Published in final edited form as:
J Phys Chem A. 2010 February 4; 114(4): 2038–2044. doi:10.1021/jp911376p.

Non-Covalent Interactions of a Benzo[*a*]pyrene Diol Epoxide with DNA Base Pairs: Insight into the Formation of Adducts of (+)-B*a*P DE-2 with DNA

Jacqueline C. Hargis[‡], **Henry F. Schaefer III**[‡], **K. N. Houk**[§], and **Steven E. Wheeler**[§,*]

[‡]Center for Computational Quantum Chemistry, University of Georgia, Athens, GA 30602

[§]Department of Chemistry and Biochemistry, University of California, Los Angeles, CA 90095

Abstract

Non-covalent complexes of a tumorigenic benzo[*a*]pyrene diol epoxide with the guanine-cytosine and adenine-thymine base pairs have been examined computationally. (+)-B*a*P DE-2 forms covalent adducts with DNA via nucleophilic attack on the (+)-B*a*P DE-2 epoxide. Computational results predict five thermodynamically accessible complexes of AT with (+)-B*a*P DE-2 that are compatible with intact DNA. Among these, two are expected to lead to adenine adducts. In the lowest energy AT···(+)-B*a*P DE-2 complex, which has a gas-phase interaction energy of −20.9 kcal mol^{-1}, the exocyclic NH_2 of adenine is positioned for backside epoxide attack and formation of a *trans* adduct. The most energetically favorable complex leading to formation of a *cis* ring-opened adduct lies only 0.6 kcal mol^{-1} higher in energy. For GC···(+)-B*a*P DE-2, there are only two thermodynamically accessible complexes. The higher-lying complex, bound in the gas phase by 24.4 kcal mol^{-1} relative to separated GC and (+)-B*a*P DE-2, would lead to a *trans* ring opened N^2-guanine adduct. In the global minimum energy GC···(+)-B*a*P DE-2 complex, bound by 27.3 kcal mol^{-1}, the exocyclic NH_2 group of cytosine is positioned for *cis* epoxide addition. However, adducts of (+)-B*a*P DE-2 with cytosine are rarely observed experimentally. The paucity of cytosine adducts, despite the predicted thermodynamic stability of this GC···(+)-B*a*P DE-2 complex, is attributed to the electrostatic destabilization of the benzylic cation intermediate thought to precede *cis* addition.

I. Introduction

Polycyclic aromatic hydrocarbons (PAHs), a main component of soot, are widely distributed in the environment due primarily to the combustion of biomass and fossil fuels.[1] These PAHs form in flames via well-characterized pathways starting with the formation of an initial aromatic ring (usually benzene),[2-4] and are known to be carcinogenic.[3-7] Similarly, components of tobacco smoke have been shown to cause cancer, though the complexity of the underlying mechanisms has rendered a complete atomic-level understanding of the process elusive. Among the carcinogenic and tumorigenic components of tobacco smoke, PAHs are known to play a major role.[8] Benzo[*a*]pyrene (B*a*P) is one of the most thoroughly studied carcinogenic PAHs in tobacco smoke, and also one of the most abundant.[9]

In mammals, B*a*P is metabolized to a bay region diol epoxide, B*a*P DE. Four stereoisomers of B*a*P DE are formed (Fig. 1a), of which (+)-B*a*P DE-2[10] is both the most abundant and most tumorigenic.[11] Intercalation of (+)-B*a*P DE-2 into DNA is followed by nucleophilic attack on

*swheele2@chem.ucla.edu

Supporting Information Available. Cartesian coordinates; absolute energies.

the epoxide by the exocyclic NH_2 of adenine or guanine.[12] Guanine adducts predominate, although adenine adducts can also be formed in significant quantities.[13-15] Yields of cytosine adducts are much smaller,[13] and the structures of these adducts were not definitively characterized until recently.[16] Nucleophilic attack on the epoxide leads to a covalently bound DNA adduct via either *cis* or *trans* ring opening.[17] *Cis* addition is postulated to occur via a resonance-stabilized benzylic cation intermediate,[18] while *trans* ring-opening can proceed either from this cationic intermediate or by direct backside attack on the epoxide. Covalent DNA adducts inhibit enzymes such as helicase[19] and topoisomerase I.[20] These DNA adducts exhibit diverse, sequence-dependent mutagenic behavior, a feature that has been attributed to the presence of different stable adduct conformations.[21-26] Tobacco-smoke-related cancers are typically due to mutations in the *p53* tumor suppressor gene arising from the preferential addition of (+)-B*a*P DE-2 to specific sequences in this gene.[6]

Most previous investigations of the effects of (+)-B*a*P DE-2 on DNA, both experimental[6,27,28] and theoretical,[29-34] have focused on the covalent adducts with DNA oligomers or single nucleotides. Of particular recent interest has been the characterization of the conformational behavior of (+)-B*a*P DE-2–DNA adducts. NMR solution data have confirmed the presence of multiple stable conformations depending on the flanking nucleobases.[35,36] In 2004, the first crystal structures of B*a*P DE-2–DNA adducts were published[27,37] providing, along with previously published NMR structures,[37-40] invaluable details regarding the structure of these adducts. A crystal structure of an adduct with B*a*P DE bound to N^2-deoxyguanosine enabled the assignment of absolute configurations to the four optically active B*a*P DE isomers and the eight associated dG adducts resulting from the *cis* and *trans* ring opening of the epoxide.[37] A second crystal structure of a B*a*P DE-2–adenine adduct in a ternary complex with DNA polymerase was published shortly thereafter.[27]

While information about (+)-B*a*P DE-2···DNA complexes can be inferred from the structures of these adducts, a complete understanding of the addition of (+)-B*a*P DE-2 to DNA will require explicit examination of the complexes that precede epoxide attack. Harvey and co-workers[41] studied the formation of non-covalent intercalative complexes of (+)-B*a*P DE-2 and (−)-B*a*P DE-1 with DNA via kinetic flow linear dichroism experiments. Structural differences between these non-covalent complexes were highlighted that presumably underlie the different mutagenic and tumorigenic activities of these diastereomers. Computational methods are ideally suited to provide additional details regarding non-covalent (+)-B*a*P DE-2···DNA complexes, because it is possible to directly probe their structures and thermochemistry and to quantify the role of π-stacking and individual hydrogen bonding interactions. Previously, the inability of popular density functional theory (DFT) functionals[42,43] to accurately describe the π-stacking interactions that drive the intercalation of (+)-B*a*P DE-2 into DNA hampered high-level computational studies of such systems. Recent advances in the development of DFT functionals[44,45] have led to methods that may accurately describe π-stacking, opening the door for quantum mechanical studies of stacking phenomena in myriad biological contexts.

Several DNA intercalators have previously been examined by *ab initio* and DFT methods,[46-48] with a focus on intercalators that are utilized in antitumor chemotherapy. In 2002, Hobza and co-workers[47] studied the π-stacking interactions of ethidium, daunomycin, ellipticine, and 4,6′-diaminide-2-phenylindole with DNA base pairs. It was shown that in each of these cases the net attraction arises from the competing effects of electrostatic and dispersion interactions and short-range exchange repulsion. In 2006, Leszczynski and co-workers[48] examined the nature of interactions between ethidium and proflavine with DNA bases starting from published crystal structures. A more recent study[46] of the interaction of ellipticine and proflavine with DNA base pairs showed that stabilizing interactions are maximized when the main axis of the intercalator is nearly aligned with the main axis of the bases. In contrast to (+)-B*a*P DE-2, these

previously studied intercalating agents interact with DNA solely through π-stacking interactions, rather than a combination of intermolecular hydrogen bonds and π-stacking.

Despite decades of study of the effects of B*a*P DE on DNA, there have been few detailed explorations of the non-covalent (+)-B*a*P DE-2···DNA complexes that precede adduct formation.[41] Consequently, there are details concerning the complexation of (+)-B*a*P DE-2 with DNA base pairs that warrant further exploration. A lingering conundrum concerns the scarcity of observed cytosine adducts,[16] despite the prevalence of adducts of cytosine with diol epoxides derived from dibenz[*a,j*]anthracene and benz[*a*]anthracene.49-51 Presumably, this could arise from disfavored non-covalent interactions between cytosine and (+)-B*a*P DE-2, an insurmountable free energy barrier for epoxide attack by cytosine, or some combination of these two factors. Similarly, the energetically favored orientation of (+)-B*a*P DE-2 relative to a given base pair prior to adduct formation has not been established. It has been proposed21-26 that the existence of multiple conformations of (+)-B*a*P DE-2–DNA adducts is responsible for their divergent mutagenic behavior. The geometry of the complex preceding adduct formation could presumably play a role in the formation of these different conformers. Finally, understanding the role of individual π-stacking and hydrogen bonding interactions in these complexes will help further unravel the unique features of (+)-B*a*P DE-2 that lead to its pronounced tumorigenicity. As a first step toward addressing these issues, we have examined the complexes of (+)-B*a*P DE-2 with the AT and GC base pairs to gain insight into the factors governing the addition of (+)-B*a*P DE-2 to DNA.

II. Methods

Accurate *ab initio* descriptions of π-stacking interactions require rigorous correlated theoretical methods paired with large basis sets.[52] However, (+)-B*a*P DE-2···DNA base pair complexes, which comprise 66 and 67 atoms with GC and AT, respectively, are too large to treat with such rigorous approaches. Unfortunately, many DFT functionals, which are the most popular methods for computational investigations of systems of this size, fail to accurately describe the dispersion effects that underlie stacking interactions.[42,43] However, the M05-2X and M06-2X functionals have been shown to provide accurate interaction energies for stacked dimers.[44, 53-56] Hohenstein, Chill, and Sherrill[57] recently showed that M06-2X in particular performs well for a standard benchmark set of stacked complexes. Also, among DFT functionals including empirical dispersion corrections, PBE-D has been shown[57,58] to yield stacking energies of comparable quality to M06-2X. Similarly, Gu *et al*.[59] recently demonstrated that the M06-2X functional paired with a double zeta basis set yields interaction energies in good agreement with reliable CCSD(T) results for stacked nucleic acid bases.

Geometry optimizations were performed for two conformers of (+)-B*a*P DE-2 with each of the DNA base pairs using the M05-2X functional. The potential energy surface of each base pair with each conformer of (+)-B*a*P DE-2 was explored by executing optimizations from several initial geometries for each relative orientation of the stacked system. Starting structures were generated by varying the position of (+)-B*a*P DE-2 relative to the base pair in ~2 Å increments while keeping the molecular planes of the two species parallel, to ensure that all low-lying configurations were sampled. Preliminary optimizations were carried out using the M05-2X functional paired with the 3-21G basis set. Structures lying within three kcal mol^{-1} of the predicted global minimum were then further refined using the 6-31+G(d) basis set. To verify the M05-2X optimized structures, M06-2X/6-31+G(d) optimizations were also carried out on the minimum energy and second lowest-lying AT and GC complexes. Differences between the M05-2X and M06-2X geometries were minor. Presented gas-phase interaction energies are M06-2X/6-31+G(d) electronic energies evaluated at M05-2X/6-31+G(d) geometries, and are given relative to separated base pair and conformer **I** of (+)-B*a*P DE-2. PBE-D/aug-cc-pVDZ single point energies[45] were also evaluated at M05-2X geometries for

the two lowest lying AT and GC complexes. The PBE-D/aug-cc-pVDZ interaction energies for the AT complexes are similar to the M06-2X/6-31+G(d) data, differing by ±0.6 kcal mol^{-1}. On the other hand, the PBE-D interaction energies for the two lowest-lying GC complexes are smaller than the M06-2X values by 3.5 and 1.9 kcal mol^{-1}. In both cases, the energy ordering of the complexes is unchanged.

All computations were performed using NWChem.[60,61] For the M05-2X and M06-2X computations a fine DFT integration grid was used, consisting of 70 radial shells and 590 angular points, since these functionals are known to be sensitive to the choice of integration grid.[42,62-64] Standard atomic labels are utilized for the nucleobases, which are depicted in Fig. 2. Relevant atom designations for (+)-B*a*P DE-2 are shown in Fig. 1a and denoted by subscripts in the text.

III. Results and Discussion

A. Conformers of Free (+)-B*a*P DE-2

In the gas phase, there are two low-lying conformers of (+)-B*a*P DE-2 (**I** and **II**, Fig. 3a). In the higher-lying conformer (**II**), the hydroxyl groups occupy pseudo-axial positions, whereas the OH groups in **I** are equatorial. The M06-2X/6-31+G(d)//M05-2X/6-31+G(d) predicted energy difference is 2.1 kcal mol^{-1}, with a conformational barrier of 7.8 kcal mol^{-1}. Associated with each of these conformers are other low-lying minima, connected to **I** and **II** via changes in OH orientations. Both conformers **I** and **II** are stabilized by intramolecular hydrogen bonds that must be broken during the change in ring conformation. Such interactions will be less important in an aqueous environment, and the conformational barriers will be smaller in solution. Most importantly, both conformers **I** and **II** will be thermodynamically accessible and rapidly inter-converting at biologically relevant temperatures.

B. (+)-B*a*P DE-2···DNA Base Pair Complexes

Gas-phase interaction energy surfaces for (+)-B*a*P DE-2 with both the GC and AT base pairs have been examined as a model for the interaction of (+)-B*a*P DE-2 with DNA. When (+)-B*a*P DE-2 approaches a given base pair,[65] there are four possible relative orientations: the epoxide functionality can be directed toward or away from the base pair and the functionalized end of (+)-B*a*P DE-2 can extend into the minor or major groove of intact DNA. Additional variation arises from the complexation of (+)-B*a*P DE-2 with AT or TA and GC or CG. For each of these arrangements, multiple relative positions of the centers of mass of the base-pair and (+)-B*a*P DE-2 were considered to ensure that the minimum energy complex has been obtained for each relative orientation. In total, 48 distinct optimizations were carried out for each conformer.

The optimized structures (Figs. 3b and 4) are named according to the relative orientation of (+)-B*a*P DE-2 and the base pair as follows: the orientation of the epoxide towards or away from the base pair is indicated by **E** or **e**, respectively. **M** or **m** indicates that the functionalized end of (+)-B*a*P DE-2 extends into the major or minor groove, respectively. Finally, the ordering of the base pair (*e.g.*: GC versus CG) is explicitly given. For example, the global minimum GC···(+)-B*a*P DE-2 complex is labeled **CG(I-EM)**, signifying that (+)-B*a*P DE-2 is complexed with the epoxide functionality directed towards cytosine-guanine, with the functionalized end extending into the major groove of DNA. It should be noted that complexes in which the epoxide is facing the base pair (denoted by **E**) can only lead to *cis* adducts, while structures in which the epoxide is facing away from the base pair (**e**) are only compatible with *trans* addition.

One of the primary goals of the current work is to gain insight into the non-covalent complexes of (+)-B*a*P DE-2 with DNA that precede covalent adduct formation. Stacking of a single DNA

base pair with (+)-B*a*P DE-2 serves as the simplest possible model of the interaction with DNA. Rather than constraining optimizations to be compatible with intact DNA, unconstrained optimizations were executed and final structures incompatible with DNA were eliminated. Since the 2-deoxyribose of DNA was replaced with a hydrogen atom in our model, eliminated structures include five complexes in which the purine N^9 or pyrimidine N^1 act as hydrogen bond donors. Also excluded were two structures in which the base pair undergoes significant distortions that would be improbable in intact DNA due to backbone constraints and the presence of flanking base pairs. One such structure [**AT(I-eM)**] is depicted in Fig. 5. Some of the excluded structures were otherwise competitive energetically with the global minimum complexes.

C. Guanine-Cytosine···(+)-B*a*P DE-2 Complexes

The present results indicate that (+)-B*a*P DE-2 will preferentially form complexes with the guanine-cytosine base pair; the global minimum GC···(+)-B*a*P DE-2 complex (Fig. 3b) lies 6.4 kcal mol^{-1} lower in energy than the global minimum energy AT···(+)-B*a*P DE-2 complex (Fig. 4) at the M06-2X/6-31+G(d) level of theory. PBE-D/aug-cc-pVDZ predicts a smaller energy difference of 2.4 kcal mol^{-1}. In the global minimum GC structure [**CG(I-EM)**], conformer **I** of (+)-B*a*P DE-2 is complexed with the epoxide functionality directed towards CG. The gas-phase interaction energy relative to separated GC and conformer **I** of (+)-B*a*P DE-2 is 27.3 kcal mol^{-1}. This complex is stabilized by two somewhat strained hydrogen bonds joining the epoxide oxygen with the exocyclic NH_2 of cytosine and the C_7 hydroxyl group with O^6 on guanine. In order to maintain these intermolecular hydrogen bonds and the favorable stacking interaction of cytosine with the aromatic core of (+)-B*a*P DE-2, the base pair is distorted slightly from planarity. In intact DNA there will be an associated energetic cost due to interactions with flanking base pairs and this structure will lie higher in energy in a more complete DNA model. The oxidized end of (+)-B*a*P DE-2 interacts with the major groove; as a result, formation of a covalent guanine adduct is impossible, since the exocyclic NH_2 of guanine is directed away from the epoxide. Instead, this complex is compatible with formation of an N^4-cytosine adduct via *cis* ring opening of the epoxide, which only form in small quantities.[16]

There is only one additional thermodynamically accessible GC···(+)-B*a*P DE-2 complex [**CG (II-em)**], lying 2.9 kcal mol^{-1} above the global minimum. In this case, it is the higher-lying conformer (**II**) that is complexed with CG, with the epoxide functionality directed away from the base-pair and the functionalized end in the minor groove. As such, this complex is poised for back-side nucleophilic attack and formation of the frequently observed *trans* N^2-guanine adducts. This complex is stabilized by two cooperative hydrogen bonds involving the C_7 OH group and the exocyclic NH_2 and N^7 of guanine.

D. Adenine-Thymine···(+)-B*a*P DE-2 Complexes

Even though the minimum-energy AT···(+)-B*a*P DE-2 complex lies 6.4 kcal mol^{-1} higher in energy than the global minimum GC···(+)-B*a*P DE-2 complex, it is important to consider the AT case since covalent adenine adducts frequently form. In contrast to the GC···(+)-B*a*P DE-2 complexes, for which there was only one structure within three kcal mol^{-1} of the global minimum, for AT···(+)-B*a*P DE-2 there are five thermodynamically accessible complexes that are compatible with intact DNA.

The two lowest-lying AT···(+)-B*a*P DE-2 complexes, **AT(II-eM)** and **TA(II-em),** involve conformer **II** of (+)-B*a*P DE-2 with the epoxide functionality directed away from the base pair. These complexes are bound by 20.9 and 20.4 kcal mol^{-1} in the gas phase, respectively. In **AT (II-eM)**, (+)-B*a*P DE-2 is complexed with AT with the functionalized end directed into the major groove, while in **TA(II-em)** the complex involves TA and the functionalized end of (+)-B*a*P DE-2 is directed towards the minor groove. Consequently, in **AT(II-eM)** the exocyclic

NH_2 is positioned for backside attack and *trans* adduct formation, while no covalent adduct could form from **TA(II-em)**.

A complex of conformer **I** with AT was also optimized **[AT(I-eM)]**, and it is 0.5 kcal mol^{-1} lower than **AT(II-eM)**. However, as seen in Fig. 5, when complexed with **I**, the AT base pair is drastically distorted from planarity to maintain the cyclic hydrogen bonding arrangement between the C_7 OH group on (+)-B*a*P DE-2 and atoms N^6 and N^7 of adenine and stacking interaction between thymine and the pyrene. As mentioned above, these extreme distortions of the AT base pair are incompatible with the structure of intact DNA and such complexes are not expected to occur in intact DNA.

There are three other structures that are within 3 kcal mol^{-1} of **AT(II-eM)**, with the most favorable of these [**AT(I-EM)**] higher in energy by only 0.6 kcal mol^{-1}. This is the lowest-lying AT complex with conformer **I** that is compatible with intact DNA. In **AT(I-EM),** (+)-B*a*P DE-2 is oriented with the epoxide facing the base pair and the functionalized end extending into the major groove. In this complex, both N^6 and N^7 are poised for front-side attack on the epoxide, leading to either *cis* N^6-adenine adduct or the less frequently observed N^7 adduct. The next higher-lying structure, **TA(II-EM),** is 2.2 kcal mol^{-1} above the global minimum and features (+)-B*a*P DE-2 bound with the epoxide towards TA and the functionalized end extending into the major groove. This complex is not expected to lead to adduct formation, since N^6 is not near the epoxide carbon. The final structure, **TA(I-Em)**, also involves conformer **I**, this time complexed with TA with the epoxide towards the base pair and the functionalized end in the minor groove. No adduct can be formed from **TA(I-Em)**.

Structures **CG(II-em)** and **AT(II-eM)** provide a demonstration of the differential stacking avidities of (+)-B*a*P DE-2 with the GC and AT base pairs. These two structures exhibit similar hydrogen bonding interactions. Both include a cyclic hydrogen bonding arrangement of the C_7 OH group on (+)-B*a*P DE-2 with the exocyclic NH_2 and ring nitrogen of the purine base. The hydrogen bonds in the GC complex are slightly weaker, yet the GC complex is more strongly bound by 3.5 kcal mol^{-1}.[66] This difference is attributed to a stronger π-stacking interaction with GC over AT, which is consistent with previous findings for proflavin.[46] This difference can be understood qualitatively in terms of simple electrostatic effects. Electrostatic potential surfaces for conformer **II** of (+)-B*a*P DE-2 and the GC and AT base pairs are shown in Fig. 6. These plots provide a simple tool for understanding the electrostatic component of non-covalent interactions.[67-69] The primary difference in ESPs between the AT and GC base pairs is the sign of the ESP surrounding cytosine-N^4 compared to O^4 on thymine. In **AT(II-eM)** there will be an unfavorable electrostatic interaction between the negative ESP surrounding O^4 on thymine and the negative ESP above the pyrene. The corresponding electrostatic interaction with N^4 on cytosine will be favorable. In other words, the AT···(+)-B*a*P DE-2 complex is destabilized by a direct electrostatic interaction[70] between thymine-O^4 and the pyrene, while the analogous direct interaction with the cytosine NH_2 stabilizes **CG (II-em)**.

E. Implications for DNA Adduct Formation

The present computations employ a simple model of the interactions of B*a*P DE-2 with DNA consisting of gas-phase complexes between (+)-B*a*P DE-2 and a single base pair. Regardless, some insight into the formation of covalent adducts between (+)-B*a*P DE-2 and DNA can be gleaned from the computed structures. The present results predict that the higher-lying conformer of (+)-B*a*P DE-2 (conformer **II**) leads to formation of *trans* adducts with both guanine and adenine, via complexes **CG(II-em)** and **AT(II-eM)**, respectively. Conformer **I** is predicted to undergo *cis* addition to yield an adenine adduct via structure **AT(I-EM)**. The lowest-lying GC complex leading to *cis* addition of guanine (Fig. 3b) lies 6.4 kcal mol^{-1} above the global minimum, which is isoenergetic with the minimum energy AT complex. The global

minimum GC structure [**CG(I-EM)**] will not lead to guanine adduct formation, and is separated from the next lowest-lying structure by 2.9 kcal mol^{-1}. This is in contrast to the AT complexes, for which both the lowest-lying structure [**AT(II-eM)**] and a second complex 0.6 kcal mol^{-1} higher in energy [**AT(I-EM)**] are expected to both lead to covalent adducts.

The prediction that the thermodynamically preferred complex, **CG(I-EM)**, is positioned for formation of an N^4-cytosine adduct is inconsistent with the paucity of observed cytosine adducts.[16] As noted above, in this structure there is some distortion of the base pair from planarity and it would likely lie slightly higher in energy in a more complete model of DNA. Additionally, the effects of solvent, the sugar-phosphate backbone, or the presence of a sandwiching base pair could all alter the predicted energetic ordering in a more realistic model of DNA. On the other hand, there could be some mechanistic origin preventing cytosine attack. One possibility arises from differences in the electrostatic potential surrounding cytosine compared to guanine or adenine. The ESP surrounding cytosine is mostly positive (see Fig. 6), in contrast to the more negative ESP surrounding the purine bases. Front-side epoxide attack is postulated to occur through a fleeting benzylic cation intermediate.[18] The positive ESP surrounding cytosine will destabilize this incipient cation, raising the energy of this intermediate and associated reaction barriers and preventing cytosine addition to (+)-B*a*P DE-2.

The two low-lying conformers of (+)-B*a*P DE-2 lead to qualitatively different intermolecular hydrogen bonds with the GC and AT base pairs. More importantly, the complexes leading to *trans* adduct formation [**CG(II-em)** and **AT(II-eM)**] involve conformer **II**, while conformer **I** is present in the complexes compatible with *cis* adenine and guanine addition [**AT(I-EM)** and **CG(I-Em)**]. These latter complexes feature a hydrogen bond between the exocyclic NH_2 of the purine base and the epoxide oxygen. Additional stabilization is achieved via hydrogen bond contacts with the OH connected to C_7, which are only possible for conformer **I**. Conversely, for structures leading to *trans* adducts, the exocyclic NH_2 is hydrogen bonded to the C_7 hydroxyl group. Forming this hydrogen bond while simultaneously maintaining favorable stacking interactions with the pyrene and a planar base pair arrangement requires conformer **II**.

The *trans* arrangement in (+)-B*a*P DE-2 enables this OH group to position the AT base pair with the NH_2 group ideally oriented for backside attack on the epoxide. Overall, the tendency for formation of a strong hydrogen bond with the this OH group combined with the maximization of π-stacking interactions between the pyrene moiety and the DNA base pairs drive the complexes to adopt arrangements that are pre-organized for epoxide attack and DNA-adduct formation.

IV. Summary and Conclusions

(+)-B*a*P DE-2 is a tumorigenic metabolite of benzo[*a*]pyrene, a polycyclic aromatic hydrocarbon found in soot and tobacco smoke. (+)-B*a*P DE-2 forms covalent DNA adducts by intercalation into DNA followed by the nucleophilic attack of the (+)-B*a*P DE-2 epoxide by the exocyclic NH_2 group of either adenine or guanine. These covalent adducts interfere with key DNA processes, leading to sequence-specific mutations.

Complexes of the AT and GC DNA base pairs with two low-lying conformers of (+)-B*a*P DE-2 have been examined using DFT methods to gain insight into the factors governing adduct formation. Formation of complexes with the GC base pair is favored over AT, due to stronger π-stacking interactions between the pyrene core of (+)-B*a*P DE-2 and guanine-cytosine. Only one GC···(+)-B*a*P DE-2 complex is predicted to lie within 3 kcal mol^{-1} of the global minimum energy GC···(+)-B*a*P DE-2 structure, which is bound in the gas phase by 27.3 kcal mol^{-1}. In

this global minimum complex [**CG(I-EM)**], the exocyclic NH_2 group of cytosine is positioned for *cis* addition to the epoxide of (+)-B*a*P DE-2. The scarcity of experimentally observed cytosine adducts arising from this thermodynamically favorable complex are explained based on a simple electrostatic model; the region surrounding cytosine in the GC base pair has a positive electrostatic potential, which will destabilize the cationic intermediate leading to *cis* addition to the epoxide. The second lowest lying GC complex is compatible with backside attack on the epoxide by guanine, leading to the observed *trans* ring-opened N^2-guanine adducts. We predict five thermodynamically accessible AT···(+)-B*a*P DE-2 complexes that are compatible with intact DNA. The most energetically favored complex [**AT(II-eM)**], lies 6.4 kcal mol^{-1} higher in energy than the global minimum GC···(+)-B*a*P DE-2 complex. Among these low-lying AT complexes, two structures are compatible with adenine adduct formation, accounting for both the *cis* and *trans* adducts.

There are two stable conformers of free (+)-B*a*P DE-2, and when bound to either the AT or GC base pair these are competitive energetically. Complexes involving the lower-lying conformer (**I**) lead to *cis* ring-opened adducts, while *trans* adducts are predicted to form from complexes featuring the higher-lying conformer (**II**).

Stable non-covalent complexes between DNA base pairs and (+)-B*a*P DE-2 arise from a combination of intermolecular hydrogen bonds and π-stacking interactions. Because of the *trans* arrangement of the C_7 OH group and epoxide on (+)-B*a*P DE-2, in many of the low-lying complexes the base pair is oriented with the exocyclic NH_2 group pre-organized for epoxide attack. The energetic tendency to maximize π-stacking interactions further orients the base pair for epoxide attack.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the Department of Energy Combustion Program and the National Science Foundation (J.C.H. and H.F.S.), National Institute of General medical Sciences (NIGMS), National Institutes of Health (NIH) grant 1R01GM075962 (K.N.H), and NIGMS, NIH grant 5F32GM082114-02 (S.E.W.). S.E.W. would also like to thank E. C. P. Hohenstein for stimulating discussions and J. M. Sayer for helpful suggestions regarding cytosine adducts. J.C.H. thanks K. Morris for comments on the manuscript. Molecular figures were generated using CYLview.[71] Computer time was provided in part by the National Energy Research Scientific Computing Center (NERSC), which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231, and the UCLA Institute for Digital Research and Education (IDRE).

References

(1). Edwards NT. J. Environ. Qual 1983;12:427.

(2). Miller JA, Pilling MJ, Troe J. Proc. Combust. Inst 2005;30:43.

(3). Wheeler SE, Allen WD, Schaefer HF. J. Chem. Phys 2004;121:8800. [PubMed: 15527344]

(4). Wheeler SE, Robertson KA, Allen WD, Schaefer HF, Bomble YJ, Stanton JF. J. Phys. Chem. A 2007;111:3819. [PubMed: 17402717]

(5). Durant JL, Busby WF, Lafleur AL, Penman BW, Crespi CL. Mutat. Res 1996;371:123. [PubMed: 9008716]

(6). Denissenko MF, Pao A, Tang MS, Pfeifer GP. Science 1996;274:430. [PubMed: 8832894]

(7). Pullman A, Pullman B. Adv. Cancer Res 1955;3:117. [PubMed: 13248740]

(8). Hecht SS. J. Natl. Cancer I 1999;91:1194.

(9). Hoffmann D, Djordjevic MV, Hoffmann I. Prev. Med 1997;26:427. [PubMed: 9245661]

(10). Adamson JD, Morter CL, DeSain JD, Glass GP, Curl RF. J. Phys. Chem 1996;100:2125.

(11). Buening MK, Wislocki PG, Levin W, Yagi H, Thakker DR, Akagi H, Koreeda M, Jerina DM, Conney AH. Proc. Natl. Acad. Sci. USA 1978;75:5358. [PubMed: 281685]

(12). Szeliga J, Dipple A. Chem. Res. Toxicol 1998;11:1. [PubMed: 9477220]

(13). Sayer JM, Chadha A, Agarwal SK, Yeh HJC, Yagi H, Jerina DM. J. Org. Chem 1991;56:20.

(14). Wei S-JC, Chang RL, Wong CQ, Bhachech N, Cui XX, Hennig E, Yagi H, Sayer JM, Jerina DM, Preston BD, Conney AH. Proc. Natl. Acad. Sci 1991;88:11227. [PubMed: 1763036]

(15). Wei S-JC, Chang RL, Bhachech N, Cui XX, Merkler KA, Wong C-Q, Hennig E, Yagi H, Jerina DM, Conney AH. Cancer Res 1993;53:3294. [PubMed: 8324741]

(16). Wolfe AR, Smith TJ, Meehan T. Chem. Res. Toxicology 2004;17:476.

(17). Geacintov NE, Cosman M, Hingerty BE, Amin S, Broyde S, Patel DJ. Chem. Res. Toxicol 1997;10:111. [PubMed: 9049424]

(18). Jerina, DM.; Chadha, A.; Cheh, AM.; Schurdak, ME.; Wood, AW.; Sayer, JM. Biological Reactive Intermediates IV. In: Witmer, CM.; Snyder, R.; Jollow, DJ.; Kalf, GF.; Kocsis, JJ.; Sipes, IG., editors. Biological Reactive Intermediates IV. Plenum Press; New York: 1991. p. 533

(19). Yong Y, Romano LJ. Chem. Res. Toxicol 1996;9:179. [PubMed: 8924589]

(20). Pommier Y, Laco GS, Kohlhagan G, Sayer JM, Kroth H, Jerina DM. Proc. Natl. Acad. Sci. USA 2000;97:10739. [PubMed: 10995470]

(21). Hanrahan CJ, Bacolad MD, Vyas RR, Liu T, Geacintov NE, Loechler EL, Basu AK. Chem. Res. Toxicol 1997;10:369. [PubMed: 9114972]

(22). Alekseyev YO, Romano LJ. Biochemistry 2002;41:4467. [PubMed: 11914095]

(23). Shukla R, Liu T, Geacintov NE, Loechler EL. Biochemistry 1997;36:10256. [PubMed: 9254624]

(24). Shukla R, Geacintov NE, Loechler EL. Carcinogenesis 1999;20:261. [PubMed: 10069463]

(25). Page JE, Zajc B, Oh-hara T, Lakshman MK, Sayer JM, Jerina DM, Dipple A. Biochemistry 1998;37:9127. [PubMed: 9636059]

(26). Seo KY, Jelinsky SA, Loechler EL. Mutat. Res 2000;463:215. [PubMed: 11018743]

(27). Ling H, Sayer JM, Plosky BS, Yagi H, Boudsocq F, Woodgate R, Jerina DM, Yang W, Davies DR. P. Natl. Acad. Sci. USA 2004;101:2265.

(28). Chiapperino D, Cai M, Sayer JM, Yagi H, Kroth H, Masutani C, Hanaoka F, Jerina DM, Cheh AM. J. Biol. Chem 2005;280:39684. [PubMed: 16188888]

(29). Lee HM, Chae Y-H, Kwon C, Kim SK. Biophys. Chem 2007;125:151. [PubMed: 16962698]

(30). Chandani S, Lee CH, Loechler EL. Chem. Res. Toxicol 2005;18:1108. [PubMed: 16022503]

(31). Yan SF, Wu M, Geacintov NE, Broyde S. Biochemistry 2004;43:7750. [PubMed: 15196018]

(32). Lee CH, Loechler EL. Mutat. Res 2003;529:59. [PubMed: 12943920]

(33). Rao SN, Lybrand T, Michaud D, Jerina DM, Kollman PA. Carcinogenesis 1989;10:27. [PubMed: 2910528]

(34). Lee CH, Chandani S, Loechler EL. Chem. Res. Toxicol 2002;15:1429. [PubMed: 12437334]

(35). Fountain MA, Krugh TR. Biochemistry 1995;34:3152. [PubMed: 7880810]

(36). Xu R, Mao B, Amin S, Geacintov NE. Biochemistry 1998;37:769. [PubMed: 9425101]

(37). Karle IL, Yagi H, Sayer JM, Jerina DM. Proc. Natl. Acad. Sci. USA 2004;101:1433. [PubMed: 14757823]

(38). Schwartz JL, Rice JS, Luxon BA, Sayer JM, Xie G, Yeh HJC, Liu X, Jerina DM, Gorenstein DG. Biochemistry 1997;36:11069. [PubMed: 9333324]

(39). Pradhan P, Tirumala S, Liu X, Sayer JM, Jerina DM, Yeh HJC. Biochemistry 2001;40:5870. [PubMed: 11352722]

(40). Lakshman MK, Keeler JC, Ngassa FN, Hilmer JH, Pradhan P, Zajc B, Thomasson KA. J. Am. Chem. Soc 2007;129:68. [PubMed: 17199284]

(41). Shahbaz M, Geacintov NE, Harvey RG. Biochem 1986;25:3290. [PubMed: 3089275]

(42). Johnson ER, Wolkow RA, DiLabio GA. Chem. Phys. Lett 2004;394:334.

(43). Swart M, van der Wijst T, Guerra CF, Bickelhaupt FM. J. Mol. Model 2007;13:1245. [PubMed: 17874150]

(44). Zhao Y, Truhlar DG. Theo. Chem. Acc 2008;120:215.

(45). Grimme S. J. Comp. Chem 2006;27:1787. [PubMed: 16955487]

(46). Li S, Cooper VR, Thonhauser T, Lundqvist BI, Langreth DC. J. Phys. Chem. B 2009;113:11166. [PubMed: 19719266]

(47). Řeha D, Kabeláč M, Ryjáček F, Šponer J, Šponer JE, Elstner M, Suhai S, Hobza P. J. Am. Chem. Soc 2002;124:3366. [PubMed: 11916422]

(48). Langner KM, Kedzierski P, Sokalski WA, Leszczynski J. J. Phys. Chem. B 2006;110:9720. [PubMed: 16686524]

(49). Chadha A, Sayer JM, Yeh HJC, Yagi H, Cheh AM, Pannell LK, Jerina DM. J. Am. Chem. Soc 1989;111:5456.

(50). Cheh AM, Chadha A, Sayer JM, Yeh HJC, Yagi H, Pannell LK, Jerina DM. J. Org. Chem 1993;58:4013.

(51). Chadha, A.; Sayer, JM.; Agarwal, SK.; CHeh, AM.; Tagi, H.; Yeh, HJC.; Jerina, DM. In: Cook, M.; Loening, K.; Merritt, J., editors. Polynuclear Aromatic Hydrocarbons: Measurements, Means and Metabolism (Eleventh International Symposium); Batelle Press: Columbus, OH. 1991;

(52). Sinnokrot MO, Sherrill CD. J. Phys. Chem. A 2006;110:10656. [PubMed: 16970354]

(53). Zhao Y, Schultz NE, Truhlar DG. J. Chem. Theory Comput 2006;2:364.

(54). Zhao Y, Truhlar DG. Phys. Chem. Chem. Phys 2005;7:2701. [PubMed: 16189582]

(55). Zhao Y, Truhlar DG. J. Chem. Theory Comput 2007;3:289.

(56). Zhao Y, Truhlar DG. Acc. Chem. Res 2008;41:157. [PubMed: 18186612]

(57). Hohenstein EG, Chill ST, Sherrill CD. J. Chem. Theory Comput 2008;4:1996.

(58). Sherrill CD, Takatani T, Hohenstein EG. J. Phys. Chem. A 2009;113:10146. [PubMed: 19689152]

(59). Gu J, Wang J, Leszczynski J, Xie Y, Schaefer HF. Chem. Phys. Lett 2008;459:164.

(60). Bylaska, EJ.; de Jong, WA.; Govind, N.; Kowalski, K.; Straatsma, TP.; Valiev, M.; Wang, D.; Aprà, E.; Windus, TL.; Hammond, J.; Nichols, P.; Hirata, S.; Hackler, MT.; Zhao, Y.; Fan, P-D.; Harrison, RJ.; Dupuis, M.; Smith, DMA.; Nieplocha, J.; Tipparaju, V.; Krishnan, M.; Wu, Q.; Van Voorhis, T.; Auer, AA.; Nooijen, M.; Brown, E.; Cisneros, G.; Fann, GI.; Fruchtl, H.; Garza, J.; Hirao, K.; Kendall, R.; Nichols, JA.; Tsemekhman, K.; Wolinski, K.; Anchell, J.; Bernholdt, D.; Borowski, P.; Clark, T.; Clerc, D.; Dachsel, H.; Deegan, M.; Dyall, K.; Elwood, D.; Glendening, E.; Gutowski, M.; Hess, A.; Jaffe, J.; Johnson, B.; Ju, J.; Kobayashi, R.; Kutteh, R.; Lin, Z.; Littlefield, R.; Long, X.; Meng, B.; Nakajima, T.; Niu, S.; Pollack, L.; Rosing, M.; Sandrone, G.; Stave, M.; Taylor, H.; Thomas, G.; van Lenthe, J.; Wong, A.; Zhang, Z. NWChem, A Computational Chemistry Package for Parallel Computers, Version 5.1. Pacific Northwest National Laboratory; Richland, Washington 99352-0999, USA: 2007.

(61). Kendall RA, Aprá E, Bernholdt DE, Bylaska EJ, Dupuis M, Fann GI, Harrison RJ, Ju J, Nichols JA, Nieplocha J, Straatasma TP, Windus TL, Wong AT. Computer Phys. Commun 2000;128:260.

(62). Johnson ER, Becke AD, Sherrill CD, DiLabio GA. J. Chem. Phys 2009;131:034111. [PubMed: 19624185]

(63). Gräfenstein J, Cremer D. J. Chem. Phys 2007;127:164113. [PubMed: 17979325]

(64). Wheeler SE, Houk KN. J. Chem. Theory Comput. 2009 Submitted.

(65). Miller JA, Melius CF. Combust. Flame 1992;91:21.

(66). PBE-D predicts a much smaller separation of 1 kcal/mol.

(67). Politzer, P.; Murray, JS. Computational Medical Chemistry for Drug Discovery. In: Bultinck, P.; De Winter, H.; Langenaeker, W.; Tollenaere, JP., editors. Computational Medical Chemistry for Drug Discovery. Marcel Dekker, Inc.; New York: 2004. p. 213

(68). Wheeler SE, Houk KN. J. Chem. Theory Comput 2009;5:2301. [PubMed: 20161573]

(69). Politzer, P.; Murray, JS. Chemical Reactivity Theory: A Density Functional View. Chattaraj, PK., editor. CRC Press; Boca Raton, FL: 2009.

(70). Wheeler SE, Houk KN. J. Am. Chem. Soc 2008;130:10854. [PubMed: 18652453]

(71). Legault, CY. CYLview, 1.0b. Université de Sherbrooke; 2009.

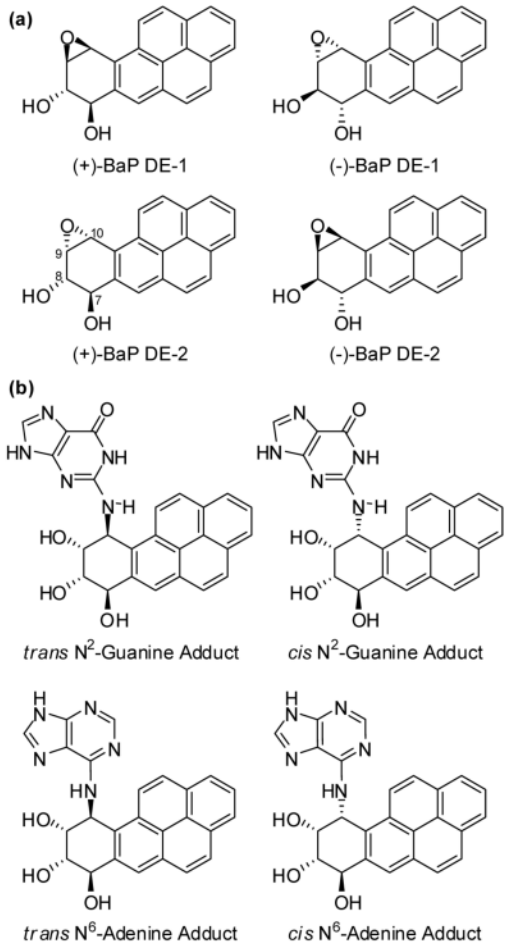

Figure 1.
(a) Four stereoisomers of the benzo[*a*]pyrene diol epoxide; (b) major *trans* and *cis* guanine and adenine (+)-B*a*P DE-2 adducts

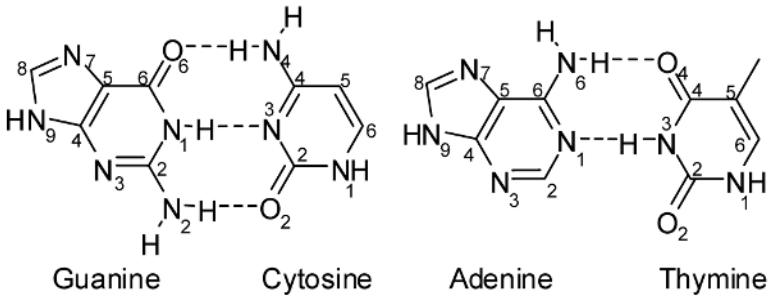

Figure 2.
Canonical atomic labels for GC and AT base pairs.

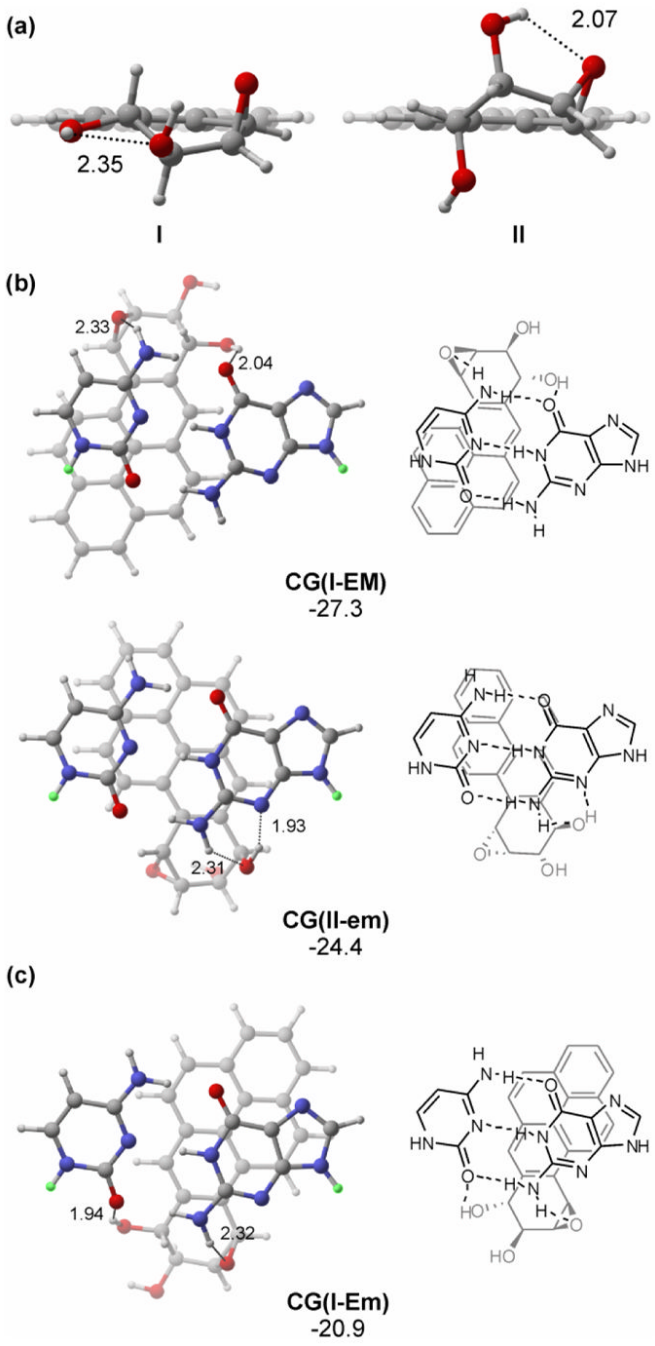

Figure 3.

(a) Two low-lying gas-phase conformers of (+)-B*a*P DE-2. At the M06-2X/6-31+G(d)// M05-2X/6-31+G(d) level of theory, conformer **II** lies 2.1 kcal mol^{-1} higher than **I**, separated by a conformational barrier of 7.8 kcal mol^{-1}. (b) Low-lying complexes of (+)-B*a*P DE-2 with the GC base pair. (c) Lowest-lying GC complex compatible with *cis* guanine addition. In (b) and (c), gas-phase interaction energies relative to separated GC and conformer **I** of (+)-B*a*P DE-2 are given in kcal mol^{-1}. Hydrogens connected to N^1 and N^9 are highlighted in green. Hydrogen bond distances are in Angstroms.

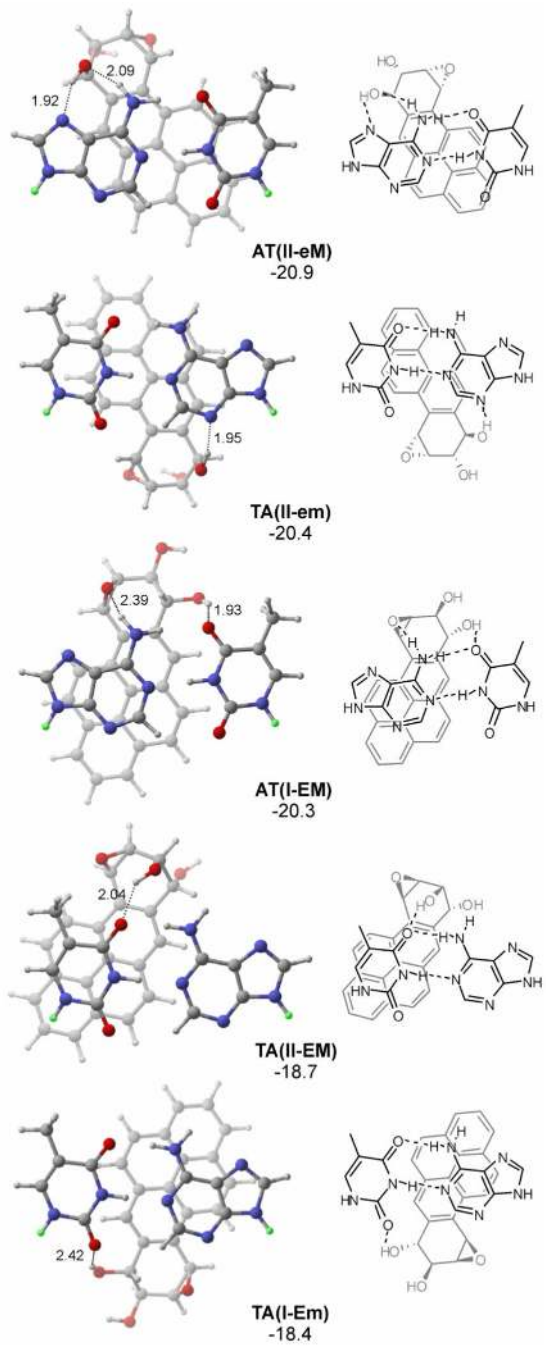

Figure 4.

Low-lying complexes of (+)-B*a*P DE-2 with the AT base pair. Gas-phase interaction energies relative to separated AT and conformer **I** of (+)-B*a*P DE-2 are given in kcal mol^{-1}. Hydrogens connected to N^1 and N^9 are highlighted in green. Hydrogen bond distances are in Angstroms.

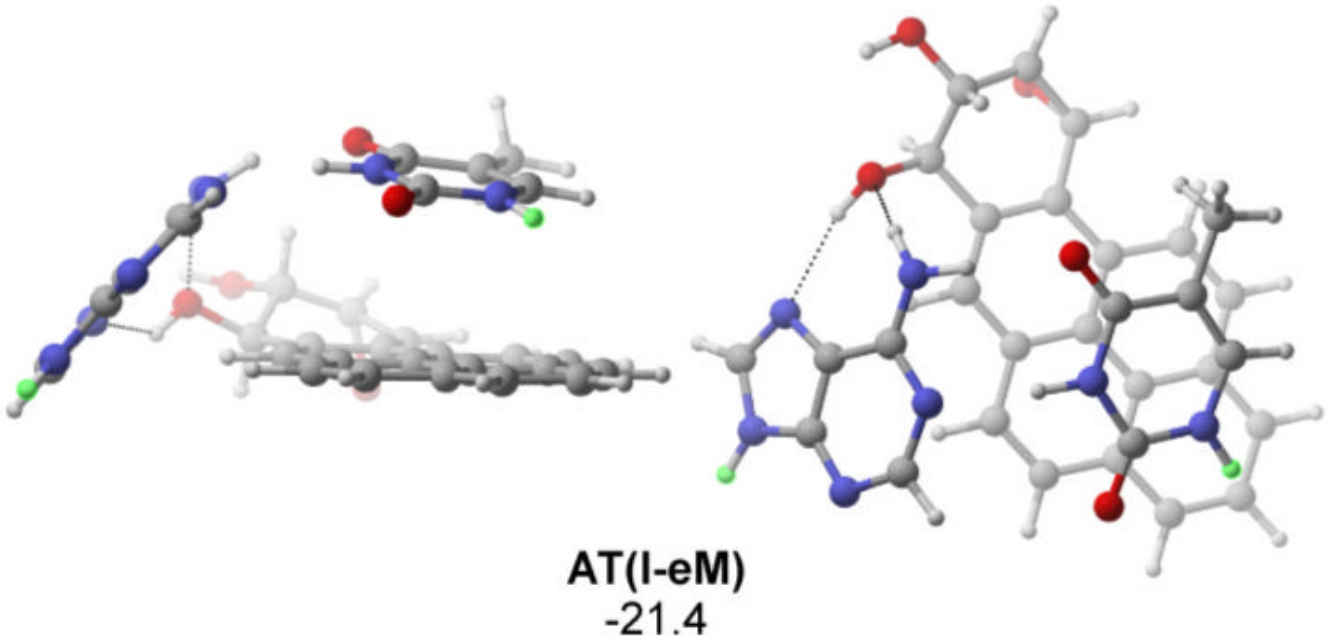

Figure 5.
Two views of a low-lying complex of conformer **I** of (+)-B*a*P DE-2 with the AT base pair. This structure is not compatible with intact DNA, since the base-pair undergoes significant distortion in order to form a strong hydrogen bond with conformer **I** of (+)-B*a*P DE-2 while maintaining favorable stacking interactions with the pyrene. The gas-phase interaction energy relative to separated AT and (+)-B*a*P DE-2 is given in kcal mol^{-1}. Hydrogens connected to N^1 and N^9 are highlighted in green.

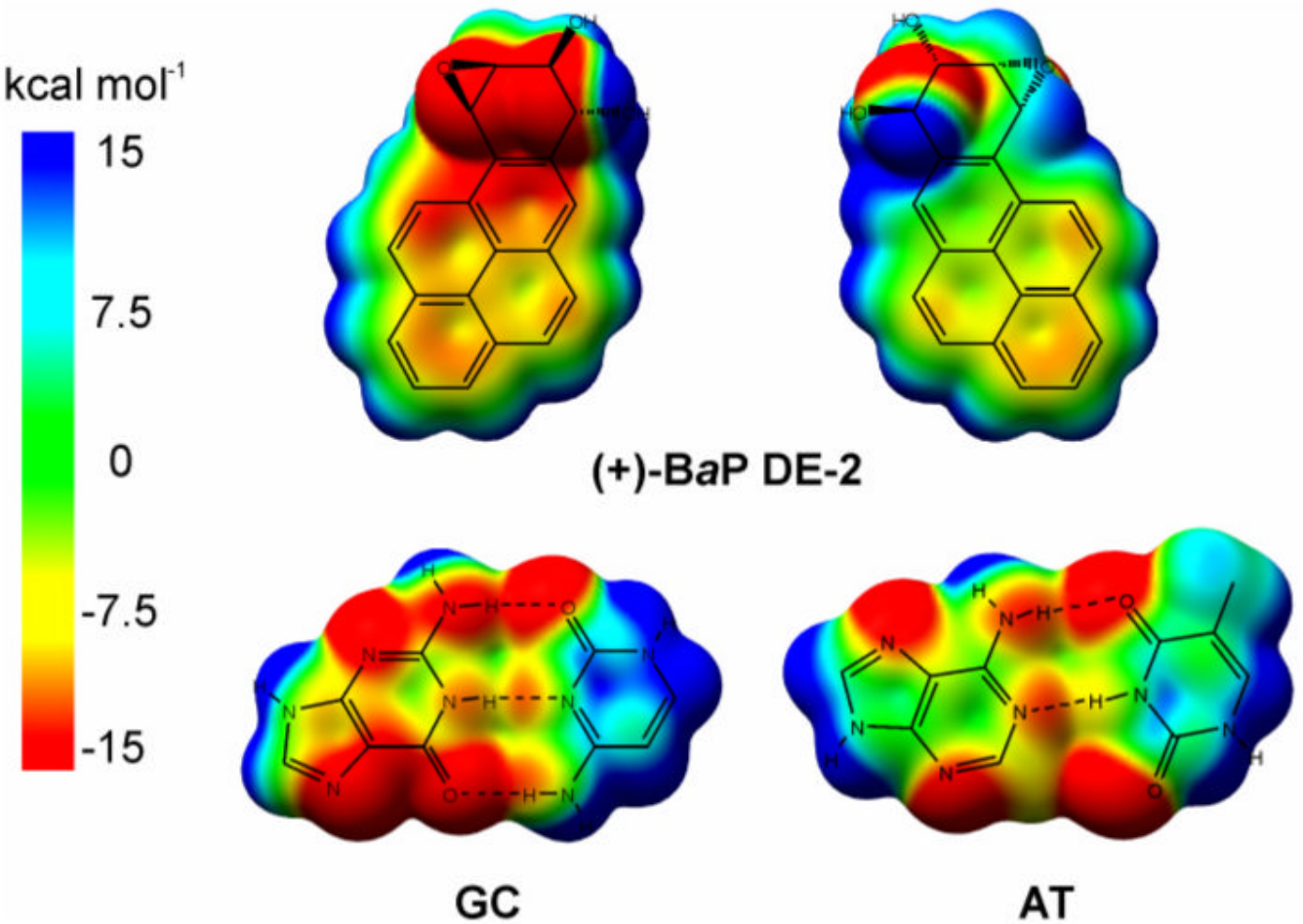

Figure 6.
Electrostatic potentials of conformer **II** of (+)-B*a*P DE-2, guanine-cytosine, and adenine-thymine, mapped onto electron density isosurfaces ($\rho = 0.001$ e/au^3).