy change associated with this is marked $\Delta G_{WC^*}{}^{A}$. **Unpaired-A** then binds to the AAG protein, forming a complex, **DNA-AAG-A**, and the associated free energy change is $\Delta G_{DNA\text{-}AAG}{}^{A}$. The free energy change associated with the entire binding process is $\Delta G_{bind}{}^{A}$. The analogous process, but for the damaged DNA containing the lesion, **L**, is shown in the bottom row in Fig. 2A: **WC-L → unpaired-L → DNA-AAG-L**, and the corresponding free energy changes are $\Delta G_{WC^*}{}^{L}$, $\Delta G_{DNA\text{-}AAG}{}^{L}$, and $\Delta G_{bind}{}^{L}$. We will also use the hydration processes (Fig. 2B), **unpaired-A/L(g) → unpaired-A/L**, and the associated changes in free energy, $\Delta G_{hydration}{}^{A/L}$, as a reference point for the performance of our methodology. In the shown thermodynamic cycles, there are also changes in free energy associated with the mutation of **A** to **L**, in the context of the WC pair, the DNA with a flipped base in the gas phase and in solution, and in the DNA-protein complex. In Fig. 2, they are named ΔG_{AL}^{WC}, $\Delta G_{AL}{}^{unpaired}$, $\Delta G_{AL}{}^{unpaired\text{-}g}$, and $\Delta G_{AL}{}^{AAG}$, respectively. These quantities close the thermodynamic cycles. Since free energy is a state function, the following relations are true:

$$\Delta G_{WC^*}{}^{A} + \Delta G_{AL}{}^{unpaired} = \Delta G_{AL}{}^{WC} + \Delta G_{WC^*}{}^{L} \quad (1)$$

$$\Delta G_{DNA-AAG}{}^{A} + \Delta G_{AL}{}^{AAG} = \Delta G_{AL}{}^{unpaired} + \Delta G_{DNA-AAG}{}^{L} \quad (2)$$

$$\Delta G_{bind}{}^{A} + \Delta G_{AL}{}^{AAG} = \Delta G_{AL}{}^{WC} + \Delta G_{bind}{}^{L} \quad (3)$$

$$\Delta G_{hydration}{}^{A} + \Delta G_{AL}{}^{unpaired} = \Delta G_{AL}{}^{unpaired-g} + \Delta G_{hydration}{}^{L}. \quad (4)$$

From the Eqs. (1)–(4), (5)–(8) follow, respectively:

$$\Delta\Delta G_{WC^*} = \Delta G_{WC^*}{}^{A} - \Delta G_{WC^*}{}^{L} = \Delta G_{AL}{}^{WC} - \Delta G_{AL}{}^{unpaired} \quad (5)$$

$$\Delta\Delta G_{DNA-AAG} = \Delta G_{DNA-AAG}{}^{A} - \Delta G_{DNA-AAG}{}^{L} = \Delta G_{AL}{}^{unpaired} - \Delta G_{AL}{}^{AAG} \quad (6)$$

$$\Delta\Delta G_{bind} = \Delta G_{bind}{}^{A} - \Delta G_{bind}{}^{L} = \Delta G_{AL}{}^{WC} - \Delta G_{AL}{}^{AAG} \quad (7)$$

$$\Delta\Delta G_{hydration} = \Delta G_{hydration}{}^{A} - \Delta G_{hydration}{}^{L} = \Delta G_{AL}{}^{unpaired-g} - \Delta G_{AL}{}^{unpaired}. \quad (8)$$

Expressions (5) and (6) provide the means for estimating the desired quantities, $\Delta\Delta G_{WC^*}$ and $\Delta\Delta G_{DNA\text{-}AAG}$. Also, Eq. (7) allows for the comparison of the entire free energy of binding for damaged and undamaged DNA, $\Delta\Delta G_{bind}$. Eq. (8) provides the relative free energy of hydration of DNA containing **A** versus **L**, and serves the purpose of checking the reasonableness of our calculations.

Thus, instead of performing the computationally prohibitive calculations of ΔG_{WC^*} and $\Delta G_{DNA\text{-}AAG}$, we can compute the free energy changes associated with chemical mutations of **A** to **L**, $\Delta G_{AL}{}^{WC}$, $\Delta G_{AL}{}^{unpaired}$, and $\Delta G_{AL}{}^{AAG}$, and also $\Delta G_{AL}{}^{unpaired\text{-}g}$. These quantities can be obtained using the FEP methodology, where the mutation of **A** to **L** consists of the gradual change of the OPLS-AA parameters on atoms, and a few internal degrees of freedom, corresponding to transforming **A** to **L** through a series of unphysical states. The perturbation was conducted using 20 windows of double-wide sampling, with the value of the coupling parameter λ gradually changing from 0 to 1 (0.05, 0.10, 0.15… 1.0). A window refers to a MC simulation at one point along the mutation coordinate λ, which interconverts two ligands as λ goes from 0 to 1. For each step in the mutation, the quantity

$$\Delta G(1\rightarrow 2) = G_2 - G_1 = -k_b T \ln\left\langle \exp\left(-\frac{E_2 - E_1}{k_b T}\right)\right\rangle 1 \quad (9)$$

was computed, and the total ΔG_{AB} was obtained by summation over the entire range of the parameter λ.

The reported uncertainties ($\pm 1\sigma$) for the free energy changes were obtained from the fluctuation in separate averages over batches of 2×10^6 configurations. Eq. (10) was used, where m is the number of batches, θ_i is the average of property θ for the ith batch, and $<\theta>$ is the overall average for θ [22].

$$\sigma^2 = \sum_i^m (\theta_i - \theta)^2 / m(m-1) \quad (10)$$

Scheme 1 shows the OPLS-AA atom types, and atomic charges for lesions used in this work are presented in Table 1. Existing OPLS-AA atom types were used for all atoms, and only the charges on C2 and H2 were adjusted slightly to keep the total charge on the nucleotide intact. The chosen OPLS-AA charges were compared to those obtained for the lesion in the gas phase, with the Natural Bond Order Analysis (NBO) [23–26] code, at B3LYP [27–29]/6–31G* [30,31], using the *Gaussian Development Version* [32]. A good agreement found between OPLS-AA and NBO charges (Table 1) serves as a justification of our model. No change in the total charge of the nucleotide occurred in the

Table 1
OPLS-AA charges, and NPA B3LYP/6–31G* charges for **L**.

Atom	OPLS-AA	NPA
N9	−0.50	−0.575
C8	+0.20	+0.170
H8	+0.20	+0.236
N7	−0.49	−0.423
C4	+0.38	+0.331
C5	+0.12	−0.047
N3	−0.55	−0.504
C6	+0.50	+0.592
O6	−0.50	−0.509
N1	−0.50	−0.633
H1	+0.30	+0.432
C2	+0.255	+0.255
H2	+0.235	+0.233

perturbation, which led to fairly fast convergence of FEP/MC simulations, and smooth free energy curves. For the **A** to **L** mutation, H atoms on NH_2 were eliminated, and the distances from N6 to the forming dummy atoms were perturbatively shrunk down to 0.1 Å. N6 was mutated to O6. The H1 atom was perturbatively formed, and the N1-H1 distance was perturbatively increased from 0.1 Å to the equilibrium N1-H1 distance from OPLS-AA. Additionally, C6–O6 and C6–N1 bond lengths were included in the perturbation.

Overall, our FEP MC computational approach entails a number of approximations associated with the empirical molecular mechanics force field and water model that lack polarization. Furthermore, the accuracy of FEP simulations depends on the number of windows in the perturbation, and the amount of MC sampling [33]. However, the TIP4P water model is known to reproduce the phase diagram of water in a good agreement with the experiment [34]. Our choice for the OPLS-AA force field, and reliable parameters in MC simulations was guided by the previous successful studies, where the FEP simulations predicted relative binding affinities of potential drugs and inhibitors to proteins, in good agreement with experimentally measured activities [35]. Additionally, it is noteworthy that classical simulations and FEP calculations on DNA in solution, including DNA with perturbed hydrogen bonding within WC pairs have been reported [36a], and provided encouragement in the reliability of the FEP methodology for this type of system. The errors associated with predicted free energy differences for the mutation of T to difluorotoluene in canonical DNA, for example, were found to be within 1 kcal/mol from the experimental values [36a].

The changes in the bond order and degree of charge transfer (CT), at the active site upon substrate binding to AAG were assessed via NBO//LC-ωPBE [37–40]/6–311++G** analysis. The LC-ωPBE functional contains a separation of the exchange component into short-range and long-range parts. The long-range portion of the approximate exchange is replaced by the Hartree–Fock counterpart, and this long-range correction was shown to work particularly well in combination with the short-range variant of the PBE exchange functional. The LC-ωPBE method is known to be remarkably accurate for a broad range of molecular properties, including long-range charge transfer. Additionally, in order to assess the electronic effects at the binding site, and, in particular, π-stacking interactions between the bound base and the Tyr residues, we performed the energy decomposition analysis (EDA) [41,42], as implemented in *Q-Chem 3.2* [43]. The method used in these calculations was B3LYP/6–311++G**. Since dispersion is expected to play a major role in these interactions, but known to be poorly represented with DFT, we performed an additional *a posteriori* DFT-D calculation that utilizes Grimme's empirical dispersion corrections [44]. The performance of DFT-D for noncovalently bound systems including many pure van der Waals complexes were reported to be exceptionally good, reaching, on average, CCSD(T) accuracy [44]. For these analyses, 10 representative low-energy structures from the final stage of MC relaxation were taken for each complex. Each system was truncated so that the remaining model complexes consisted of the 4′H-nucleoside, two phenol molecules representing Tyr127 and Tyr159, a full His amino acid in place of His136, and acetamide in place of the side chain of Asn169 (Fig. 3).

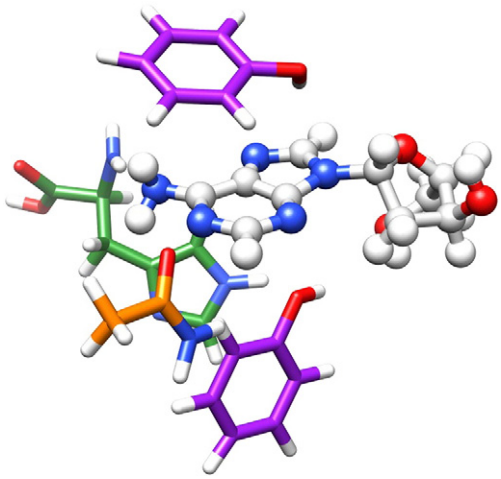


Fig. 3. A typical truncated structure of a DNA-AAG complex used in DFT calculations. Tyr residues are represented with phenol molecules (in purple), Asn is represented with acetamide (in orange), and the full His residue in shown in green.

3. Results and discussion

3.1. Initial Monte Carlo relaxation

Fig. 4 shows the representative configurations taken from the end of the MC relaxation runs for the DNA-AAG complexes. In Fig. 4A, the DNA molecule contains the flipped lesion, **L**. In Fig. 4B, the DNA contains a flipped undamaged **A**. Importantly, in both cases, the geometries of the MC relaxed complexes are similar, and closely resemble the crystal structure. First of all, consistent with the crystal structure, the integrity and length of the glycosidic bond are preserved in the complexes. As an averaged over 25 representative structures from the averaging stage of the MC simulations, the length of the glycosidic bond is 1.48 ± 0.02 Å for **L**, and 1.49 ± 0.02 Å for **A**, which is fairly close to its bond length in adenine alone (1.44 Å, in OPLS-AA). Also, there is almost no difference in the glycosidic bond length between **L** and **A**. Hence, no bond stretching specific to bound **L** occurs at the binding site. In order to provide an additional assessment of the potential antibonding effect on the glycosidic bond at the binding site, we performed NBO//LC-ωPBE/6–311++G** calculations on truncated (as in Fig. 3) structures of the complexes, and on isolated 4′H-nucleosides of optimized gas phase structures. NBO provides the electronic populations of localized bonding and antibonding molecular orbitals, corresponding to the glycosidic bond. And the LC-ωPBE method has a correction for the long-range interactions that is essential for the assessment of the antibonding effect. The results are presented in Table 2. Remarkably, there is almost no difference in the population of the analogous localized orbitals, for the two nucleosides bound to AAG and in the gas phase. Hence, the binding site indeed does not impose any specific antibonding effect on the glycosidic bond of the bound base.

All the important contacts pointed out by Lau et al. for the crystal structure are also found in the MC relaxed structures. Tyr127 forms a parallel π-stacking interaction with the bound nucleotide. On the opposite side, Tyr159 interacts with the π-system of the bound base in a perpendicular orientation.

His136 forms a hydrogen bond to the backbone phosphate. The imidazole fragment of His136 mainly forms a tilted π-stacking contact with Tyr159. Both substrates also form a short hydrogen bond with the backbone of His136, where the lone pair on N is the H-bond acceptor in **A**, and that on carbonyl O is the acceptor in **L**. Noticeably, due to this contact with His136, the NH_2 group in **A** deviates significantly from its equilibrium, almost planar geometry, and clearly stops participating in the resonance with aromatic systems of the base. We estimated the energy cost of this geometric constraint, by taking the base out of the DAN-AAG complex and allowing the NH_2 group to relax to its equilibrium geometry in vacuum, using OPLS-AA. From this estimation, binding of the NH_2 group disfavors the binding of A to AAG by ca. 8 kcal/mol. Additionally, the hydrogen bond to the backbone of His136 is shorter for **L** than for **A**: the distances between the N atom of His136 and N6/O6 atom of the substrate are 3.02 ± 0.11 Å and 2.84 ± 0.07 Å, for **A** and **L**, respectively (averaged over 25 structures). Hence, this

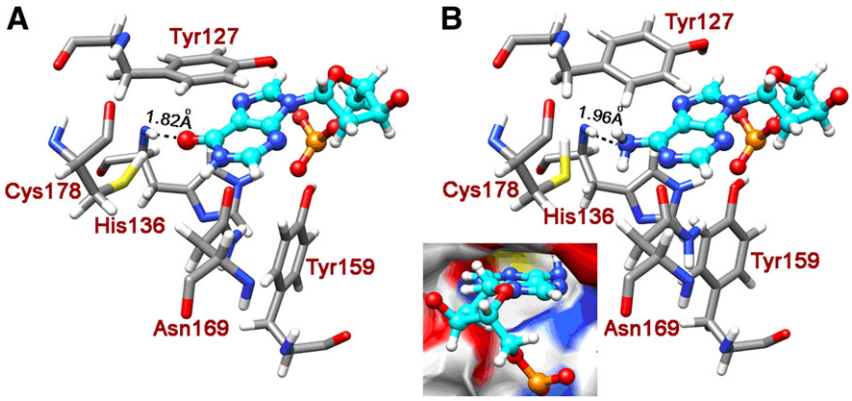


Fig. 4. Representative structures of the DNA-AAG complexes resulting from the Monte Carlo relaxation: only the bound nucleotide (in cyan) and the residues in immediate contact with it (Tyr127, Tyr159, His136, Asn169, and Cys178) are shown. (A) DNA containing **L**, and (B) DNA containing undamaged **A**, and the inset showing the tight fit of **A** into the binding site.

hydrogen bond seems to play a differentiating role in the binding affinity. A more quantitative assessment of the importance of this contact is provided in the next Section.

Asn169 and Cys178 also come into relatively close contact with the substrates. However, Cys178 does not interact with the substrates directly, and, for Asn169, the only noticeable interaction is a long hydrogen bond with the NH group of **L** (Fig. 4A).

From the first glance, the binding of undamaged **A** to the active site looks tight (Fig. 4B, and the inset). No specific positioning of **L**, altered from that of **A**, can be seen. Most of the important contacts with the residues at the active site are retained. **A** did not drift out of the binding site during the MC run, even though the flexibility of its backbone in principle allowed for that. The positioning of **A** and **L** in the binding site are very similar. This structural analysis is suggestive of that AAG is not capable of positioning the lesion so that its excision would be selectively facilitated.

In order to get an additional insight into more subtle electronic effects at the binding site, and to see if those favor one bound base over the other, we performed NBO and EDA analyses (results are presented in Table 3). In these calculations, for each bound base, we used 10 representative structures of the complexes like that shown in Fig. 3. These analyses revealed the components of the binding energy that originated from partial charge transfer (CT), induction, and dispersion. Assessing the π-stacking interactions: according to both NBO and EDA, CT happens both from the bases to Tyr residues, and from Tyr residues to the bases. The net flow of charge is tiny for both substrates; however, the gain in the binding energy due to this delocalization is not negligible. NBO analysis systematically predicts larger net CT than EDA does, for two reasons: 1) NBO was performed with LC-ωPBE, a DFT functional that has a long-range component, and should be more accurate for assessing weakly bound complexes. EDA, on the other hand, was done with B3LYP, because LC-ωPBE is unavailable for the use with EDA. When NBO was done with B3LYP, CT effects were approximately a factor of 2 smaller. 2) When CT is calculated in EDA, the basis set superposition error is subtracted, whereas in NBO it is not. Despite these complementary deficiencies, however, both methods predict that CT contributes similarly to the binding energies, for **L** and **A**, with NBO predicting only a slight favoring of the bound **L** versus **A**. The effect of induction on the binding energy is reported only for the entire complex, since EDA does not permit the pair-wise decomposition of this effect. The energy contributions from induction are small, and similar in magnitude, for **A** and **L**. Dispersion appears to be the main contributor to the enthalpy of binding, as predicted by empirical DFT-D. The dispersion interactions slightly favor the binding of **A** over the binding of **L** (Table 3). Overall, the resulting joint contributions of electronic effects to the enthalpies of binding are similar for the two bases, and do not seem to be a major factor in the selectivity of AAG.

The overlay of base pairs of interest from the representative structures of relaxed canonical DNA double helices containing **A** and **L** are shown in Fig. 5A. Importantly, upon MC relaxation, **L** remains within the double helix and does not flip out. Therefore, there is a barrier to the base flipping process for the lesion. The pairing for **L** is not as perfect as that for **A**. **L** shifts, and extends more into the solvent. However, it still forms two H-bonds to T, and the π-stacking interactions with the bases above and below the **L**-T pair is only slightly disrupted, as compared to undamaged **A**. Hence, the DNA double helix containing **L** has a sound structure. This might be one of the explanations for the highly miscoding nature of the **L** lesion.

For DNA containing flipped bases, the overlaid structures are shown in Fig. 5B. In this case, the resemblance between the structures containing **A** and **L** is obvious. This indicates that in both cases there are minima on the potential energy surfaces corresponding to the

Table 2
The population of the glycosidic bond in model complexes of the nucleosides with AAG and nucleosides alone, from NBO//LC-WPBE/6–311++G* (averaged over 10 configurations).

Base	Pop(C–N), electrons	Pop(C–N)*, electrons
$A_{complex}$	1.986 ± 0.000	0.057 ± 0.003
$A_{isolated}$	1.987	0.051
$L_{complex}$	1.986 ± 0.001	0.060 ± 0.006
$L_{isolated}$	1.988	0.059

Table 3
Averaged electronic contributions to the binding energies between AAG and the bound base, assessed with NBO//LC-ωPBE/6–311++G** and EDA//B3LYP/6–311++G**. Total energy contributions refer to all inter-fragment interactions in considered complexes (Fig. 3); otherwise, contributions from specific interactions between the base and particular residue(s) named in parentheses, but in the context of the full complex, are shown. The empirical contributions from dispersion are assessed with DFT-D. All shown values are in kcal/mol.

Contribution	$A_{complex}$	$L_{complex}$
CT_{NBO} (Tyr127)	−7.47 ± 2.88	−11.75 ± 2.00
CT_{EDA} (Tyr127)	−2.21 ± 0.33	−0.71 ± 0.04
CT_{NBO} (Tyr159)	−2.26 ± 0.84	−3.89 ± 1.12
CT_{EDA} (Tyr159)	−0.66 ± 0.34	−0.89 ± 0.18
Total $induction_{EDA}$	−5.75 ± 1.98	−5.91 ± 0.27
$Dispersion_{EDA}$ (Tyr127,159)	−15.79 ± 1.18	−13.99 ± 0.90
Total $Dispersion_{EDA}$	−25.98 ± 1.07	−21.58 ± 1.07

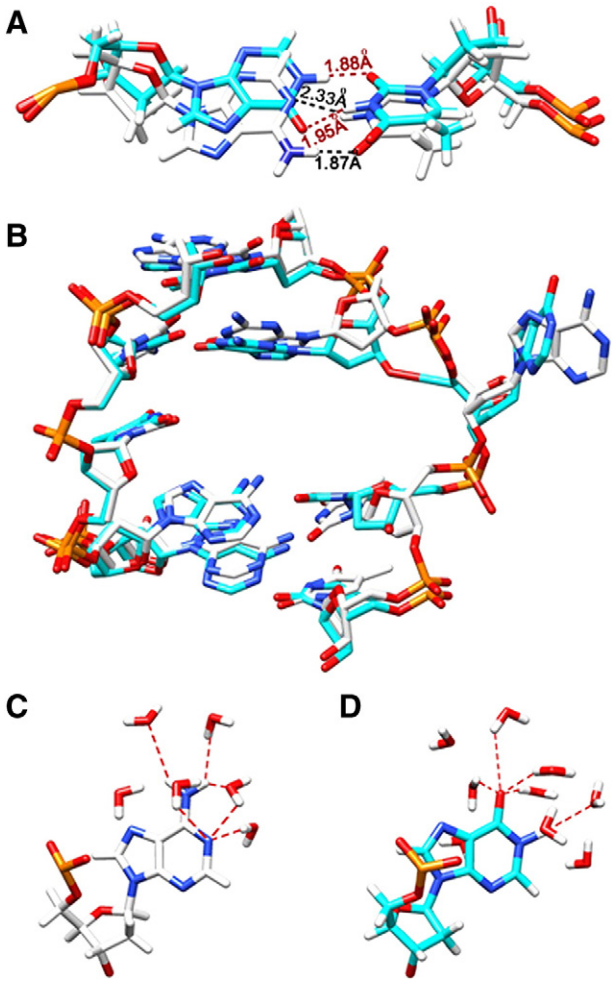


Fig. 5. The representative structures from the MC simulations for the **A**→**L** mutation: systems containing **A** are in white, and systems containing **L** are in cyan. (A) The overlay of the two WC pairs, found in the context of the canonical DNA double helix, with the marked H-bond distances between the bases, showing that in both cases the WC pairs are well-structured. (B) The overlay of the structures of DNA with flipped bases. (C and D) Illustrations of the solvation patterns for flipped **A** and **L**, respectively; structures correspond to those in (B).

Table 4
Relative free energy changes computed using Eqs. (5)–(8), in kcal/mol.

$\Delta\Delta G_{WC^*}$	$\Delta\Delta G_{DNA\text{-}AAG}$	$\Delta\Delta G_{bind}$	$\Delta\Delta G_{DNA\text{-}AAG(H136P)}$	$\Delta\Delta G_{bind(H136P)}$	$\Delta\Delta G_{hydration}$
-2.9 ± 1.1	35.3 ± 0.8	32.3 ± 0.9	9.2 ± 1.2	6.2 ± 1.3	-0.1 ± 0.1

flipped base conformations. Fig. 5C and D shows the solvation patterns for the flipped bases, and illustrate that the solvent network is well-structured in both cases.

3.2. FEP simulations

FEP results for mutation of **A** to **L** in the context of DNA, DNA with flipped base, and DNA-AAG complex are illustrated in Fig. 6 and Table 4. Smooth free energy curves were obtained for the mutation in all structural contexts (Fig. 6A). The convergence of these results is illustrated in Fig. 6B. The plots for the evolution of the error associated with computed ΔG, for the first (curves marked with squares) and last (curves marked with triangles) FEP windows, in each structural context, are shown as a function of the number of configurations in MC. Noticeably, the error increases as the perturbation progresses from **A** to **L**. Nevertheless, the sampling was sufficient for the convergence of the errors to stable values (curves on the plot level up at right). To check the performance of FEP simulations, we also tested the mutation in the reverse direction: from **L** to **A**, and found that the absolute values of the resulting ΔGs are within 1 kcal/mol from the ones reported here.

The initial and final structures in the perturbations were always very close to those resulting from MC relaxation (described in the previous section). Consistent with the MC relaxation simulations, all three structural contexts appear to correspond to minima on the potential energy surface for both bases.

Using Eqs. (5)–(8), relative changes in free energies associated with the base flip, $\Delta\Delta G_{WC^*}$, complexation, $\Delta\Delta G_{DNA\text{-}AAG}$, entire binding event, $\Delta\Delta G_{bind}$, and hydration of DNA with flipped bases, $\Delta\Delta G_{hydration}$, were computed (Table 4). They indicate the ease with which these processes occur in the presence of the mutation, relative to the systems containing undamaged **A**. The difference between $\Delta\Delta G_{WC^*}$ and $\Delta\Delta G_{DNA\text{-}AAG}$ for the lesion shows whether base flipping or complexation with AAG is thermodynamically dominant in the binding event.

$\Delta\Delta G_{WC^*}$ appears to be slightly negative (-2.9 ± 1.1 kcal/mol). This means that the base flipping is a little more facile for DNA containing **A**

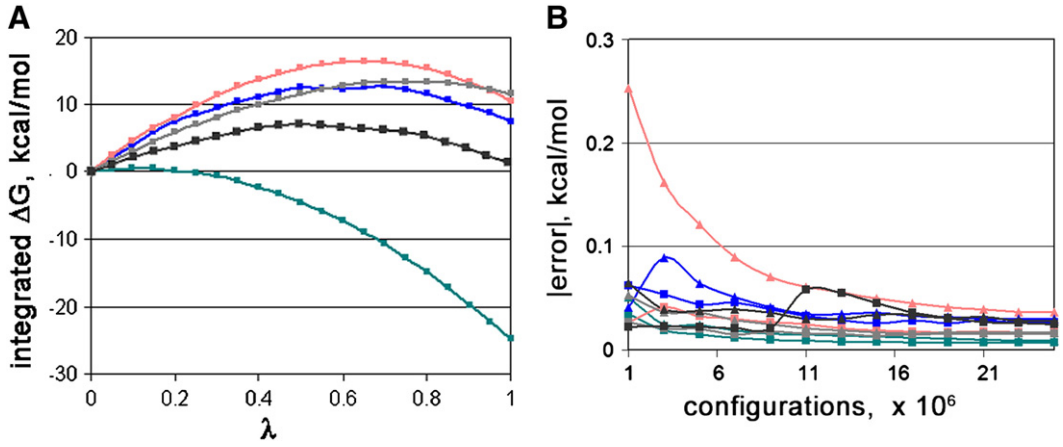


Fig. 6. FEP results for the **A**→**L** mutation: blue curves — DNA double helix in solution, pink — DNA containing a flipped base, grey — DNA with a flipped base in the gas phase, green — in the context of DNA-AAG complex, and black — in the context of DNA-AAG His136Pro mutant complex. (A) FEP curves, (B) The convergence plots for the error in computed ΔG as a function of the number of MC configurations at the stage of averaging: the two curves of each respective color belong to the beginning of the perturbation (marked with squares), and at the end (marked with triangles).

than for DNA containing **L**. Two reasons could be suggested for this: firstly, the hydrogen bonding in the WC pair is changed but not completely disrupted upon mutation (Fig. 5A). The **L** base forms two hydrogen bonds to T. The base is also slightly more exposed to the solvent, and thus better solvated, in the context of canonical DNA. Thus, for both bases, the canonical DNA seems to be a stable minimum on the potential energy surface. Secondly, both bases in the flipped conformation are quite efficiently solvated (Fig. 5B). Hence, the thermodynamics of base flipping is expected to be similar in the two cases. The barriers to base flipping are probably different for **L** and **A**, but the difference might be small, considering that the number of H-bonds in the WC pairs is the same. Our simulations do not address the kinetics, however. Importantly, the fact that **A** has a greater propensity to flip out of canonical DNA than **L** does indicates that thermodynamics of base flipping cannot contribute to the selectivity of AAG. We would like to point out that this conclusion is likely not universal for all known lesions. For example, *εA*, which is bound to AAG in the crystal structure that we are using, is larger than **L**, and does not have the capability of forming H-bonds to the opposing T in canonical DNA. Hence, the thermodynamics of its flipping from DNA is expected to be different.

$\Delta\Delta G_{DNA\text{-}AAG}$ is 35.0 ± 0.8 kcal/mol, where the positive sign indicates a greater binding affinity between AAG and **L** than between AAG and **A**. Importantly, MC equilibration of complexes between AAG and the flipped bases (Fig. 4A and C) showed a comfortable docking of both **L** and **A**, as previously described. Only the FEP simulations were capable of revealing this qualitative difference in the binding affinity. The overall $\Delta\Delta G_{bind}$ is 32.3 ± 0.9 kcal/mol, which indicates that the entire process of binding to AAG is much more favorable for **L** than for undamaged **A**. Hence, thermodynamically, AAG is capable of discriminating against **A**, and preferentially binds **L**. In comparing the values of $\Delta\Delta G_{WC^*}$ and $\Delta\Delta G_{DNA\text{-}AAG}$, it is the latter that defines the favorable thermodynamics of binding **L**, versus **A**. Hence, summarizing the FEP results, if the thermodynamics governs the selective binding of the lesion, the selectivity should be driven by the greater affinity of the AAG binding site to **L**, rather than its unfavorable WC pairing.

In order to gain further insight into this big difference in the binding affinity, we performed additional FEP simulations allowing us to explore the role of H-bond to the backbone of His136. Specifically, we mutated His136 to Pro, a residue that is unable to form the aforementioned hydrogen bond to the substrates, and mutated **A** to **L** in the context of the mutated binding site. The FEP results are presented in Fig. 7 and Table 4. Both bases are accommodated by the mutated binding site (Fig. 7). The π-stacking interactions with Tyr127 and Tyr159 are preserved in the two complexes. Remarkably, however, $\Delta\Delta G_{DNA\text{-}AAG}$ for the mutant is 9.2 ± 1.2 kcal/mol (compare to 35.3 ± 0.8 kcal/mol in the original protein). $\Delta\Delta G_{bind}$ drops from 32.3 ± 0.9 kcal/mol in AAG to 6.2 ± 1.3 kcal/mol in the His136Pro mutant. The relative binding affinity of AAG changes dramatically upon mutation, and no longer favors **L** over **A** as strongly. These simulations do not assess the exact contribution of this hydrogen bond to the free energy of binding, since other interactions at the binding site also get perturbed upon mutation. However, qualitatively, these FEP results indicate the significance of this hydrogen bond for the binding affinity. Importantly, AAG operates on products of deamination and alkylation of **A**, and therefore, all lesions can be expected to form shorter H-bonds to His136 than undamaged **A**. **A** also undergoes a structural distortion, pyramidalization of the amino group, due to this hydrogen bond, whereas for deaminated lesions this should not be an issue. This suggests that the His136 hydrogen bond, as a major reason for the binding specificity of the AAG active site, could be valid for all lesions. However, this remains to be addressed in future studies.

The $\Delta\Delta G_{hydration}$ value (Table 4) is used here as a check for the performance of the methodology, since hydration is the process for which we have the most intuition and literature available. The $\Delta\Delta G_{hydration}$ value is small, and its sign indicates that DNA containing a flipped undamaged **A** base is solvated slightly better than that containing **L**. This is an intuitive result illustrated by the solvation patterns for the flipped bases shown in Fig. 5C and D. Reasonable magnitudes of $\Delta\Delta G_{hydration}$ indicate that the FEP simulations are well-behaved.

4. Conclusions

Recognition of damaged sites in DNA, and their successful excision by enzymes like AAG are the key events in cellular resistance to cancer. Understanding the mechanisms for the specificity of AAG toward binding damaged DNA versus undamaged DNA is, therefore, critical for the development of means of cancer treatment and prevention. There is some question as to whether AAG is in fact capable of discriminating against undamaged adenine. This study sheds light on the process of binding of damaged (**L**, Scheme 1) and undamaged adenine (**A**) to AAG. The specific studied **L** is a product of oxidative deamination of **A**, known to be highly miscoding. We address various stages of the binding process, and elucidate that the actual docking to the active site provides a great thermodynamic drive for the selectivity toward binding the lesion. Specifically,

(1) we show that the canonical DNA double helix containing the lesion is well-structured: there are two H-bonds between **L** and the opposing T, and the π-stacking interactions with the bases below and above the base of interest are structurally sound, for both **A**, and **L**. The structure and solvation of DNA containing flipped bases are also similar.

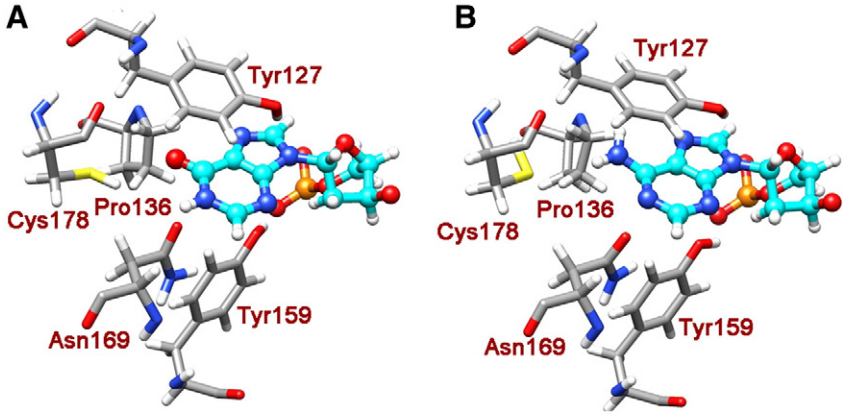


Fig. 7. The typical structures of the Hi136Pro variant of AAG binding (A) **L**, and (B) undamaged **A**, from the end points of the **A** → **L** mutation. One may see that the NH_2 group on A is not distorted, as in the case of the original protein.

(2) FEP/MC simulations suggest that base flipping out of the double helix is thermodynamically slightly more facile for **A** than for **L**. Thus, the thermodynamics of base flipping does not constitute the mechanism for selectivity. Importantly, not all lesions are expected to form sound WC structures, and therefore the energetics of base flipping might also be different for other lesions.

(3) Both studied bases, **A**, and **L**, can tightly dock into the active site of AAG. The conformation at which they bind is very similar in both cases, and also reminiscent of that in the crystal structure. There is no specific positioning of the substrate, such that it would selectively facilitate the cleavage binding of **L**, and not of **A**. NBO and EDA analyses on the bound complexes showed that the electronic effects at the binding site upon binding **A** and **L** are similar. There is no specific antibonding effect on the glycosidic bond. CT, induction, and dispersion interactions similarly contribute to the binding affinity of AAG to the two bases. The only significant difference in the bound complexes is the shorter hydrogen bond to His136 for **L** than for **A**, and the associated costly geometric distortion of **A**, as compared to the base in the gas phase or solution.

(4) The thermodynamic propensity for complexation of flipped bases with AAG is dramatically greater for **L** versus **A**. We further show that the thermodynamic specificity is majorly due to the H-bond between the base and the backbone of His136, in agreement with an earlier report by O'Brien and Ellenberger [15]. Upon mutation of His136 to Pro, the binding specificity is dramatically reduced. Since AAG operates on deaminated and alkylated **A**, all lesions may be able to display this effect. However, this issue needs to be further investigated. There is also an entropic contribution to the found thermodynamic propensity, which is likely to favor **L** versus **A**, due to the smaller size of **L**. The entropic contributions are hard to estimate, however.

Thus, AAG is capable of discriminating against undamaged adenine at least by virtue of the more favorable thermodynamics of binding the lesion, **L**. The kinetic contribution to the specific binding is not assessed in the present study. Also, this work focuses on the binding event only, and does not provide an insight into the catalytic bond cleavage. Further mechanistic investigations are required to address this.

Acknowledgements

This work was supported by the American Cancer Society, PF-08-165-01GMC, and University of California, Los Angeles. We thank Drs. Julian Tirado-Rives and Ryan P. Steele for a helpful discussion.

References

[1] A. Sancar, DNA excision repair, Annu. Rev. Biochem. 65 (1996) 43–81.
[2] A.Y. Lau, M.D. Wyatt, B.J. Glassner, L.D. Samson, T. Ellenberger, Proc. Natl. Acad. Sci. USA 97 (2000) 13573.
[3] R.J. Roberts, X. Cheng, Base flipping, Annu. Rev. Biochem. 67 (1998) 181–198.
[4] R.D. Wood, Nucleotide excision repair in mammalian cells, J. Biol. Chem. 272 (1997) 23465–23468.
[5] R.J. Roberts, On base flipping, Cell 82 (1995) 9–12.
[6] A.Y. Lau, O.D. Scharer, L. Samson, G.L. Verdine, T. Ellenberger, Crystal structure of a human alkylbase-DNA repair enzyme complexed to DNA, Cell 95 (1998) 249–258.
[7] J.T. Stivers, Y.L. Jiang, A mechanistic perspective on the chemistry of DNA repair glycosylases, Chem. Rev. 103 (2003) 2729–2760.
[8] K.G. Berdal, R.F. Jhansen, E. Seeberg, Release of normal bases from intact DNA by a native DNA repair enzyme, EMBO J. 17 (1998) 363–367.
[9] P.J. O'Brien, T.J. Ellenberger, DNA: replication, repair, and recombination: the *Escherichia coli* 3-methyladenine DNA glycosylase AlkA has a remarkably versatile active site, J. Biol. Chem. 279 (2004) 26876–26884.
[10] M.E. Branum, J.T. Reardon, A. Sancar, DNA repair excision nuclease attacks undamaged DNA: a potential source of spontaneous mutations, J. Biol. Chem. 276 (2001) 25421–25426.
[11] D.J. Krosky, F.P. Schwarz, J.T. Stivers, Linear free energy correlations for enzymatic base flipping: how do damaged base pairs facilitate specific recognition? Biochemistry 43 (2004) 4188–4195.
[12] D.J. Krosky, F. Song, J.T. Stivers, The origins of high-affinity enzyme binding to an extrahelical DNA base, Biochemistry 44 (2005) 5949–5959.
[13] D.M. Lyons, P.J. O'Brien, Efficient recognition of an unpaired lesion by a dna repair glycosylase, J. Am. Chem. Soc. 131 (2009) 17742–17743.
[14] M.D. Wyatt, L.D. Samson, Influence of DNA structure on hypoxanthine and 1, N6-ethenoadenine removal by murine 3-methyladenine DNA glycosylase, Carcinogenesis 21 (2000) 901–908.
[15] P.J. O'Brien, T.J. Ellenberger, Dissecting the broad substrate specificity of human 3-methyladenine-DNA glycosylase, J. Biol. Chem. 279 (2004) 9750–9757.
[16] E.E. Connor, M.D. Wyatt, Active-site clashes prevent the human 3-methyladenine DNA glycosylase from improperly removing bases, Chem. Biol. 9 (2002) 1033–1041.
[17] M. Yasui, E. Suenaga, N. Koyama, C. Masutani, F. Hanaoka, P. Gruz, S. Shibutani, T. Nohmi, M. Hayashi, M. Honma, Miscoding properties of 2′-deoxyinosine, a nitric oxide-derived DNA adduct, during translesion synthesis catalyzed by human DNA polymerases, J. Mol. Biol. 377 (2008) 1015–1023.
[18] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, W. Impey, M.L. Klein, Comparison of simple potential functions for simulating liquid water, J. Chem. Phys. 79 (1983) 926–935.
[19] H. Park, K. Zhang, Y. Ren, S. Nadji, N. Sinha, J.-S. Taylor, C. Kang, Polymerization of calsequestrin. Implications for Ca^{2+} regulation, Proc. Natl. Acad. Sci. USA 99 (2002) 15965–15970.
[20] W.L. Jorgensen, D.S. Maxwell, J. Tirado-Rives, Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids, J. Am. Chem. Soc. 118 (1996) 11225–11236.
[21] W.L. Jorgensen, J. Tirado-Rives, Molecular modeling of organic and biomolecular systems using *BOSS* and *MCPRO*, J. Comput. Chem. 26 (2005) 1689–1700.
[22] M.P. Allen, D.J. Tildesley, Computer Simulations of Liquids, Clarendon, Oxford, 1987.
[23] J.E. Carpenter, F. Weinhold, Analysis of the geometry of the hydroxymethyl radical by the "different hybrids for different spins" natural bond orbital procedure, J. Mol. Struct. (Theochem) 169 (1988) 41–62.
[24] J.P. Foster, F. Weinhold, Natural hybrid orbitals, J. Am. Chem. Soc. 102 (1980) 7211–7218.
[25] A.E. Reed, F. Weinhold, Natural bond orbital analysis of near-Hartree–Fock water dimer, J. Chem. Phys. 78 (1983) 4066–4073.
[26] A.E. Reed, L.A. Curtiss, F. Weinhold, Intermolecular interactions from a natural bond orbital, donor-acceptor viewpoint, Chem. Rev. 88 (1988) 899–926.
[27] R.G. Parr, W. Yang, Density-functional Theory of Atoms and Molecules, Oxford Univ. Press, Oxford, 1989.
[28] A.D. Becke, Density-functional thermochemistry. III. The role of exact exchange, J. Chem. Phys. 98 (1993) 5648–5652.
[29] J.P. Perdew, J.A. Chevary, S.H. Vosko, K.A. Jackson, M.R. Pederson, D.J. Singh, C. Fiolhais, Atoms, molecules, solids, and surfaces: applications of the generalized gradient approximation for exchange and correlation, Phys. Rev. B 46 (1992) 6671–6687.
[30] A.D. McLean, G.S. Chandler, Contracted Gaussian basis sets for molecular calculations. I. Second row atoms, Z = 11–18, J. Chem. Phys. 72 (1980) 5639–5648.
[31] M.J. Frisch, J.A. Pople, J.S. Binkley, Self-consistent molecular orbital methods 25. Supplementary functions for Gaussian basis sets, J. Chem. Phys. 80 (1984) 3265.
[32] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, J.A. Montgomery, T. Vreven, K.N. Kudin, J.C. Burant, J.M. Millam, S.S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G.A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J.E. Knox, H.P. Hratchian, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, P.Y. Ayala, K. Morokuma, G.A. Voth, P. Salvador, J.J. Dannenberg, V.G. Zakrzewski, S. Dapprich, A.D. Daniels, M.C. Strain, O. Farkas, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J.V. Ortiz, Q. Cui, A.G. Baboul, S. Clifford, J. Cioslowski, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, M. Challacombe, P.M.W. Gill, B. Johnson, W. Chen, M.W. Wong, C. Gonzalez, J.A. Pople, Gaussian Development Version, Revision G.03 Wallingford, CT, 2008.
[33] W.L. Jorgensen, L.L. Thomas, Perspective on free-energy perturbation calculations for chemical equilibria, J. Chem. Theory Comput. 4 (2008) 869–876.
[34] (a) E. Sanz, C. Vega, J.L.F. Abascal, L.G. MacDowell, Tracing the phase diagram of the four-site water potential (TIP4P), J. Chem. Phys. 121 (2004) 1165–1166;
(b) D. Paschek, Temperature dependence of the hydrophobic hydration and interaction of simple solutes: an examination of five popular water models, J. Chem. Phys. 120 (2004) 6674–6690.
[35] (a) W.L. Jorgensen, Efficient drug lead discovery and optimization, Acc. Chem. Res. 42 (2009) 724–733 and references therein;
(b) J.G. Zeevaart, W. Ligong, V.V. Thakur, C.S. Leung, J. Tirado-Rives, C.M. Bailey, R.A. Domaoal, K.S. Anderson, W.L. Jorgensen, Optimization of azoles as anti-human immunodeficiency virus agents guided by free-energy calculations, J. Am. Chem. Soc. 130 (2008) 9492–9499;
(c) C.S. Leung, J.G. Zeevaart, R.A. Domaoal, M. Bollini, V.V. Thakur, K.A. Spasov, K.S. Anderson, W.L. Jorgensen, Eastern extension of azoles as non-nucleoside inhibitors of HIV-1 reverse transcriptase; cyano group alternatives, Bioorg. Med. Chem. Lett. 20 (2010) 2485–2488;

(d) C.R.W. Guimaraes, D.J. Kopecky, J. Mihalic, S. Shen, S. Jeffries, S.T. Thibault, X. Chen, N. Walker, M. Cardozo, Thermodynamic analysis of mRNA cap binding by the human initiation factor eIF4E via free energy perturbations, J. Am. Chem. Soc. 131 (2009) 18139–18146;
(e) W.L. Jorgensen, J. Ruiz-Caro, J. Tirado-Rives, A. Basavapathruni, K.S. Andersonb, A.D. Hamilton, Computer-aided design of non-nucleoside inhibitors of HIV-1 reverse transcriptase, Bioorg. Med. Chem. Lett. 16 (2006) 663–667.

[36] (a) E. Cubero, C.A. Laughton, F.J. Luque, M.J. Orozco, Molecular dynamics study of oligonucleotides containing difluorotoluene, J. Am. Chem. Soc. 122 (2000) 6891–6899;
(b) J. Florián, M.F. Goodman, A. Warshel, Free-energy perturbation calculations of DNA destabilization by base substitutions: the effect of neutral guanine·thymine, adenine·cytosine and adenine·difluorotoluene mismatches, J. Phys. Chem. B 104 (2000) 10092–10099.

[37] Y. Tawada, T. Tsuneda, S. Yanagisawa, T. Yanai, K. Hirao, A long-range-corrected time-dependent density functional theory, J. Chem. Phys. 120 (2004) 8425–8433.

[38] O.A. Vydrov, G.E. Scuseria, Assessment of a long-range corrected hybrid functional, J. Chem. Phys. 125 (2006) 234109.

[39] O.A. Vydrov, J. Heyd, A. Krukau, G.E. Scuseria, Importance of short-range versus long-range Hartree-Fock exchange for the performance of hybrid density functionals, J. Chem. Phys. 125 (2006) 074106.

[40] O.A. Vydrov, G.E. Scuseria, J.P. Perdew, Tests of functionals for systems with fractional electron number, J. Chem. Phys. 126 (2007) 154109.

[41] R.Z. Khaliullin, E.A. Cobar, R.C. Lochan, A.T. Bell, M. Head-Gordon, Unravelling the origin of intermolecular interactions using absolutely localized molecular orbitals, J. Phys. Chem. A 111 (2007) 8753–8765.

[42] R.Z. Khaliullin, A.T. Bell, M. Head-Gordon, Analysis of charge transfer effects in molecular complexes based on absolutely localized molecular orbitals, J. Chem. Phys. 128 (2008) 184112.

[43] Y. Shao, L. Fusti-Molnar, Y. Jung, J. Kussmann, C. Ochsenfeld, S.T. Brown, A.T.B. Gilbert, L.V. Slipchenko, S.V. Levchenko, D.P. O'Neill, R.A. DiStasio Jr., R.C. Lochan, T. Wang, G.J.O. Beran, N.A. Besley, J.M. Herbert, C.Y. Lin, T. Van Voorhis, S.H. Chien, A. Sodt, R.P. Steele, V.A. Rassolov, P.E. Maslen, P.P. Korambath, R.D. Adamson, B. Austin, J. Baker, E.F.C. Byrd, H. Dachsel, R.J. Doerksen, A. Dreuw, B.D. Dunietz, A.D. Dutoi, T.R. Furlani, S.R. Gwaltney, A. Heyden, S. Hirata, C.-P. Hsu, G. Kedziora, R.Z. Khaliullin, P. Klunzinger, A.M. Lee, M.S. Lee, W. Liang, I. Lotan, N. Nair, B. Peters, E.I. Proynov, P.A. Pieniazek, Y.M. Rhee, J. Ritchie, E. Rosta, C.D. Sherrill, A.C. Simmonett, J.E. Subotnik, H.L. Woodcock III, W. Zhang, A.T. Bell, A.K. Chakraborty, D.M. Chipman, F.J. Keil, A. Warshel, W.J. Hehre, H.F. Schaefer III, J. Kong, A.I. Krylov, P.M.W. Gill, Z. Gan, Y. Zhao, N.E. Schultz, D. Truhlar, E. Epifanovsky, M. Oana, R. Baer, B.R. Brooks, D. Casanova, J.-D. Chai, C.-L. Cheng, C. Cramer, D. Crittenden, A. Ghysels, G. Hawkins, E.G. Hohenstein, C. Kelley, W. Kurlancheek, D. Liotard, E. Livshits, P. Manohar, A. Marenich, D. Neuhauser, R. Olson, M.A. Rohrdanz, K.S. Thanthiriwatte, A.J.W. Thom, V. Vanovschi, C.F. Williams, Q. Wu, Z.-Q. You, M. Head-Gordon, Q-Chem, Version 3.2, Q-Chem, Inc, Pittsburgh, PA, 2007.

[44] S. Grimme, Semiempirical GGA-type density functional constructed with a long-range dispersion correction, J. Comput. Chem. 27 (2006) 1787–1799.