JCTC Journal of Chemical Theory and Computation

Remarkably Strong T-Shaped Interactions between Aromatic Amino Acids and Adenine: Their Increase upon Nucleobase Methylation and a Comparison to Stacking

Lesley R. Rutledge and Stacey D. Wetmore*

Department of Chemistry and Biochemistry, University of Lethbridge, 4401 University Drive, Lethbridge, Alberta, Canada T1K 3M4

Received June 19, 2008

Abstract: T-shaped geometries and interaction energies between select DNA nucleobases (adenine or 3-methyladenine) and all aromatic amino acids (histidine, phenylalanine, tyrosine, or tryptophan) were examined using BSSE-corrected MP2/6−31G*(0.25) potential energy surface scans, which determined the preferred nucleobase (face)–amino acid (edge) and nucleobase (edge)–amino acid (face) interactions. The energies of dimers with the strongest interactions were further studied at the CCSD(T)/CBS level of theory, which suggests that the T-shaped interactions in adenine dimers are very strong (up to −35 kJ mol^{-1}). Nucleobase methylation to form a cationic damaged base (3-methyladenine) plays a large role in the relative monomer orientations and magnitude of the interactions, which increase by 17−125%. Most importantly, this study is the first to compare the stacking and T-shaped interactions between all aromatic amino acids and select (natural and damaged) DNA nucleobases where the differences between stacking and T-shaped interactions at the CCSD(T)/CBS level are small. Therefore, our results indicate that T-shaped interactions cannot be ignored when studying biological processes, and this manuscript discusses the importance of these interactions in the context of DNA repair.

1. Introduction

Noncovalent interactions play a major role in determining structures and properties of molecular assemblies in biology, chemistry, and materials science.[1] These interactions, which include hydrogen bonding, π-π, cation-π, and X–H$\cdots\pi$ (where X=N, O, C) among others, control the design of molecular devices and govern the self-assembly of natural and artificial systems.[2,3] In biology, the dynamic interactions between phospholipid bilayers and proteins, the double helical structure of DNA, and the three-dimensional structure of proteins are all dependent on noncovalent interactions.[4,5] Many pharmaceutical ligand–protein interactions are also noncovalent. For example, some drugs, including anticancer agents, use π-π interactions to intercalate into DNA.[6]

Noncovalent interactions between protein and DNA building blocks also play vital roles in the development of pharmaceuticals and biochemical techniques as well as in nature. For example, these interactions are essential for DNA replication,[7] transcription,[7] and DNA repair.[8,9] We are specifically interested in the role of these interactions in DNA base excision repair (BER). BER is perhaps the most important natural repair mechanism, which utilizes multiple enzymes.[8,9] Specifically, the first step in the BER process, where damaged bases are removed by enzymes (DNA glycosylases) that cleave the (glycosidic) bond connecting the base to the deoxyribose sugar,[10−13] likely relies on several noncovalent DNA−protein interactions for substrate identification and removal.

There are many classes of DNA glycosylases, which each act on damaged bases formed through different pathways (such as deamination, oxidation, or alkylation) and each use different catalytic mechanisms. For example, uracil DNA glycosylase (UDG) and formamidopyrimidine [fapy]-DNA glycosylase (FPG) use hydrogen bonds to bind and remove (neutral) damaged nucleobases formed via deamination and oxidation, respectively.[10,12] Alternatively, DNA glycosylases that repair alkylation damage, which include 3-methyladenine

* Corresponding author fax: (403)329-2057; e-mail: stacey.wetmore@uleth.ca.

10.1021/ct8002332 CCC: $40.75
Published on Web 09/03/2008

DNA glycosylase II (AlkA, *E. coli*) and human alkyladenine DNA glycosylase (AAG, human), possess no obvious polar groups that can form strong hydrogen bonds to the cationic damaged base in the active site pockets.[12] Instead, crystal structures[14,15] show stacking (face-to-face) and T-shaped (edge-to-face) interactions between the aromatic amino acids and nucleobases bound at the active site. Therefore, it has been proposed that stacking and T-shaped interactions are responsible for substrate recognition and stabilization.[12] Specifically, these enzymes may use aromatic π-cation interactions to attract cationic damaged bases into the active site and to differentiate between damaged and natural (undamaged) bases.[16]

Although crystal structures are highly informative about biomolecular geometries, they leave many unanswered questions about the nature or magnitude of discrete interactions between substrates and enzymes. Furthermore, in the case of DNA glycosylases that repair alkylation damage, no structure of a cationic nucleotide bound to the active sites exists. This is due to difficulties with experimental synthesis of stable cationic substrates,[12] which undergo spontaneous hydrolysis more rapidly than their neutral counterparts.[12,17,18] Therefore, unanswered questions related to DNA alkylation repair enzymes include the following: What is the nature of the attractive forces between active site residues and the substrates? What is the magnitude of these interactions? How does structure affect the magnitude of these interactions? What orientations between amino acids and bases are preferred? How do the interactions differ between neutral and cationic nucleobases? These lead to more general questions such as the following: Do these interactions affect the catalytic power of the enzyme?

Computational chemistry can address some of these unanswered questions regarding the nature of active site interactions. Previously, we used computational chemistry to systematically examine the MP2/6−31G*(0.25) stacking interactions between all aromatic amino acids (histidine (HIS), phenylalanine (PHE), tyrosine (TYR), tryptophan (TRP)) and adenine (A) or 3-methyladenine (3MeA, Figure 1),[19] which is the second most common alkylation product and has been shown to stop DNA replication.[20−23] This selection of molecules allowed us to separately characterize π-π and π-cation interactions and thereby determine the effects of nucleobase alkylation as well as the amino acid on the magnitude of stacking interactions.

Although our previous calculations revealed important information about biologically relevant stacking (face-to-face) interactions, crystal structures of AlkA and AAG show that the amino acid and substrate molecular planes are not always above/below each other in a perfectly parallel alignment. Instead, many T-shaped (edge-to-face) interactions also exist. Furthermore, these interactions have not been well characterized, and the influence of (cationic) charge on these interactions is even less understood. To fully understand the implications of these T-shaped interactions, we must also investigate the natural bases. Therefore, the T-shaped interactions between all aromatic amino acids and adenine or 3-methyladenine are systematically investigated in the present study. We consider both amino acid (edge)–nucleobase (face) and nucleobase (edge)–amino acid (face) interactions. High-level calculations (CCSD(T)) and extrapolation to the complete basis set limit were performed on dimer orientations that yield the strongest T-shaped interactions as well as the strongest stacking interactions previously reported.[24,25]

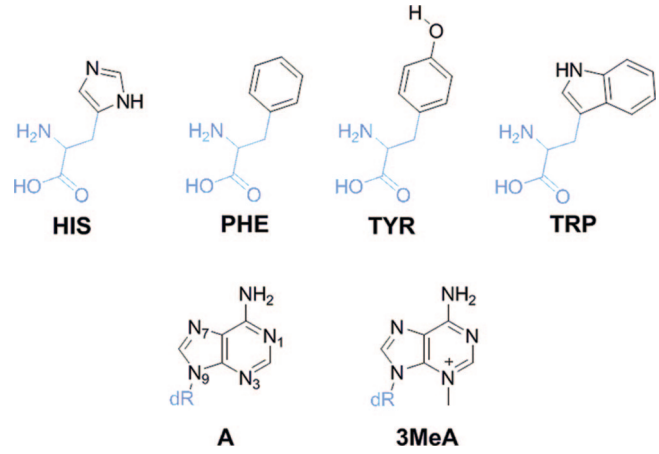

Figure 1. The structure of the amino acids (histidine (HIS), phenylalanine (PHE), tyrosine (TYR) and tryptophan (TRP)) and nucleobases (adenine (A) and 3-methyladenine (3MeA)) considered in this study, where blue fragments were replaced with a hydrogen atom in our models.

In addition to revealing important information (geometries and magnitudes) about biologically relevant T-shaped interactions, this work represents the most accurate comparison in the literature of the stacking and T-shaped interactions between natural or damaged nucleobases and the aromatic amino acids. Indeed, our work will reveal the magnitude of, and differences between, stacking and T-shaped interactions that occur within the active sites of DNA repair enzymes. This study also has more general implications due to the use, and significance, of noncovalent protein–DNA interactions in a variety of biological processes. Since information about the strengths of these interactions is difficult to extract from experiment alone, it is important to study these interactions using the highest levels of theory possible.

2. Computational Methods

Computational chemistry has been widely used to examine noncovalent interactions. Although an abundance of studies on hydrogen bonding exist, recent emphasis has been placed on anticipated weaker stacking (face-to-face) interactions between two aromatic rings. For example, there is a wealth of literature that examines the stacking interactions of benzene or substituted benzenes[26] as well as the stacking interactions between the natural nucleobases.[27−31] Stacking interactions in amino acid dimers have also been considered.[32] However, a comparatively limited number of studies have examined nucleobase–amino acid dimers. Rooman et al. investigated π-π or cation-π nucleobase–amino acid (Arg, Lys, Asn, and Gln) interactions[33] and more recently investigated the stacking interactions between (neutral or protonated) histidine and adenine or phenylalanine.[34] In addition, the stacking interactions between the four aromatic amino acids and the natural nucleobases[24,35] as well as the 10 most

common (cationic) methylated nucleobases[25] have been investigated. T-shaped (X–H···π, where X=N, O, C) interactions between small molecules or molecular fragments and various aromatic rings[36−42] or biomolecular (DNA or protein) π-systems have also been examined.[43−46] However, to our knowledge, very few studies have been completed on T-shaped (edge-to-face) interactions between two aromatic rings.[26,34,47] Indeed, these studies have been primarily limited to the T-shaped dimers between benzene and substituted benzenes.[26] In terms of biological systems, Cauët et al. investigated select T-shaped orientations between adenine and phenylalanine or histidine (neutral and protonated).[34] Despite their importance, a full systematic study of interactions between the aromatic amino acids and the nucleobases has not yet been performed.

The vast literature discussed above has revealed that stacking and T-shaped interactions are sensitive to the level of theory and basis set implemented, where appropriate levels of theory to study these interactions in biological systems have been identified. We use an approach similar to that used by Hobza and Šponer to study stacking interactions between DNA nucleobases[27] and our group to study stacking interactions between aromatic amino acids and nucleobases.[19,24,25] Specifically, monomers (Figure 1) were optimized in fixed planar geometries using MP2/6−31G(d). The nucleotide monomers were modeled by replacing the sugar−phosphate backbone with a hydrogen atom, while the protein backbone and β-carbon of the amino acids were replaced by a hydrogen atom. Therefore, HIS was modeled as imidazole, PHE as benzene, TRP as indole, and TYR as phenol. The gas-phase potential energy surfaces between monomers were scanned using basis set superposition error (BSSE)-corrected MP2 single-point calculations with the 6−31G*(0.25) basis set, which replaces the standard d-exponent for second-row atoms (0.8) with 0.25.[27] MP2/6−31G*(0.25) has been previously justified for use to study 'weak' interactions and has been shown to produce the same trends as, and recover approximately 80% of, the CCSD(T) stacking energies for natural nucleobase dimers calculated at the complete basis set (CBS) limit.[29] This manuscript will show an even better agreement between MP2 and CCSD(T) for nucleobase–amino acid stacking and T-shaped binding energies. During the potential energy surface scans, the relative orientations of the monomers were varied as a function of different variables as outlined in the following sections.

2.1. Stacking Interactions. Previously, we scanned the BSSE-corrected MP2/6−31G*(0.25) gas-phase potential energy surface of nucleobase–amino acid dimers by stacking monomers (face-to-face) with respect to their centers of mass.[19,24,25] Two relative orientations of the molecular planes were considered, where the first is defined by stacking the amino acid and nucleobase in the orientation shown in Figure 1 and the second, denoted as flipped using the subscript f, is obtained by flipping the amino acid relative to Figure 1 prior to stacking with the nucleobase. Three variables were investigated that define the relative orientation between the nucleobases and amino acids (Figure 2a): vertical separation (R_1), angle of rotation (α), and horizontal displacement (R_2). Our methodology for considering these variables is discussed

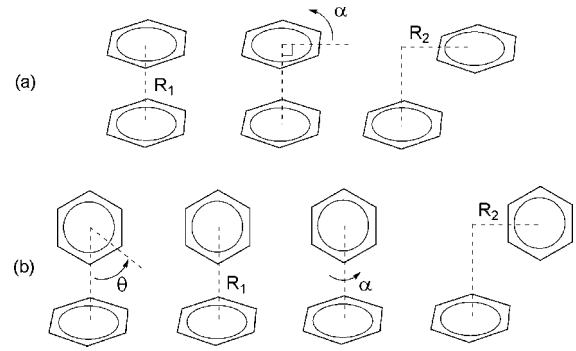

Figure 2. The definition of the variables considered in (a) previous stacking potential energy surface scans (vertical separation (R_1), angle of rotation (α), and horizontal displacement (R_2))[24,25] and (b) the present T-shaped potential energy surface scans (angle of 'edge' rotation (θ), vertical separation (R_1), angle of rotation (α), and horizontal 'edge' displacement (R_2)).

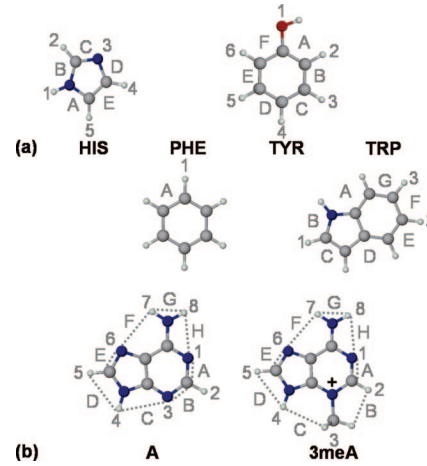

Figure 3. The definition of θ for (a) amino acid edges and (b) nucleobase edges considered in potential energy surface scans.

in detail in our previous publications.[24,25] In the present study, the strongest (most negative) MP2/6−31G*(0.25) stacking energies are reported, and the corresponding geometries were used for higher-level calculations (discussed below).

2.2. T-Shaped Interactions. To characterize a different part of the same potential energy surface examined in our previous study on nucleobase–amino acid stacking interactions, we used a series of gas-phase BSSE-corrected MP2/6−31G*(0.25) single-point calculations to identify the strongest T-shaped interactions between each amino acid and A or 3MeA. Four variables were considered (Figure 2b). First, the angles of 'edge' rotation (θ) were chosen, which define the ring edge (monomer edge) directed toward the center of mass of the π-system (monomer face). Figure 3 shows the amino acid and base edges considered. Our nomenclature uses numbers to indicate the atom directed toward the center of mass of the π-system and letters to indicate a bridged structure involving more than one atom directed toward the π-system. For example, in dimers involving a PHE edge, θ=1 indicates a hydrogen from the PHE ring is directed at the center of mass of the nucleobase,

while θ=A indicates a C−C bond of the benzene ring is parallel to the nucleobase molecular plane.[48] In total, 2 different edges were considered for PHE,[48] 10 for HIS and TRP, 12 for TYR, and 16 for A and 3MeA. Therefore, 196 different monomer orientations were considered (i.e., 34 amino acid edges directed toward each of two nucleobases and 32 nucleobase edges directed toward each of four amino acids). We emphasize that in addition to identifying the strongest T-shaped interactions between each amino acid and A or 3MeA, we have characterized interactions that may not correspond to the global minimum but may be important in biological systems due to natural dynamics or structural constraints of proteins or DNA.

The initial structures for dimers involving an amino acid edge were obtained by aligning the centers of mass of the amino acid and nucleobase and setting the molecular planes perpendicular. For dimers involving a nucleobase edge, the edge is sometimes located off the amino acid π-system when initial structures with aligned centers of mass are considered due to the smaller size of the amino acids. Therefore, when a nucleobase atom edge was directed toward the amino acid face (θ = number, Figure 3b), the atom was placed directly on top of the center of mass of the amino acid in the initial structure. Alternatively, when a nucleobase bond was set parallel to the amino acid face (θ = letter, Figure 3b), a dummy atom was placed at the midpoint between the atoms linked by the dotted lines in Figure 3b, and the dummy atom was aligned with the amino acid center of mass. In all T-shaped calculations, α=0° was defined as the structure with the monomer face in the XY plane, where the model glycosidic (nucleobase face) or β-carbon and peptide backbone (amino acid face) bond is parallel to the Y-axis. The monomer edge is placed in the YZ plane with the molecular plane parallel to the glycosidic (nucleobase face) or β-carbon and peptide backbone (amino acid face) bond.

Once the edges (θ) were chosen and α=0° was defined, the vertical separation distance (R_1) was altered by 0.1 Å increments along the Z-axis. For dimers involving the amino acid edge, R_1 is the distance between the center of mass of the two monomers. For the nucleobase edge dimers, R_1 is the distance between the amino acid center of mass and the nucleobase atom (θ = number) or the dummy atom at the midpoint of the line connecting the two atoms that define the monomer edge (θ = letter). Once the preferred vertical separation was determined, R_1 was held fixed in the remaining calculations.

Next, the angle of rotation (α) was altered by rotating the monomer edge in 30° increments in the right-hand sense. For dimers involving the amino acid edge, the rotation axis passes through the centers of mass of both monomers. For the nucleobase edge dimers, the rotational axis passes through the center of mass of the amino acid and the nucleobase atom (θ = number) or the dummy atom defining the midpoint of the line connecting the two relevant base atoms (θ = letter).

Finally, the horizontal 'edge' displacement (R_2) was considered. Due to the large number of calculations required to completely scan the monomer faces (81 calculations for HIS face and up to 225 calculations for 3MeA face), only the edges (up to 7) that lead to the strongest interactions after varying R_1 and α for each monomer pair were considered in R_2 scans. To perform the R_2 shift, the center of mass of the monomer face (in the XY plane) was defined as the origin (0,0). For all dimers, the Y-axis was defined to be parallel to the glycosidic (nucleobase face) or β-carbon and peptide backbone (amino acid face) bond, and this bond lies in quadrant III. The monomer edge was shifted by 0.5 Å along the X and Y axis, where single-point calculations were completed at each increment over the entire monomer face. Thus, despite the reduction in the number of edges considered for R_2 scans, dimers involving an amino acid edge still required approximately 225 (PHE edge) to 900 (TRP edge) calculations per nucleobase–amino acid pair, while dimers involving a nucleobase edge required approximately 425 (PHE face) to 700 (TRP face) calculations per pair.

2.3. Higher-Level ab Initio Methods and Extrapolation Techniques. To verify our computational approach, higher-level calculations were performed on all dimers yielding the strongest (most negative) interaction energies as identified from the MP2/6−31G*(0.25) potential energy surface scans. Since extrapolation techniques have been shown to accurately estimate the stacking interactions between the natural DNA nucleobases at the limit of large basis sets and high levels of correlation,[30] a similar approach was used in the present study to approximate CCSD(T)/CBS results. Specifically, the Helgaker basis set extrapolation technique was implemented,[49] which has been used previously for many different systems and has been specifically shown to work well for T-shaped interactions in general[38,39,45,47] as well as stacking interactions between the DNA nucleobases.[29e,f,30a,b,31c,32d] In our work, extrapolation from the aug-cc-pVDZ and aug-cc-pVTZ basis sets was used to estimate the MP2/CBS level. Previous results for hydrogen bonded and stacked DNA and RNA nucleobases showed that this extrapolation scheme was only improved by 2 kJ mol^{-1} when increased to the aug-cc-pVTZ and aug-cc-pVQZ extrapolation,[32d] and calculations at the aug-cc-pVQZ were not feasible given our current computer resources for the size of the complexes examined in the present study. A Δ(CCSD(T)–MP2) correlation correction factor was subsequently evaluated using the 6−31G*(0.25) basis set and added to the MP2/CBS binding strengths to yield estimated CCSD(T)/CBS results. Previous work supports this extrapolation approach for the correlation effects, where the 6−31G*(0.25) basis set was determined to yield a satisfactory Δ(CCSD(T)–MP2) correction for the natural DNA nucleobases due to the basis set insensitivity of this correction.[29e,f,30a,b,31c,32d]

All energy calculations include basis set superposition error (BSSE) corrections.[50] All MP2 calculations were performed using Gaussian 03,[51] while CCSD(T) calculations were completed using MOLPRO.[52]

3. Results and Discussion

3.1. T-Shaped Interactions. As mentioned in the Introduction, one of the driving forces of the present study is to address unanswered questions regarding the nature of T-shaped interactions between the aromatic amino acids and

natural versus damaged nucleobases. Our approach for scanning the potential energy surface as a function of four variables allows us to understand the dependence of the interaction energy on the relative monomer orientations, and therefore we will begin by discussing these dependencies in the following section. Subsequently, the geometry of the preferred T-shaped complexes for adenine and 3-methyladenine will be discussed and compared, and important structural characteristics that optimize T-shaped interaction energies will be summarized. Next, the dependence of T-shaped interaction energies on the cationic charge introduced upon nucleobase methylation will be highlighted, where results from our MP2/6−31G*(0.25) scans will be validated using CCSD(T) calculations extrapolated to the complete basis set limit.

3.1.1. Dependence of T-Shaped Interactions on Monomer Orientations. Since a large number of data points (up to 900) were considered in the potential energy surface scans for each monomer pair, we focus our discussion on major conclusions and use select examples to illustrate our findings. Data tables in the Supporting Information summarize the optimal MP2/6−31G*(0.25) T-shaped interactions after consideration of each variable (Figure 2b), and the overall strongest (most negative) interaction energies for each dimer after consideration of all variables are presented in Table 1. For nucleobase edge dimers, two interaction energies are reported in Table 1: (1) the overall strongest interaction energy, which involves the model N−H glycosidic bond for all dimers, and (2) the strongest interaction that does not involve the model glycosidic bond (denoted as 'enzymatic'). The latter energies are also reported since the hydrogen atom in the glycosidic bond is replaced with a sugar in DNA, and part of our goal is to understand the magnitude of these interactions in DNA repair enzymes. For amino acid edge dimers, only one interaction energy is reported in Table 1 since the strongest interaction does not involve the bond where the β-carbon and protein backbone are attached.

A direct comparison of the R_1 distances (Supporting Information) for dimers involving each amino acid or nucleobase edge (θ, Figure 3) is not meaningful since the definition of R_1 changes depending on the type of dimer examined. Instead, the most important conclusion from our R_1 scans is that the interaction energies are not largely dependent upon changes in the vertical separation, which was also previously reported for stacking interactions in the same nucleobase–amino acid systems.[19,24,25] Specifically, the interaction energy changes by less than 1.4 kJ mol^{-1} when R_1 deviates from the optimum interaction orientation by 0.1 Å.

Using the optimal vertical separations, Figure 4 illustrates the dependence of T-shaped interactions on α for HIS complexes as a representative example. For complexes involving a HIS edge, the dependence on α was found to range between 2.6 kJ mol^{-1} (θ=1) and 8.4 kJ mol^{-1} (θ=B) for (neutral) A complexes (Figure 4a) and between 2.7 kJ mol^{-1} (θ=1) and 20.0 kJ mol^{-1} (θ=C) for (cationic) 3MeA complexes (Figure 4b). Similarly, when an A edge interacts with the HIS face, the largest effect of α is 20 kJ mol^{-1} (θ=8, Figure 4c), while the largest effect for a cationic 3MeA

Table 1. Strongest MP2 and CCSD(T) Interaction Energies (kJ mol^{-1}) between Adenine or 3-Methyladenine and the Four Aromatic Amino Acids Calculated with a Variety of Basis Sets[a,b]

	adenine						3-methyladenine					
	MP2/6−31G*(0.25)	MP2/aug-cc-pVDZ	MP2/aug-cc-pVTZ	MP2/CBS	CCSD(T)/6−31G*(0.25)	CCSD(T)/CBS	MP2/6−31G*(0.25)	MP2/aug-cc-pVDZ	MP2/aug-cc-pVTZ	MP2/CBS	CCSD(T)/6−31G*(0.25)	CCSD(T)/CBS
HIS edge	−22.5	−23.8	−25.1	−25.7	−19.2	−22.4	−61.6	−63.2	−64.2	−64.6	−58.7	−61.6
PHE edge	−14.1	−17.1	−18.1	−18.5	−10.8	−15.2	−16.8	−20.0	−21.2	−21.7	−13.0	−17.8
TYR edge	−21.9	−22.9	−23.7	−24.1	−20.3	−22.4	−33.9	−35.2	−36.5	−37.0	−32.2	−35.2
TRP edge	−23.2	−25.3	−26.6	−27.2	−18.9	−22.9	−28.0	−31.4	−32.7	−33.1	−23.0	−28.2
HIS face[c]	−33.6 (−22.6)	−34.4 (−23.8)	−35.8 (−25.2)	−36.4 (−25.9)	−30.5 (−20.3)	−33.3 (−23.6)	−64.5 (−43.1)	−63.6 (−43.1)	−66.6 (−44.5)	−67.9 (−45.1)	−61.6 (−41.4)	−65.1(−43.5)
PHE face[c]	−25.6 (−16.0)	−27.5 (−17.8)	−30.3 (−19.3)	−31.5 (−20.0)	−20.5 (−13.5)	−26.4 (−17.4)	−46.4 (−37.2)	−47.1 (−36.0)	−49.3 (−38.2)	−50.3 (−39.2)	−41.7 (−33.3)	−45.6 (−35.4)
TYR face[c]	−27.8 (−18.4)	−29.8 (−20.5)	−31.9 (−21.8)	−32.8 (−22.4)	−23.3 (−16.0)	−28.3 (−20.0)	−54.1 (−39.2)	−54.6 (−40.8)	−57.2 (−42.9)	−58.3 (−43.8)	−48.7 (−36.2)	−52.9 (−40.8)
TRP face[c]	−34.8 (−23.1)	−37.5 (−25.3)	−39.8 (−26.6)	−40.8 (−27.1)	−28.8 (−20.0)	−34.8 (−24.0)	−72.0 (−52.6)	−72.8 (−54.6)	−75.8 (−57.0)	−77.2 (−58.1)	−64.8 (−48.4)	−70.1 (−53.9)
HIS stacked	−27.2[d]	−33.5	−35.6	−36.4	−18.5	−27.8	−52.7[e]	−55.2	−57.4	−58.3	−43.6	−49.3
HIS_t stacked	−29.7[d]	−35.7	−37.7	−38.5	−20.9	−29.8	−51.7[e]	−54.4	−56.6	−57.6	−42.6	−48.4
PHE stacked	−24.3[d]	−31.5	−33.7	−34.6	−13.0	−23.3	−48.6[e]	−52.8	−56.1	−57.5	−35.3	−44.2
TYR stacked	−30.7[d]	−38.2	−40.2	−41.1	−19.6	−30.0	−55.0[e]	−59.0	−62.3	−63.7	−42.2	−50.9
TYR_t stacked	−28.9[d]	−36.2	−38.4	−39.3	−18.0	−28.4	−53.9[e]	−57.8	−61.0	−62.4	−41.1	−49.6
TRP stacked	−35.0[d]	−42.4	−44.1	−44.8	−23.6	−33.5	−69.7[e]	−73.6	−76.9	−78.3	−55.7	−64.2
TRP_t stacked	−32.0[d]	−39.9	−41.7	−42.4	−20.2	−30.6	−71.5[e]	−75.5	−78.6	−79.9	−57.1	−65.5

[a] The strongest interactions for both T-shaped and stacked complexes were determined through MP2/6−31G*(0.25) potential energy surface scans (see Figure 2 for variables considered in scans). [b] Extrapolation to the complete basis set limit was completed by the Helgaker scheme (see Computational Methods). [c] The nucleobase edge interactions involving the model glycosidic bond with 'enzymatic' interactions not involving the model glycosidic bond, which are more relevant to biological processes involving nucleosides and nucleotides, provided in parentheses. [d] Reference 24. [e] Reference 25.

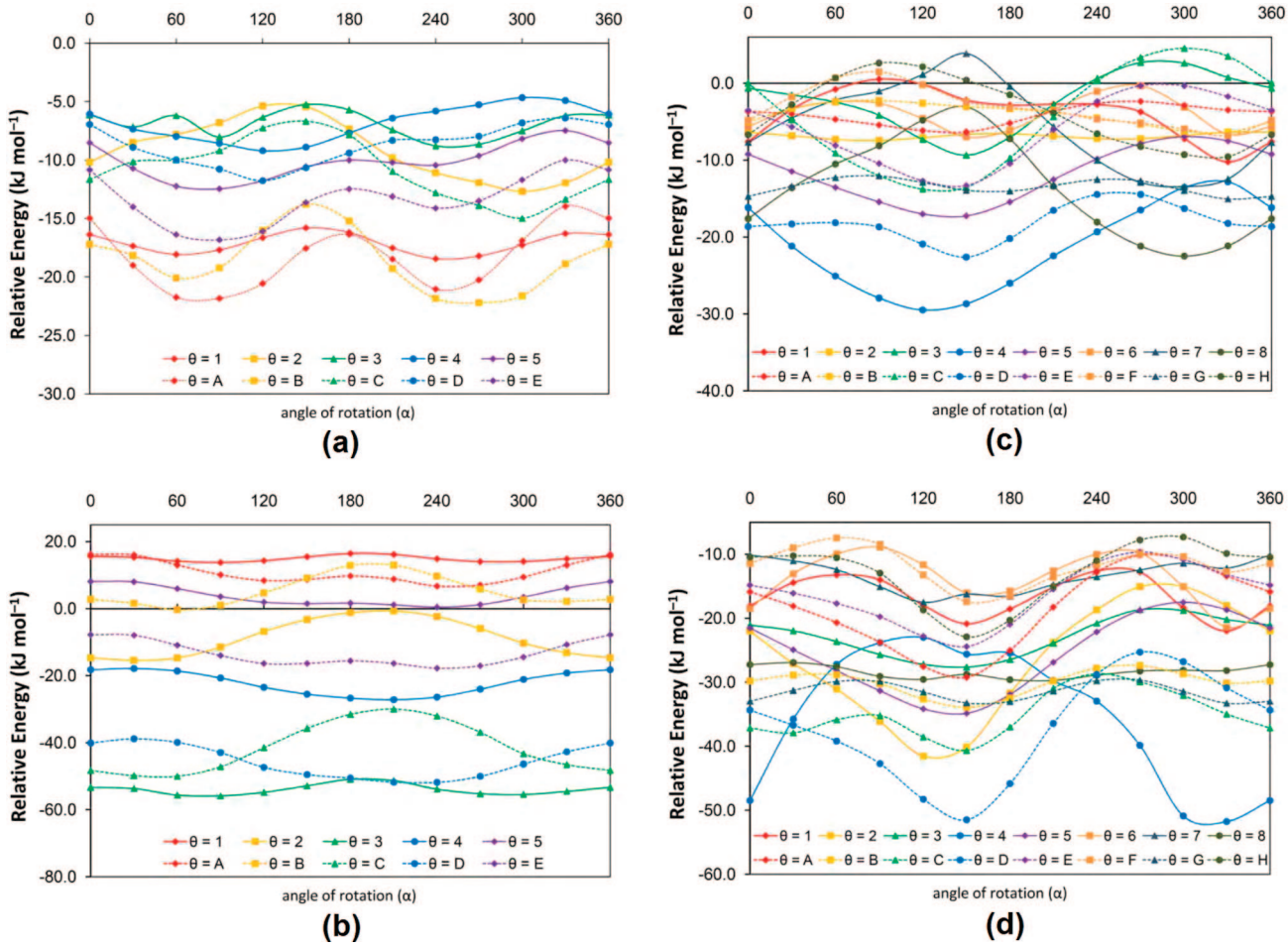

Figure 4. Interaction energy between HIS and (a) adenine (face), (b) 3-methyladenine (face), (c) adenine (edge), and (d) 3-methyladenine (edge) as a function of the angle of rotation (α) for different edges (θ) (see Figure 2b and Figure 3 for variable and edge definitions, respectively).

edge is 29 kJ mol^{-1} (θ=4, Figure 4d). Therefore, although the potential energy surface for rotation about α is shallow for many dimers, the interaction energy has a larger dependence on the angle of rotation (α) compared to the vertical separation.

For any given θ, the dependence on the angle of rotation (α) is due to the strength of secondary intramolecular interactions, where the strongest interaction arises when the acid−base interactions between monomer edge and monomer face are maximized. To illustrate this point, Figure 5 shows the orientations with the strongest interactions (after considering R_1 and α) for HIS complexes, where the secondary intramolecular interactions that govern the optimal α alignment are highlighted with dotted lines. In HIS edge complexes (Figure 5a,b), the preferred orientation aligns HIS protons toward N1 and N7 of the nucleobase, which are the sites with the largest proton affinity.[53] Alternatively, in nucleobase (edge) complexes, the best α aligns the electron-rich N atoms and/or the acidic N−H of HIS with strong nucleobase proton donors and/or acceptors, respectively. In all cases, the strongest interaction also occurs for the α that best directs the monomer edge across the entire π-system of the monomer face and thereby maximizes monomer overlap.

Although the interaction energies change with the angle of rotation (α) by up to 29 kJ mol^{-1}, the interaction energy has a larger dependence on θ. This can be seen in Figure 4 since the separation between lines (dependence on θ) is larger

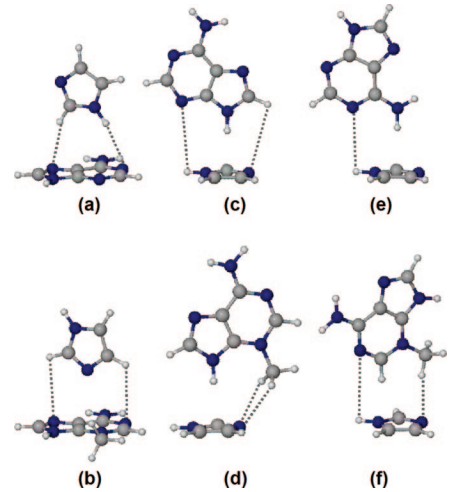

Figure 5. HIS T-shaped dimers with (a) adenine (face), (b) 3-methyladenine (face), (c) adenine (edge involving the model glycosidic bond), (d) 3-methyladenine (edge involving the model glycosidic bond), (e) adenine ('enzymatic' edge, not involving model glycosidic bond), and (f) 3-methyladenine ('enzymatic' edge, not involving model glycosidic bond) for optimal θ after considering R_1 and α.

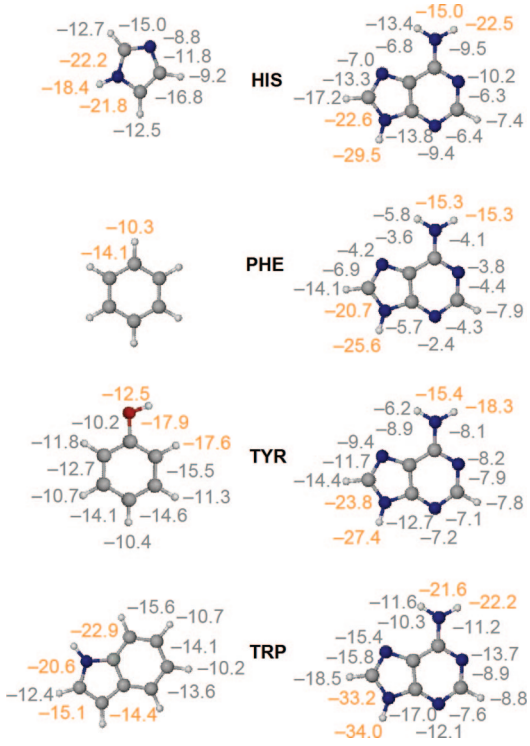

Figure 6. The strongest interactions for each θ in adenine (face)–amino acid (edge) (left) and adenine (edge)–amino acid (face) (right) dimers after considering R_1 and α. Edges highlighted in orange were considered for R_2 shifts.

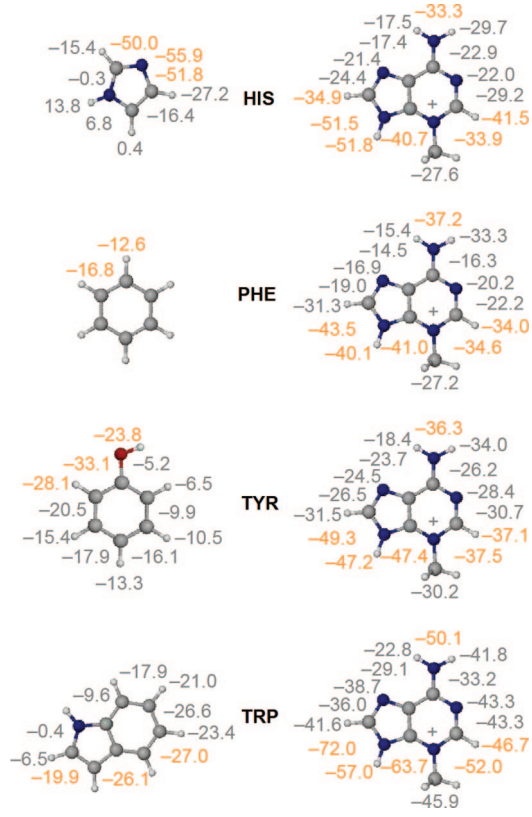

Figure 7. The strongest interactions for each θ in 3-methyladenine (face)–amino acid (edge) (left) and 3-methyladenine (edge)–amino acid (face) (right) dimers after considering R_1 and α. Edges highlighted in orange were considered for R_2 shifts.

than variations within a line (dependence on α). Figures 6 and 7 report the strongest (most negative) interaction energy (after considering R_1 and α) for each θ in adenine and 3-methyladenine complexes, respectively. The range in the interaction energy as a function of θ falls between 3.3 and 55.5 kJ mol^{-1} for each monomer. This large effect is partially due to variations in the properties of the monomer edges. The dependence of the interaction energy on θ will be discussed in greater detail in the following section.

Since we are in part searching the potential energy surface to determine the strongest interaction, the edges with the optimal interactions after consideration of R_1, α, and θ (highlighted in Figures 6 and 7 with orange) were subsequently examined by varying the R_2 horizontal displacement. We find that the horizontal displacement generally does not strengthen the T-shaped interactions by a considerable amount, and performing R_2 shifts does not change the preferred θ edge. These results justify our decision to consider R_2 effects on only select θ for each complex. More specifically, for the 64 complexes considered, most of the interaction energies increase (become more negative) by less than 1 kJ mol^{-1} and only 22 increase by 2−13 kJ mol^{-1}. Furthermore, the R_2 shifts that yield the strongest interaction energies are very small, where only 13 structures involve an R_2 shift greater than 1.0 Å in any direction.

Figure 8 shows how the potential energy surface changes with R_2 for HIS complexes. The lowest (most negative) energy region (yellow) is typically centered on or close to the center of mass (purple circles). The preference for small R_2 shifts and the resulting small strengthening in the T-shaped interaction energy arise since θ and α already optimize secondary intramolecular interactions that lead to the strongest T-shaped interactions in most dimers. In cases where the R_2 shift is larger (1.0 Å), the shift better optimizes these interactions. For example, in the HIS(edge):3MeA(face) complex (Figure 8b), the R_2 shift moves the HIS lone pair across the face of 3-methyladenine toward the atom with the largest positive charge (C4). Similarly, for both the A or 3MeA(edge):HIS(face) dimer (Figure 8c,d,f), the R_2 shift moves acidic nucleobase bonds toward the basic N atom of the HIS ring.

In summary, among the four geometrical variables considered, the T-shaped interactions are most dependent on the monomer edge (θ). Therefore, this key structural feature will be discussed in more detail in the following section.

3.1.2. Dependence of Optimal T-Shaped Structure on Nucleobase Methylation. The first important conclusion about the geometries of T-shaped complexes is that, regardless of the nucleobase methylation (charged) state, a stronger interaction is generally observed for dimers with the monomer edge bridging the π-system (θ = letter) compared to edges involving a single atom (θ = number) due to greater overlap. Although early studies on the benzene dimer have also identified bridged structures to be the most stable T-shaped orientations,[26a] the majority of recent studies on T-shaped complexes have only considered interactions that involve a particular atom directed at the aromatic π-system.[34,38,46] In cases where bridged structures involving small molecules directed toward an aromatic ring are considered,[45] only select

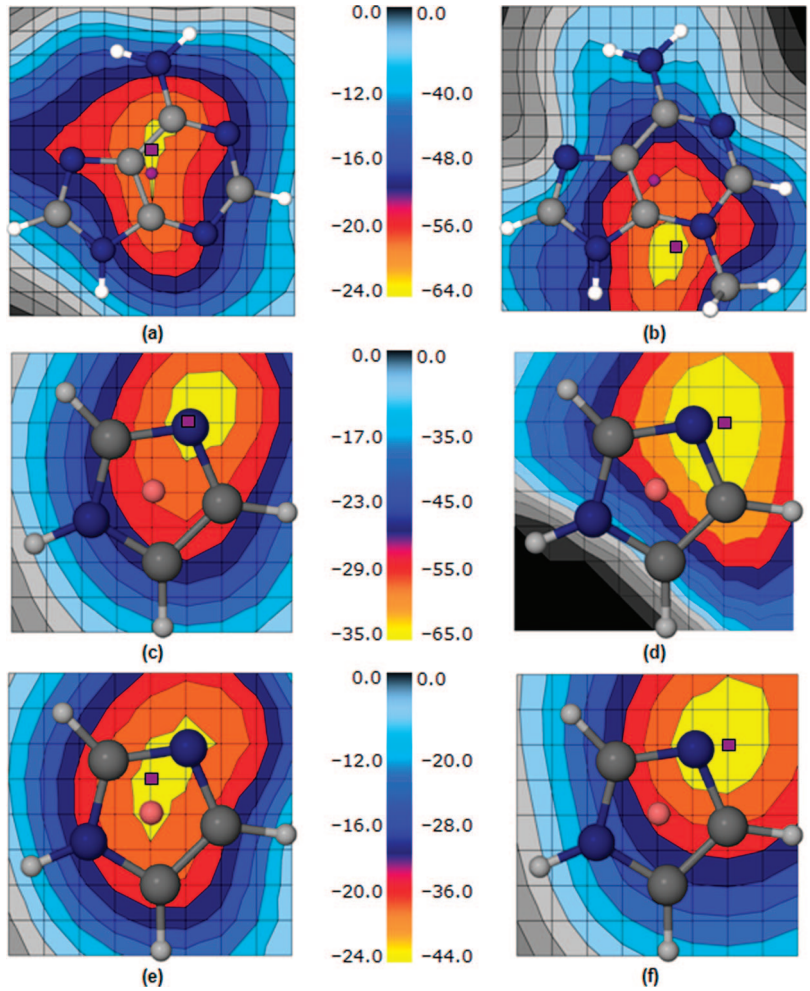

Figure 8. Interaction energy (kJ mol^{-1}) as a function of R_2 shift for optimal R_1, α and θ orientations of HIS T-shaped dimers with (a) adenine (face), (b) 3-methyladenine (face), (c) adenine (edge involving model glycosidic bond), (d) 3-methyladenine (edge involving model glycosidic bond), (e) adenine ('enzymatic' edge, not involving model glycosidic bond), and (f) 3-methyladenine ('enzymatic' edge, not involving model glycosidic bond). Purple circles indicate the origin (center of mass) and purple squares indicate the point with the strongest interaction.

combinations of molecules are examined, and no systematic investigation has been done. Furthermore, studies of T-shaped interactions involving amino acids have not considered bridged structures.[34] Our results clearly indicate that to identify global minima, and fully understand these interactions, the bridged structures must be examined.

The second important conclusion regarding the T-shaped structures is that the favored bridged orientation depends on both the properties of the monomer edge and the monomer face. For dimers with an amino acid edge interacting with the face of A, Figure 6 reveals that the most acidic (or positive) edge of the amino acid prefers to be directed toward the electron-rich face of adenine. For example, in the HIS edge complex, the optimal structure directs the acidic N−H bond and a C−H bond toward adenine (θ=B). This is consistent with T-shaped structures reported by Tsuzuki et al. for benzene–pyridine complexes.[47] Similarly, TYR bridges an acidic O−H bond and a C−H bond (θ=A) about the center of mass of adenine. This result is consistent with the preferred structure of the freely optimized phenol–benzene T-shaped dimer,[26m] which also validates our potential energy surface scans.

For dimers with an amino acid edge interacting with the face of 3MeA, we find that the most basic (or negative) edge of the amino acid prefers to be directed toward the cationic nucleobase. The best example is the HIS complex, where directing the HIS lone pair toward the cationic face (θ=3) results in the largest (most negative) interaction energy (−61.6 kJ mol^{-1}). Alternatively, when an acidic edge of HIS (θ=1) is directed toward 3MeA, the interaction is extremely repulsive (+13.8 kJ mol^{-1}). This is opposite to the trend discussed for adenine, where the θ=1 acidic edge leads to a stronger interaction (−18.4 kJ mol^{-1}) than the θ=3 basic edge (−8.8 kJ mol^{-1}). Indeed, the optimal orientation for HIS edge interacting with adenine (θ=B) leads to a very small interaction with 3-methyladenine (−0.3 kJ mol^{-1}). Similarly, for the TYR edge, the most stable complexes involve the hydroxyl lone pair directed toward 3MeA (θ=F) but the acidic hydroxyl hydrogen directed toward A (θ=A). These results show that the cationic charge of the damaged nucleobase dictates the relative orientation of the amino acid and base and therefore plays a large role in the nature of T-shaped interactions.

Due to the differences in charge between adenine and 3-methyladenine, it is not surprising that the interaction energies have a different dependence on the amino acid edge. Indeed, the strongest interactions for adenine face dimers increase as PHE < TYR < HIS < TRP, which is due to the relative dipole moments and size of the π-system of the various aromatic amino acids, while the strengths of dimers involving 3-methyladenine face increase as PHE < TRP < TYR < HIS, which is due to the increasing basicity of the amino acid. Since the favored monomer orientation in nucleobase face dimers depends on the relative acidity and basicity of the amino acid edge, a larger amino acid dipole moment causes a larger variation in the interaction energies as a function of θ,[54] which suggests that electrostatics play a very important role in these T-shaped contacts.

For complexes involving adenine or 3-methyladenine edges, the optimal interaction occurs with the most acidic edge of the nucleobase, the model glycosidic bond (θ=4 (adenine and 3MeA(edge):HIS(face) dimer) or θ=D (3-methyladenine)), directed toward the electron-rich amino acid face. Our observations for 3-methyladenine edge dimers are consistent with previous research on the benzene (face)–pyridinium cation (edge) interactions,[47] where the strongest T-shaped complex directs the N−H bond of pyridinium toward the π-system of benzene.[55] For the 'enzymatic' interactions, the strongest binding occurs between the amino group of A (θ=8) and the amino acid π-system. Although the amino group of 3MeA (θ=G) interacts with PHE in the most stable complex, two C−H$\cdots\pi$ interactions (θ=B) yield the strongest binding with the remaining amino acid faces. Due to the similarity in the structures of adenine and 3-methyladenine edges, the optimal interactions for amino acid face dimers increase with the dipole moments as well as the size of the π-system of the aromatic amino acids (PHE < TYR < HIS < TRP) in the 'strongest' and 'enzymatic' complexes for both nucleobases.

3.1.3. Dependence of T-Shaped Interaction Energies on Nucleobase Methylation. The discussion in the previous section reveals that the optimal T-shaped structures of dimers involving an amino acid edge are highly dependent upon the nature of the nucleobase. These differences in key structural features have large implications for the relative strengths of nucleobase–amino acid dimers. Indeed, adenine interactions range between −14 and −35 kJ mol^{-1}, while 3-methyladenine interactions are even larger, ranging between −17 and −72 kJ mol^{-1}.

To further examine the magnitude of the nucleobase–amino acid T-shaped interactions and validate our MP2/6−31G*(0.25) results, the T-shaped interactions for structures with the strongest (most negative) interactions as found in the MP2/6−31G*(0.25) potential energy surface scans were estimated at the CCSD(T)/CBS limit. Table 1 displays the CCSD(T)/CBS estimates as well as the binding strengths calculated at all levels of theory required to perform the extrapolation.

Previous studies on stacking interactions show that MP2 binding strengths increase with the basis set size,[26,27] and we see the same trend for T-shaped interactions. Thus, in comparison to 6−31G*(0.25), MP2/CBS leads to a 8−32% increase in the T-shaped interaction energies for adenine and 3−30% for 3-methyladenine. Nevertheless, MP2 is known to overestimate the stacking interaction energies of the natural nucleobases,[29e,f,30a,b,31c,32d] and MP2/6−31G*(0.25) overestimates the CCSD(T)/6−31G*(0.25) correlation energy of the T-shaped interactions examined in this study by 1−8 kJ mol^{-1}. However, CCSD(T)/CBS and MP2/6−31G*(0.25) results deviate by only 0−2 kJ mol^{-1} for all T-shaped dimers. Therefore, the MP2/6−31G*(0.25) calculations account for 91−105% of the CCSD(T)/CBS T-shaped interaction energies. This is an even better agreement between MP2/6−31G*(0.25) and CCSD(T)/CBS T-shaped energies than previously reported for the stacking interactions between the natural nucleobases, where only 80% of the correct stacking interaction is recovered at the MP2/6−31G*(0.25) level.[29] This justifies our choice of MP2/6−31G*(0.25) for the potential energy surface scans as a balance between cost and accuracy. Furthermore, our MP2/6−31G*(0.25) results are reliable for understanding the relative magnitude and importance of a wide range of nucleobase–amino acid interactions that do not necessarily correspond to global minima but may be imposed in biological systems.

For adenine complexes, the strongest CCSD(T)/CBS amino acid edge interactions range between −15 and −23 kJ mol^{-1}. When the edges of adenine are considered, the strongest T-shaped interactions range between −26 and −35 kJ mol^{-1}, while the 'enzymatic' interactions range between −17 and −24 kJ mol^{-1}. Most importantly, the magnitude of the T-shaped interactions between adenine and the amino acids are up to −35 kJ mol^{-1}, which is much larger than anticipated.

As mentioned for the MP2/6−31G*(0.25) results, methylation has a large effect on T-shaped interactions at the CCSD(T)/CBS level. Specifically, the strongest CCSD(T)/CBS amino acid edge interactions with 3-methyladenine range between −17 and −62 kJ mol^{-1}. Any given amino acid edge complex increases in strength by 17−176% upon methylation of adenine at N3. The largest methylation effect occurs for TYR and HIS complexes (57% and 176%, respectively) due to a strong interaction between an amino acid lone pair and the 3-methyladenine face. Indeed, the HIS lone pair–π(cation) interaction is the strongest T-shaped interaction involving HIS and is consistent with a previously published pyridinium (edge)–benzene (face) interaction (−61.7 kJ mol^{-1}), which was described as a hydrogen bond since the interaction energy is stronger than that of the stacked structure and the origin of the attraction is mainly electrostatic.[47] The magnitude of this HIS lone pair–π(cation) interaction is also consistent with the magnitude of similar interactions calculated by Egli et al. for various biological systems.[56]

The strongest (most negative) CCSD(T)/CBS T-shaped interactions involving a 3-methyladenine edge range between −45 and −70 kJ mol^{-1}, while the 'enzymatic' interactions range between −35 and −54 kJ mol^{-1}. Although there are not large changes in the preferred nucleobase orientation upon methylation as discussed for amino acid edge dimers, the cationic charge increases the acidity of the nucleobase bonds interacting with the amino acid face, which results in

stronger interaction energies. Indeed, the effect of methylation in these complexes corresponds to a 63−125% increase, which is larger than discussed for amino acid edge complexes. Thus, although the adenine interactions are strong, methylation has a large effect on the T-shaped interactions.

3.2. Comparison of the Magnitude of T-Shaped and Stacking Interactions between Adenine or 3-Methyladenine and the Aromatic Amino Acids. Due to the remarkable magnitude of T-shaped interactions between the amino acids and natural or damaged nucleobases, a direct comparison of these interactions to the previously calculated stacking interactions between the same molecules is necessary to understand their relative impact on biological processes. However, we must first extrapolate the previously reported MP2/6−31G*(0.25) stacking energies[24,25] to the CCSD(T)/CBS level of theory, where the extrapolated binding strengths and results from all intermediate calculations are provided in Table 1. We find that MP2/6−31G*(0.25) generally overestimates the stacking interaction energy by 0.5−6 kJ mol^{-1} (or up to 10%) when compared to CCSD(T)/CBS, where only 2 dimers decrease in strength by 0.1 and 0.5 kJ mol^{-1}. Therefore, MP2/6−31G*(0.25) is useful for scanning the potential energy surfaces of stacked complexes between the amino acids and natural or damaged nucleobases. However, the Δ(CCSD(T)−MP2) difference is larger for stacking (0−6 kJ mol^{-1}) than T-shaped (0−2 kJ mol^{-1}) interactions. Furthermore, CCSD(T)/CBS strengthens the T-shaped interactions (by up to 5%) and weakens the stacking (by up to 10%). These differences indicate that the strength of the stacking and T-shaped interactions are even more similar at the CCSD(T)/CBS level of theory compared with MP2/6−31G*(0.25), and therefore it is crucial to compare the T-shaped and stacking interactions of nucleobase–amino acid dimers at the CCSD(T)/CBS level.

CCSD(T)/CBS interaction energies for adenine show that the largest (most negative) T-shaped interactions are generally stronger than stacking interactions by 1.4 (TRP) − 3.5 (HIS) kJ mol^{-1}. In the case of TYR, the stacking interaction is only slightly larger (more negative by 1.7 kJ mol^{-1}) than the strongest T-shaped interaction. Furthermore, adenine 'enzymatic' interactions are only 21% (HIS) − 34% (TYR) weaker than stacking interactions. Similarly, adenine face T-shaped interactions are 25% (HIS) − 35% (PHE) smaller than the stacking interactions. These results emphasize that T-shaped interactions between the aromatic amino acids and the natural nucleobases are equivalent in magnitude or only slightly weaker than stacking interactions.

As discussed for adenine, the strongest 3-methyladenine T-shaped interactions are larger (more negative) than stacking interactions by 1.4 (PHE) − 5.0 (HIS) kJ mol^{-1}. 'Enzymatic' interactions involving a 3MeA edge are slightly weaker than stacking interactions, where the largest decrease is 20% (for PHE and TYR). Since 3-methyladenine amino acid edge interactions are weaker than the corresponding nucleobase edge interactions, they are also weaker than the corresponding stacking interactions (by up to 60% for PHE). However, the HIS edge interaction in the HIS:3MeA complex is 12 kJ mol^{-1} stronger than the corresponding stacking interaction. The implications and general importance of the relative magnitude of T-shaped and stacking interactions will be discussed in the following section.

3.3. Summary and Importance of Protein–DNA Noncovalent Interactions. By systematically examining the potential energy surfaces for protein–DNA noncovalent stacking and T-shaped interactions at very high levels of theory, we have found that although T-shaped interactions are often believed to be weak and relatively insignificant, these interactions can be very close in magnitude to π-stacking interactions between the same monomers. This statement is further justified by surveys of the protein data bank. For example, Rooman et al. identified several different HIS:A contacts in a range of X-ray crystal structures, and approximately 40% of these correspond to T-shaped arrangements.[34] Furthermore, although only two crystal structures of DNA repair enzymes that remove alkylated nucleobases are available with bound substrates, both crystal structures show a range of active site stacking and T-shaped interactions.[14,15]

Both stacking and T-shaped interactions are close in magnitude to biologically relevant hydrogen bonds. Indeed, lone pair–π interactions involving a positively charged nucleobase can exceed the strength of a single hydrogen bond, which has been noted previously for a range of biological systems.[56] For example, the adenine-thymine Watson-Crick hydrogen-bond strength, which involves at least two strong hydrogen bonds, is estimated to be −70 kJ mol^{-1} at the CCSD(T)/CBS level.[32d] Even T-shaped structures that do not correspond to the strongest interaction for a given amino acid–nucleobase dimer can be quite strong, and these structures may sometimes be more relevant to biological processes where dynamics or geometrical constraints prohibit optimal monomer orientations. Thus, our results indicate that the T-shaped interactions between nucleobases and aromatic amino acids can provide stability to many different enzymatic systems, including those involved in DNA transcription, replication, and repair, and these interactions cannot be ignored when studying biological processes.

In addition to gaining a better understanding of the magnitude of T-shaped interactions, our study reveals how the relative monomer orientation governs these interactions and thereby leads to extremely attractive forces. Specifically, we find that the strongest interactions between neutral monomers occur when the most acidic bond is directed toward the electron-rich π-system, while the strongest interactions for cationic monomers occur when a basic monomer edge is directed toward the positively charged π-system. These results, in conjunction with crystallographic data, can help clarify the interactions observed in the active sites of enzymes. Since we are ultimately interested in the role of noncovalent interactions in the DNA repair process facilitated by DNA glycosylases, we will illustrate this point by considering the active site of AAG.[15] Specifically, the protonation state of an active site HIS is not clear from crystallographic data, and our study suggests that the preferred HIS edge interaction will direct the lone pair of the basic N atom in HIS toward the cationic damaged nucleobase. Furthermore, our calculations indicate that this orientation is less favorable for the undamaged base and

therefore may help the enzyme differentiate between damaged and natural bases. Similarly, our calculations suggest that the hydroxyl group hydrogen of a TYR in the AAG active site is directed away from the face of the damaged nucleobase to maximize a lone pair–π interaction.[15] Thus, our calculations provide clues about how DNA repair enzymes can use T-shaped interactions to govern their activity by, for example, selectively removing cationic damaged nucleobases. These examples confirm the hypothesis of Egli et al. that lone pair–π interactions in particular may serve as a reporter of unusual protonation states of nucleobases[56] and also suggest that other T-shaped interactions involving cationic nucleobases can behave in a similar way.

Our calculations show that both stacking and T-shaped interactions are dependent on the amino acid. We previously reported that the stacking interactions of adenine and 3-methyladenine increase with the dipole moment of the amino acid, which induces an electrostatic interaction and suggests the importance of long-range dispersion interactions.[19,24,25] Similarly, this study reveals that T-shaped adenine and 3-methyladenine dimers also have strong electrostatic interactions. Therefore, PHE was found to have the weakest stacking and T-shaped interactions among all amino acids. Furthermore, whether a stacking or T-shaped interaction is favored depends on the amino acid. Therefore, we can speculate about the most beneficial interactions in, for example, the active sites of DNA glycosylases for substrate identification. Specifically, we can compare the magnitude of stacking, amino acid edge, and nucleobase 'enzymatic' edge interactions. Our calculations suggest that stacking interactions will be slightly more favorable in the case of PHE, TRP, and TYR, while T-shaped interactions with the HIS edge are stronger than the corresponding stacking interactions. Based on these global conclusions, it is interesting to note that close examination of the crystal structure of AlkA[14] with neutral hypoxanthine bound in the active site reveals that TRP and TYR are stacked with the substrate. Furthermore, the crystal structure of AAG[15] shows the substrate held in the active site by a T-shaped interaction with the HIS edge and stacking interactions with TYR, as well as T-shaped interactions with the TYR edge (which our calculations predict are only slightly weaker than the TYR stacking interactions).

One of the main goals of our study was to understand the effects of methylation on the stacking and T-shaped interactions. We find that stacking interactions increase by 60−115% upon methylation, while T-shaped interactions increase by 17−176% for amino acid edge dimers and 84−125% for nucleobase edge dimers. Indeed, since the differences between T-shaped and stacking interactions are generally small, our results suggest that alkylation has a greater effect on the interaction energies than the type of noncovalent interaction occurring between the substrate and active site residues or the amino acid involved. This general conclusion is especially true for nucleobase edge and stacking interactions. Although the strength of amino acid edge interactions is more dependent on the amino acid, the nucleobase face alkylation state still drastically affects the interaction energy. Thus, this paper shows that both the geometry and the magnitude of these noncovalent interactions are greatly affected by nucleobase alkylation. Therefore, our work suggests that alkylation must play a key role in dictating the active site interactions used by DNA repair enzymes for substrate recognition and binding and validates the range of amino acids found in the active sites of these enzymes.

4. Conclusions

T-shaped interactions between all aromatic amino acids and adenine or 3-methyladenine were systematically investigated to gain information about the geometric constraints that govern these interactions as well as their magnitude. To our knowledge, this work represents the most detailed study of T-shaped interactions between two different aromatic systems. Our study reveals that T-shaped interactions are highly dependent on the nature of the monomer edge, where the strongest T-shaped interaction generally occurs when the monomer edge bridges the π-system using two atoms rather than directing a single atom at the aromatic system. This result is crucial for future studies that wish to determine the global minimum for T-shaped monomer orientations, where the majority of past studies neglect these orientations. Furthermore, the favored monomer edge depends on the properties of the monomer face, where the most acidic edge is directed toward the (neutral) π-systems of adenine and the amino acids, while the most basic edge is directed toward (cationic) 3-methyladenine. Thus, the methylation state governs the preferred orientation of the amino acid and base, which provides clues about interactions within the active sites of enzymes that repair DNA alkylation damage and reveals ways to identify the alkylation (or protonation) state of nucleobases in DNA−protein systems.

The most significant result of the present work revolves around the magnitude of T-shaped interactions. After extrapolation to the highest level of theory possible for these systems, we find that T-shaped interactions involving adenine are up to -35 kJ mol^{-1}, which is comparable to the strength of stacking interactions between the corresponding monomers and also comparable to biologically relevant hydrogen bonds evaluated at the same high level of theory. Furthermore, T-shaped interactions involving 3-methyladenine are up to -70 kJ mol^{-1}, which is also similar to the stacking interactions between the corresponding monomers but much larger than the corresponding adenine interactions (up to 176% increase upon methylation).

To the best of our knowledge, this is the first study to compare noncovalent stacking and T-shaped interactions of biological systems at the CCSD(T)/CBS level with an extensive potential energy search, where our results emphasize the importance of comparing these interactions at the highest level of theory possible. Due to the magnitude of these interactions, our calculations suggest that it is crucial to examine T-shaped interactions to understand biological processes, where we have specifically applied our findings to better understand DNA repair enzymes. Future work must consider environmental effects on our findings as well as a greater range of natural and damaged nucleobases.

Acknowledgment. We thank the Natural Sciences and Engineering Research Council (NSERC), the Canada Research Chair program, and the Canada Foundation for Innovation (CFI) for financial support. We also thank the Upscale and Robust Abacus for Chemistry in Lethbridge (URACIL) for computer resources. L.R.R. thanks NSERC and the Alberta Ingenuity Fund (AIF) for student scholarships.

Supporting Information Available: MP2/6−31G*-(0.25) interaction energies as well as optimal R_1, α, and R_2 values for each θ in all adenine and 3-methyladenine complexes. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Hunter, C. A. *Angew. Chem., Int. Ed. Engl.* **1993**, *32*, 1584–1586.

(2) Lehn, J.-M. *Supramolecular Chemistry: Concepts and Perspectives;* VCH: New York, 1995; pp 89−195..

(3) Meyer, E. A.; Castellano, R. K.; Diederich, F. *Angew. Chem., Int. Ed.* **2003**, *42*, 1210–1250.

(4) Dziubek, K.; Podsiadło, M.; Katrusiak, A. *J. Am. Chem. Soc.* **2007**, *129*, 12620–12621.

(5) Nishio, M. *Cryst. Eng. Comm.* **2004**, *6*, 130–158.

(6) Brana, M. F.; Cacho, M.; Gradillas, A.; Pascual-Teresa, B.; Ramos, A. *Curr. Pharm. Des.* **2001**, *7*, 1745–1780.

(7) Chen, Y. C.; Wu, C. Y.; Lim, C. *Proteins: Struct., Funct., Bioinformat.* **2007**, *67*, 671–680.

(8) Seeburg, E.; Eide, L.; Bjoras, M. *Trends Biochem. Sci.* **1995**, *20*, 391–397.

(9) Wood, R. D.; Mitchell, M.; Sgouros, J.; Lindahl, T. *Science* **2001**, *291*, 1284–1289.

(10) Berti, P. J.; McCann, J. A. B. *Chem. Rev.* **2006**, *106*, 506–555.

(11) Stivers, J. T.; Drohat, A. C. *Arch. Biochem. Biophys.* **2001**, *396*, 1–9.

(12) Stivers, J. T.; Jiang, Y. L *Chem. Rev.* **2003**, *103*, 2729–2759.

(13) David, S. S.; Williams, S. D. *Chem. Rev.* **1998**, *98*, 1221–1261.

(14) Teale, M.; Symersky, J.; DeLucas, L. *Bioconjugate Chem.* **2002**, *13*, 403–407.

(15) Lau, A. Y.; Wyatt, M. D.; Glassner, B. J.; Samson, L. D.; Ellenberger, T. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 13573–13578.

(16) Labahn, J.; Scharer, O. D.; Long, A.; EzazNikpay, K.; Verdine, G. L.; Ellenberger, T. E. *Cell* **1996**, *86*, 321–329.

(17) Fujii, T.; Saito, T.; Nakasaka, T. *Chem. Pharm. Bull.* **1989**, *37*, 2601–2609.

(18) Fujii, T.; Itaya, T. *Heterocycles* **1988**, *48*, 1673–1724.

(19) Rutledge, L. R.; Campbell-Verduyn, L. S.; Hunter, K. C.; Wetmore, S. D. *J. Phys. Chem. B* **2006**, *110*, 19652–19663.

(20) Robertson, K. D.; Jones, P. A. *Carcinogenesis* **2000**, *21*, 461–467.

(21) Drabløs, F.; Feyzi, E.; Aas, P. A.; Vaagbø; C. B.; Kavli, B.; Bratlie, M. S.; Pena-Diaz, J.; Otterlei, M.; Slupphaug, G.; Krokan, H. E. *DNA Repair* **2004**, *3*, 1389–1407.

(22) Mishina, Y.; Duguid, E. M.; He, C. *Chem. Rev.* **2006**, *106*, 215–232.

(23) Wyatt, M. D.; Allan, J. M.; Lau, A. Y.; Ellenberger, T. E.; Samson, L. D. *Bioessays* **1999**, *21*, 668–676.

(24) Rutledge, L. R.; Campbell-Verduyn, L. S.; Wetmore, S. D. *Chem. Phys. Lett.* **2007**, *444*, 167–175.

(25) Rutledge, L. R.; Durst, H. F.; Wetmore, S. D. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2801–2812.

(26) (a) Hobza, P.; Selzle, H. L.; Schlag, E. W. *J. Am. Chem. Soc.* **1994**, *116*, 3500–3506. (b) Hobza, P.; Selzle, H. L.; Schlag, E. W. *J. Phys. Chem.* **1996**, *100*, 18790–18794. (c) Williams, V. E.; Lemieux, R. P.; Thatcher, G. R. J. *J. Org. Chem.* **1996**, *61*, 1927–1933. (d) Sinnokrot, M. O.; Valeev, E. F.; Sherrill, C. D. *J. Am. Chem. Soc.* **2002**, *124*, 10887–10893. (e) Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M.; Tanabe, K. *J. Am. Chem. Soc.* **2002**, *124*, 104–112. (f) Sinnokrot, M. O.; Sherrill, C. D. *J. Phys. Chem. A* **2003**, *107*, 8377–8379. (g) Sinnokrot, M. O.; Sherrill, C. D. *J. Am. Chem. Soc.* **2004**, *126*, 7690–7697. (h) Sinnokrot, M. O.; Sherrill, C. D. *J. Phys. Chem. A* **2004**, *108*, 10200–10207. (i) Riley, K. E.; Merz, K. M., Jr. *J. Phys. Chem. B* **2005**, *109*, 17752–17756. (j) Lee, E. C.; Hong, B. H.; Lee, J. Y.; Kim, J. C.; Kim, D.; Kim, Y.; Tarakeshwar, P.; Kim, K. S. *J. Am. Chem. Soc.* **2005**, *127*, 4530–4537. (k) Sinnokrot, M. O.; Sherrill, C. D. *J. Phys. Chem. A* **2006**, *110*, 10656–10668. (l) Waller, M. P.; Robertazzi, A.; Platts, J. A.; Hibbs, D. E.; Williams, P. A. *J. Comput. Chem.* **2006**, *27*, 491–504. (m) Lee, E. C.; Kim, D.; Jureèka, P.; Tarakeshwar, P.; Hobza, P.; Kim, K. S. *J. Phys. Chem. A* **2007**, *111*, 3446–3457. (n) Arnstein, S. A.; Sherrill, C. D. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2646–2655. (o) Quiñonero, D.; Frontera, A.; Escudero, D.; Ballester, P.; Costa, A.; Deyà, P. M. *Theor. Chem. Acc.* **2008**, *120*, 385–393.

(27) (a) Hobza, P.; Šponer, J. *Chem. Rev.* **1999**, *99*, 3247–3276. (b) Šponer, J.; Leszczynski, J.; Hobza, P. *J. Mol. Struct. (Theochem)* **2001**, *573*, 43–53. (c) Šponer, J.; Leszczynski, J.; Hobza, P. *Biopolymers (Nucleic Acid Sci.)* **2002**, *61*, 3–31. (d) Hobza, P. *Annu. Rep. Prog. Chem., Sec. C: Phys. Chem.* **2004**, *100*, 3–27.

(28) (a) Hobza, P.; Šponer, J.; Polasek, M. *J. Am. Chem. Soc.* **1995**, *117*, 792–798. (b) Šponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem. A* **1997**, *101*, 9489–9495. (c) Kratochvil, M.; Engkvist, O.; Šponer, J.; Jungwirth, P.; Hobza, P. *J. Phys. Chem. A* **1998**, *102*, 6921–6926.

(29) (a) Šponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem.* **1996**, *100*, 5590–5596. (b) Šponer, J.; Hobza, P. *Chem. Phys. Lett.* **1997**, *267*, 263–270. (c) Jureèka, P.; Nachtigall, P.; Hobza, P. *Phys. Chem. Chem. Phys.* **2001**, *3*, 4578–4582. (d) Leininger, M. L.; Nielsen, I. M. B.; Colvin, M. E.; Janssen, C. L. *J. Phys. Chem. A* **2002**, *106*, 3850–3854. (e) Jureèka, P.; Šponer, J.; Hobza, P. *J. Phys. Chem. B* **2004**, *108*, 5466–5471. (f) Šponer, J.; Jureèka, P.; Marchan, I.; Javier Luque, F.; Orozco, M.; Hobza, P. *Chem. Eur. J.* **2006**, *12*, 2854–2865.

(30) (a) Hobza, P.; Šponer, J. *J. Am. Chem. Soc.* **2002**, *124*, 11802–11808. (b) Jureèka, P.; Hobza, P. *J. Am. Chem. Soc.* **2003**, *125*, 15608–15613. (c) Šponer, J.; Jureèka, P.; Hobza, P. *J. Am. Chem. Soc.* **2004**, *126*, 10142–10151.

(31) (a) Šponer, J.; Florian, J.; Ng, H.-L.; Šponer, J. E.; Spadkova, N. A. *Nucleic Acids Res.* **2000**, *28*, 4893–4902. (b) Hill, G.; Forde, G.; Hill, N.; Lester, W. A.; Sokalski, W. A.; Leszczynski, J. *Chem. Phys. Lett.* **2003**, *381*, 729–732. (c) Dabkowska, I.; Gonzalez, H. V.; Jureèka, P.; Hobza, P. *J. Phys. Chem. A* **2005**, *109*, 1131–1136. (d) Cysewski, P.;

Czyznikowska-Balcerak, Z. *J. Mol. Struct. (Theochem)* **2005**, *757*, 29–36. (e) Matta, C. F.; Castillo, N.; Boyd, R. J. *J. Phys. Chem. B* **2006**, *110*, 563–578. (f) Kabelac, M.; Sherer, E. C.; Cramer, C. J.; Hobza, P. *Chem.-Eur. J.* **2007**, *13*, 2067–2077.

(32) (a) Mitchell, J. B.; Nandi, C. L.; McDonald, I. K.; Thornton, J. M.; Price, S. L. *J. Mol. Biol.* **1994**, *239*, 315–331. (b) Chelli, R.; Gervasio, F. L.; Procacci, P.; Schettino, V. *J. Am. Chem. Soc.* **2002**, *124*, 6133–6143. (c) Morozov, A. V.; Misura, K. M. S.; Tsemekhman, K.; Baker, D. *J. Phys. Chem. B* **2004**, *108*, 8489–8496. (d) Jureèka, P.; Šponer, J.; Cerny, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985–1993.

(33) (a) Rooman, M.; Lievin, J.; Buisine, E.; Wintjens, R. *J. Mol. Biol.* **2002**, *319*, 67–76. (b) Biot, C.; Buisine, E.; Kwasigroch, J. M.; Wintjens, R.; Rooman, M. *J. Biol. Chem.* **2002**, *277*, 40816–40822. (c) Biot, C.; Buisine, E.; Rooman, M. *J. Am. Chem. Soc.* **2003**, *125*, 13988–13994. (d) Wintjens, R.; Biot, C.; Rooman, M.; Lievin, J. *J. Phys. Chem. A* **2003**, *107*, 6249–6258. (e) Biot, C.; Wintjens, R.; Rooman, M. *J. Am. Chem. Soc.* **2004**, *126*, 6220–6221. (f) Marsili, S.; Chelli, R.; Schettino, V.; Procacci, P. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2673–2685.

(34) Cauët, E.; Rooman, M.; Wintjens, R.; Lievin, J.; Biot, C. *J. Chem. Theory Comput.* **2005**, *1*, 472–483.

(35) Cysewski, P. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2636–2645.

(36) Zhang, R. B.; Somers, K. R. F.; Kryachko, E. S.; Nguyen, M. T.; Zeegers-Huyskens, T.; Ceulemans, A. *J. Phys. Chem. A* **2005**, *109*, 8028–8034.

(37) Vaupel, S.; Brutschy, B.; Tarakeshwar, P.; Kim, K. S. *J. Am. Chem. Soc.* **2006**, *128*, 5416–5426.

(38) Shibasaki, K.; Fujii, A.; Mikami, N.; Tsuzuki, S. *J. Phys. Chem. A* **2006**, *110*, 4397–4404.

(39) Bendová, L.; Jureèka, P.; Hobza, P.; Vondrášek, J. *J. Phys. Chem. B* **2007**, *111*, 9975–9979.

(40) Mishra, B. K.; Sathyamurthy, N. *J. Phys. Chem. A* **2007**, *111*, 2139–2147.

(41) Cheney, B. V.; Schulz, M. W.; Cheney, J.; Richards, W. G. *J. Am. Chem. Soc.* **1988**, *110*, 4195–4198.

(42) Tsuzuki, S.; Honda, K.; Fujii, A.; Uchimaru, T.; Mikami, M. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2860–2865.

(43) Gervasio, F. L.; Chelli, R.; Marchi, M.; Procacci, P.; Schettino, V. *J. Phys. Chem. B* **2001**, *105*, 7835–7846.

(44) Scheiner, S.; Kar, T.; Pattanayak, J. *J. Am. Chem. Soc.* **2002**, *124*, 13257–13264.

(45) Ringer, A. L.; Figgs, M. S.; Sinnokrot, M. O.; Sherrill, C. D. *J. Phys. Chem. A* **2006**, *110*, 10822–10828.

(46) Gil, A.; Branchadell, V.; Bertran, J.; Oliva, A. *J. Phys. Chem. A* **2007**, *111*, 9372–9379.

(47) Tsuzuki, S.; Mikami, M.; Yamada, S. *J. Am. Chem. Soc.* **2007**, *129*, 8656–8662.

(48) We note that initial calculations were completed where the PHE edge was rotated about its center of mass every 5° to give 5 dimers between θ=1 and θ=A. Since maximum and minimum structures were found at θ=1 and A, respectively, only these two extremes were modeled for the remaining dimers.

(49) (a) Halkier, A.; Helgaker, T.; Jorgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. *Chem. Phys. Lett.* **1998**, *286*, 243–252. (b) Halkier, A.; Helgaker, T.; Jorgensen, P.; Klopper, W.; Olsen, J. *Chem. Phys. Lett.* **1999**, *302*, 437–446.

(50) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553–566.

(51) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria,G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03, Revision D.02*; Gaussian Inc.: Wallingford, CT, 2004.

(52) Werner, H. J.; Knowles, P. J.; Lindh, R.; Manby, F. R.; Schutz, M.; Celani, P.; Korona, T.; Rauhut, G.; Amos, R. D.; Bernhardsson, A.; Berning, A.; Cooper, D. L.; Deegan, M. J. O.; Dobbyn, A. J.; Eckert, F.; Hampel, C.; Hetzer, G.; Lloyd, A. W.; McNicholas, S. J.; Meyer, W.; Mura, M. E.; Nicklass, A.; Palmieri, P.; Pitzer, R.; Schumann, U.; Stoll, H.; Stone, A. J.; Tarroni, R.; Thorsteinsson, T. *MOLPRO, Version 2006.1*; University College Cardiff Consultants Ltd.: Cardiff, U.K., 2006.

(53) McConnell, T. L.; Wheaton, C. A.; Hunter, K. C.; Wetmore, S. D. *J. Phys. Chem. A* **2005**, *109*, 6351–6362.

(54) The MP2/6−31G(d) dipole moments of the amino acids decrease as HIS (3.949 D) > TRP (1.926 D) > TYR (1.480 D).

(55) We note that Tsuzuki et al. (ref 47) did not find a bridged structure similar to our θ=D edge despite free optimizations. However, their input geometries also did not consider a bridged structure. Our results suggest that geometry optimizations starting from the bridged structure would lead to a minimum on the potential energy surface that is lower in energy.

(56) Egli, M.; Sarkhel, S. *Acc. Chem. Res.* **2007**, *40*, 197–205.

CT8002332