JCTC Journal of Chemical Theory and Computation

# Evidence for Stabilization of DNA/RNA−Protein Complexes Arising from Nucleobase−Amino Acid Stacking and T-Shaped Interactions

Lesley R. Rutledge, Holly F. Durst, and Stacey D. Wetmore*

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

Received December 19, 2008

**Abstract:** The stacking and T-shaped interactions between the natural DNA or RNA nucleobases (adenine, cytosine, guanine, thymine, uracil) and all aromatic amino acids (histidine, phenylalanine, tyrosine, tryptophan) were investigated using ab initio quantum mechanical calculations. We characterized the potential energy surface of nucleobase−amino acid dimers using the MP2/6-31G*(0.25) method. The stabilization energies in dimers with the strongest interactions were further examined at the CCSD(T)/CBS level of theory. Results at the highest level of theory possible for these systems indicate that both stacking and T-shaped interactions are very close in magnitude to biologically relevant hydrogen bonds. Additionally, T-shaped interactions are as strong, if not stronger, than the corresponding stacking interactions. Our systematic consideration of the interaction energies in 485 possible combinations of monomers shows that a variety of these contacts are essential when considering the role of aromatic amino acids in the binding of proteins to DNA or RNA. This work also illustrates how our calculated binding strengths can be used by biochemists to estimate the magnitude of these noncovalent interactions in a variety of DNA/RNA−protein active sites.

## Introduction

Interactions between DNA/RNA and protein building blocks are essential for fundamental cellular processes such as transcription, DNA replication, DNA damage repair, apoptosis, and gene expression.[1,2] Thus, understanding how proteins recognize specific DNA sequences or certain nucleobases is crucial for understanding the role of these contacts in a variety of biological processes and medical applications. For example, gene expression is regulated by protein switches that bind to specific DNA sequences,[3,4] which has led to proposals that DNA−protein interactions can be used to target numerous genetic diseases through rational drug design.[3,5]

In attempts to understand how proteins recognize nucleobase sequences, structural analyses of a large number of DNA/RNA−protein complexes have been performed.[6−8] These studies reveal that it is not possible to establish a simple set of rules for predetermining interactions between DNA and protein building blocks.[9,10] Additionally, proteins often readily undergo structural changes to accommodate different nucleobases,[10−12] where each substrate exploits unique active site interactions to promote binding.[10] Because of the complexity, noncovalent interactions have been proposed to govern these biologically crucial contacts.[3,13]

There has been debate in the literature regarding the relevance of aromatic amino acids (Figure 1) in base recognition,[13,14] which may be attributed to a poor understanding of the nature of aromatic−aromatic interactions. Additionally, identifying important aromatic−aromatic contacts in experimental crystal structures can be more challenging than other intermolecular interactions, such as (conventional) hydrogen bonding, which has well-defined structural characteristics.[13] However, a recent study searched 141 DNA−protein and 61 RNA−protein complexes found in the Protein Data Bank (PDB) and revealed that aromatic amino acids appear in nucleic acid binding sites more frequently than expected.[15] Indeed, 532 and 242 aromatic−

* Corresponding author. Fax: (403) 329-2057. E-mail: stacey.wetmore@uleth.ca.

10.1021/ct800567q CCC: $40.75 
Published on Web 03/30/2009

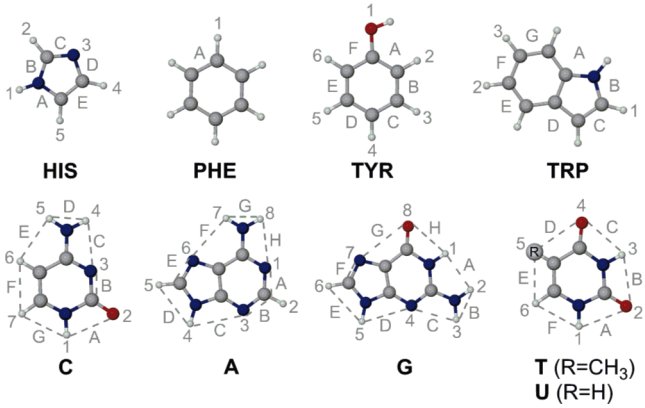


***Figure 1.*** Structures and definition of edges of the aromatic amino acids (histidine (HIS), phenylalanine (PHE), tyrosine (TYR), tryptophan (TRP)) and the natural nucleobases (cytosine (C), adenine (A), guanine (G), thymine (T), uracil (U)).

***Table 1.*** The Number of Aromatic−Aromatic Contacts between Amino Acids and DNA or RNA Bases Identified in the Protein Data Bank (PDB) by Baker and Grant[15]

| | HIS | | PHE | | TYR | | TRP | |
|---|---|---|---|---|---|---|---|---|
| | DNA | RNA | DNA | RNA | DNA | RNA | DNA | RNA |
| A | 41 | 12 | 48 | 14 | 28 | 15 | 8 | 7 |
| G | 65 | 35 | 28 | 9 | 30 | 18 | 6 | 14 |
| C | 41 | 22 | 15 | 16 | 36 | 24 | 11 | 19 |
| T | 66 | - | 60 | - | 41 | - | 8 | - |
| U | - | 12 | - | 6 | - | 16 | - | 3 |

aromatic nucleobase−amino acid contacts were identified in DNA and RNA containing structures, respectively (Table 1).[15] Since these complexes involve all four aromatic amino acids, it can be proposed that all aromatic side chains play a vital role in binding nucleic acid bases.

Although identifying a large number of nucleic acid−protein aromatic−aromatic interactions is the first step for demonstrating their importance in biological processes, very little is understood about these contacts. For example, key unanswered questions include the following: how does the relative nucleobase−side chain orientation affect the strength, what is the nature behind these attractive forces, how strong are the contacts, how can these interactions be used in nature, do the aromatic amino acids play a role in protein complex stability and/or DNA/RNA recognition, and is there a simple way to predict the strength of these interactions in DNA/RNA−protein complexes observed in experimental crystal structures?

Computational modeling of aromatic−aromatic interactions can provide valuable insights into the above questions. Unfortunately, few accurate ab initio calculations have been conducted on stacked[16−20] or T-shaped[16,21] orientations between nucleobases and aromatic side chains. Adenine has been investigated by several groups since it is a fundamental building block commonly used to study protein structure.[22] For example, the Rooman group characterized T-shaped and stacked orientations between adenine and phenylalanine or histidine (neutral and protonated) using crystal structure geometries, as well as identified optimal geometries using MP2/6-31G*(0.8,0.2) calculations.[16] The Hu group also

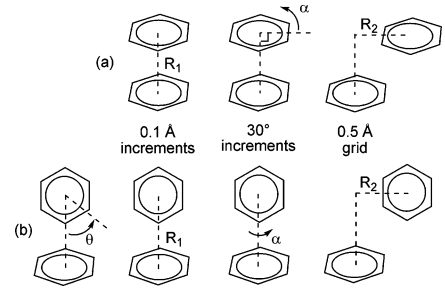


***Figure 2.*** The definition of the geometric variables considered and increments used in (a) stacking potential energy surface scans and (b) T-shaped potential energy surface scans (angle of 'edge' rotation ($\theta$), vertical separation ($R_1$), angle of rotation ($\alpha$), and horizontal displacement ($R_2$)). See Figure 1 for $\theta$ edges considered.

carried out large scale data mining of the PDB to identify crystal structures involving adenine stacked with the aromatic amino acids (phenylalanine, tyrosine, and tryptophan), where their MP2/6-311+G(d) calculations provide evidence for strongly attractive forces between these $\pi-\pi$ systems.[19] Recently, the Tschumper group investigated adenine−phenylalanine stacked dimers in experimental crystal structure orientations, as well as fully optimized geometries, where CCSD(T)/CBS interaction energies were reported.[20] Despite their potential importance, to our best knowledge, a full study of the heterodimers of the aromatic amino acids and nucleobases in both stacked and T-shaped geometries has not yet been performed.

In this work, we present the first systematic investigation of both stacked and T-shaped interactions between all aromatic amino acids (histidine, phenylalanine, tyrosine, tryptophan) and natural nucleobases (adenine, cytosine, guanine, thymine, uracil) at the highest levels of accuracy possible for these systems. Specifically, we use MP2/6-31G*(0.25) to scan the potential energy surface of the nucleobase−amino acid dimers. Subsequently, the dimer orientations that yield the strongest interactions were further examined at the CCSD(T)/CBS level. Our unique methodology for scanning the potential energy surface reveals more information than simply identifying the important minima on these dimer surfaces. For example, this study also reveals how geometrical changes in the relative orientations of stacked and T-shaped dimers dictate the strength and importance of aromatic−aromatic contacts. This methodology also reveals significant information about the magnitude of these biologically relevant contacts by providing the most accurate comparison in the literature to date of the stacking and T-shaped interactions between all aromatic amino acids and DNA and RNA nucleobases.

## Computational Methods

A series of MP2/6-31G*(0.25) single-point calculations were used to scan the potential energy surface for the stacked and T-shaped orientations of nucleobase−amino acid dimers according to several geometric variables (Figure 2), which is a similar approach to that used by Hobza and Šponer to study natural nucleobase dimers,[23−26] as well as our group to study dimers of the aromatic amino acids and natural or

damaged nucleobases.[18,21,27] Initially, $C_s$ symmetric, MP2/6-31G(d) optimized monomers were placed so that the nucleobase and amino acid molecular planes were parallel (stacked) or perpendicular (T-shaped) to one another. In stacked dimers involving HIS, TYR, and TRP, two relative orientations of the molecular planes were considered. The first was obtained by stacking the amino acid and nucleobase orientations as shown in Figure 1, while the second conformation was obtained by mirror flipping the amino acid relative to Figure 1 before stacking with the nucleobase (indicated using a prime (′) throughout the manuscript). These orientations yield a total of 35 stacked complexes. T-Shaped dimers involving either amino acid edges or nucleobase edges were considered. To define the ring edge (monomer edge) directed toward the $\pi$-system (monomer face), the geometric variable $\theta$ (Figure 2b) is introduced. Our edge nomenclature (Figure 1) uses a number to indicate an atom directed toward the monomer face, and a letter to indicate a bridged structure involving more than one atom directed toward the $\pi$-system. For example, in dimers involving PHE edges, $\theta = 1$ indicates that a hydrogen atom from the PHE ring is directed at the $\pi$-system, while $\theta$ = A indicates that a C−C bond of the PHE ring is parallel to the nucleobase molecular plane. This choice of edges led to 170 amino acid edge and 280 nucleobase edge dimers. A full description of initial structures is provided in the computational methods in the Supporting Information.

From these initial structures, three geometric variables were considered in our potential energy surface scans (Figure 2): the vertical separation ($R_1$), the angle of rotation ($\alpha$), and the horizontal displacement ($R_2$) between monomers. The increments used in our scans are indicated in Figure 2, and full details of our initial structures, as well as the ($R_2$) shift directions, are provided in the Supporting Information. We note that our $R_2$ calculations shift one monomer relative to the other monomer in two directions that are perpendicular to the $R_1$ axis. As a consequence of the large number of calculations involved in $R_2$ scans, we only considered the strongest 124 (out of 450) T-shaped dimers, while all 35 stacked dimers were further investigated.

The MP2 binding strengths for the orientations with the strongest (most negative) interactions as identified in potential energy surface scans were extrapolated to the CCSD(T) level at the complete basis set (CBS) limit. We considered all stacked dimers, the strongest amino acid edge dimers, the strongest nucleobase edge dimers that involve the model N−H glycosidic bond (relevant for nucleobases), and the strongest nucleobase edge dimers that do not involve the model glycosidic bond (relevant for nucleotides and nucleosides). MP2/aug-cc-pVDZ and MP2/aug-cc-pVTZ energies were used to extrapolate to the MP2 complete basis set limit (CBS) using the procedure of Helgaker.[28,29] To account for errors in MP2 electron correlation, the difference between the MP2/6-31G*(0.25) and CCSD(T)/6-31G*(0.25) energies ($\Delta$(CCSD(T)–MP2)) was added to the MP2/CBS results to give an estimated CCSD(T)/CBS interaction energy. Previous work indicates that the $\Delta$(CCSD(T)–MP2) correction is insensitive to basis set effects,[30] where the 6-31G*(0.25) basis set was previously found to yield satisfactory results for nucleobase or amino acid dimers.[31−35] We note that despite recent literature denoting these extrapolated values as CBS(T) results (see, for example, refs 33, 34, and 36), we will refer to these final interaction energies as CCSD(T)/CBS values to be consistent with our previous work,[21] as well as the majority of previous literature on stacking interactions (see, for example, refs 20, 35, and 37).

All reported interaction energies include basis set superposition error corrections.[38] MP2 and CCSD(T) calculations were performed with Gaussian 03[39] and MolPro,[40] respectively.

## Results and Discussion

**Dependence of Aromatic−Aromatic Interactions on Structural Characteristics.** In contrast to the optimizations typically performed on noncovalently bound dimers, our method for scanning the MP2/6-31G*(0.25) potential energy surface provides more information than simply the optimal stacked and T-shaped orientations. Specifically, our calculations also provide information on how adjusting the relative orientations of monomers in different ways affects the strength of each interaction. Along with our minima structures, this knowledge can be used to understand the relative strength of interactions between nucleobases and amino acid side chains in a variety of orientations. For example, our data can be used to understand contacts observed in nature as demonstrated in the final Results and Discussion section, using aromatic−aromatic DNA−protein interactions found in the PDB.

As a consequence of the large number of data points and dimers considered in this study, select examples will be discussed in this section to illustrate our major findings of the dependence of aromatic−aromatic interactions on structural characteristics. However, all MP2/6-31G*(0.25) binding strengths for all dimers considered are provided in the Supporting Information, which allows the reader to gain a more complete understanding of any given dimer.

For both stacked and T-shaped geometries, we find that the interaction energies are not highly dependent upon changes in the vertical separation between the monomers ($R_1$). Specifically, for all dimers examined in this study, the interaction energy changes by less than 1.3 kJ mol$^{-1}$ when $R_1$ deviates from the optimal vertical separation distance by 0.1 Å. Stacked dimers typically have a larger dependence on the vertical separation between monomers than T-shaped dimers, where this trend is generally independent of the monomers involved.

The interaction energy generally has a much larger dependence on the relative angle of rotation ($\alpha$) than on the vertical separation ($R_1$). Furthermore, the magnitude of this dependence is related to the type of interaction, as well as the monomers involved. In stacked dimers, $\alpha$ rotations affect the dipole−dipole interactions between monomers, where the dipole moments of the nucleobase and amino acid are antialigned in preferred structures. Therefore, the degree that the interaction energy is dependent on $\alpha$ is largely governed by the magnitude of the dipole moments of the monomers. For instance, since PHE (modeled as benzene) has no dipole moment, the stacking interaction energy has little to no

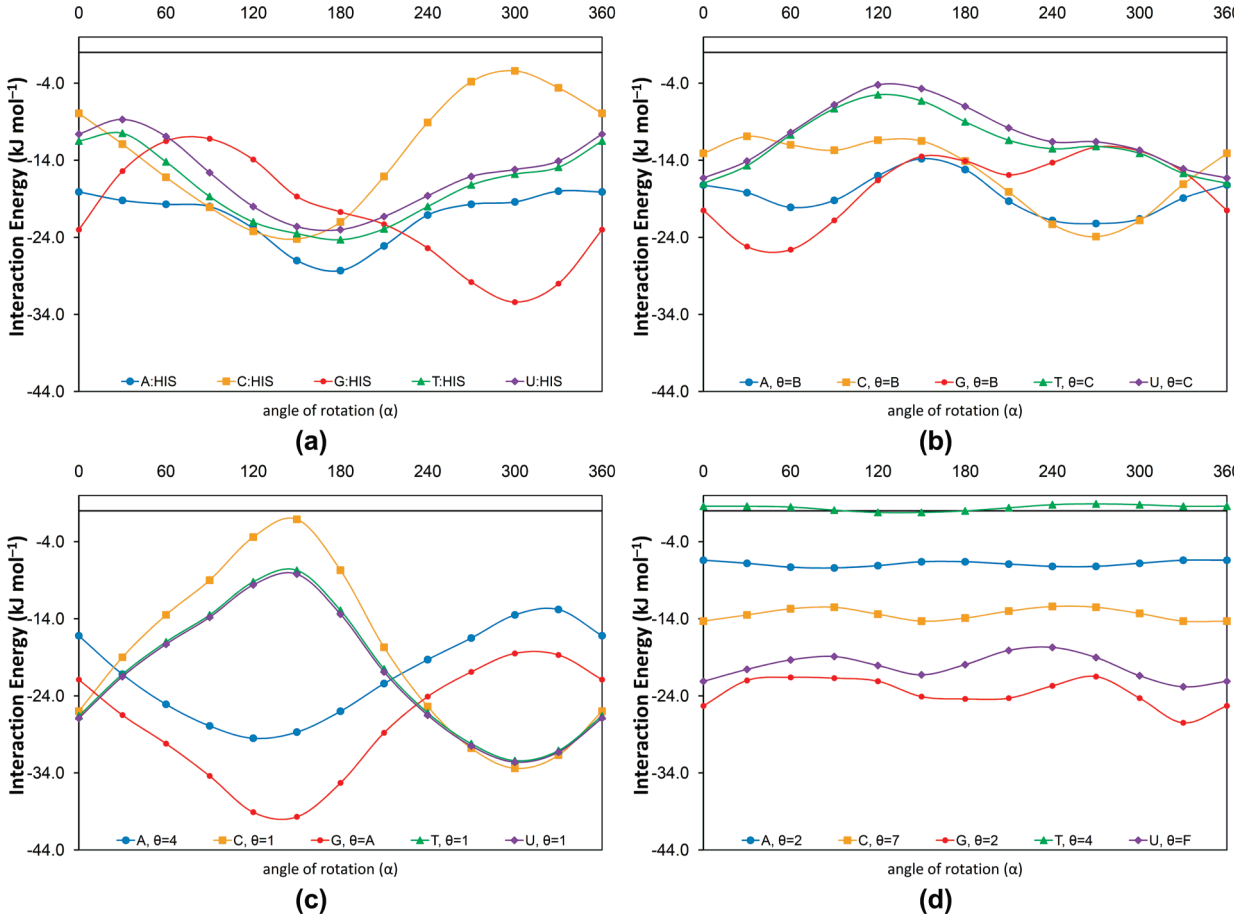


***Figure 3.*** Interaction energy as a function of α rotation for (a) HIS stacked dimers, (b) HIS (edge) T-shaped dimers with the optimal $\theta$, (c) HIS (face) T-shaped dimers with optimal $\theta$, and (d) HIS (face) T-shaped dimers for $\theta$ with the weakest α dependence. In all graphs, dimers with adenine are large blue circles, cytosine are orange squares, guanine are small red circles, thymine are green triangles, and uracil are purple diamonds. See Figure 4 for dimer geometries.

dependence on α. However, HIS has the largest dependence on the dipole−dipole alignment (Figures 3a and 4a) since it is the amino acid with the largest dipole moment. Indeed, the α dependency of HIS dimers ranges from 10.3 kJ mol$^{-1}$ for A (which has the smallest nucleobase dipole moment) to 21.8 kJ mol$^{-1}$ for C (which has the largest nucleobase dipole moment).

In T-shaped dimers, the geometric variable α affects how the monomer edge aligns with the monomer face, where the preferred orientation arises when acid/base interactions or secondary intermolecular hydrogen bonds are maximized. For example, in HIS edge interactions, acidic HIS protons align with basic nucleobase sites (see Figure 4b, where the preferred alignment of secondary intermolecular hydrogen bonds are indicated with dotted lines). Therefore, the relative strengths of the intermolecular hydrogen bonds govern the α dependence. Furthermore, since these secondary hydrogen-bonding interactions in amino acid edge dimers are weaker than the dipole−dipole interactions in stacking interactions, the α dependency is smaller in amino acid edge dimers. For example, the α dependence for optimal HIS edge dimers (Figure 3b) ranges from 5.1 (G) to 13.0 (C) kJ mol$^{-1}$.

Similar to amino acid edge dimers, α rotations in nucleobase edge dimers are affected by the strength of secondary intermolecular hydrogen bonds. For example, Figures 4c and 4d illustrate how nucleobase edges prefer to align basic or acidic edge atoms with the acidic N−H or basic N atom of the HIS face, respectively. When the intermolecular bonds are strong, as found for nucleobase edges with the strongest interactions (see, for example, HIS face dimers in Figure 4c), the α dependence is very large (up to 32.3 kJ mol$^{-1}$ for HIS dimers in Figure 3c). In fact, the α dependence for these nucleobase edge dimers is greater than that observed for stacked dimers. However, other dimers have much weaker intermolecular hydrogen bonds, and therefore a very small dependence on α. For example, the HIS face dimers in Figure 4d lead to a small (1.0 to 6.0 kJ mol$^{-1}$) dependence on α (Figure 3d).

Only small changes (less than 2.0 kJ mol$^{-1}$ for 98 of the 159 dimers considered) in the interaction energies are observed when one monomer is shifted in the molecular plane relative to the other ($R_2$ scans) from (0 Å, 0 Å) to the preferred $R_2$. Additionally, 104 of the 159 $R_2$ scans reveal that the monomers prefer to be shifted by 1.0 Å or less. These relatively small energetic and geometrical effects indicate that optimal interaction energies are generally observed when the centers of mass of the nucleobase and amino acid are aligned (stacked and amino acid edge interactions) or when the nucleobase edge interacts with the amino acid center of mass (see full computational methods in the Supporting Information). When $R_2$ shifts strengthen stacked dimers, the preferred shift allows for more $\pi-\pi$ overlap and/or decreases

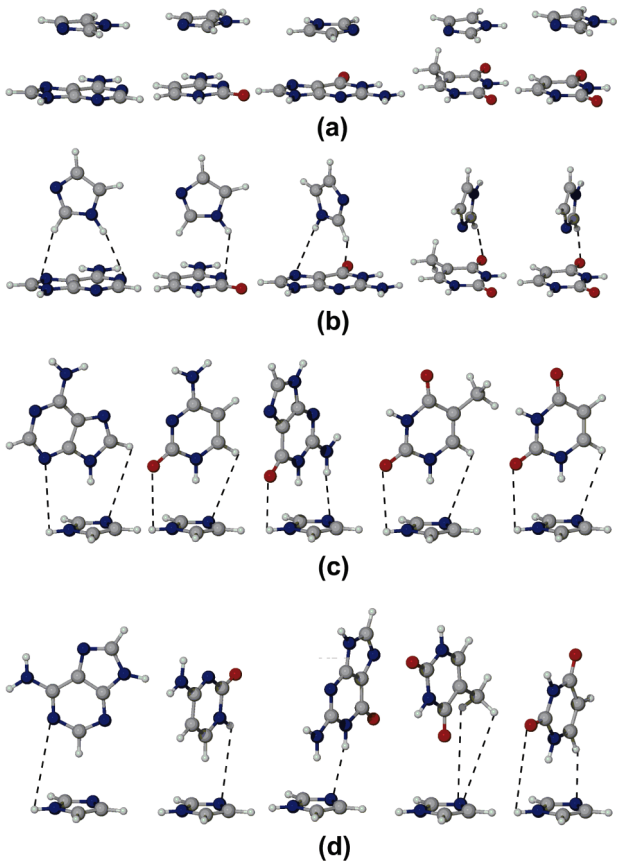


***Figure 4.*** Stacked and T-shaped $R_2$ = (0 Å, 0 Å) dimers for (left to right) adenine, cytosine, guanine, thymine, and uracil with (a) HIS, (b) HIS (edge) with optimal $\theta$, (c) HIS (face) with optimal $\theta$, and (d) HIS (face) for $\theta$ with the weakest α dependence.

repulsion between ring atoms. However, in T-shaped dimers, the preferred $R_2$ moves the acidic and/or basic regions in the monomer edge closer to the corresponding basic and/or acidic sites in the monomer face to maximize electrostatic interactions.

As previously mentioned, a large number of both amino acid and nucleobase edges ($\theta$) were considered for T-shaped dimers. All T-shaped interactions examined in this study have a very large dependence on the nature of the monomer edge ($\theta$) interacting with the $\pi$-system. This large dependence is due to variations in the acid−base properties of different monomer edges, where monomers with large dipole moments (and therefore a large variance in the acid−base properties of its edges) led to a larger dependence on $\theta$. The strongest interactions occur when the monomer edges with the largest (Lewis) acidity are directed toward the $\pi$-system. For example, in nucleobase edge dimers, the strongest interactions generally involve the most acidic model (N−H) glycosidic bond of the nucleobases (Figure 4c). However, in the case of guanine, the strongest interaction occurs when two strongly acidic N1−H and N2−H bonds[41] interact with the amino acid face ($\theta$ = A, Figure 1) rather than one strongly acidic bond[41] ($\theta$ = 5, Figure 1) or one strongly and one weakly acidic bond[41] ($\theta$ = E, Figure 1). Similarly, in amino acid edge dimers, the acidic N−H bond of TRP is included in the preferred edge ($\theta$ = A), while in HIS edge dimers with A, G, or C (Figure 4b), the optimal T-shaped interaction involves the acidic N−H bond ($\theta$ = B, Figure 4b). The only exception is the HIS edge dimer with T or U, where the basic N atom of HIS ($\theta$ = C) interacts with the nucleobase $\pi$-system in the optimal T-shaped dimer (Figure 4b). This exception occurs since T and U have more positive $\pi$-systems and therefore have a greater (Lewis) acidity, due to the two electron-withdrawing carbonyl groups. Recent electron affinity data supports this statement since T and U have the largest electron affinities among all DNA and RNA nucleobases.[42] However, it should also be mentioned that the acidic HIS edge preferred by A, C, and G ($\theta$ = B) binds by only 0.4 and 1.1 kJ mol$^{-1}$ weaker to T and U, respectively, than the basic ($\theta$ = C) edge.

Since our scans reveal how nucleobase−amino acid aromatic−aromatic contacts depend on several geometric variables, we can draw some general conclusions about the nature of these attractive interactions. First, although stacking interactions are known to be highly dependent on dispersion, it has been shown that additional special $\pi-\pi$ (dispersion) interactions are also significant for $\pi$-systems larger than those considered in the present study.[43] We find that our stacking interactions are also highly dependent on the electrostatic contribution. Specifically, our results indicate that the strongest interactions occur when the nucleobase and amino acid dipole moments are anti-aligned, and the interaction energies vary by up to 24 kJ mol$^{-1}$ when one monomer is rotated (by α, Figure 2) relative to the other. Additionally, we find that electrostatics play an even bigger role in T-shaped interactions compared with stacking. For example, the binding strengths of T-shaped dimers vary by up to 56 kJ mol$^{-1}$ by changing the edge ($\theta$, Figure 2) interacting with the $\pi$-system. Thus, our potential energy surface scans show that, although dispersion interactions are very important for the attractive nature of these interactions, electrostatics also play an essential role in determining the relative orientations of monomers.

**Magnitude of the Strongest Aromatic−Aromatic Interactions.** As previously mentioned, monomer orientations with the strongest interactions were further examined by extrapolating the corresponding interaction energies to the CCSD(T) level at the complete basis set (CBS) limit. All stacked dimers, the strongest amino acid edge dimers, the strongest nucleobase edge dimers that involve the model N−H glycosidic bond (relevant for nucleobases), and the strongest nucleobase edge dimers that do not involve the model glycosidic bond (relevant for nucleotides and nucleosides) were examined, where the MP2/6-31G*(0.25) and CCSD(T)/CBS interaction energies are reported in Table 2.

Our calculations reveal that MP2/6-31G*(0.25) interaction energies account for 80−110% of the CCSD(T)/CBS binding strengths in all nucleobase−amino acid dimers (Table 2). Interestingly, the Δ(CCSD(T)–MP2) correction is repulsive for all interaction energies examined, while the Δ(MP2/CBS–MP2/6-31G*(0.25)) correction is attractive. Therefore, our MP2/6-31G*(0.25) interaction energies are very close to the CCSD(T)/CBS results because of the cancelation of errors between basis set incompleteness and the missing electron correlation effects. Nevertheless, the remarkable agreement between these methods verifies that MP2/6-

***Table 2.*** MP2/6-31G*(0.25) and CCSD(T)/CBS Stacked and T-Shaped Interaction Energies (kJ mol$^{-1}$) between the Aromatic Amino Acids and the Natural DNA or RNA Nucleobases[a]

| | A[g] | | | C | | | G | | | T | | | U | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\theta^b$ | MP2[c] | CCSD(T)[d] | $\theta^b$ | MP2[c] | CCSD(T)[d] | $\theta^b$ | MP2[c] | CCSD(T)[d] | $\theta^b$ | MP2[c] | CCSD(T)[d] | $\theta^b$ | MP2[c] | CCSD(T)[d] |
| HIS stacked[f] | | −29.7 | −29.8 | | −26.9 | −29.5 | | −35.3 | −37.8 | | −25.0 | −27.1 | | −23.6 | −26.0 |
| HIS′ stacked[f] | | −27.2 | −27.8 | | −26.0 | −29.1 | | −31.4 | −32.7 | | −26.8 | −29.3 | | −26.4 | −29.1 |
| PHE stacked[f] | | −24.3 | −23.3 | | −18.4 | −20.6 | | −25.3 | −25.3 | | −22.4 | −24.5 | | −20.1 | −21.7 |
| TYR stacked[f] | | −28.9 | −28.4 | | −24.2 | −27.1 | | −32.8 | −35.2 | | −26.1 | −28.3 | | −25.2 | −27.6 |
| TYR′ stacked[f] | | −30.7 | −30.0 | | −22.7 | −25.0 | | −33.4 | −34.1 | | −25.5 | −27.5 | | −25.7 | −27.7 |
| TRP stacked[f] | | −32.0 | −30.6 | | −33.4 | −35.4 | | −42.4 | −42.1 | | −36.0 | −37.8 | | −34.5 | −36.8 |
| TRP′ stacked[f] | | −35.0 | −33.5 | | −32.9 | −35.4 | | −42.5 | −43.1 | | −36.4 | −38.5 | | −34.0 | −36.1 |
| HIS (edge) | B | −22.5 | −22.4 | B | −26.4 | −27.5 | B | −27.7 | −27.2 | C | −19.1 | −23.3 | C | −18.4 | −22.8 |
| PHE (edge) | A | −14.1 | −15.2 | A | −12.1 | −14.0 | A | −14.5 | −16.0 | A | −10.8 | −12.3 | A | −9.7 | −11.7 |
| TYR (edge) | A | −21.9 | −22.4 | A | −22.5 | −23.7 | A | −21.8 | −21.8 | A | −15.7 | −16.6 | 1 | −14.9 | −18.0 |
| TRP (edge) | A | −23.2 | −22.9 | A | −27.4 | −28.8 | A | −28.7 | −28.9 | A | −20.9 | −22.1 | A | −18.3 | −19.2 |
| HIS (face)[e] | 4 (8) | −33.6 (−22.6) | −33.3 (−23.6) | 1 (4) | −33.6 (−31.2) | −36.1 (−33.7) | 5 (A) | −26.7 (−46.4) | −26.7 (−48.4) | 1 (3) | −34.2 (−22.5) | −34.2 (−24.4) | 1 (3) | −34.7 (−23.0) | −36.9 (−24.9) |
| PHE (face)[e] | 4 (8) | −25.6 (−16.0) | −26.4 (−17.4) | 1 (5) | −22.5 (−19.5) | −25.4 (−21.8) | 5 (A) | −23.7 (−31.6) | −24.3 (−33.2) | 1 (3) | −24.7 (−18.9) | −22.2 (−21.6) | 1 (3) | −25.3 (−19.1) | −27.6 (−21.7) |
| TYR (face)[e] | 4 (8) | −27.8 (−18.4) | −28.3 (−20.0) | 1 (4) | −26.7 (−25.3) | −29.1 (−28.2) | E (A) | −24.6 (−36.2) | −24.9 (−37.8) | 1 (3) | −27.6 (−20.6) | −29.4 (−22.9) | 1 (3) | −27.8 (−20.8) | −29.7 (−23.1) |
| TRP (face)[e] | 4 (8) | −34.8 (−23.1) | −34.8 (−24.0) | 1 (4) | −34.5 (−32.2) | −36.6 (−33.9) | E (A) | −32.5 (−47.3) | −32.4 (−49.5) | F (3) | −37.0 (−25.4) | −37.9 (−27.2) | F (3) | −36.7 (−25.7) | −37.7 (−27.4) |

[a] See full computational methods provided in the Supporting Information. [b] See Figures 1 and 2 for definitions of $\theta$ and the edges considered in this study. [c] MP2/6-31G*(0.25) interaction energies. [d] CCSD(T)/CBS interaction energies. [e] Nucleobase edge interactions involving the model glycosidic bond and not involving the glycosidic bond (in parentheses). [f] Reference 18. [g] Reference 21.

31G*(0.25) very accurately describes the potential energy surfaces of nucleobase−amino acid dimers in both stacked and T-shaped orientations and suggests that we are correctly identifying the preferred orientations for these aromatic−aromatic contacts. This also suggests that the MP2/6-31G*(0.25) interaction energies reported in the Supporting Information for all relative monomer orientations are reliable and can be used to characterize interactions observed in nature (to be discussed in a subsequent section).

One of our most important findings is the magnitude of these aromatic−aromatic interactions. At the most accurate level of theory used to date for nucleobase−aromatic amino acid stacked and T-shaped dimers (CCSD(T)/CBS), we find that stacking interactions range between −20 and −43 kJ mol$^{-1}$, while the corresponding T-shaped interactions range between −12 and −50 kJ mol$^{-1}$ at optimal monomer orientations. Thus, both the stacking and T-shaped interactions approach the adenine–thymine Watson−Crick hydrogen-bond strength (−70 kJ mol$^{-1}$) calculated at the same level of theory,[35] which involves at least two strong hydrogen bonds. Additionally, we find that the binding strengths for stacking and T-shaped interactions are very similar in magnitude for all nucleobases, where the amino acid edge interactions are slightly weaker than the nucleobase edge interactions. Indeed, the most favorable nucleobase edge interactions are as strong, if not stronger, than the corresponding stacking interactions in the same dimer. Therefore, stacking arrangements are not the only orientations that are significantly attractive, but rather T-shaped interactions can also play stabilizing roles in nucleobase−amino acid complexes found in nature.

Comparison of the largest interaction energy for each dimer reveals interesting information about the trends in the binding energy with respect to the amino acid or nucleobase. Specifically, the strongest stacking interactions are found to increase as PHE ≪ TYR ≈ HIS < TRP, where similar stabilization energies are observed for HIS, TYR, and TRP. This trend is dictated by the relative dipole moments and size of the $\pi$-system of the amino acids. The same trend is also found for amino acid edge and nucleobase edge T-shaped interactions, which depend on the strength of secondary intermolecular hydrogen bonds between the acidic and basic sites in the amino acid and the nucleobase. However, when the trend with respect to the nucleobase is considered, no clear affinity is observed. The implications of these findings will be further discussed in the next section.

**The Role of DNA−Protein Aromatic−Aromatic Interactions in Nature.** A previous statistical survey of DNA−protein and RNA−protein interactions suggests that some aromatic amino acids could be involved in nucleic acid bases recognition because of a large number of observed contacts (Table 1).[15] For example, HIS forms the greatest number of contacts with G and T and therefore may selectively bind to these bases, while PHE forms the greatest number of contacts with T and A. Conversely, it was concluded that TYR provides strength to both DNA and RNA−protein complexes rather than having recognition capabilities since there is no clear affinity for one nucleobase.

Even though stacking and T-shaped interactions between the nucleobases and aromatic amino acids are strong, they can only play a role in DNA (RNA) recognition if there are large deviations in the binding energies with a change in the nucleobase. However, for a specific amino acid, we find that the differences between the binding strengths of all nucleobases are less than 10.5 kJ mol$^{-1}$. Additionally, in the nucleobase trend for a given amino acid, the difference in the binding strengths for any two sequential bases is generally very small (less than 3 kJ mol$^{-1}$ for 82% of the cases). Therefore, our results indicate that none of the aromatic amino acids preferentially bind to any of the nucleobases through stacking or T-shaped contacts. Furthermore, comparison of our calculated binding strengths (Table 2) to the number of observed contacts (Table 1) indicates that there is no significant correlation between the frequency of a particular nucleobase−amino acid contact and the magnitude of the interaction energy of the dimer.

We note that the lack of significant correlation between the frequency of a particular nucleobase−amino acid contact (Table 1)[15] and our calculated interaction energies may arise for many reasons. First, the original study of Baker and Grant[15] searched the PDB for DNA−protein structures involving only double stranded DNA, where many other relevant contacts may not fit this restriction. Second, the observed frequency of these interactions in the PDB may not reflect the natural occurrence of these contacts since crystal structures are not available for many DNA−protein complexes because of, for example, difficulties crystallizing native forms of the complexes. Third, classifying each aromatic−aromatic interaction as a specific type of contact is not easy and therefore determining their overall frequency and importance is difficult. Additionally, our results in Table 2 report the magnitude of these interactions in the preferred stacked or T-shaped orientations, where the geometries (and therefore magnitude) in PDB structures may be quite different. We also only considered dimers and interactions with additional residues may affect the magnitude of the binding interaction. Finally, since biomolecular interactions are so complex, the strongest interaction is not always the interaction that leads to optimal biological function. Therefore, there may not be a simple justification for expecting a correlation between the frequency of a particular nucleobase−amino acid contact and our strongest interaction energies (see, for example, comments in refs 36, 44, and 45).

All of the factors discussed above likely have a large effect on the correlation between the frequency and magnitude of DNA/RNA−protein interactions. Therefore, it is difficult to determine whether aromatic−aromatic interactions are selective enough to be used for recognition of specific DNA nucleobases. Nevertheless, our calculations at the highest level of theory possible for these systems demonstrate that aromatic−aromatic stacking and T-shaped interactions can be very stabilizing. Therefore, our calculations suggest that these interactions are most certainly important for providing stability to DNA/RNA−protein complexes.

**Additional T-Shaped Contacts That May Stabilize DNA/RNA−Protein Complexes.** As previously mentioned, our method for examining nucleobase−amino acid T-shaped interactions reveals more information than simply the identity of edges that lead to the strongest interactions. Specifically, we characterized the potential energy surfaces of all atom-directed and bridge-directed edges for these systems, where our full results (Supporting Information) show that many monomer edges lead to attractive interactions. For instance, we characterized a large number of C−H$\cdots\pi$ interactions, which have been extensively studied computationally,[46−52] and found that these nucleobase−amino acid interactions can be up to −26 kJ mol$^{-1}$. For example, in TRP edge dimers, the bridged edge involving two C−H$\cdots\pi$ interactions ($\theta$ = C) is only 0.8 (U) to 7.8 (A) kJ mol$^{-1}$ weaker than the strongest interactions observed with a bridged edge involving a N−H and C−H bond ($\theta$ = A). Additionally, in TRP face dimers, interactions involving only one C−H$\cdots\pi$ contact are up to −18.5 (A, $\theta$ = 5) or −18.9 (G, $\theta$ = 6) kJ mol$^{-1}$.

We also examined lone pair (lp)$\cdots\pi$ interactions for a variety of dimers, which have been studied computationally[50,53] and recently reviewed by Egli et al.[54] In nucleobase edge dimers, lp$\cdots\pi$ interactions involving carbonyl oxygens were found to be unstable (i.e., no minima was identified for U ($\theta$ = 2 and $\theta$ = 4) and C ($\theta$ = 2)) or very weak (i.e., less than −1 kJ mol$^{-1}$ for G ($\theta$ = 8) and T ($\theta$ = 2 and $\theta$ = 4)). However, some nitrogen lp$\cdots\pi$ interactions are as strong as −23 kJ mol$^{-1}$. For example, all HIS edge dimers involving the lone pair N ($\theta$ = 3) were found to be quite stable, where the interaction energies range from −8.2 to −15.2 kJ mol$^{-1}$. Furthermore, the interactions with N7 of G ($\theta$ = 7) can be up to −13.8 kJ mol$^{-1}$, while contacts with N3 of C ($\theta$ = 3) were found to be up to −23.0 kJ mol$^{-1}$. Therefore, our calculations suggest that even nonoptimal contacts that might arise in nature due to, for example, protein folding, can help stabilize DNA/RNA−protein complexes.

**Quantifying Experimentally Observed Nucleobase−Amino Acid Aromatic−Aromatic Interactions.** We acknowledge that the most favorable structures and interaction energies of nucleobase−amino acid dimers identified in our calculations may not appear in nature, where additional effects, such as protein folding or other (steric) constraints, come into play. However, by carefully searching the potential energy surface of stacked and T-shaped dimers, we have revealed how the interaction energies in these noncovalent complexes generally depend on different geometric variables. For example, because of the small dependence on the vertical separation between monomers ($R_1$), deviations from our reported optimal $R_1$ distances will have a small effect on the interaction energies. Conversely, large deviations from our reported preferred angle of rotation ($\alpha$) between stacked monomers will have significant effects on the interaction energy between monomers with large dipole moments. Alternatively, deviations from our optimal $\alpha$ for T-shaped dimers will affect the interaction energy in different ways, and therefore the properties of the edge involved must be carefully considered. Our calculations also suggest that changes in the optimal horizontal displacement ($R_2$) between monomers can weaken the (strongest) reported interaction energies, where the effect is dependent on the monomers involved, as well as the size and direction of the shift.

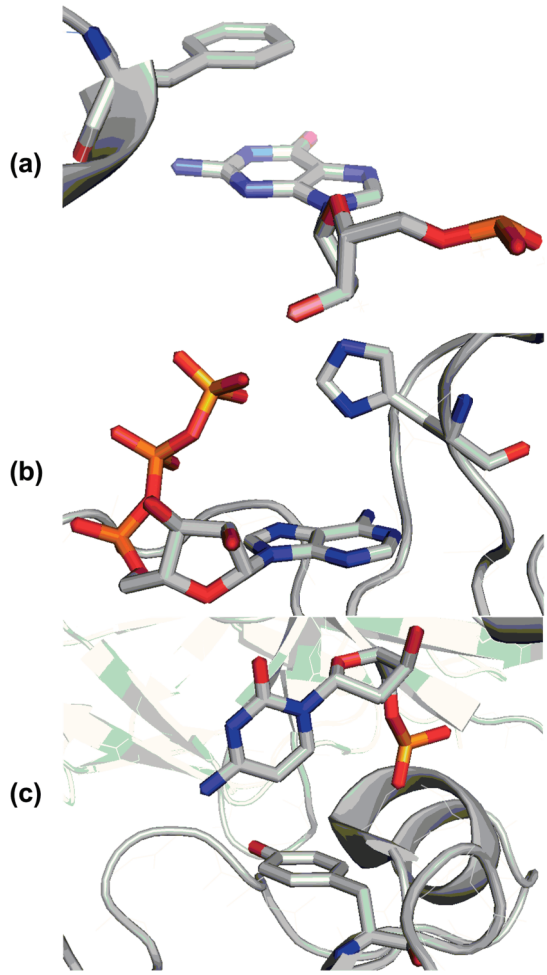


***Figure 5.*** Examples of nucleobase−amino acid interactions observed in nature (a) G:PHE stacked (PDB code: 1ckt with DG109 and F37 highlighted)[55] (b) A(face):HIS(edge) T-shaped (PDB code: 1b8a with ATP1500 and H1223 highlighted),[56] and (c) C(edge):TYR(face) T-shaped (PDB code: 1p7h with DC11 (chain B) and Y424 (chain M) highlighted).[57]

In addition to the general conclusions mentioned above, we can use the results from our detailed potential energy surface scans in a quantitative, predictive way. Specifically, in addition to the optimal interaction energies presented in Table 2, the interaction energies calculated for a range of geometric variables are reported in the Supporting Information for all nucleobase−amino acid dimers. These results can be compared to any stacked or T-shaped contact found in nature, not just those corresponding to minima on our calculated surfaces. This complete data set will allow biochemists to determine the relative magnitude and importance of different DNA/RNA−protein contacts.

To illustrate how our data can be used in conjunction with experimental structures to gain a better understanding of DNA−protein interactions that appear in nature, we will discuss three crystal structures containing representative examples of nucleobase−amino acid stacking (Figure 5a),[55] amino acid edge T-shaped (Figure 5b),[56] and nucleobase edge T-shaped (Figure 5c)[57] interactions. In this section, our Supporting Information will be used to estimate the interaction energy of each dimer, which will be compared to the interaction energy calculated using the exact crystal structure orientation (see crystal structure overlay section in the computational methods of the Supporting Information for details of how the interaction energies were calculated in crystal structure orientations).

**Stacking Interaction between Guanine and Phenylalanine.** The crystal structure of HMG1 domain A bound to a cisplatin-modified DNA duplex contains a stacked orientation between guanine (DG109) and phenylalanine (F37).[55] Table 3 reports relevant geometric parameters measured from the crystal structure that define this stacking orientation (column "Crystal"). These parameters can subsequently be compared to the closest corresponding point on our scanned potential energy surface (column "Scan"). For example, in this crystal structure, the orientation between G and PHE is most closely described by the point on our calculated potential energy surface with a vertical separation ($R_1$) of 3.5 Å, an angle of rotation ($\alpha$) of 30°, and a horizontal displacement ($R_2$) of (1.5 Å, 1.5 Å). Our calculations (Table SI-2, guanine–phenylalanine) reveal that an $R_2$ shift of (1.5 Å, 1.5 Å) has a corresponding interaction energy of −22.9 kJ mol$^{-1}$. However, this interaction was calculated for a ($R_1$) vertical separation distance of 3.4 Å and $\alpha = 0°$. Nevertheless, corrections based on our scans can be added to this interaction energy. Specifically, our $R_1$ scan reveals that increasing the vertical separation from 3.4 to 3.5 Å weakens the interaction energy by 0.6 kJ mol$^{-1}$. Additionally, an $\alpha$ rotation from 0 to 30° weakens the interaction energy by 0.1 kJ mol$^{-1}$. Therefore, the interaction between G and PHE in the crystal structure geometry of HMG1 domain A is estimated from our potential energy surface scans to be −22.2 kJ mol$^{-1}$ (i.e., (−22.9 + 0.6 + 0.1) kJ mol$^{-1}$). Interestingly, when optimized monomers are overlaid onto this crystal structure, the calculated interaction energy is also found to be −22.2 kJ mol$^{-1}$. This is in astounding agreement with our estimate from the calculated potential energy surfaces, especially since we have assumed additivity of the effects of different geometric variables.

**Amino Acid Edge Interaction between Histidine and Adenine.** In the crystal structure of aspartyl-tRNA synthetase, the edge of histidine (H1223) interacts with the face of adenine (ATP1500).[56] Since protein crystal structures are not able to determine the positions of hydrogen atoms, it is not clear whether the HIS N atom or N−H bond is directed toward A. Therefore, both HIS edge interactions ($\theta = 3$ and $\theta = 1$, Figure 2) are considered, where Table 3 outlines the geometric information obtained from the crystal structure and identifies the closest corresponding point on our potential energy surface scan for each histidine orientation ($R_1 = 4.5$ Å, $R_2 = (0.5$ Å, 2.5 Å), and $\alpha = 60°$ ($\theta = 3$) or 300° ($\theta = 1$), see Table SI-2, adenine–histidine).

When the nitrogen lone pair of HIS is directed toward A ($\theta = 3$), the interaction energy is calculated to be −6.2 kJ mol$^{-1}$ for the $\alpha$ value (60°) closest to that in the crystal structure, $R_1 = 4.4$ Å and $R_2 = (0$ Å, 0 Å). To determine the binding in the crystal structure orientation, our potential energy surface scans reveal that the binding strength is weakened by 8.1 kJ mol$^{-1}$ when HIS is shifted across the face of A from (0 Å, 0 Å) to (0.5 Å, 2.5 Å), and weakened

***Table 3.*** Structural Characteristics Measured for Three Representative Nucleobase−Amino Acid Interactions Observed in Nature (Crystal) and the Closest Corresponding Point from Our Potential Energy Surface Scans (Scan)

| | G:PHE stacked[51] | | A(face):HIS(edge)[52] | | C(edge):TYR(face)[53] | |
|---|---|---|---|---|---|---|
| | crystal | scan | crystal | scan | crystal | scan |
| angle (°) | 4.44[a] | 0 | 88.90[a] | 90 | 82.72[a] | 90 |
| vertical separation distance ($R_1$, Å) | 3.51[b] | 3.5 | 4.48[b] | 4.5 | 4.02[c] | 3.0[d] |
| angle of rotation (α, °) | 18.8[e] | 30 | 67.8[e] | 60 ($\theta = 3$), 300 ($\theta = 1$) | 66.3[e] | 60, 300 |
| horizontal shift distance ($R_2$, Å and (Å,Å)) | ∼2.2[f] | (1.5, 1.5) | ∼3.1[f] | (0.5, 2.5) | ∼2.7[f] | (2.5, 0) |

[a] Angle measured between molecular planes formed by aromatic ring atoms of nucleobase and amino acid. [b] Vertical separation measured between the molecular planes (stacked orientation), or the center of the ring (monomer edge) and monomer face molecular plane (T-shaped orientations). [c] Vertical separation measured between C5 of C and molecular plane of TYR. [d] The $R_1$ distance from our potential energy surface scans measured between the midpoint of the H4−C5 edge of C ($\theta$ = E) and the center of mass of TYR, which has the closest corresponding vertical separation measured between C5 of C and the center of the TYR ring. See Supporting Information for computational methods and definitions of $R_1$ used in this study. [e] The dihedral angle measured between the glycosidic bond and the peptide backbone bond of the amino acid (stacked orientation), the glycosidic bond and C1′ plane and the molecular plane of the amino acid (amino acid edge), or the peptide backbone bond plane of the amino acid and the molecular plane of the nucleobase (nucleobase edge). [f] The distance estimated between the centers of monomer rings in two dimensions.

by 0.1 kJ mol$^{-1}$ when the vertical separation is increased to 4.5 Å. These additive effects result in a weakly (+2 kJ mol$^{-1}$) repulsive interaction for the $\theta$ = 3 HIS(edge):A(face) interaction, while crystal structure overlay calculations determine this interaction to be only slightly more favorable (−2.1 kJ mol$^{-1}$).

When the N−H bond of HIS is directed toward A ($\theta$ = 1, Figure 1), our potential energy surface scan determines the interaction energy to be −17.3 kJ mol$^{-1}$ for the angle of rotation (α = 300°) and vertical separation distance ($R_1$ = 4.5 Å) observed in the crystal structure. Although this interaction energy was calculated for $R_2$ = (0 Å, 0 Å), accounting for the correct horizontal displacement yields an interaction energy of −9.2 kJ mol$^{-1}$. In comparison, the crystal structure overlay calculations determine this interaction to be −10.0 kJ mol$^{-1}$. This example again provides strong evidence that our calculated potential energy surfaces accurately describe the interaction energies in a range of nucleobase−amino acid dimers.

**Nucleobase Edge Interaction between Cytosine and Tyrosine.** The crystal structure of NFAT1 bound to the HIV-1 LTR kB element reveals the edge of C (DC11 (chain B)) interacting with the face of TYR (Y424 (chain M).[57] Because of an uncertainty in the orientation of the hydroxyl hydrogen of TYR, we consider both possible (in-plane) orientations, namely OH directed toward and away from C. The nucleobase edge interaction in the crystal structure with the hydroxyl group directed away from C ($\theta$ = E) best corresponds to α = 60°, which has a calculated interaction energy of −18.8 kJ mol$^{-1}$ for $R_1$ = 2.6 Å and $R_2$ = (0 Å, 0 Å) (Table SI-2, cytosine–tyrosine). When a (∼10 kJ mol$^{-1}$) correction for the horizontal shift and a (3.2 kJ mol$^{-1}$) correction for the vertical separation are taken into account, the interaction energy is predicted to be −5.6 kJ mol$^{-1}$, which is very consistent with the interaction energy calculated from a crystal structure overlay (−4.7 kJ mol$^{-1}$). Alternatively, when the hydroxyl group of TYR is directed toward C ($\theta$ = E, α = 300°), the interaction energy is estimated to be −3.5 kJ mol$^{-1}$ (i.e., −16.7 (α = 300°, $R_1$ = 2.6 Å, $R_2$ = (0 Å, 0 Å)) + 3.2 ($R_1$ correction) + 10.0 kJ mol$^{-1}$ ($R_2$ correction)), while our overlay calculation determines an interaction energy of −5.5 kJ mol$^{-1}$.Therefore, our predictions using the data set generated from our potential energy surface scans are very similar to those determined from crystal structure overlay calculations.

In summary, this section illustrates how our data reported in the Supporting Information can help predict the stabilizing and functional role of nucleobase−amino acid aromatic−aromatic interactions in a variety of biological structures. Our results can therefore be used in the future by biochemists in a similar fashion to accurately estimate the magnitude of these interactions in DNA/RNA−protein active sites identified through experimental structure determination. Thus, our data complements experimental structural data and helps reveal the relative importance of aromatic−aromatic contacts in a variety of DNA/RNA−protein systems.

## Conclusion

This is the first extensive study of both stacking and T-shaped interactions between all aromatic amino acids and natural nucleobases using our unique methodology for identifying important contacts on the potential energy surface. The variables used in this study allow us to understand how a variety of geometric criteria determine the strength of these contacts. The interaction energies were generally found to have a small dependence on the vertical separation between the monomers, while the dependence on the angle of rotation, as well as horizontal displacement, was found to be much greater. Additionally, stacked complexes prefer to have monomer dipole moments anti-aligned, while preferred T-shaped orientations direct the most (Lewis) acidic edge of one monomer toward the $\pi$-system of the second monomer.

Our ab initio calculations reveal that the strongest possible aromatic−aromatic interaction for each nucleobase−amino acid combination is very large. Indeed, at the highest level of theory possible for these systems, we find that both stacking and T-shaped interactions are very close in magnitude to biologically relevant hydrogen bonds. Additionally, T-shaped interactions are as strong, if not stronger, than the corresponding stacking interactions. Most importantly, we have not only fully characterized structures with the strongest interactions, but we have investigated a wide variety of other nucleobase−amino acid interactions, including C−H$\cdots\pi$, N−H$\cdots\pi$, and lp$\cdots\pi$ interactions, and our calculations reveal that these contacts are also very favorable.

On the basis of our calculations, we stress that most aromatic−aromatic nucleobase−amino acid interactions are large enough to play important roles in biological processes and therefore cannot be ignored. However, we do not find a correlation between the maximum strength of a specific interaction and its frequency in nature, nor do we find a large difference between the binding strengths with the natural bases for any given amino acid. Therefore, it is difficult to determine whether these interactions play a role in selective nucleobase recognition. Nevertheless, because of the large magnitude of their $\pi$-interactions, we suggest that the aromatic amino acids can take advantage of their $\pi$-clouds to add to the stability of DNA/RNA−protein complexes already provided by hydrogen-bonding interactions with these or other protein side chains.

Our high accuracy results (CCSD(T)/CBS) indicate that the MP2/6-31G*(0.25) binding strengths provided in the Supporting Information are very precise for all 485 dimers investigated in this study. Therefore, the MP2 results can provide reliable estimates of the size of interactions present in a variety of experimental DNA/RNA−protein crystal structures. Indeed, we have demonstrated how this additional, detailed data set can be used to estimate dimer interaction energies for contacts found in the Protein Data Bank, where our estimates are very close to interaction energies calculated using exact crystal structure orientations. Thus, this manuscript illustrates how our results can be used by biochemists to accurately approximate the magnitude and relative importance of $\pi$-interactions in DNA/RNA−protein actives sites identified in their experiments.

Although our results correspond to contacts within binding sites of low polarity,[58] solvation effects will be important to consider when extending our work to highly polar environments, such as those found in common thermodynamic experiments, where electrostatic forces between monomers are diminished.[59] Therefore, future work will consider environmental effects on the magnitude of these interactions, which will also extend the applicability of our results to all DNA/RNA−protein binding sites. Future work will also analyze the strength of specific contacts found in a greater number of crystal structures, as well as the effects of more than one simultaneous interaction on the noncovalent contacts between the nucleobases and aromatic amino acids.

**Acknowledgment.** We thank the Natural Sciences and Engineering Research Council of Canada (NSERC), the Canadian Research Chair Program, the Canadian Foundation for Innovation (CFI), and the University of Lethbridge Research Fund for financial support. L.R.R acknowledges NSERC, Alberta Ingenuity Fund (AIF), the Alberta Government, and the University of Lethbridge for student scholarships.

**Supporting Information Available:** Full computational methods, MP2/6-31G*(0.25) interaction energies for all points considered in potential energy surface scans, optimal $R_1$, $\alpha$, and $R_2$ values for each $\theta$ examined, and all high-level interaction energies required to estimate CCSD(T)/CBS binding strengths. Coordinates for all dimers in Table 2 are available from the authors upon request. This material is available free of charge via the Internet at http://pubs.acs.org.

## References

(1) Gromiha, M. M.; Siebers, J. G.; Selvaraj, S.; Kono, H.; Sarai, A. *Gene* **2005**, *364*, 108–113.

(2) Chan, L. L.; Pineda, M.; Heeres, J. T.; Hergenrother, P. J.; Cunningham, B. T. *ACS Chem. Biol.* **2008**, *3*, 437–448.

(3) Höglund, A.; Kohlbacher, O. *Proteome Sci.* **2004**, *2*, 3.

(4) Ptashne, M. *Nature (London)* **1967**, *214*, 232–234.

(5) Bartsevich, V. V.; Miller, J. C.; Case, C. C.; Pabo, C. O. *Stem Cells* **2003**, *21*, 632–637.

(6) Luscombe, N. M.; Thornton, J. M. *J. Mol. Biol.* **2002**, *320*, 991–1009.

(7) Nadassy, K.; Wodak, S. J.; Janin, J. *Biochemistry* **1999**, *38*, 1999–2017.

(8) Pailard, G.; Lavery, R. *Structure* **2004**, *12*, 113–122.

(9) Pabo, C. O.; Nekludova, L. *J. Mol. Biol.* **2000**, *301*, 597–624.

(10) Matthews, B. W. *Nature (London)* **1988**, *335*, 294–295.

(11) Hogan, M. E.; Austin, R. H. *Nature (London)* **1987**, *329*, 263–266.

(12) Olson, W. K.; Gorin, A. A.; Lu, X. J.; Hock, L. M.; Zhurkin, V. B. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 11163–11168.

(13) Luscome, N. M.; Laskowski, R. A.; Thornton, J. M. *Nucleic Acids Res.* **2001**, *29*, 2860–2874.

(14) Lejeune, D.; Delsaux, N.; Charloteaux, B.; Thomas, A.; Brasser, R. *Proteins: Struct. Funct. Bioinf.* **2005**, *61*, 258–271.

(15) Baker, C. M.; Grant, G. H. *Biopolymers* **2007**, *85*, 456–470.

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

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

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

(19) Mao, L.; Wang, Y.; Liu, Y.; Hu, X. *J. Mol. Biol.* **2004**, *336*, 787–807.

(20) Copeland, K. L.; Anderson, J. A.; Farley, A. R.; Cox, J. R.; Tschumper, G. S. *J. Phys. Chem. B* **2008**, *112*, 14291–14295.

(21) Rutledge, L. R.; Wetmore, S. D. *J. Chem. Theory Comput.* **2008**, *4*, 1768–1780.

(22) Biot, C.; Buisine, E.; Rooman, M. *J. Am. Chem. Soc.* **2003**, *125*, 13988–13994.

(23) Hobza, P.; Šponer, J. *Chem. Rev.* **1999**, *99*, 3247–3276.

(24) Šponer, J.; Leszczynski, J.; Hobza, P. *J. Mol. Struct. (THEOCHEM)* **2001**, *573*, 43–53.

(25) Šponer, J.; Leszczynski, J.; Hobza, P. *Biopolymers* **2002**, *61*, 3–31.

(26) Hobza, P. *Annu. Rep. Prog. Chem., Sect. C: Phys. Chem.* **2004**, *100*, 3–27.

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

(28) Halkier, A.; Helgaker, T.; Jorgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. *Chem. Phys. Lett.* **1998**, *286*, 243–252.

(29) Halkier, A.; Helgaker, T.; Jorgensen, P.; Klopper, W.; Olsen, J. *Chem. Phys. Lett.* **1999**, *302*, 437–446.

(30) Sinnokrot, M. O.; Sherrill, C. D. *J. Phys. Chem. A* **2004**, *108*, 10200–10207.

(31) Hobza, P.; Šponer, J. *J. Am. Chem. Soc.* **2002**, *124*, 11802–11808.

(32) Jurečka, P.; Hobza, P. *J. Am. Chem. Soc.* **2003**, *125*, 15608–15613.

(33) Jurečka, P.; Šponer, J.; Hobza, P. *J. Phys. Chem. B* **2004**, *108*, 5466–5471.

(34) Šponer, J.; Jurečka, P.; Marchan, I.; Javier Luque, F.; Orozco, M.; Hobza, P. *Chem.—Eur. J.* **2006**, *12*, 2854–2865.

(35) Jurečka, P.; Šponer, J.; Černý, J; Hobza, P. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985–1993.

(36) Šponer, J.; Riley, K. E.; Hobza, P. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2595–2610.

(37) Pitoňák, M.; Neogrády, P.; Řezáč, J.; Jurečka, P.; Urban, M.; Hobza, P. *J. Chem. Theory Comput.* **2008**, *4*, 1829–1834.

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

(39) 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.

(40) 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, UK, 2006.

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

(42) Roca-Sanjuán, D.; Merchán, M.; Serrano-Andrés, L.; Rubio, M. *J. Chem. Phys.* **2008**, *129*, 095104.

(43) Grimme, S. *Angew. Chem., Int. Ed.* **2008**, *47*, 3430–3434.

(44) Kopitz, H.; Živković, A.; Engels, J. W.; Gohlke, H. *ChemBioChem* **2008**, *9*, 2619–2622.

(45) Luo, R.; Gilson, H. S. R.; Potter, M. J.; Gilson, M. K. *Biophys. J.* **2001**, *80*, 140–148.

(46) Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M.; Tanabe, K. *J. Am. Chem. Soc.* **2000**, *122*, 3746–3753.

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

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

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

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

(51) Tsuzuki, S.; Fujii, A. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2584–2594.

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

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

(54) Egli, M.; Sarkhel, S. *Acc. Chem. Res.* **2007**, *40*, 195–205.

(55) Ohndorf, U. M.; Rould, M. A.; He, Q.; Pabo, C. O.; Lippard, S. J. *Nature (London)* **1999**, *399*, 708–712. PDB Code: 1ckt.

(56) Schmitt, E.; Moulinier, L.; Fujiwara, S.; Imanaka, T.; Thierry, J. C.; Moras, D. *EMBO J.* **1998**, *17*, 5227–5237. PDB Code: 1b8a.

(57) Giffin, M. J.; Stroud, J. C.; Bates, D. L.; von Koenig, K. D.; Hardin, J.; Chen, L. *Nat. Struct. Biol.* **2003**, *10*, 800–806. PDB Code: 1p7h.

(58) Černý, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2007**, *9*, 5291–5303.

(59) Florián, J.; Šponer, J.; Warshel, A. *J. Phys. Chem. B* **1999**, *103*, 884–892.

CT800567Q