Title

Relative genotoxicity of polycyclic aromatic hydrocarbons inferred from free energy perturbation approaches

Journal

Proceedings of the National Academy of Sciences of the United States of America, 121(37)

ISSN

0027-8424

Authors

Urwin, Derek J
Tran, Elise
Alexandrova, Anastassia N

Publication Date

2024-09-10

DOI

10.1073/pnas.2322155121

PNAS RESEARCH ARTICLE | BIOPHYSICS AND COMPUTATIONAL BIOLOGY

Relative genotoxicity of polycyclic aromatic hydrocarbons inferred from free energy perturbation approaches

Derek J. Urwin[a], Elise Tran[a], and Anastassia N. Alexandrova[a,1]

Edited by Chris Jarzynski, University of Maryland, College Park, MD; received January 8, 2024; accepted June 27, 2024

Utilizing molecular dynamics and free energy perturbation, we examine the relative binding affinity of several covalent polycyclic aromatic hydrocarbon - DNA (PAH-DNA) adducts at the central adenine of NRAS codon-61, a mutational hotspot implicated in cancer risk. Several PAHs classified by the International Agency for Research on Cancer as probable, possible, or unclassifiable as to carcinogenicity are found to have greater binding affinity than the known carcinogen, benzo[a]pyrene (B[a]P). van der Waals interactions between the intercalated PAH and neighboring nucleobases, and minimal disruption of the DNA duplex drive increases in binding affinity. PAH-DNA adducts may be repaired by global genomic nucleotide excision repair (GG-NER), hence we also compute relative free energies of complexation of PAH-DNA adducts with RAD4-RAD23 (the yeast ortholog of human XPC-RAD23) which constitutes the recognition step in GG-NER. PAH-DNA adducts exhibiting the greatest DNA binding affinity also exhibit the least RAD4-RAD23 complexation affinity and are thus predicted to resist the GG-NER machinery, contributing to their genotoxic potential. In particular, the fjord region PAHs dibenzo[a,l]pyrene, benzo[g]chrysene, and benzo[c]phenanthrene are found to have greater binding affinity while having weaker RAD4-RAD23 complexation affinity than their respective bay region analogs B[a]P, chrysene, and phenanthrene. We also find that the bay region PAHs dibenzo[a,j]anthracene, dibenzo[a,c]anthracene, and dibenzo[a,h]anthracene exhibit greater binding affinity and weaker RAD4-RAD23 complexation affinity than B[a]P. Thus, the study of PAH genotoxicity likely needs to be substantially broadened, with implications for public policy and the health sciences. This approach can be broadly applied to assess factors contributing to the genotoxicity of other unclassified compounds.

polycyclic aromatic hydrocarbons | DNA adducts | genotoxicity | cancer risk | free energy perturbation

Significance

Polycyclic aromatic hydrocarbons (PAHs) in ambient air pollution contribute to cancer risk in exposed populations and occupational exposures are a major factor in the International Agency for Research on Cancer (IARC) classification of firefighting, coke production, and iron/steel founding as Group 1 carcinogens. While PAHs constitute a large class of compounds, only benzo[a]pyrene (B[a]P) is classified as Group 1. Utilizing computational approaches, we examine the relative binding and genomic repair affinities of several PAH-DNA adducts in the NRAS codon-61 sequence context, a mutational hotspot that contributes to cancer risk. We find that several PAHs classified by IARC as Group 2A, 2B, and 3 have greater binding and weaker repair affinities than B[a]P, contributing to their genotoxic potential.

Polycyclic aromatic hydrocarbons (PAHs) are a large class of compounds produced by processes that involve incomplete combustion of organic substances, many of which are classified as known, probable, or possible human carcinogens (Group 1, 2A, or 2B) by the International Agency for Research on Cancer (IARC) (1). Human exposure to PAHs via inhalation, dermal absorption, and ingestion is pervasive and results from ubiquitous sources that range from grilled food and automotive exhaust to industrial emissions and catastrophic wildfires. Environmental and occupational exposures to PAHs are known to be associated with an elevated incidence of cancer in affected populations and occupational exposures in certain fields of work such as firefighting can be extreme (2–7). Many PAHs are genotoxic, mutagenic, and ultimately carcinogenic owing to the formation of covalent PAH-DNA adducts at mutational hotspots in the genome (6–13).

Genotoxicity is a broad term, encompassing mutagenicity, that includes a chemical's propensity to induce DNA damage that may be reversed by genomic repair mechanisms (7, 8, 14). In human cells, bay region PAHs such as benzo[a]pyrene (B[a]P) and fjord region PAHs such as dibenzo[a,l]pyrene (DB[a,l]P) are enantioselectively and diastereoselectively metabolized to form genotoxic PAH-diol-epoxides (PAH-DEs) by cytochrome P450s and epoxide-hydrolases that operate in the lipid bilayer of the endoplasmic reticulum (6–8). These enzymes are heavily expressed in the liver as well as most extrahepatic tissues, with the metabolic process yielding the major products: (7R,8S,9S,10R)-B[a]P-DE and (11R,12S,13S,14R)-DB[a,l]P-DE, respectively (Fig. 1*A*), while stereoisomers of these compounds are only produced in small quantities (7, 8). Covalent adenine adducts form at the exocyclic amino group via preferential trans-opening of the (-R,-S,-S,-R)-PAH-DE's epoxide ring yielding (-R,-S,-R,-S)-trans-anti-PAH-DE-N6-dA adducts (Fig. 1*B*) while adducts formed by cis opening of the epoxide ring are only formed in small quantities (7, 8). Although adenine N6-dA adducts

Author affiliations: [a]Department of Chemistry and Biochemistry, University of California Los Angeles, Los Angeles, CA 90095

Author contributions: D.J.U. and A.N.A. designed research; D.J.U. and E.T. performed research; D.J.U., E.T., and A.N.A. analyzed data; and D.J.U. and A.N.A. wrote the paper.

Competing interest statement: NSF grant CHE-2203366.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: ana@chem.ucla.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2322155121/-/DCSupplemental.

Published September 3, 2024.

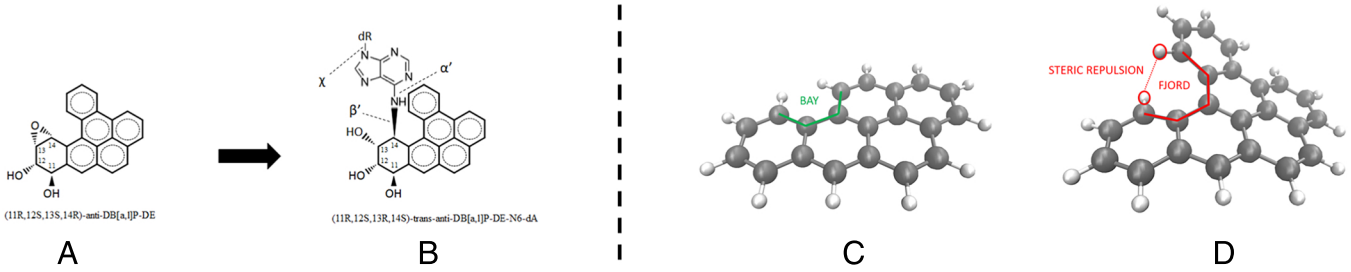

Fig. 1. (*A*) Fjord region (11R,12S,13S,14R)-DB[a,l]P-DE (*B*) resulting (11R,12S,13R,14S)-trans-anti-DB[a,l]P-DE-N6-dA adduct, glycosidic torsion angle χ, and PAH linkage torsion angles α' and β' (see *SI Appendix*, Fig. S.3 for linkage angle definitions by atom name). (*C*) Planar bay region benzo[a]pyrene (B[a]P) (*D*) fjord region dibenzo[a,l]pyrene (DB[a,l]P) where the nonplanar structure is caused by steric repulsion between hydrogens on opposite ends of the fjord region.

are less abundant than guanine N2-dG adducts, N6-dA adducts are thought to have a disproportionate impact on cancer risk and are of particular interest (13, 15–17). When PAH-DNA adducts such as these evade genomic repair mechanisms, they can induce nucleotide misincorporation in the complementary DNA strand during replication, leading to cancer-promoting mutations (6–8, 15–19).

Mutations in the RAS family of proto-oncogenes are found in approximately one-third of human cancers (12) and a well-known mutational hotspot is the central adenine (henceforth dA*) of codon 61 in NRAS, which normally codes for glutamine with nucleotide sequence CAA [henceforth NRAS(Q61)] (6–8, 13, 15–17, 19–21). NRAS mutations are found in 27.7% of human melanomas and 88.1% of these mutations are found in NRAS(Q61) (12). Single nucleotide variants (SNVs) of NRAS(Q61) abrogate the catalytic activity of the NRAS enzyme, locking it in an active GTP-bound conformation (12). In particular, NRAS(Q61L) variants that code for leucine rather than glutamine exhibit elevated mitogen-activated protein kinase signaling, resulting in overrepresentation of the protein kinase CK2α which is associated with cellular proliferation in primary human melanocytes (12). While exposure to ultraviolet (UV)-light is a known source of DNA damage associated with melanoma (22, 23), firefighters who experience significant dermal exposure to PAHs in the line of duty have an increased incidence of melanoma as compared to the general public (5, 24, 25). Studies both in vitro and in vivo have shown that PAH-DNA adducts such as the (11R,12S,13R,14S)-trans-anti-DB[a,l]P-DE-N6-dA* adduct at the central adenine of NRAS(Q61) that evade repair tend to induce dA* $\rightarrow$ dT transversions (7, 26–29) that correspond to a CAA$\rightarrow$CTA SNV that codes for leucine, indicating that such PAH-DNA adducts are a possible source of NRAS(Q61L) mutations in human cancers such as melanoma.

While humans are known to be exposed to both B[a]P and DB[a,l]P from products of incomplete combustion, it has been noted that DB[a,l]P is the most tumorigenic compound known to date, with tumorigenicity estimated to be approximately 100-fold that of B[a]P due to its nonplanar aromatic ring structure (6, 7) (Fig. 1 *C* and *D*). Despite this, B[a]P is classified as an IARC Group 1 known human carcinogen while DB[a,l]P is classified as a Group 2A probable human carcinogen (1). Furthermore, B[a]P appears on the list of 16 Environmental Protection Agency (EPA) priority PAHs utilized for environmental risk factor assessments while DB[a,l]P and other PAHs known to be more genotoxic and mutagenic than B[a]P do not (1, 6–11, 30, 31). Because the process of carcinogenesis may be initiated by the clonal expansion of a single cell through the heritable abrogation of cellular processes that regulate cell division (7), risk factor assessments that do not include the most genotoxic and mutagenic PAHs paint an incomplete picture. Hence examining factors that contribute to the genotoxicity of PAHs other than B[a]P and elucidating the associated molecular mechanisms is essential for accurate risk factor assessments.

The process of examining the genotoxic and carcinogenic potential of a particular compound either in vitro or in vivo is both slow and expensive, with results developed on the order of years (32). For example, mutagenesis assays utilizing transgenic rodents are widely accepted as an effective approach that includes the biology of an entire organism, but requires extensive infrastructure over several generations of animals to complete (32). As a result, this field of study can benefit from in silico examination of factors that contribute to the relative genotoxicity of a collection of structurally diverse and largely unstudied PAHs. Such in silico studies can inform in vivo and in vitro research efforts by prioritizing compounds with the potential to be the most genotoxic in an approach similar to lead optimization in computational drug design (33).

While several factors such as DNA sequence context, PAH metabolism, resultant PAH-DE stereochemistry, and dose contribute to the relative genotoxicity of a given PAH, the likelihood of the PAH-DE forming a PAH-DNA adduct and the likelihood of that PAH-DNA adduct evading genomic repair mechanisms such as global genomic nucleotide excision repair (GG-NER) contribute to the PAH's genotoxic potential and subsequent likelihood of inducing a cancer-promoting mutation (6–8, 15–17, 34). In this work, we calculate the relative differences in free energy of binding of classically intercalated covalent (-R,-S,-R,-S)-trans-anti-PAH-DE-N6-dA$^*_6$ adducts (henceforth PAH-DNA adduct) as compared to a (7R,8S,9R,10S)-trans-anti-B[a]P-DE-N6-dA$^*_6$ adduct at the central adenine dA$^*_6$ of NRAS(Q61) in a DNA 11-mer consisting of NRAS codons 60 to 62 with terminal G-C base pairs modeled on PDB: 1DL4 (17) (Fig. 2*A*) for several IARC Group 2A, 2B, and 3 PAHs (Table 1). This approach allows us to identify those PAHs that are either more or less likely to form covalent PAH-DNA adducts as compared to the IARC Group 1 carcinogen B[a]P.

We then calculate the relative differences in free energy of complexation for these PAH-DNA adducts in the productive complex with RAD4-RAD23 (henceforth productive complex), which constitutes the recognition step of GG-NER. Note that RAD4-RAD23 is the often studied yeast ortholog of human XPC-RAD23 (10, 35–37). The structure of the RAD4-RAD23 protein consists of a transglutaminase-homology domain (TGD, blue in Fig. 2*B*) and three beta hairpin domains (BHD 1 to 3, green, cyan, and yellow respectively in Fig. 2*B*) (10, 35–37). Similar to the RAD4-RAD23/PAH-DNA adduct complex examined by Mu et al. (36), the productive complex examined in this work is characterized by the TGD and BHD1 domains forming

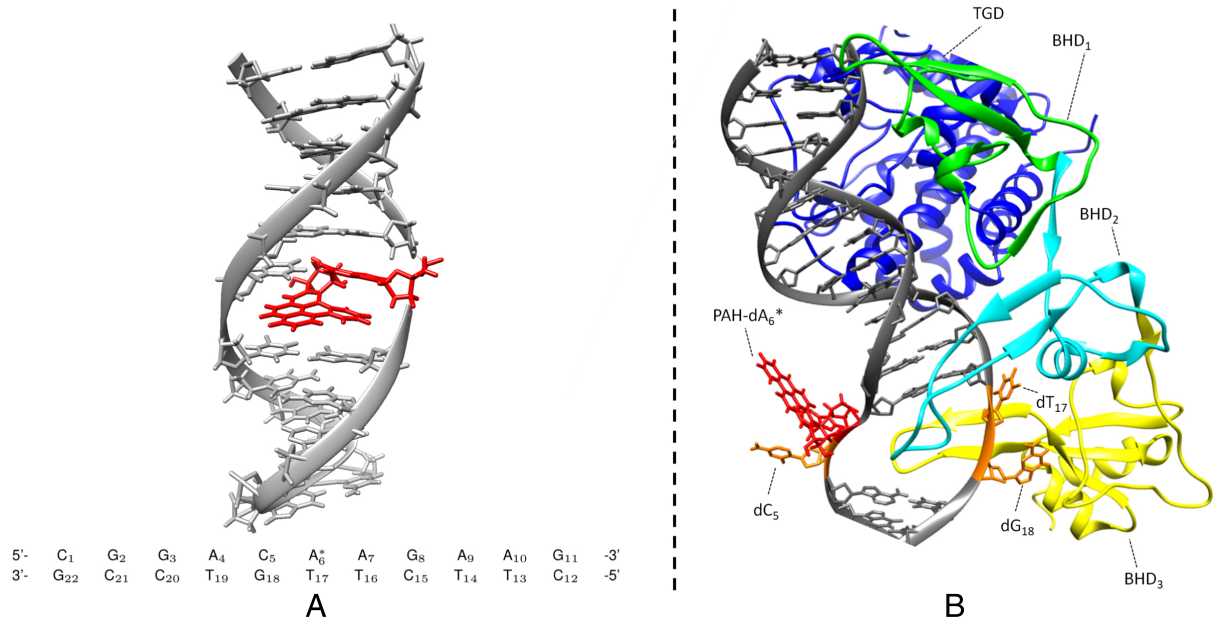

Fig. 2. (*A*) PAH-DNA adduct based on the NMR solution structure of a stereochemically analogous (1S,2R,3S,4R)-trans-anti-B[a]A-DE-N6-dA$^*_6$ adduct in the NRAS(Q61) sequence context [note the reverse order of atom numbering vs. B[a]P, PDB: 1DL4 (17)]. (*B*) RAD4-RAD23 in the productive complex with a PAH-DNA adduct based on the X-ray crystal structure of RAD4-RAD23 bound to a 6-4 photoproduct UV lesion [PDB: 6CFI (35)].

a clamp-like structure as seen in Fig. 2*B* and is modeled on PDB: 6CFI (35). The adducted central adenine dA$^*_6$ of NRAS(Q61) (red in Fig. 2*B*) and its 5′ neighboring dC$_5$ (orange in Fig. 2*B*) are extruded from the hydrophobic core of the DNA duplex and exposed to solvent. The partner dG$_{18}$ and dT$_{17}$ nucleobases in the complementary strand (orange in Fig. 2*B*) are also extruded and bound by the BHD2 and BHD3 domains of RAD4-RAD23. The BHD2 domain contacts the minor groove and the BHD3 domain inserts via the major groove occupying the space left in the DNA duplex by the extruded nucleotides (10, 35, 36).

In order to elucidate relationships between calculated relative binding and complexation affinities and the aromatic ring structure of the PAHs studied in this work, we examine van der Waals interactions resulting from intercalation of PAH-DNA adducts, the intercalated conformations assumed by these adducts, the resultant distortions in DNA rigid body parameters, and disruptions in hydrogen bonding between canonical DNA base pairs in the NRAS(Q61) 3-mer. Our approach establishes a workflow for in silico determination of relative binding and complexation affinities of PAHs as compared to the known carcinogen B[a]P in the NRAS(Q61) sequence context, providing insight into their potential genotoxicity. It is transferable to other PAHs or other classes of compounds that form covalent DNA adducts in other DNA sequence contexts. This in turn allows for rapid identification of compounds of interest for either environmental risk factor assessments or in vitro and in vivo genotoxicity assays.

Table 1. PAHs examined in this work

PAH	Abbreviation	IARC group	Bay/Fjord
Phenanthrene	PHE	3	Bay
Chrysene	CHR	2B	Bay
Benzo[a]pyrene	B[a]P	1	Bay
Benzo[e]pyrene	B[e]P	3	Bay
Dibenzo[a,h]pyrene	DB[a,h]P	2B	Bay
Dibenzo[a,i]pyrene	DB[a,i]P	2B	Bay
Dibenzo[a,e]pyrene	DB[a,e]P	3	Bay
Dibenzo[e,l]pyrene	DB[e,l]P	3	Bay
Benz[a]anthracene	B[a]A	2B	Bay
Dibenzo[a,c]anthracene	DB[a,c]A	3	Bay
Dibenzo[a,h]anthracene	DB[a,h]A	2A	Bay
Dibenzo[a,j]anthracene	DB[a,j]A	3	Bay
Benzo[b]chrysene	B[b]C	3	Bay
Benzo[c]phenanthrene	B[c]P	2B	Fjord
Benzo[g]chrysene	B[g]C	3	Fjord
Dibenzo[a,l]pyrene	DB[a,l]P	2A	Fjord

Methods

Calculating absolute free energies in systems consisting of biological macromolecules is largely intractable due to the size of the systems of interest and the challenges posed by quasi-nonergodicity, where such systems may be formally ergodic but computational simulations of such systems do not properly sample phase space (38, 39). This may result in calculated statistical averages being strongly dependent on initial simulation conditions, yielding inaccurate results. Alternative simulation approaches designed to calculate the change in free energy due to binding ($\Delta G_{Binding}$) or complexation ($\Delta G_{Complex}$) such as umbrella sampling are often hampered by the need to sample the entire binding and unbinding processes along a reaction coordinate while utilizing biasing potentials to overcome potential barriers (38, 39). In order to circumvent these challenges, we employ the alchemical free energy perturbation (FEP) approach over closed thermodynamic cycles.

Alchemical Free Energy Perturbation Calculations. Simulations meant to simulate systems of biological macromolecules under physiological conditions are carried out under constant temperature and pressure, resulting in an *NPT* ensemble with partition function $Q(N, P, T)$. The Gibbs free energy is then defined as

$$G = -\frac{1}{\beta}\ln[Q(N, P, T)], \quad [1]$$

where $\beta = 1/k_BT$. From this, we can determine the difference in free energy between an initial state A and final state B of the system utilizing the FEP approach of Zwanzig: (38, 39)

$$\Delta G_{A\to B} = G_B - G_A = -\frac{1}{\beta}\ln\frac{Q_B}{Q_A}$$
$$= -\frac{1}{\beta}\ln\left\langle e^{-\beta[\mathbf{H}_B(\mathbf{p},\mathbf{q})-\mathbf{H}_A(\mathbf{p},\mathbf{q})]}\right\rangle_A, \quad [2]$$

where $\mathbf{H}_A$ and $\mathbf{H}_B$ are the Hamiltonians for the initial state A and final state B respectively, $\mathbf{p}$ is the momentum vector, $\mathbf{q}$ is the coordinate vector, and the quantity in brackets is an ensemble average.

Note however that FEP calculations carried out with this approach will only provide accurate estimates of the free energy difference between states if the target state B is sufficiently similar to the reference state A (38, 39). Similarity of target and reference states can be gauged in terms of overlap of important regions in the phase space of each state. Important regions in phase space are those containing configurations with highly probable energies that make the largest contribution to the estimated free energy in Eq. **2**. These important regions must be sufficiently sampled in both the reference and target states to obtain an accurate estimate of the difference in free energy between the two (38, 39). If there is insufficient overlap of the phase space of the reference and target states, configurations generated in the reference state A will be high energy states with low probability when evaluated using the Hamiltonian of the target state $\mathbf{H}_B$ and will thus make a minimal contribution to the ensemble average in Eq. **2** leading to inaccurate results (38, 39).

To address the key issue of overlap of important regions in phase space when calculating differences in free energy of binding and complexation for pairs of PAHs with different aromatic ring structures, the coupled Hamiltonian/dual topology approach (38–40) is employed. In this approach, the topologies of the initial state PAH_A and the final state PAH_B are defined concurrently, and an unphysical path from PAH_A to PAH_B consisting of a discrete set of N intermediate states is utilized, exploiting the fact that free energy is a state function and thus independent of path. Along this discrete set of intermediate states, moieties from PAH_A and PAH_B incrementally fade out or fade in. To this end, the parameter $\lambda_i \in [0, 1]$, where $i = 0, 1, \ldots, N$ is introduced, and the coupled Hamiltonian is then defined as (38–40):

$$\mathbf{H}_{\lambda_i} = (1-\lambda_i)\mathbf{H}_A + \lambda_i\mathbf{H}_B. \quad [3]$$

For $\lambda_0 = 0$, $\mathbf{H}_{\lambda_0} = \mathbf{H}_A$ and for $\lambda_N = 1$, $\mathbf{H}_{\lambda_N} = \mathbf{H}_B$. For $\lambda_i \in (0, 1)$, the system topology is in an unphysical intermediate state between PAH_A and PAH_B, hence the term "alchemical transformation" from the initial to the final state. The free energy difference in Eq. **2** can then be estimated as the sum of the free energy differences between intermediate states λ_i and λ_{i+1}:

$$\Delta G_{A\to B} = -\frac{1}{\beta}\sum_{i=0}^{N-1}\ln\left\langle e^{-\beta[\mathbf{H}_{\lambda_{i+1}}(\mathbf{p},\mathbf{q})-\mathbf{H}_{\lambda_i}(\mathbf{p},\mathbf{q})]}\right\rangle_i. \quad [4]$$

Equilibrium sampling in the reference state λ_i is carried out, and for each configuration, the energy is evaluated using the Hamiltonian $\mathbf{H}_{\lambda_i}$ and then evaluated again using the Hamiltonian $\mathbf{H}_{\lambda_{i+1}}$. For each configuration, the energy difference is evaluated, and a corresponding ensemble average is computed to estimate the free energy (38–40). This is carried out for both the forward alchemical transformation $PAH_A \to PAH_B$ and the backward alchemical transformation $PAH_B \to PAH_A$ so that individual states serve as both reference and target states (38–40). This approach ensures sufficient overlap of important regions in phase space and has the advantage of being computationally tractable because the alchemical transformation from PAH_A to PAH_B requires a much smaller perturbation of the system than approaches such as umbrella sampling (38, 39).

Closed Thermodynamic Cycles. We utilize the alchemical FEP approach described above over closed thermodynamic cycles to calculate the relative difference in free energy of binding ($\Delta\Delta G_{Binding}$) and the relative difference in free energy of formation of the productive complex ($\Delta\Delta G_{Complex}$) for a given pair of PAHs. In order to describe the approach, we utilize DB[a,l]P (PAH_A) and B[a]P (PAH_B) as an illustrative pair in the thermodynamic cycles depicted in Figs. 3 and 4.

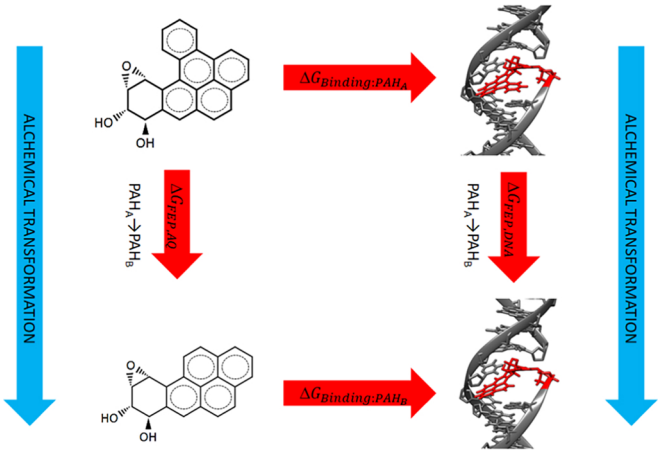

Fig. 3. Closed thermodynamic cycle examining $\Delta\Delta G_{Binding}$.

To calculate $\Delta\Delta G_{Binding}$ for a PAH_A-DNA adduct as compared to a PAH_B-DNA adduct at dA^*_6 in NRAS(Q61), we employ a closed thermodynamic cycle where two alchemical FEP calculations are performed (Fig. 3). The first is an alchemical transformation from PAH_A-DE $\to$ PAH_B-DE in solution that yields $\Delta G_{FEP,AQ:PAH_A\to PAH_B}$ (Fig. 3, *Left* leg). The second is an alchemical transformation from PAH_A-DNA $\to$ PAH_B-DNA in the NRAS(Q61) centered DNA 11-mer that yields $\Delta G_{FEP,DNA:PAH_A\to PAH_B}$ (Fig. 3, *Right* leg). Because free energy is a state function and thus independent of path, we have that:

$$\Delta G_{Binding:PAH_A} + \Delta G_{FEP,DNA:PAH_A\to PAH_B} = \Delta G_{FEP,AQ:PAH_A\to PAH_B} + \Delta G_{Binding:PAH_B} \quad [5]$$

and hence that:

$$\Delta\Delta G_{Binding} = \Delta G_{Binding:PAH_B} - \Delta G_{Binding:PAH_A} = \Delta G_{FEP,DNA:PAH_A\to PAH_B} - \Delta G_{FEP,AQ:PAH_A\to PAH_B} \quad [6]$$

Although we have not directly calculated $\Delta G_{Binding:PAH_A}$ and $\Delta G_{Binding:PAH_B}$ (Fig. 3, *Top* and *Bottom* legs), we are able to calculate $\Delta\Delta G_{Binding}$ using values obtained from the alchemical FEP calculations in order to determine whether PAH_A or PAH_B is more likely to form a PAH-DNA adduct.

In order to calculate $\Delta\Delta G_{Complex}$ for a PAH_A-DNA adduct as compared to a PAH_B-DNA adduct, we employ a closed thermodynamic cycle analogous to the one described above. The first is the alchemical transformation described above from PAH_A-DNA $\to$ PAH_B-DNA in the NRAS(Q61) centered DNA 11-mer that yields $\Delta G_{FEP,DNA:PAH_A\to PAH_B}$ (Fig. 4, *Left* leg). The second is an alchemical transformation from PAH_A-DNA $\to$ PAH_B-DNA in the productive complex that yields $\Delta G_{FEP,Complex:PAH_A\to PAH_B}$ (Fig. 4, *Right* leg). From this thermodynamic cycle, we have that:

$$\Delta G_{Complex:PAH_A} + \Delta G_{FEP,Complex:PAH_A\to PAH_B} = \Delta G_{FEP,DNA:PAH_A\to PAH_B} + \Delta G_{Complex:PAH_B} \quad [7]$$

and hence that:

$$\Delta\Delta G_{Complex} = \Delta G_{Complex:PAH_B} - \Delta G_{Complex:PAH_A} = \Delta G_{FEP,Complex:PAH_A\to PAH_B} - \Delta G_{FEP,DNA:PAH_A\to PAH_B} \quad [8]$$

As above, we are able to calculate $\Delta\Delta G_{Complex}$ using values obtained from the alchemical FEP calculations in order to determine whether the PAH_A-DNA adduct or the PAH_B-DNA adduct is more likely to form the productive complex, despite not having directly calculated $\Delta G_{Complex:PAH_A}$ and $\Delta G_{Complex:PAH_B}$ (Fig. 4, *Top* and *Bottom* legs).

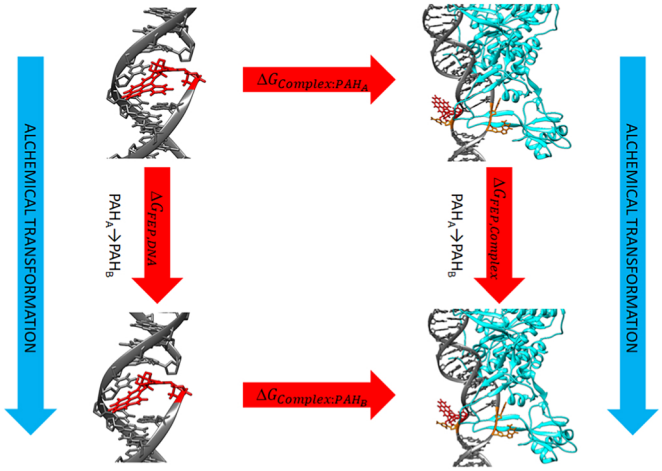

Fig. 4. Closed thermodynamic cycle examining $\Delta\Delta G_{Complex}$.

To examine factors that contribute to the relative genotoxicity of the PAHs listed in Table 1, we calculate $\Delta\Delta G_{Binding}$ and $\Delta\Delta G_{Complex}$ as compared to the IARC Group 1 carcinogen B[a]P. Alchemical FEP calculations are carried out over closed thermodynamic cycles for the pairs of PAHs connected by arrows in *SI Appendix*, Fig. S.4. Where necessary, $\Delta\Delta G_{Binding}$ and $\Delta\Delta G_{Complex}$ are calculated by concatenating thermodynamic cycles as depicted in *SI Appendix*, Fig. S.5 where for example:

$$\begin{aligned}\Delta\Delta G_{Binding} &= \Delta G_{Binding:PAH_C} - \Delta G_{Binding:PAH_A} \\ &= \left(\Delta G_{FEP,DNA:PAH_A \rightarrow PAH_B} + \Delta G_{FEP,DNA:PAH_B \rightarrow PAH_C}\right) \\ &\quad - \left(\Delta G_{FEP,AQ:PAH_A \rightarrow PAH_B} + \Delta G_{FEP,AQ:PAH_B \rightarrow PAH_C}\right). \quad [9]\end{aligned}$$

Computational Parameters. The approach described above is implemented utilizing molecular dynamics (MD) simulations in NAMD (40–42) with the CHARMM molecular mechanics force field (43–45). Force field parameterization, computational parameters for MD and FEP simulations, and the structure of the dual topology model systems developed for this work are described in *SI Appendix*, Sections 1 and 2. Briefly, force field parameters for PAH-DNA adducts were assigned and optimized utilizing quantum mechanical target data, the CHARMM General Force Field/ParamChem.com (45–47), and the VMD-Force Field Toolkit (48). PAH-DNA linkage dihedral parameters were optimized utilizing truncated singular value decompositions (49). Systems were solvated in explicit TIP3P (50) water with counterions, relaxed with harmonic constraints, and equilibrated for 100 ns of MD prior to FEP calculations. Alchemical FEP transformations between PAHs were conducted over 20 or 40 intermediate states, depending on the particular pair of PAHs, in order to achieve sufficient phase space overlap between intermediate states. Each intermediate MD window consisted of 200 ps of equilibration and 800 ps of production. Further details of the alchemical stratification approach are described in *SI Appendix*. The VMD ParseFEP plugin (51) was utilized to calculate free energy differences using the Bennett Acceptance Ratio (52) and to evaluate convergence of FEP calculations.

Results

Relative Free Energies of Binding and Complexation. $\Delta\Delta G_{Binding}$ and $\Delta\Delta G_{Complex}$ for PAH-DNA adducts as compared to a B[a]P-DNA adduct were calculated for the PAHs listed in Table 1 as described above. Results of free energy calculations are listed in Table 2 and plotted in Fig. 5 for each PAH from *Left* to *Right* in order of increasing $\Delta\Delta G_{Binding}$ (blue bars), $\Delta\Delta G_{Complex}$ is plotted in orange bars, and PAH structures are depicted at the top of the figure. ParseFEP probability distribu-

Table 2. Average PAH linkage torsion angles α'_{avg} and β'_{avg} and average dA^*_6 glycosidic torsion angle χ_{avg}; $E_{vdW:dT_{16}|dT_{17}}$, $E_{vdW:dA^*_6|dA_7}$: average van der Waals interactions between the aromatic rings of the PAH and the nucleobases of the primary and secondary intercalation pockets, respectively (SDs in parenthesis); $E_{vdW:Intercalation} = E_{vdW:dT_{16}|dT_{17}} + E_{vdW:dA^*_6|dA_7}$: total van der Waals interactions from PAH intercalation; $\Delta\Delta G_{Binding}$: relative free energy of binding for a (-R,-S,-R,-S)-trans-anti-PAH-DE-N6-dA^*_6 adduct as compared to a (7R,8S,9R,10S)-trans-anti-B[a]P-DE-N6-dA^*_6 adduct at the central dA^*_6 in NRAS(Q61); $\Delta\Delta G_{Complex}$: relative free energy of complexation for the corresponding productive complex

	PAH-DNA adduct	α'_{avg} degrees	β'_{avg} degrees	χ_{avg} degrees	$E_{vdW:dT_{16}\|dT_{17}}$ kcal/mol	$E_{vdW:dA^*_6\|dA_7}$ kcal/mol	$E_{vdW:Interc.}$ kcal/mol	$\Delta\Delta G_{Binding}$ kcal/mol	$\Delta\Delta G_{Complex}$ kcal/mol
Strongly preferred	DB[a,l]P	56.59 (8.43)	84.52 (7.67)	−114.91 (21.66)	−13.38 (1.13)	−14.11 (1.33)	−27.49	−14.34	3.46
	B[g]C	56.55 (8.61)	81.25 (8.32)	−108.73 (19.28)	−12.17 (1.42)	−13.07 (1.60)	−25.24	−13.82	2.98
	B[c]P	7.34 (12.43)	144.42 (13.36)	−79.4 (21.63)	−12.74 (1.18)	−4.08 (2.13)	−16.82	−9.38	1.96
	DB[a,j]A	39.12 (11.42)	112.31 (8.29)	−91.33 (19.83)	−13.2 (0.99)	−10.72 (1.34)	−23.92	−6.96	4.31
	DB[a,c]A	45.77 (11.68)	111.69 (11.47)	−94.76 (23.96)	−14.8 (1.21)	−7.42 (1.73)	−22.22	−6.17	4.75
	DB[a,h]A	44.46 (11.17)	102.04 (9.67)	−90.63 (19.11)	−13.12 (1.03)	−8.38 (1.62)	−21.50	−5.68	3.80
Weakly preferred	B[a]A	41.39 (12.85)	107.55 (12.00)	−84.86 (17.79)	−12.49 (1.05)	−6.63 (1.54)	−19.12	−2.23	0.59
	DB[a,e]P	116.85 (16.91)	4.68 (21.69)	−140.24 (10.82)	−8.23 (0.95)	−9.88 (1.78)	−18.11	−2.14	1.68
	DB[a,i]P	74.78 (42.3)	69.0 (48.95)	−122.28 (25.21)	−13.31 (2.14)	−8.70 (2.75)	−22.01	−2.00	−0.26
	B[b]C	49.86 (18.49)	91.76 (18.42)	−99.64 (26.25)	−12.19 (1.39)	−7.70 (1.84)	−19.89	−1.65	1.32
Equally preferred	DB[a,h]P	73.51 (34.78)	66.19 (40.42)	−133.2 (20.31)	−11.83 (1.76)	−10.13 (1.99)	−21.96	−0.53	1.91
	PHE	16.2 (13.47)	146.05 (13.70)	−64.72 (16.63)	−10.14 (0.99)	−1.16 (1.36)	−11.30	−0.51	0.88
	B[a]P	5.40 (11.86)	154.1 (22.27)	−64.32 (9.31)	−13.69 (1.01)	−2.04 (1.32)	−15.73	0.00	0.00
	CHR	25.28 (24.09)	127.71 (33.36)	−75.96 (20.93)	−12.32 (1.10)	−3.94 (2.72)	−16.26	0.68	−1.08
Non preferred	DB[e,l]P	−0.14 (9.53)*	−33.89 (9.11)	−61.88 (8.63)	−7.73 (1.21)	−1.22 (1.09)	−8.95	1.88	−1.54
	B[e]P	28.53 (20.10)	137.48 (14.00)	−71.42 (23.42)	−12.92 (1.37)	−2.81 (1.25)	−15.73	4.11	−3.49
	Unmodified	N/A	N/A	−108.79 (12.06)	−5.62 (0.64)	−7.17 (0.72)	−12.79	N/A	N/A

*See *SI Appendix*, section 3.8 for description of α' measurement in DB[e,l]P.

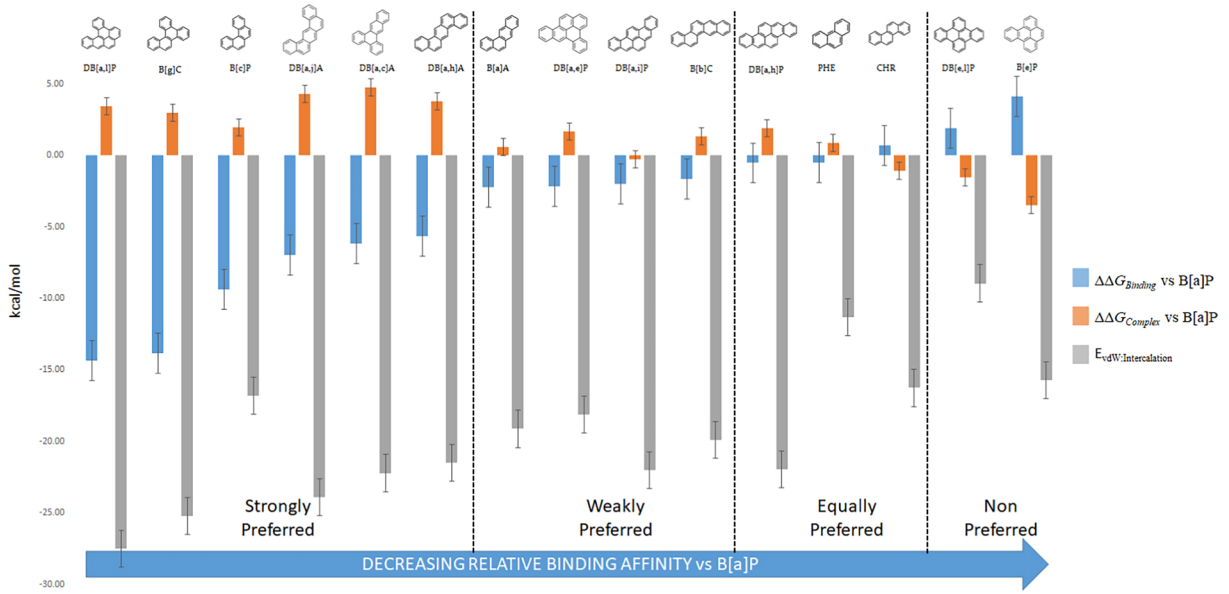

Fig. 5. Relative free energies of binding and complexation ($\Delta\Delta G_{Binding}$ – blue bars & $\Delta\Delta G_{Complex}$ – orange bars) for (-R,-S,-R,-S)-trans-anti-PAH-DE-N6-dA$^*_6$ adducts as compared to a (7R,8S,9R,10S)-trans-anti-B[a]P-DE-N6-dA$^*_6$ adduct at the central dA$^*_6$ in NRAS(Q61) and total van der Waals interactions from PAH-DNA adduct intercalation ($E_{vdW:Intercalation} = E_{vdW:dT_{16}\,|\,dT_{17}} + E_{vdW:dA^*_6\,|\,dA_7}$ – gray bars). Plots show SEs. ParseFEP computed errors are listed in *SI Appendix*, Table S.7.

tion, convergence plots, and computed errors for all alchemical FEP calculations are included in *SI Appendix*, Appendix A and Table S.7. Note that larger values of $\Delta\Delta G_{Binding}$ and $\Delta\Delta G_{Complex}$ correspond to lower relative binding and complexation affinities. For discussion purposes, we group PAHs by relative binding affinity into four categories as compared to B[a]P:

Strongly Preferred	$-14.34\text{kcal/mol} < \Delta\Delta G_{Binding} < -5.68\text{kcal/mol}$
Weakly Preferred	$-2.23\text{kcal/mol} < \Delta\Delta G_{Binding} < -1.65\text{kcal/mol}$
Equally Preferred	$-0.53\text{kcal/mol} < \Delta\Delta G_{Binding} < 0.68\text{kcal/mol}$
Nonpreferred	$1.88\text{kcal/mol} < \Delta\Delta G_{Binding} < 4.11\text{kcal/mol}$

Notably, the strongly preferred PAHs, those with the greatest relative binding affinity, were also found to have the lowest relative complexation affinity with $1.96\text{ kcal/mol} < \Delta\Delta G_{Complex} < 4.75\text{ kcal/mol}$, indicating that those PAHs more likely to form PAH-DNA adducts as compared to B[a]P were also less likely to form the productive complex that constitutes the GG-NER recognition step. Meanwhile, the weakly, equally, and nonpreferred PAHs, those with lower relative binding affinities, were found to have greater relative complexation affinities than the strongly preferred PAHs with $-3.49\text{ kcal/mol} < \Delta\Delta G_{Complex} < 1.91\text{ kcal/mol}$, indicating that those PAHs less likely to form PAH-DNA adducts than the strongly preferred systems are also more likely to form the productive complex.

In particular, we find that the fjord region DB[a,l]P system exhibits the greatest relative binding affinity ($\Delta\Delta G_{Binding} = -14.34$ kcal/mol) and among the weakest relative complexation affinities ($\Delta\Delta G_{Complex} = 3.46$ kcal/mol). These findings are consistent with studies examining the efficiency of GG-NER of PAH-DNA adducts at the central adenine dA$^*_6$ of NRAS(Q61) in human HeLA cell extracts which found that fjord region (11R,12S,13R,14S)-trans-anti-DB[a,l]P-DE-N6-dA$^*_6$ adducts were almost entirely repair resistant while stereochemically analogous bay region (7R,8S,9R,10S)-trans-anti-B[a]P-DE-N6-dA$^*_6$ adducts were repaired with high efficiency (7, 8). This is largely due to the nonplanar structure of DB[a,l]P that results from the steric repulsion between hydrogens on opposite ends of the fjord region (Fig. 1*D*) which allows DB[a,l]P-DNA adducts to intercalate in energetically favorable conformations that minimize distortions of the DNA duplex and result in stabilizing van der Waals interactions between the DB[a,l]P moiety and neighboring nucleobases (i.e. enhanced π-stacking), which in turn results in an adduct that evades the GG-NER recognition step (6–8, 11, 13, 53–58). These features are examined in further detail for DB[a,l]P and the other PAHs examined in this work in the sections below.

In addition to DB[a,l]P, the fjord region B[g]C and B[c]P and bay region DB[a,j]A, DB[a,c]A, and DB[a,h]A (henceforth DB[a,j/c/h]A) systems were also strongly preferred while exhibiting the weakest relative complexation affinities. Meanwhile, the bay region B[a]A, DB[a,e]P, DB[a,i]P, and B[b]C systems were weakly preferred followed by the DB[a,h]P, PHE, and CHR systems which were equally preferred and the DB[e,l]P and B[e]P systems which were nonpreferred. These systems also exhibited the greatest relative complexation affinities.

van der Waals Interactions. π-stacking between nucleobases is of fundamental importance to the conformational stability of DNA and it is known that intercalated PAH-DNA adducts can either increase or decrease the melting point of DNA depending on the PAH and sequence context (6, 15–17). With the exception of the nonpreferred DB[e,l]P system (discussed in *SI Appendix*, section 3.8), all PAH-DNA adducts examined in this work intercalate on the 3′ side of the modified dA$^*_6$ from the major groove with the aromatic rings of the PAH situated in a primary intercalation pocket formed by dT_{16} and dT_{17} in the complementary strand (Fig. 6 *A* and *B* blue box) and a secondary intercalation pocket formed by dA$^*_6$ and dA_7 in the adducted strand (Fig. 6*B* green box) that is accessible to varying degrees dependent upon the glycosidic conformation of dA$^*_6$ (i.e. syn or anti) and the intercalated conformation of the given PAH-DNA adduct. In order to examine associations between relative free energies of binding/complexation and van der Waals interactions from PAH-DNA adduct intercalation, average van der Waals between the aromatic rings of the PAH-DNA adduct and the

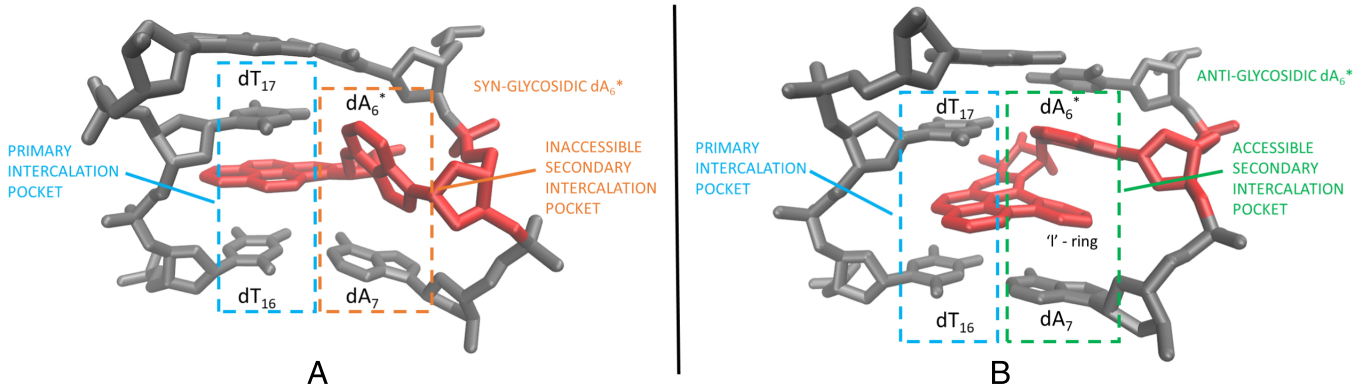

Fig. 6. Average structure of (*A*) syn-glycosidic (7R,8S,9R,10S)-trans-B[a]P-DE-N6-dA$^*_6$ adduct and (*B*) anti-glycosidic (11R,12S,13R,14S)-trans-DB[a,l]P-DE-N6-dA$^*_6$ adduct.

nucleobases that form each intercalation pocket ($E_{vdW:dT_{16}\,|\,dT_{17}}$ and $E_{vdW:dA^*_6\,|\,dA_7}$ respectively) were calculated utilizing Python scripts and VMD as described in *SI Appendix*, Section 2. Total van der Waals from PAH-DNA adduct intercalation ($E_{vdW:Intercalation} = E_{vdW:dT_{16}\,|\,dT_{17}} + E_{vdW:dA^*_6\,|\,dA_7}$) are plotted with $\Delta\Delta G_{Binding}$ and $\Delta\Delta G_{Complex}$ in Fig. 5 (gray bars) and are listed in Table 2.

With the exception of B[c]P (discussed below), the strongly preferred PAHs generally exhibited the greatest total van der Waals with -27.49 kcal/mol $< E_{vdW:Intercalation} < -21.50$ kcal/mol, and total van der Waals among the strongly preferred PAHs decreased monotonically with decreasing relative binding affinity, demonstrating an enthalpic drive. Furthermore, as the productive complex is characterized by rupture of the dA$^*_6$:dT$_{17}$ base pair and extrusion of the corresponding nucleobases, strong van der Waals in the primary and secondary intercalation pocket contribute to formation of the productive complex being enthalpically unfavorable. In contrast, B[a]P and the equally preferred PHE, CHR, and nonpreferred B[e]P systems which exhibit greater relative complexation affinity also exhibit the least total van der Waals with -16.26 kcal/mol $< E_{vdW:Intercalation} < -11.30$ kcal/mol, making formation of the productive complex more favorable in these systems.

Strong total van der Waals are not driven only by the number of aromatic rings in the given PAH, but also their configuration which in turn impacts the intercalated conformation of the PAH-DNA adduct as discussed below. The strongly preferred PAHs have aromatic ring structures that span both the primary dT$_{16}$ | dT$_{17}$ and secondary dA$^*_6$ | dA$_7$ intercalation pockets, maximizing total van der Waals and avoiding steric clashes that result in unfavorable distortions of the DNA duplex. Meanwhile, the weakly, equally, and nonpreferred PAHs have suboptimal structures due to a lack of aromatic rings and/or aromatic ring configurations that fail to effectively span both intercalation pockets thus limiting total van der Waals (Fig. 7*A*) or resulting in unfavorable structural distortions of the DNA duplex to avoid steric clashes.

In particular, we find that the strongly preferred fjord region PAHs DB[a,l]P and B[g]C exceed their respective bay region analogs B[a]P and CHR (equally preferred) in total van der Waals. This is due to the flexible and nonplanar aromatic ring structure that is characteristic of fjord region PAHs and which facilitates optimal intercalated conformations of the PAH-DNA adduct described below. In contrast bay region PAHs have rigid and planar aromatic ring structures that in many cases hinder optimal intercalated conformations of the PAH-DNA adduct (Fig. 7*A*). However, as described below, we find that the strongly preferred bay region DB[a,j/c/h]A systems which also exhibit the least relative complexation affinity among the PAHs examined in this work, have aromatic ring structures that achieve intercalated conformations resulting in strong total van der Waals.

Discussion

Intercalated Conformations. Equilibrated PAH-DNA adduct systems were used to calculate average glycosidic torsion angles of the adducted dA$^*_6$ (χ_{avg}) and average PAH-DNA adduct linkage torsion angles (α'_{avg} and β'_{avg}) (Fig. 1*B* and Table 2) utilizing Python scripts and VMD following an approach similar to Cai et al. (53). Average structures were calculated utilizing the NAFlex server (59). Plots of MD trajectories are included in *SI Appendix*, B. In order to illustrate the impact of aromatic ring structure on intercalated conformation and resultant van der Waals interactions we begin by comparing and contrasting the structural and conformational features of B[a]P and DB[a,l]P.

The rigid and planar bay aromatic ring structure of B[a]P results in the adducted dA$^*_6$ assuming an average syn-glycosidic conformation ($\chi_{avg} = -64.32°$) accompanied by the PAH linkage torsion angle β' assuming a particularly wide obtuse angle

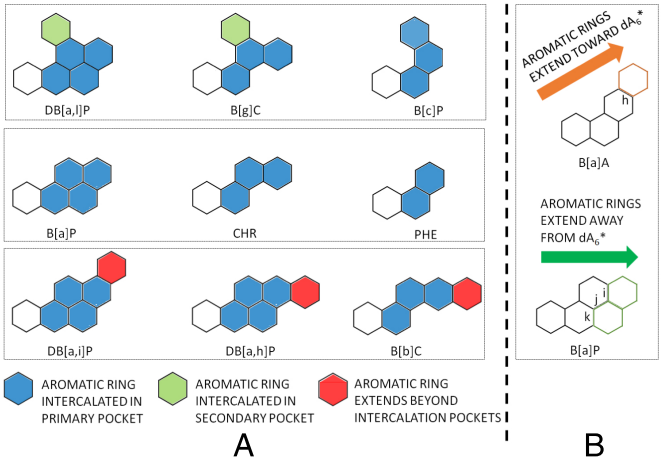

Fig. 7. (*A*) Impact of aromatic ring structure on total van der Waals where optimal structures span both intercalation pockets and maximize total van der Waals. (*B*) Distinct bay region aromatic ring structures of B[a]A and B[a]P built upon a PHE root compound.

(β'_{avg} = 154.10°) in order to avoid steric clashes between the B[a]P and dA^*_6 moieties. In this conformation, the plane of dA^*_6 is nearly perpendicular to dA_7 and to the plane of the aromatic rings of B[a]P in the average structure, resulting in the secondary intercalation pocket becoming totally inaccessible (Fig. 6*A* orange box). The wide angle assumed by β' serves to swing the B[a]P moiety away from the dA^*_6 moiety, resulting in the aromatic rings of B[a]P intercalating solely in the primary intercalation pocket (Fig. 6*A* blue box). The PAH linkage torsion angle α' assumes an average value of α'_{avg} = 5.40° in order to compensate for the syn-glycosidic conformation of dA^*_6 by leveling the aromatic rings of B[a]P in the primary intercalation pocket. As a result, $E_{vdW:dT_{16}\,|\,dT_{17}} = -13.69$ kcal/mol in the primary intercalation pocket, which is strong as compared to unmodified DNA, where van der Waals interactions between dT_{16} and dT_{17} in the absence of an intercalated PAH average −5.62 kcal/mol (Table 2). However, in the secondary intercalation pocket, $E_{vdW:dA^*_6\,|\,dA_7} = -2.04$ kcal/mol, which is weak as compared to unmodified DNA where dA^*_6 maintains an average anti-glycosidic conformation and van der Waals interactions between dA^*_6 and dA_7 in the absence of an intercalated PAH average −7.17 kcal/mol. Total van der Waals in the B[a]P system come to $E_{vdW:Intercalation} = -15.73$ kcal/mol which is markedly less than the DB[a,l]P system described below. This intercalated conformation, resultant limitations in stacking interactions, and disruption of DNA rigid body parameters described below are consistent with those described in the works of Zegar et al. (15, 16) which examined NMR solution structures of (7S,8R,9S,10R)-trans-anti-B[a]P-DE-N6-dA^*_6 adducts in NRAS(Q61) which intercalate from the major groove on the 5′ side of the modified dA^*_6 due to the 10R-stereochemistry of the adduct, as well as the work of Yan et al. that compared and contrasted the stereoisomeric 10(S) and 10(R) B[a]P-DNA adducts in NRAS(Q61) (13). The equally preferred PHE and CHR and the nonpreferred B[e]P systems assume similar intercalated conformations with correspondingly diminished total van der Waals, weak relative binding affinity, and greater relative complexation affinity as compared to the strongly preferred systems (Table 2 and *SI Appendix*, Figs. S.20, S.22, and S.24).

In contrast to B[a]P, the flexible and nonplanar fjord aromatic ring structure of DB[a,l]P allows the modified dA^*_6 to maintain an average anti-glycosidic conformation (χ_{avg} = −114.91°) characteristic of unmodified DNA and the PAH linkage torsion angle β' assumes an acute angle (β'_{avg} = 84.52°) without steric clashes between the DB[a,l]P and dA^*_6 moieties. In this conformation, the planes of the dA^*_6 and dA_7 nucleobases remain approximately parallel in the average structure, making the secondary intercalation pocket fully accessible. The narrow angle assumed by β' swings the aromatic rings of DB[a,l]P back toward the dA^*_6 moiety and the nonplanar aromatic ring structure allows the fjord aromatic ring to intercalate unimpeded in the secondary intercalation pocket formed by dA^*_6 and dA_7 (Fig. 6*B* green box) while the nonfjord aromatic rings intercalate in the primary intercalation pocket (Fig. 6*B* blue box). The PAH linkage torsion angle α' assumes an average value of α'_{avg} = 56.59° in order to level the aromatic rings of DB[a,l]P in the primary and secondary intercalation pockets. As a result, DB[a,l]P exhibits strong van der Waals in both intercalation pockets with $E_{vdW:dT_{16}\,|\,dT_{17}} = -13.38$ kcal/mol and $E_{vdW:dA^*_6|dA_7} = -14.11$ kcal/mol, resulting in $E_{vdW:Intercalation} = -27.49$ kcal/mol which is markedly greater than B[a]P and the strongest total van der Waals among the PAHs examined in this work. Concomitantly, DB[a,l]P exhibits the greatest relative binding affinity with $\Delta\Delta G_{Binding} = -14.34$ kcal/mol and among the weakest relative complexation affinities with $\Delta\Delta G_{Complex} = 3.46$ kcal/mol. The fjord region B[g]C system assumes a similar intercalated conformation with correspondingly strong van der Waals interactions, strong relative binding affinity, and weak relative complexation affinity (Table 2). This is in contrast to its bay region analog CHR described above.

The other strongly and weakly preferred bay region PAHs examined in this work assume average intercalated conformations that are intermediate between those of DB[a,l]P and B[a]P described above, with comparatively strong average van der Waals in the primary intercalation pocket while average van der Waals in the secondary intercalation pocket generally diminish with decreasing relative binding affinity owing to varied availability of the secondary intercalation pocket that is driven by the aromatic ring structure of the given PAH and its intercalated conformation. This is well illustrated by considering the strongly preferred bay region DB[a,j/c/h]A systems which each have one additional aromatic ring on a B[a]A root compound and assume intercalated conformations that are very similar to one another. The aromatic ring structure of these systems differs from that of B[a]P when viewed as built upon a PHE root compound. In the DB[a,j/c/h]A systems, the aromatic rings build "diagonally" off of the "h" side of PHE while the aromatic rings in B[a]P build "laterally" off of the "i," "j," and "k" sides of PHE (Fig. 7*B*). The DB[a,j/c/h]A systems can thus be viewed as each having at least one additional aromatic ring extending toward the dA^*_6 nucleobase that the B[a]P system does not have. Thus steric clashes cannot be avoided by χ and α' assuming values that orient dA^*_6 perpendicular to the PAH with concomitant opening of the β' torsion angle as seen in B[a]P, because the distinct aromatic ring structure of the DB[a,j/c/h]A systems would still clash with the dA^*_6 moiety unless β' was markedly wider, which would likely result in an unfavorable conformation. Thus in the DB[a,j/c/h]A systems, the modified dA^*_6 assumes a glycosidic conformation that fluctuates about the −90° anti/syn threshold during equilibration, assuming average values of χ_{avg} = −91.33°, −94.76°, and −90.63° respectively while the PAH linkage torsion angle β' assumes average values of β'_{avg} = 112.31°, 111.69°, and 102.04° and α' assumes average values of α'_{avg} = 39.12°, 45.77°, and 44.46° respectively. These average conformational parameters are intermediate between those of DB[a,l]P and B[a]P.

In this conformational motif, the plane of dA^*_6 is subparallel to that of dA_7 (i.e. approximately diagonal) in the average structure, creating a partially accessible secondary intercalation pocket (*SI Appendix*, Figs. S.12–S.14). The average angle assumed by β' in these systems is closer to that assumed by DB[a,l]P, allowing the aromatic rings of the DB[a,j/c/h]A systems to interact with the partially accessible secondary intercalation pocket. This results in the aromatic rings of the DB[a,j/c/h]A systems being situated for strong van der Waals interactions in the primary intercalation pocket that are comparable to or stronger than the DB[a,l]P system while van der Waals interactions in the secondary intercalation pocket are generally limited by comparison but stronger than those observed in B[a]P (Table 2). As a result, total van der Waals in the DB[a,j/c/h]A systems are less than those observed in the DB[a,l]P system while still being markedly greater than B[a]P with $E_{vdW:Intercalation}$ = −23.92, −22.22, and −21.50 kcal/mol respectively. In turn, these systems are strongly preferred with $\Delta\Delta G_{Binding}$ = −6.96, −6.17, and −5.68 kcal/mol respectively. Relative free energies of complexation for these systems are $\Delta\Delta G_{Complex}$ = 4.31, 4.75, and 3.80 kcal/mol respectively, indicating that these systems also have weak relative complexation affinities that accompany their

strong relative binding affinities and total van der Waals. The details of van der Waals interactions in the DB[a,j/c/h]A systems are nuanced because of the unique location of the additional aromatic ring in each system and these are discussed further in *SI Appendix*, section 3.2. Similar intercalated conformations are assumed by the weakly preferred B[a]A and B[b]C systems, resulting in total van der Waals of −19.12 and −19.89 kcal/mol respectively which are less than those observed in the strongly preferred systems described above. Correspondingly, the B[a]A and B[b]C systems exhibit weaker relative binding affinities and greater relative complexation affinities (Table 2). These systems are discussed further in *SI Appendix*, sections 3.6 and 3.7, but we note here that the intercalated conformation assumed by B[a]A is consistent with the NMR solution structure of the (1S,2R,3S,4R)-trans-anti-B[a]A-DE-N6-dA^*_6 adduct in the NRAS(Q61) sequence context examined by Li et al. [PDB: 1DL4 (17)], and upon which the PAH-DNA adduct systems in this work were modeled.

The strongly preferred fjord region B[c]P system is nuanced, assuming an average intercalated conformation that is similar to the bay region B[a]P (*SI Appendix*, Fig. S.15) rather than the fjord region DB[a,l]P system, with $\chi_{avg} = -79.40°$ (syn-glycosidic), $\alpha'_{avg} = 7.34°$, and $\beta'_{avg} = 144.42°$. Like the B[a]P system, B[c]P's aromatic rings, including the fjord aromatic ring, are situated solely in the primary intercalation pocket resulting in $E_{vdW:dT_{16}\,|\,dT_{17}} = -12.74$ kcal/mol and $E_{vdW:dA^*_6\,|\,dA_7} = -4.08$ kcal/mol. Total van der Waals are comparable to B[a]P with $E_{vdW:Intercalation} = -16.82$ kcal/mol which is markedly weaker than the other fjord region DB[a,l]P and B[g]C systems where the fjord aromatic ring intercalates in the secondary intercalation pocket. When comparing the aromatic ring structures of DB[a,l]P and B[g]C to B[c]P, it is evident that this difference is due to B[c]P having fewer nonfjord aromatic rings available to intercalate in the primary intercalation pocket as compared to DB[a,l]P and B[g]C (Fig. 7*A*). In the absence of these nonfjord aromatic rings, the aromatic rings of B[c]P are unable to span both the primary and secondary intercalation pockets to achieve strong van der Waals interactions in both, resulting in a conformation that maximizes van der Waals interactions in the primary intercalation pocket alone. This intercalated conformation, along with disruption of DNA rigid body parameters described below, is consistent with the NMR solution structure of a stereochemically identical B[c]P-DNA adduct in a CA*C sequence context described by Cosman et al. (60). The strongly preferred bay region DB[a,j/c/h]A systems discussed above are further illustrative of this concept, whereby the aromatic ring configurations of these systems are better suited to span both intercalation pockets as evidenced by the greater van der Waals observed in each intercalation pocket as compared to B[c]P (Table 2). Consequently, the relative binding affinity of the B[c]P system is markedly weaker than the DB[a,l]P and B[g]C systems with $\Delta\Delta G_{Binding} = -9.38$ kcal/mol, and it exhibits the greatest relative complexation affinity among the strongly preferred systems with $\Delta\Delta G_{Complex} = 1.96$ kcal/mol. We note however that B[c]P exhibits greater relative binding affinity than the strongly preferred DB[a,j/c/h]A systems despite having weaker total van der Waals. This is likely due to an entropic effect that will be examined in future work.

The weakly preferred DB[a,e]P and DB[a,i]P and the equally preferred DB[a,h]P systems each have one additional aromatic ring ("e," "i," and "h") on a B[a]P root compound and assume unique intercalated conformations in which dA^*_6 assumes an average anti-glycosidic conformation accompanied by varied α' and β' PAH linkage torsion angles that accommodate the aromatic ring structures of these PAHs (Table 2). The conformational details of these systems are examined further in *SI Appendix*, sections 3.3–3.5; however, it is of key importance to note that total van der Waals in these systems ($E_{vdW:Intercalation}$ = −18.11, −22.01, and −21.96 kcal/mol respectively) are weaker than the −25.24 and −23.92 kcal/mol observed in the B[g]C and DB[a,j]A systems and are also comparable to or weaker than the −22.22 and −21.50 kcal/mol observed in DB[a,c]A and DB[a,h]A systems, all of which have one less aromatic ring. The location of the "e," "i," and "h" aromatic rings in these systems result in structures and intercalated conformations that fail to effectively span both intercalation pockets, thus limiting total van der Waals (Figs. 7*A* and *SI Appendix*, Figs. S.18, S.19, and S.21). This highlights the impact of optimal aromatic ring structures that facilitate intercalated conformations that span both intercalation pockets and maximize total van der Waals.

DNA Rigid Body Parameters and Hydrogen Bonding. Being that the productive complex is characterized by rupture and extrusion of the dC_5:dG_{18} and dA^*_6:dT_{17} base pairs, examining the impact of PAH-DNA adduct intercalation on DNA rigid body parameters (*SI Appendix*, Section 4) and hydrogen bond occupancies provides further critical insight into the mechanisms that facilitate formation of the productive complex. Utilizing the equilibration trajectories of the PAH-DNA adduct systems, DNA rigid body parameters were calculated in x3DNA (61–63) (Fig. 8 and *SI Appendix*, Section 4) and hydrogen bond occupancies for the canonical dC_5:dG_{18}, dA^*_6:dT_{17}, and dA_7:dT_{16} base pairs in NRAS(Q61) were computed based on bond distances (*SI Appendix*, Table S.5) and angles following an approach similar to Cai et al. (53) utilizing Python scripts and VMD (Fig. 9). Plots of MD trajectories (hydrogen bond distances/angles and DNA rigid body parameters) are included in *SI Appendix*, B. The impact of aromatic ring structure on distortions in DNA rigid body parameters, resultant disruptions in hydrogen bond occupancy as compared to unmodified DNA, and associations with relative binding and complexation affinity are once again well illustrated by comparing and contrasting the B[a]P and DB[a,l]P systems.

As compared to unmodified DNA, the B[a]P system exhibits marked distortions in buckle (+40.16°), propeller (−28.65°), and stretch (+0.22 Å) of the dA^*_6:dT_{17} base pair that are not observed in the DB[a,l]P and B[g]C systems (Fig. 8). These are a result of the average syn-glycosidic conformation assumed by dA^*_6 in the B[a]P system described above and result in an increase of the average dA^*_6-N1:dT_{17}-N3 hydrogen bond distance to 3.02 Å as compared to 2.96 Å in unmodified DNA (*SI Appendix*, Table S.5). This results in a decrease of −12.00 percentage points in the dA^*_6-N1:dT_{17}-N3 hydrogen bond occupancy in the B[a]P system (Fig. 9). Similar distortions in buckle, propeller, and stretch accompanied by disruptions in the dA^*_6-N1:dT_{17}-N3 hydrogen bond are observed in the equally preferred CHR and PHE and the nonpreferred B[e]P systems which each assume an intercalated conformation similar to B[a]P (*SI Appendix*, Figs. S.32 and S.34 and Fig. 9).

In contrast, the average anti-glycosidic conformation maintained by dA^*_6 in the strongly preferred fjord region DB[a,l]P and B[g]C systems results in comparatively small distortions in buckle (+10.02° and +13.09°), propeller (−0.34° and +0.60°), and stretch (+0.08 Å and +0.07 Å) (Fig. 8), resulting in no disruptions of the dA^*_6-N1:dT_{17}-N3 hydrogen bond (Fig. 9). Recalling that the DB[a,l]P and B[g]C systems exhibit the greatest relative binding affinities, weak relative complexation affinities, and the greatest total van der Waals, while the B[a]P, CHR, PHE, and B[e]P systems exhibit weak relative binding affinities,

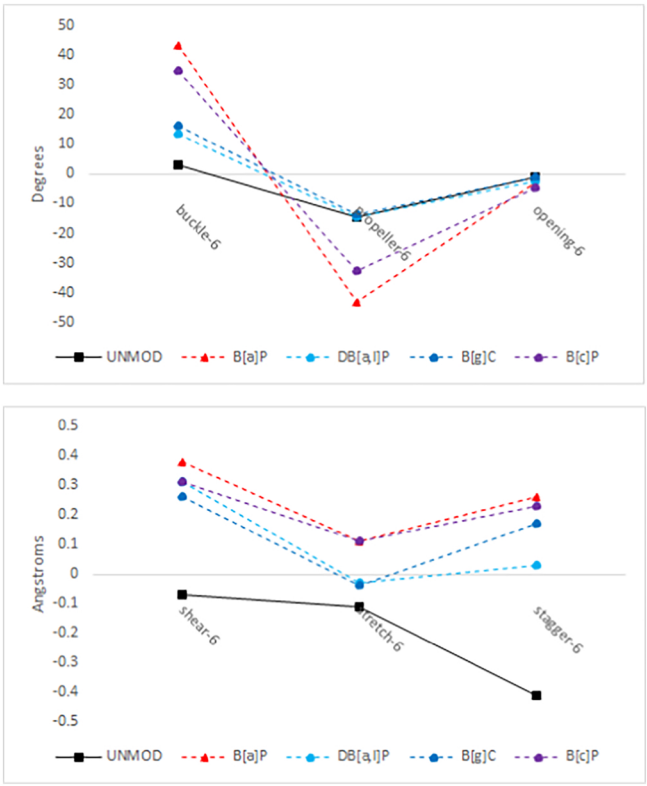

Fig. 8. DNA rigid body parameters of the dA^*_6:dT_{17} base pair for unmodified DNA and B[a]P, DB[a,l]P, B[g]C, and B[c]P DNA adducts.

greater relative complexation affinities, and the weakest total van der Waals by comparison, it is evident that disruption of the dA^*_6-N1:dT_{17}-N3 hydrogen bond in conjunction with weak total van der Waals facilitates formation of the productive complex while the absence of hydrogen bond disruption coupled with strong total van der Waals in the DB[a,l]P and B[g]C systems hampers it.

Recalling that the B[c]P system also assumes an intercalated conformation very similar to B[a]P despite being a strongly preferred fjord region PAH, we observe distortions in buckle (+31.52°), propeller (−18.32°), and stretch (+0.22 Å) of the dA^*_6:dT_{17} base pair that are similar to those observed in B[a]P (Fig. 8). Accordingly, the average dA^*_6-N1:dT_{17}-N3 bond distance increases to 3.05 Å, accompanied by a decrease of −14.70 percentage points in hydrogen bond occupancy (Fig. 9). Recalling that B[c]P exhibits the greatest relative complexation affinity and weakest total van der Waals among the strongly preferred PAHs, this finding is consistent with the notion that disruption of the dA^*_6-N1:dT_{17}-N3 hydrogen bond in conjunction with weak total van der Waals facilitates formation of the productive complex.

Turning our attention to the dA^*_6-N6:dT_{17}-O4 hydrogen bond, there are decreases of −1.68°, −2.00°, and −3.80° in opening of the dA^*_6:dT_{17} base pair in the DB[a,l]P, B[a]P, and B[c]P systems (Fig. 8 and *SI Appendix*, Fig. S.27) that compensate for rigid body distortions in the dA^*_6:dT_{17} base pair by maintaining the average dA^*_6-N:dT_{17}-O4 hydrogen bond distance at 2.94 Å, 2.96 Å, and 2.89 Å respectively as compared to 2.94 Å in unmodified DNA (*SI Appendix*, Table S.5). As a result, there are no disruptions of the dA^*_6-N6:dT_{17}-O4 hydrogen bond in these systems (Fig. 9). In contrast, there is only a −0.16° decrease in opening of the dA^*_6:dT_{17} base pair in the B[g]C system which results in an average 2.98 Å dA^*_6-N6:dT_{17}-O4 hydrogen bond distance and a small −3.95 percentage point decrease in hydrogen bond occupancy. Similar compensating decreases in opening of the dA^*_6:dT_{17} base pair are observed in the CHR and B[e]P systems, but in the PHE system a larger increase of +0.81 Å in stagger of the dA^*_6:dT_{17} base pair that is not seen in the CHR and B[e]P systems results in an increase in the dA^*_6-N6:dT_{17}-O4 hydrogen bond distance to 2.97 Å and an accompanying −9.21 percentage point decrease in hydrogen bond occupancy (Fig. 9 and *SI Appendix*, Figs. S.32 and S.34). This phenomenon and its impact on formation of the product complex is examined further in the DB[a,j/c/h]A systems.

The intercalated conformation assumed by the strongly preferred DB[a,j/c/h]A systems, with dA^*_6 fluctuating about the syn/anti-glycosidic threshold as described above, results in distortions in buckle (+31.00°, +29.18°, and +31.07°) and propeller (−21.60°, −15.67°, and −18.00°) that are intermediate between those seen in the B[a]P and DB[a,l]P systems while distortions in stretch (+0.21 Å, +0.21 Å, and +0.21 Å) are nearly identical to the B[a]P system (*SI Appendix*, Fig. S.28). As a result, the average dA^*_6-N1:dT_{17}-N3 hydrogen bond distance increases to 3.09 Å, 3.05 Å, and 3.08 Å, resulting in decreases of −26.90, −16.67, and −20.94 percentage points in occupancy respectively (Fig. 9). Compared to the B[a]P and B[c]P systems, there are larger compensating decreases of −8.03°, −5.06°, and −6.55° in opening of the dA^*_6:dT_{17} base pair in the DB[a,j/c/h]A systems that maintain the average dA^*_6-N6:dT_{17}-O4 hydrogen bond distances at 2.83 Å, 2.94 Å, and 2.85 Å respectively. Accordingly, occupancy in this hydrogen bond is enhanced by 15.36, 4.87, and 13.50 percentage points respectively (Fig. 9). Distortions in buckle, propeller, and stretch accompanied by compensating decreases in opening and disruptions and enhancements in hydrogen bonding are similar in the weakly preferred B[a]A and B[b]C systems which assume intercalated conformations similar to the DB[a,j/c/h]A systems (*SI Appendix*, Fig. S.30 and Fig. 9). Despite hydrogen bond disruptions and enhancements being similar between these two sets of systems, the weakly preferred B[a]A and B[b]C systems exhibit greater relative complexation affinities than the strongly preferred DB[a,j/c/h]A systems which exhibit stronger total van der Waals. Hence, strong total van der Waals are sufficient to hamper formation of the productive complex despite the observed hydrogen bond disruptions in the DB[a,j/c/h]A systems.

From this it is evident that the strongly preferred systems which exhibit the greatest relative binding affinity and least relative complexation affinity are characterized by either no disruption of hydrogen bonding in the dA^*_6:dT_{17} base pair coupled with strong total van der Waals (DB[a,l]P and B[g]C]) or disruption of the dA^*_6-N1:dT_{17}-N3 hydrogen bond occupancy accompanied by enhancement of the dA^*_6-N6:dT_{17}-O4 hydrogen bond occupancy with strong total van der Waals (DB[a,j/c/h]A).

As described in *SI Appendix*, Section 3, the weakly preferred DB[a,e]P, DB[a,i]P, and equally preferred DB[a,h]P systems assume unique intercalated conformations where dA^*_6 assumes an anti-glycosidic conformation accompanied by varied α' and β' PAH linkage torsion angles that accommodate the aromatic ring structures of the intercalated PAH-DNA adduct. These systems exhibit greater relative complexation affinities than the strongly preferred systems described above, and distortions to their DNA rigid body parameters in the dA^*_6:dT_{17} base pair result in disruptions to *both* the dA^*_6-N1:dT_{17}-N3 and the dA^*_6-N6:dT_{17}-O4 hydrogen bonds. These systems are the only three that exhibit increases in opening (+21.50°, +1.66°, and +9.62°) of the dA^*_6:dT_{17} base pair (*SI Appendix*, Figs. S.30 and S.32), which

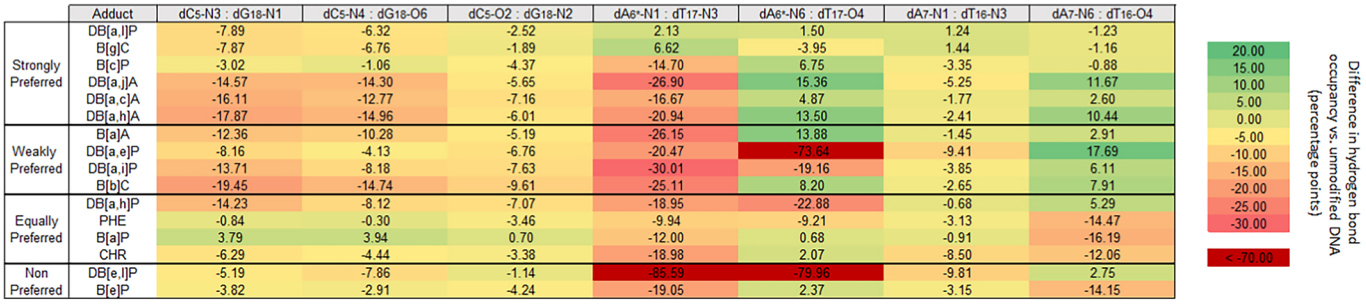

Fig. 9. Differences in hydrogen bond occupancy as compared to unmodified DNA (percentage points) for base pairs in the NRAS(Q61) 3-mer.

is accompanied by decreases in the dA^*_6-N6:dT_{17}-O4 hydrogen bond occupancy not observed in the systems discussed above (Fig. 9). DB[a,e]P is also the only system that exhibits a decrease in buckle (−2.18°) which is accompanied by a large increase of +0.42 Å in stretch and a decrease of −27.07° in propeller that is comparable to the B[a]P system. Along with a large increase of +1.21 Å in stagger, these distortions result in the dA^*_6-N1:dT_{17}-N3 hydrogen bond distance increasing to 3.06 Å and the dA^*_6-N6:dT_{17}-O4 hydrogen bond distance increasing to 4.63 Å. Accordingly, hydrogen bond occupancies decrease by −20.47 and −73.64 percentage points respectively (Fig. 9). The DB[a,i]P and DB[a,h]P systems exhibit distortions in buckle (+14.15° and +13.88°) that are similar to the B[g]C system and distortions in propeller (−16.73° and −12.96°) are similar to those observed in the DB[a,c]A system. The DB[a,i]P system exhibits a unique decrease of −0.20 Å in stretch accompanied by a large increase of +1.33 Å in shear that results in the dA^*_6-N1:dT_{17}-N3 hydrogen bond distance increasing to 3.68 Å and the dA^*_6-N6:dT_{17}-O4 hydrogen bond distance increasing to 3.89 Å accompanied by decreases in hydrogen bond occupancy of −30.01 and −19.16 percentage points respectively (Fig. 9). The DB[a,h]P system exhibits a uniquely large increase of +0.45 Å in stretch that results in the dA^*_6-N1:dT_{17}-N3 hydrogen bond distance increasing to 3.21 Å and the dA^*_6-N6:dT_{17}-O4 hydrogen bond distance increasing to 3.67 Å accompanied by decreases in hydrogen bond occupancy of −18.95 and −22.88 percentage points respectively (Fig. 9). These systems exhibit greater relative complexation affinities than the strongly preferred systems despite exhibiting comparable total van der Waals indicating that strong total van der Waals are not sufficient to hamper formation of the productive complex in the event of substantial disruption of both hydrogen bonds in the dA^*_6:dT_{17} base pair.

With the exception of B[c]P, hydrogen bonding in the dC_5:dG_{18} base pair is disrupted to varying extents in the strongly and weakly preferred systems and the equally preferred DB[a,h]P system (Fig. 9) which all exhibit comparatively strong van der Waals in both the primary and secondary intercalation pockets (Table 2). Meanwhile B[c]P and the equally preferred PHE and CHR and the nonpreferred B[e]P systems which intercalate solely in the primary intercalation pocket with minimal van der Waals in the secondary intercalation pocket exhibit minimal to no disruption of hydrogen bonding in the dC_5:dG_{18} base pair. Noting that the nucleobases of the dC_5:dG_{18} base pair are not immediately adjacent to the intercalated PAH-DNA adduct, and that the strongly preferred systems exhibit the least relative complexation affinity, it is evident that when considering formation of the productive complex, disruption of hydrogen bonding in the dC_5:dG_{18} base pair may be of secondary importance as compared to disruption of hydrogen bonding in the dA^*_6:dT_{17} base pair where the intercalated PAH is immediately adjacent to the nucleobases.

The impact of PAH-DNA adduct intercalation on the DNA rigid body parameters in the dC_5:dG_{18} and dA_7:dT_{16} base pairs, the base steps formed by the dC_5:dG_{18}/dA^*_6:dT_{17} base pairs and the dA^*_6:dT_{17}/dA_7:dT_{16} base pairs, and major and minor groove widths are outlined in *SI Appendix*, Section 4 and Appendix B and will be the subject of future work.

Conclusion

The fjord region PAHs DB[a,l]P, B[c]P, and B[g]C and the bay region PAHs DB[a,j]A, DB[a,c]A, and DB[a,h]A were found to exhibit the greatest relative binding affinity, least relative complexation affinity, and strongest total van der Waals from PAH-DNA adduct intercalation as compared to B[a]P, contributing to their genotoxic potential at the central adenine of NRAS(Q61) as compared to B[a]P. Meanwhile the bay region B[a]P, PHE, CHR, and B[e]P systems exhibited the least relative binding affinity, greatest relative complexation affinity, and weakest total van der Waals as compared to B[a]P, indicating that these factors contribute less to the genotoxic potential of these compounds as compared to B[a]P. Total van der Waals were found to not merely be a function of the number of aromatic rings in a given PAH, but rather are dependent upon the intercalated conformation of the PAH-DNA adduct which is in turn dependent upon the given PAH's aromatic ring structure. In particular, the PAHs with aromatic ring structures that effectively span the intercalation pockets formed by the dA^*_6 | dA_7 and dT_{16} | dT_{17} nucleobases achieve the greatest total van der Waals, resulting in stabilizing nonbonded interactions that increase relative binding affinity and decrease relative complexation affinity. Distortions in DNA rigid body parameters and disruptions in hydrogen bonding in the NRAS(Q61) 3-mer are also found to be dependent upon a given PAH's aromatic ring structure and result in increased relative complexation affinity when coupled with weak total van der Waals. These findings make clear that a wider swath of PAHs, including those with heteroatoms and substituent functional groups should be studied to examine the structural and thermodynamic features that contribute to their relative genotoxicities and to better inform risk factor assessments for the occupational and public health communities.

Data, Materials, and Software Availability. All study data are included in the article and/or *SI Appendix*.

ACKNOWLEDGMENTS. This research was supported in part by the California Breast Cancer Research Program (B30MP8524), using the high-performance computing resources of the University of California Los Angeles Department of Chemistry and Biochemistry. Portions of this paper were developed from the Ph.D. thesis of D.J.U.

1. International Agency for Research on Cancer – World Health Organization, IARC monographs on the identification of carcinogenic hazards to humans. https://monographs.iarc.fr/list-of-classifications. Accessed 13 February 2022.
2. E. Dybing, P. E. Schwarze, P. Nafstad, K. Victorin, T. M. Penning, "PAHs in ambient air and cancer" in *Air Pollution and Cancer*, K. Straif, A. Cohen, J. Samet, Eds. (IARC Scientific Publication and International Agency for Research on Cancer, 2013), pp. 75–94.
3. R. J. Tsai *et al*., Risk of cancer among firefighters in California, 1988–2007. *Am. J. Ind. Med.* **58**, 715–729 (2015).
4. R. D. Daniels *et al*., Mortality and cancer incidence in a pooled cohort of US firefighters from San Francisco, Chicago, Philadelphia (1950–2009). *Occup. Environ. Med.* **71**, 388–397 (2014).
5. K. W. Fent *et al*., Systemic exposure to PAHs and benzene in firefighters suppressing controlled structure fires. *Ann. Occup. Hyg.* **58**, 830–845 (2014).
6. N. E. Geacintov, S. Broyde, Repair-resistant DNA lesions. *Chem. Res. Toxicol.* **30**, 1517–1548 (2017).
7. A. Luch, "On the impact of the molecule structure in chemical carcinogenesis" in *Molecular, Clinical and Environmental Toxicology. Volume 1: Molecular Toxicology*, A. Luch, Ed. (Birkhäuser Verlag, 2009), pp. 151–178.
8. S. Broyde *et al*., "Carcinogenicity, structure, and function" in *Chemical Carcinogenesis*, T. M. Penning, Ed. (Springer, 2011), pp. 181–207.
9. H. Zheng *et al*., Base flipping free energy profiles for damaged and undamaged DNA. *Chem. Res. Toxicol.* **23**, 1868–1870 (2010).
10. J. H. Min, N. P. Pavletich, Recognition of DNA damage by the Rad4 nucleotide excision repair protein. *Nature* **449**, 570–575 (2007).
11. Y. Cai *et al*., Free energy profiles of base flipping in intercalative polycyclic aromatic hydrocarbon-damaged DNA duplexes: Energetic and structural relationships to nucleotide excision repair susceptibility. *Chem. Res. Toxicol.* **26**, 1115–1125 (2013).
12. C. Posch *et al*., Phosphoproteomic analyses of NRAS(G12) and NRAS(Q61) mutant melanocytes reveal increased CK2a kinase levels in NRAS(Q61) mutant cells. *J. Invest. Dermatol.* **136**, 2041–2048 (2016).
13. S. Yan, R. Shapiro, N. E. Geacintov, S. Broyde, Stereochemical, structural, and thermodynamic origins of stability differences between stereoisomeric benzo[a]pyrene diol epoxide deoxyadenosine adducts in a DNA mutational hot spot sequence. *J. Am. Chem. Soc.* **123**, 7054–7066 (2001).
14. International Programme on Chemical Safety, "Section 4.5 - Genotoxicity" in Principles and Methods for the Risk Assessment of Chemicals in Food (WHO Environmental Health Criteria 240, 2020), pp. 4-1–4-121.
15. I. S. Zegar *et al*., Adduction of the human N-ras codon 61 sequence with (-)-(7s,8r,9r,10s)-7,8-Dihydroxy-9, 10-epoxy-7,8,9, 10-tetrahydrobenzo[a]pyrene: Structural refinement of the intercalated SRSR (61,2) (-)-(7s,8r,9s,10r)-n^6-[10-(7,8,9, 10-tetrahydrobenzo[a]pyrenyl)]-2′ -deoxyadenosyl adduct from ^{1}H NMR. *Biochemistry* **35**, 6212–6224 (1996).
16. I. S. Zegar *et al*., Multiple conformations of an intercalated (-)-(7s,8r,9s,10r)-n^6-[10-(7,8,9, 10-tetrahydrobenzo[a]pyrenyl)]-2′-deoxyadenosyl adduct in the N-ras codon 61 sequence. *Biochemistry* **37**, 16516–16528 (1998).
17. Z. Li *et al*., Intercalation of the (1s,2r,3s,4r)-n^6-[1-(1,2,3,4-tetrahydro-2,3, 4-trihydroxybenz[a]anthracenyl)]-2′-deoxyadenosyl adduct in an oligodeoxynucleotide containing the human N-ras codon 61 sequence. *Biochemistry* **38**, 16045–16057 (1999).
18. H. Ling *et al*., Crystal structure of a blenzo[a]pyrene diol epoxide adduct in a ternary complex with a DNA polymerase. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 2265–2269 (2004).
19. B. Hwa Yun, J. Guo, M. Bellamri, R. J. Turesky, DNA adducts: Formation, biological effects, and new biospecimens for mass spectrometric measurements in humans. *Mass Spectrom. Rev.* **39**, 55–82 (2020).
20. C. Grill, L. Larue, NRAS, NRAS, Which Mutation is Fairest of Them All? *J. Invest. Dermatol.* **136**, 1936–1938 (2016).
21. I. A. Prior, P. D. Lewis, C. Mattos, A comprehensive survey of Ras mutations in cancer. *Cancer Res.* **72**, 2457–2467 (2012).
22. H. E. Kanavy, M. R. Gerstenblith, Ultraviolet radiation and melanoma. *Semin. Cutan. Med. Surg.* **30**, 222–228 (2011).
23. B. A. Gilchrest, M. S. Eller, A. C. Geller, M. Yaar, The pathogenesis of melanoma induced by ultraviolet radiation. *N. Engl. J. Med.* **340**, 1341–1348 (1999).
24. P. A. Demers *et al*., Carcinogenicity of occupational exposure as a firefighter. *Lancet Oncol.* **23**, 985–986 (2022).
25. D. M. Bolstad-Johnson *et al*., Characterization of firefighter exposures during fire overhaul. *AIHAJ Am. Ind. Hyg. Assoc.* **61**, 636–641 (2000).
26. D. Chakravarti, J. C. Pelling, E. L. Cavalieri, E. G. Rogan, Relating aromatic hydrocarbon-induced DNA adducts and c-H-ras mutations in mouse skin papillomas: The role of apurinic sites. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 10422–10426 (1995).
27. D. Chakravarti *et al*., The role of polycyclic aromatic hydrocarbon-DNA adducts in inducing mutations in mouse skin. *Mutat. Res. Toxicol. Environ. Mutagen.* **649**, 161–178 (2008).
28. B. Mahadevan *et al*., Mutations induced by (-)-anti-11R, 12S-dihydrodiol 13S,14R-epoxide of dibenzo[a,l]pyrene in the coding region of the hypoxanthine phosphoribosyltransferase (Hprt) gene in chinese hamster V79 cells. *Environ. Mol. Mutagen.* **41**, 131–139 (2003).
29. J. H. Yoon *et al*., DNA damage, repair, and mutation induction by (+)-syn and (-)-anti-dibenzo[a,l]pyrene-11, 12-diol-13, 14-epoxides in mouse cells. *Cancer Res.* **64**, 7321–7328 (2004).
30. J. T. Andersson, C. Achten, Time to say goodbye to the 16 EPA PAHs? Toward an up-to-date use of PACs for environmental purposes. *Polycycl. Aromat. Compd.* **35**, 330–354 (2015).
31. US Environmental Protection Agency, EPA priority pollutant list. https://www.epa.gov/sites/production/files/2015-09/documents/priority-pollutant-list-epa.pdf. Accessed 13February 2022.
32. C. C. Valentine *et al*., Direct quantification of in vivo mutagenesis and carcinogenesis using duplex sequencing. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 33414–33425 (2020).
33. Z. Cournia, B. Allen, W. Sherman, Relative binding free energy calculations in drug discovery. *J. Chem. Inf. Model.* **57**, 2911–2937 (2017).
34. Z. Li *et al*., Intercalation of the (1r,2s,3r,4s)-n^6-[1-(1,2,3,4-tetrahydro-2,3, 4-trihydroxybenz[a]anthracenyl)]-2′- deoxyadenosyl adduct in the N-ras codon 61 sequence: DNA sequence effects. *Biochemistry* **40**, 6743–6755 (2001).
35. D. Paul *et al*., Nar breakthrough article structure and mechanism of pyrimidine–pyrimidone. *Nucleic Acids Res.* **47**, 6015–6028 (2019).
36. H. Mu, N. E. Geacintov, J. H. Min, Y. Zhang, S. Broyde, NER lesion-recognition protein Rad4 captures a pre-flipped partner base in a benzo[a]pyrene-derived DNA lesion: How structure impacts the binding pathway. *Chem. Res. Toxicol.* **30**, 1344–1354 (2017).
37. Y. C. Lee *et al*., The relationships between XPC binding to conformationally diverse DNA adducts and their excision by the human NER system. *DNA Rep.* **19**, 55–63 (2014).
38. V. Gapsys, S. Michielssens, J. H. Peters, B. Groot, H. Leonov, "Calculation of binding free energies" in *Molecular Modeling of Proteins*, A. Kukol, Ed. (Springer, 2015), pp. 173–209.
39. C. Chipot, A. Pohorille, Free energy calculations. *Springer Ser. Chem. Phys.* **86**, 159–184 (2007).
40. J. Henin, J. Gumbart, C. Chipot, In silico alchemy: A tutorial for alchemical free-energy perturbation calculations with NAMD. https://www.ks.uiuc.edu/Training/Tutorials/namd/FEP/tutorial-FEP.pdf. Accessed 13 February 2022.
41. J. C. Phillips *et al*., Scalable molecular dynamics w/NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005).
42. J. C. Phillips *et al*., Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J. Chem. Phys.* **153**, 044130 (2020).
43. N. Foloppe, A. D. MacKerell, All-atom empirical force field for Nucleic Acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *J. Comput. Chem.* **21**, 86–104 (2000).
44. A. D. MacKerell, N. K. Banavali, All-atom empirical force field for Nucleic Acids: II. Application to MD simulations of DNA and RNA in solution. *J. Comput. Chem.* **21**, 105–120 (2000).
45. K. Vanommeslaeghe *et al*., CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **31**, 671–690 (2010).
46. K. Vanommeslaeghe, A. MacKerell Jr., Automation of the CHARMM general force field (CGenFF) I: Bond perception and atom typing. *J. Chem. Inf. Model.* **52**, 3144–3154 (2012).
47. K. Vanommeslaeghe, E. Raman, A. MacKerell Jr., Automation of the CHARMM general force field (CGenFF) II. *J. Chem. Inf. Model.* **52**, 3155–3168 (2012).
48. C. G. Mayne, J. Saam, K. Schulten, E. Tajkhorshid, J. C. Gumbart, Rapid parameterization of small molecules using the force field toolkit. *J. Comput. Chem.* **34**, 2757–2770 (2013).
49. D. J. Urwin, A. N. Alexandrova, Regularization of least squares problems in CHARMM parameter optimization by truncated singular value decompositions. *J. Chem. Phys.* **154**, 184101 (2021).
50. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
51. P. Liu, F. Dehez, W. Cai, C. Chipot, A toolkit for the analysis of free-energy perturbation calculations. *J. Chem. Theory Comput.* **8**, 2606–2616 (2012).
52. C. H. Bennett, Efficient estimation of free energy differences from monte carlo data. *J. Comput. Phys.* **22**, 245–268 (1976).
53. Y. Cai, N. E. Geacintov, S. Broyde, Nucleotide excision repair efficiencies of bulky carcinogen-DNA adducts are governed by a balance between stabilizing and destabilizing interactions. *Biochemistry* **51**, 1486–1499 (2012).
54. K. Kropachev *et al*., Adenine-DNA adducts derived from the highly tumorigenic dibenzo[a,l]pyrene are resistant to nucleotide excision repair while guanine adducts are not. *Chem. Res. Toxicol.* **26**, 783–793 (2013).
55. N. E. Geacintov *et al*., Thermodynamic and structural factors in the removal of bulky DNA adducts by the NER machinery. *Biopolym. Orig. Res. Biomol.* **65**, 202–210 (2002).
56. G. E. Menzies, I. A. Prior, A. Brancale, S. H. Reed, P. D. Lewis, Carcinogen-induced DNA structural distortion differences in the RAS gene isoforms; The importance of local sequence. *BMC Chem.* **15**, 1–13 (2021).
57. Q. Ruan *et al*., Synthesis and characterization of site-specific and stereoisomeric fjord dibenzo[a,l]pyrene diol epoxide- n^6-adenine adducts: Unusual thermal stabilization of modified DNA duplexes. *Chem. Res. Toxicol.* **15**, 249–261 (2002).
58. M. Wu, S. Yan, D. J. Patel, N. E. Geacintov, S. Broyde, Relating repair susceptibility of carcinogen-damaged DNA with structural distortion and thermodynamic stability. *Nucleic Acids Res.* **30**, 3422–3432 (2002).
59. A. Hospital *et al*., Naflex: A web server for the study of nucleic acid flexibility. *Nucleic Acids Res.* **41**, W47–W55 (2013).
60. M. Cosman *et al*., Solution conformation of the (-)-trans-anti-benzo[c]phenanthrene-dA ([BPh]dA) Adduct opposite dT in a DNA Duplex: Intercalation of the covalently attached Benzo[c]phenanthrenyl ring to the 3′-side of the adduct site and comparison with the (+)-trans-anti-[BPh]dA opposite dT stereoisomer. *Biochemistry* **34**, 1295–1307 (1995).
61. X. J. Lu, W. K. Olson, 3DNA: A software package for the analysis, rebuilding and visualization of 3D nucleic acid structures. *Nucleic Acids Res.* **31**, 5108–5121 (2003).
62. X. J. Lu, W. K. Olson, 3DNA: A versatile, integrated software system for the analysis, rebuilding and visualization of 3D nucleic-acid structures. *Nat. Protoc.* **3**, 1213–1227 (2008).
63. A. V. Colasanti, X. J. Lu, W. K. Olson, Analyzing and building nucleic acid structures with 3DNA. *J. Vis. Exp.* **74**, e4401 (2013).