

DNA–protein π -interactions in nature: abundance, structure, composition and strength of contacts between aromatic amino acids and DNA nucleobases or deoxyribose sugar

Katie A. Wilson, Jennifer L. Kellie and Stacey D. Wetmore*

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

Received February 25, 2014; Revised March 19, 2014; Accepted March 21, 2014

ABSTRACT

Four hundred twenty-eight high-resolution DNA–protein complexes were chosen for a bioinformatics study. Although 164 crystal structures (38% of those searched) contained no interactions, 574 discrete π -contacts between the aromatic amino acids and the DNA nucleobases or deoxyribose were identified using strict criteria, including visual inspection. The abundance and structure of the interactions were determined by unequivocally classifying the contacts as either π - π stacking, π - π T-shaped or sugar- π contacts. Three hundred forty-four nucleobase-amino acid π - π contacts (60% of all interactions identified) were identified in 175 of the crystal structures searched. Unprecedented in the literature, 230 DNA–protein sugar- π contacts (40% of all interactions identified) were identified in 137 crystal structures, which involve C–H... π and/or lone-pair... π interactions, contain any amino acid and can be classified according to sugar atoms involved. Both π - π and sugar- π interactions display a range of relative monomer orientations and therefore interaction energies (up to -50 (-70) kJ mol^{-1} for neutral (charged) interactions as determined using quantum chemical calculations). In general, DNA–protein π -interactions are more prevalent than perhaps currently accepted and the role of such interactions in many biological processes may yet to be uncovered.

INTRODUCTION

DNA–protein interactions are essential to life. Indeed, the genetic information contained in the sequence of DNA nucleobases (A, C, T and G) must be processed by enzymes, which transcribe the nucleobase code into RNA and sub-

sequently generate new proteins. Alternatively, proteins can bind to DNA in order to replicate the nucleobase sequence as cells grow and divide. DNA–protein interactions are also evident in other critical cellular processes, such as the repair of DNA damage caused by carcinogenic compounds or UV light (1–4). Contacts between DNA and proteins are typically noncovalent, which allows the resulting complex to perform necessary biological functions, yet readily degrade such that both biomolecules can provide additional function to the cell (5,6). The noncovalent contacts between DNA and proteins have traditionally been categorized as (direct or water-mediated) hydrogen bonding, ionic (salt bridges or DNA backbone interactions) and other forces, including van der Waals and hydrophobic interactions (7–9). Understanding each class of DNA–protein contacts will provide a greater appreciation of critical cell functions and open the door for the development of new medicinal and biological applications, including rational drug design (10–12) and the control of gene expression (13–16).

To gain an understanding of the interactions between DNA and proteins, previous work has searched crystal structures published in the protein data bank (PDB) and determined the relative frequency of different types of contacts. Early studies in this area were limited by the lack of high-resolution crystal structures of DNA–protein complexes (17–20). While this problem has been overcome in the past decade (7,21–23), more recent works disagree about the relative frequency of different types of contacts. Indeed, characterization of 129 DNA–protein complexes suggests that van der Waals interactions are more common than (direct or water-mediated) hydrogen bonding (7). In contrast, a survey of 139 DNA–protein complexes suggests that hydrogen bonding is more frequent than van der Waals, hydrophobic or electrostatic interactions (22). Such discrepancies may arise since, unlike hydrogen bonding, there are relatively undefined guidelines for the structure of van der Waals interactions, and therefore there is likely substantial variation among the interactions included in this category.

*To whom correspondence should be addressed. Tel: +1 403 329-2323; Fax: +1 403 329-2057; Email: stacey.wetmore@uleth.ca

Regardless, both studies determined that van der Waals interactions compose more than 30% of DNA–protein contacts (7,22).

In addition to traditional classifications of DNA–protein interactions, careful examination of the list of contacts identified in previous works suggests that many interactions occur between the DNA nucleobases and the aromatic amino acids (Supplementary Figure S1) (7,22). In general, interactions between aromatic rings are known to be widespread throughout chemistry and biology (24,25). Indeed, the prevalence and potential importance of interactions between aromatic side chains in proteins (26–31), as well as at protein–protein interfaces (32), have been documented through PDB searches. Furthermore, investigation of 89 RNA–protein complexes suggests that RNA–protein van der Waals interactions are more prevalent than hydrogen bonding, with the most favoured nucleotide–amino acid pairs including the aromatic amino acids (specifically, the U:Tyr, A:Phe and G:Trp pairs) (33), while a search of 61 structures revealed an abundance of interactions between Trp and the purines (8). Collectively, these studies suggest that closer investigations of DNA–protein π – π interactions are warranted.

Among the first studies to specifically consider DNA–protein π – π contacts, Mao *et al.* investigated the molecular recognition of adenosine 5'-triphosphate (ATP) by different proteins, and determined that π – π interactions between A and the aromatic amino acids are essential for substrate binding, with a 2.7:1.0 DNA–protein hydrogen bonding: π – π contact ratio (34). Subsequently, Baker and Grant identified a large number of π – π interactions between the DNA nucleobases and Tyr, Phe, His or Trp in 141 DNA–protein complexes (8). Unfortunately, the overall trends in the relative abundances of A–amino acid pairs are significantly different in these two studies. This discrepancy may arise due to differences in the structures searched, but is more likely an artefact of the (distance only) search criteria implemented. Indeed, ring proximity alone does not guarantee a suitable relative orientation of two residues, and therefore not all previously characterized interactions correspond to π – π (stacking or T-shaped) contacts (Supplementary Figure S2). Thus, the true frequency and structure of these interesting aromatic interactions between DNA and proteins remain unclear. Nevertheless, the proximity of the nucleobases and aromatic amino acids suggests that aromatic–aromatic (π – π or C/N–H... π) interactions may help stabilize DNA–protein complexes or may be involved in nucleic acid recognition.

Recent works corroborate that modern computational techniques can provide important information about π – π interactions (see, for example, references 35–39). In terms of DNA–protein contacts, quantum chemical calculations have been used to clarify the strength of π – π contacts between the nucleobases and aromatic amino acids found in experimental crystal structures (8,34,40–42). To complement this data, the preferred (lowest energy) relative monomer orientations have been identified for isolated dimers by systematically changing the relative orientations of monomers of fixed geometry (41,43–46) or fully relaxed systems (40–42). Both π – π stacking (face-to-face) (41,43–46) and π – π T-shaped (edge-to-face) (41,43–

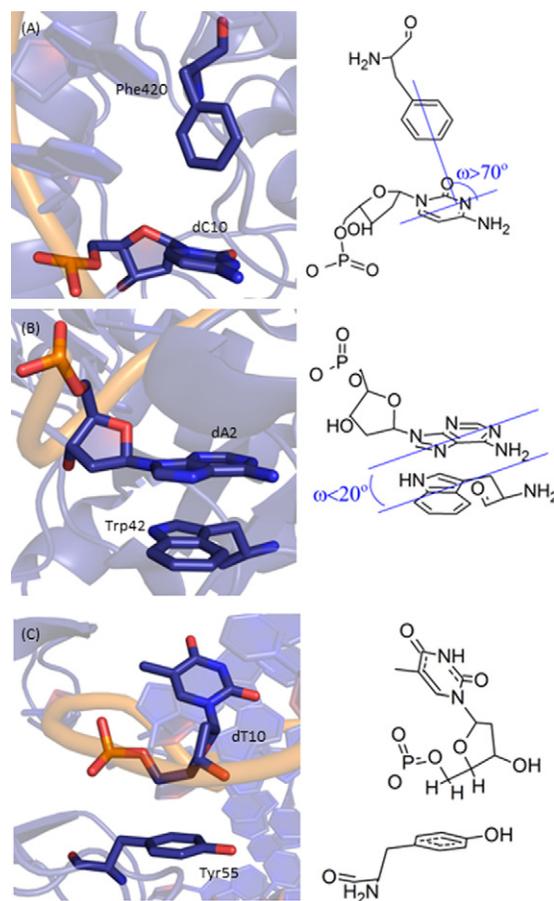


Figure 1. Examples of (A) nucleobase–amino acid π – π T-shaped interaction (PDB ID: 2WQ7), (B) nucleobase–amino acid π – π stacking interaction (PDB ID: 3MR5) and (C) deoxyribose–amino acid sugar– π interaction (PDB ID: 3BKZ).

46) contacts have been considered in these studies (Figure 1A and B). Our group has completed the most extensive investigations, where over 1000 relative monomer orientations were considered for each nucleobase–aromatic amino acid pair to determine the preferred relative monomer orientation (46–49). Our highly accurate calculations suggest that the strengths of these π – π stacking and T-shaped interactions are up to approximately -43 kJ mol^{-1} (46,50), which were calculated as the energy difference between the dimer and individual monomers. This suggests that π – π contacts can contribute to DNA–protein binding and/or stabilize DNA–protein complexes to the same extent as hydrogen bonding. Furthermore, our group investigated the enhancement in the binding energy due to charge by considering dimers involving cationic His (49) or a damaged (cationic alkylated) nucleobase (47,51,52), as well as the effects of water molecules on the stability of charged dimers (53). Although most of these studies were performed on model systems that only include aromatic rings, the extension of the computational model to include the biological backbone (54–56) or additional π – π contacts (57) has been determined to minimally affect the strength of individual contacts. Together, these works provide important details about

the preferred structure and magnitude of DNA–protein π – π interactions, and their potential biological roles.

In addition to interactions with the DNA nucleobases, analysis of crystal structures reveals a significant number of short distances between the aromatic amino acids and the DNA backbone (7,22). Although many of these likely correspond to ionic contacts or hydrogen bonding with the phosphate moiety, a significant number of interactions were deemed to specifically involve the deoxyribose sugar. Indeed, all aromatic amino acids were found to participate in these interactions in nature. Despite short distances between the sugar and the aromatic amino acids, the nature of these contacts has yet to be explicitly discussed in the literature.

In contrast to π -interactions involving the DNA sugar moiety, contacts between various carbohydrates and the aromatic amino acids have been identified in crystal structures (58–61), and the importance of these contacts has been accepted in many fields, including glycobiology (see, for example, (62)–(68) and reference therein) and nanotechnology (see, for example, (69)–(74) and references therein). The significant strength of carbohydrate– π contacts in crystal structures has been verified using computational methods (58–61). Other modeling studies have characterized the binding strengths of dimers between different carbohydrates and aromatic amino acids modeled as benzene (Phe) (73,75–79), toluene (Phe) (80–82), phenol (Tyr) (83) and/or indole (His) (80,83), or with the protein backbone included (84,85). Complexes involving naphthalene have also been considered in an effort to better understand the properties of carbohydrate C–H... π interactions (86). These works have collectively determined that the amino acid can interact with either side (face) of the carbohydrate. The strengths of the carbohydrate– π interactions are dependent on the carbohydrate, the amino acid and relative monomer orientation, and are up to approximately -50 kJ mol^{-1} , with the most stable structures containing both carbohydrate– π contacts and hydrogen bonding (with an exocyclic hydroxyl group). Interestingly, carbohydrate– π interactions involving a DNA nucleobase have also been characterized (87–90).

By analogy to the importance of carbohydrate– π interactions to glycobiology, it is reasonable to propose that π –contacts between the DNA deoxyribose moiety and the aromatic amino acids in proteins may provide stability and/or function in DNA–protein complexes. Furthermore, previous work on carbohydrate– π interactions suggests that deoxyribose contacts could involve C–H... π and/or hydrogen-bonding interactions (via the hydroxyl groups) with the amino acid π -system. From a fundamental perspective, the ring size is notably different between deoxyribose and the most widely studied carbohydrates (pyranoses), which could substantially affect the structure and energetics of the π -interactions. Although interactions predominantly involve one of the two carbohydrate faces, contacts may also occur with the sides of deoxyribose due to the relative positions of the ring hydrogen atoms.

In the current study, over 400 high-resolution DNA–protein complexes available in the PDB were searched to definitively determine the frequency and characterize the nature (structure, composition and strength) of contacts be-

tween the aromatic amino acids (including cationic His) and the DNA nucleobases (π – π contacts, Figure 1A and B) or the deoxyribose moiety (sugar– π contacts, Figure 1C). Unprecedented in the DNA–protein interaction literature, all nucleobase–aromatic amino acid dimers identified were visually inspected to unequivocally verify each contact represents a π – π interaction, and to classify the contact as either a nucleobase–amino acid stacking or T-shaped interaction (Figure 1A and B), which could involve either a nucleobase edge interacting with an amino acid π -system (face) or an amino acid edge interacting with the nucleobase face. Although experimental data can be used to identify contacts in nature, no information is obtained about the strength of these interactions. Therefore, accurate quantum chemical methods were used to evaluate the binding energy of each dimer system found in the crystal structures. Our study thereby clarifies previous literature by providing the most complete information to date on DNA–protein π – π interactions in nature. Using the same thorough approach, deoxyribose–aromatic amino acid sugar– π interactions in experimental crystal structures have been quantified for the first time, and determined to be based on many different types of noncovalent interactions that are known in structural chemistry, including C–H... π (Figure 1C) and lone-pair... π contacts. As a result, a novel classification system is developed based on the nature of the edge of the sugar. Combining data on the natural occurrence and strength of these two broad classes of DNA–protein interactions provides important information that will help unveil their potential roles in many biological systems.

MATERIALS AND METHODS

Datasets

DNA–protein complexes were identified in the PDB using similar criteria to those previously used in the literature to detect nucleobase–amino acid π – π contacts (Supplementary Figure S3) (8,30). Specifically, X-ray crystal structures published before 24 May 2011 with a resolution better than 2.0 Å and less than 90% sequence identity were chosen for analysis (428 crystal structures total).

Selecting systems for analysis

Pymol (91) was used to select all aromatic amino acids and nucleobase or deoxyribose moieties separated by less than 5.0 Å in each crystal structure. This choice of distance is supported by computational studies that determined the optimal vertical separation in DNA–protein nucleobase–aromatic amino acid dimers is typically 3.5 Å (45,46). As outlined in the Introduction, the qualifying DNA–protein dimers were then visually inspected to indisputably verify the contact is a π -interaction and classify the contact as either a nucleobase–amino acid stacking, nucleobase–amino acid T-shaped (nucleobase or amino acid edge) or deoxyribose sugar– π interaction. The PDB IDs for the crystal structures searched in the present work, as well as the type(s) of interactions identified and the nucleobase/sugar–amino acid residues involved, are provided in the SI.

Geometries used for quantum mechanical calculations

For the nucleobase–amino acid π – π interactions, the interplanar angle between the two rings, denoted as tilt (ω , Figure 1), was measured using Mercury (92), and used to further classify the π – π interaction as stacked ($\omega = 0$ – 20°), T-shaped ($\omega = 70$ – 90°) or inclined ($20^\circ < \omega < 70^\circ$). Mercury was also used to measure the closest heavy atom distance between monomers. The dimer binding strengths were determined using truncated models obtained by replacing the DNA or protein backbone with a hydrogen atom (Supplementary Figure S1). Previous research has shown that neglect of the DNA or protein backbone does not significantly affect the magnitude of the π – π contact (52,54,55). For His interactions, both a cationic (His⁺) and two neutral (His^δ and His^ε; Supplementary Figure S1) models were considered due to the unique pK_a of this amino acid, and therefore varied protonation states adopted in biological systems (93). Additionally, the hydroxyl group of Tyr was orientated in two directions, denoted as clockwise (CW) and counter-clockwise (CCW) according to the direction of the hydroxyl moiety when the dimer is oriented with Tyr below the nucleobase (see Supplementary Figure S1). The planar (C_s symmetric) monomers were aligned by overlaying MP2/6–31G(d) optimized geometries onto the crystal structure orientation according to root-mean-square (RMS) fitting of the ring heavy atoms using HyperChem 8.0.8 (94).

For all identified sugar– π interactions, the amino acid was initially overlaid (using RMS fitting) onto the crystal structure geometry as discussed for the nucleobase–amino acid interactions (94). However, due to variations in the sugar pucker throughout the crystal structures, and the anticipated effect of sugar puckering on the binding energy, a fully optimized isolated sugar could not be overlaid onto the crystal structure. Instead, the sugar moiety was first truncated by replacing the nucleobase, as well as the 5' and 3' phosphorus atoms, with hydrogen atoms (Supplementary Figure S1). Subsequently, all protons in the sugar–amino acid dimer were then optimized at the MP2/6–31G(d) level of theory, while fixing the heavy atoms. The $\angle(\text{C}_4\text{--C}_5\text{--O}_5\text{--H})$ and $\angle(\text{C}_4\text{--C}_3\text{--O}_3\text{--H})$ dihedral angles in the sugar (Supplementary Figure S1) were also frozen to the crystal structure geometry during the optimizations, in order to constrain the orientation of the hydrogen atoms at the O_{5'} and O_{3'} truncation points. This approach for sugar– π contacts is justified by studies revealing that neither structures nor binding strengths of carbohydrate– π interactions deviate significantly (< 2 kJ/mol) when crystal structures or fully optimized geometries are considered (58).

Interaction energies

Quantum chemical calculations were used to determine the strength of the intermolecular forces acting between the nucleobase and amino acid (π – π interactions) and the intermolecular forces acting between the sugar and amino acid (sugar– π interactions) based on the dimer geometries discussed in the previous section. Specifically, the interaction or binding energy (ΔE) was calculated according to Equation (1).

$$\Delta E = E^{\text{dimer}} - E^{\text{aa}} - E^{\text{nt}} \quad (1)$$

In this equation, E^{dimer} stands for the electronic energy of the π – π stacking, T-shaped or sugar– π dimer, while E^{aa} and E^{nt} stand for the electronic energies of the isolated subsystems (aromatic amino acid (aa) and nucleobase or deoxyribose subunit of the nucleotide (nt), respectively). The geometry of each monomer in the dimer is the same as the structure of the isolated monomer. The calculated interaction energy does not include zero-point vibrational or Gibbs energy correction. Furthermore, the binding energies were calculated in the gas phase and are therefore relevant to DNA–protein binding environments of low polarity (95). We acknowledge that polar environments will likely decrease the magnitude of the reported interaction energies, as well as diminish the impact of His protonation. Nevertheless, previous work has shown that π – π and $\pi_{\text{cation}}\text{--}\pi$ interactions are of significant strength in more polar environments (41,49,51). Future work should consider the effects of solvation and thereby extend our conclusions to all DNA–protein binding environments including the rarer high polarity active sites.

To identify a quantum chemical method that best balances accuracy and computational cost due to the large number of contacts identified, the binding strength of select dimers that span the range of interactions found in the PDB search was calculated with several levels of theory (Supplementary Table S1). The M06–2X density functional theory (DFT) functional was chosen (with both 6–31+G(d,p) and aug-cc-pVTZ basis sets) based on literature testing the ability of this functional to accurately describe carbohydrate– π contacts (96), as well as DNA–protein nucleobase–amino acid π –contacts (48,50). However, other DFT functionals were also considered that were originally developed to account for dispersion interactions and have proven to work well for noncovalent contacts (97,98), namely B3LYP-D3, B97-D3 and ω B97-D (with aug-cc-pVTZ basis sets). The DFT results were validated using the highly accurate CCSD(T) calculations at the complete basis set (CBS) limit. To obtain CCSD(T)/CBS estimates, MP2/CBS energies were determined using the aug-cc-pVDZ and aug-cc-pVTZ basis sets with Helgaker's extrapolation scheme (99,100), and the differences in the (counterpoise-corrected) MP2 and CCSD(T) energies were calculated with aug-cc-pVDZ and added to the MP2/CBS values. We note that these energies are denoted as CCSD(T)/CBS for consistency with our previous work on other DNA–protein interactions (46,48,50) despite some literature referring to these extrapolated values as CBS(T) (44,101–106). Furthermore, only slight changes in the interaction energies of nucleobase pairs have been reported upon considering a higher-level triple to quadruple-zeta extrapolation (107,108).

Upon changing the M06–2X basis set from 6–31+G(d,p) to aug-cc-pVTZ, the MUD (mean unsigned deviation) for the sugar– π interactions decreases (Supplementary Table S1). However, due to significant errors in the nucleobase–aromatic amino acid π – π interactions, the overall MUD increases with respect to the CCSD(T)/CBS estimate from 1.5 to 2.4 upon basis set expansion along with a substantial increase in computational time. Indeed, M06–2X has been shown to accurately describe other DNA–protein noncovalent interactions with a moderately sized basis set (48,50).

In contrast, ω B97x-D/aug-cc-pVTZ describes both broad classes of contacts as accurately as M06-2X/6-31+G(d,p), leading to the same overall MUD at an increased computational cost. Among the functionals tested, B3LYP-D3/aug-cc-pVTZ performs the best, but again this is coupled with significantly increased computational cost compared to the efficient M06-2X/6-31+G(d,p) combination. Most importantly, the trends in the interaction energies and the large magnitude of the nucleobase and sugar–aromatic amino acid π -interactions predicted by M06-2X/6-31+G(d,p) are preserved upon consideration of the CCSD(T)/CBS estimates. Thus, M06-2X/6-31+G(d,p) was confidently used in the present study to compare the strength of many different types of DNA–protein π - π interactions.

Software

All M06-2X, MP2 and CCSD(T) calculations were performed with program defaults using Gaussian 09 (revisions A.02 and C.01) (109), while all DFT-D and DFT-D3 calculations were performed using Q-Chem 4.0.1.0 (110).

RESULTS

Crystal structure analysis of nucleobase–aromatic amino acid contacts in nature

Overall distribution of contacts in DNA–protein complexes. Among the 428 crystal structures considered in the present work, 175 (41%) contain at least one nucleobase–amino acid stacking or T-shaped interaction, with 344 total nucleobase–amino acid stacking or T-shaped interactions identified. Most of the 175 crystal structures contain one or two interactions, but as many as 13 contacts can be found in a single structure (Figure 2A). These interactions occur in a wide variety of proteins, including DNA–binding and transcription proteins, with approximately 38% of the π - π contacts being identified in transferase proteins and 25% in hydrolase proteins (Figure 2B).

Occurrence of nucleobases and aromatic amino acids in contacts. Pyrimidines are involved in more π - π interactions than purines (Figure 3A), where the population trend with respect to the nucleobase decreases according to T > C > A ~ G. Specifically, 37% of the contacts involve T, with the remaining being relatively equally distributed among the other bases (~20%). When the distribution is considered as a function of the amino acid (Figure 3B), significantly more interactions are found with Phe (44%) and Tyr (32%) than either His (11%) or Trp (13%). Nevertheless, Trp is the least common amino acid (~1% abundance), which may explain the fewer contacts identified with this residue. On the other hand, Tyr, Phe and His have similar natural abundances (3–4%) and therefore our results suggest that His is less likely to form π - π stacking or T-shaped interactions with a DNA nucleobase. When all nucleobase–amino acid combinations are considered (Figure 3C), Phe, Tyr and Trp contacts decrease in abundance with respect to the nucleobase as T > C > A ~ G, while His forms the most contacts with C (the second most frequently observed interaction with respect to the nucleobase) and does not form any contacts with G.

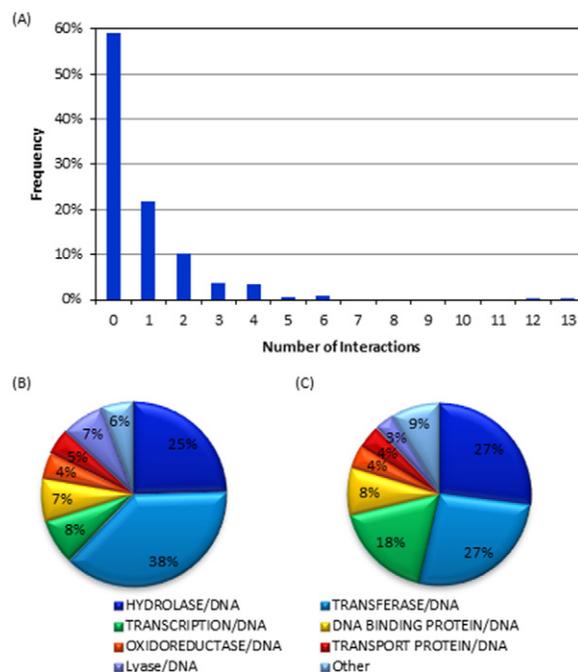


Figure 2. (A) Number of nucleobase–amino acid stacking/T-shaped interactions identified in PDB structures in the present study. (B) Types of proteins in which nucleobase–amino acid stacking/T-shaped interactions were found. (C) Overall composition of the proteins in the crystal structures considered in the present work.

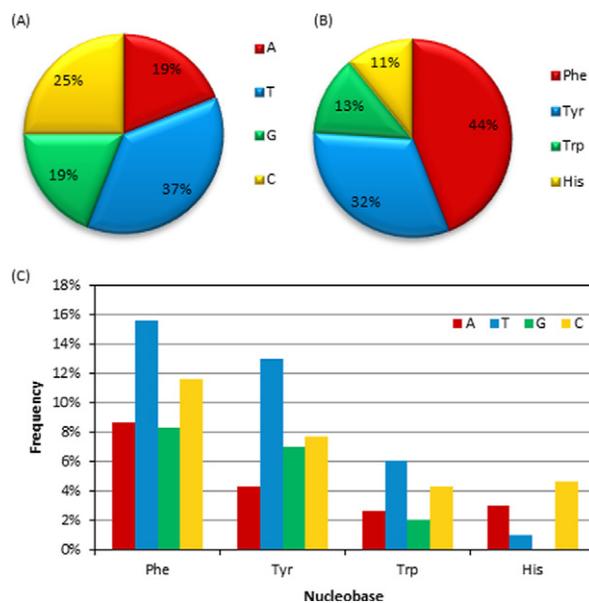


Figure 3. The proportions of (A) nucleobases, (B) amino acids and (C) nucleobase–amino acid combinations in DNA–protein π - π stacked and T-shaped orientations found in nature.

Relative abundance of face-to-face and face-to-edge π - π binding arrangements

The nucleobase–amino acid π - π contacts adopt conformations ranging from stacked ($\omega = 0$ – 20°) to T-shaped ($\omega = 70$ – 90°) orientations (Figure 4). However, the stacked orientation is substantially more common (58%) than the T-

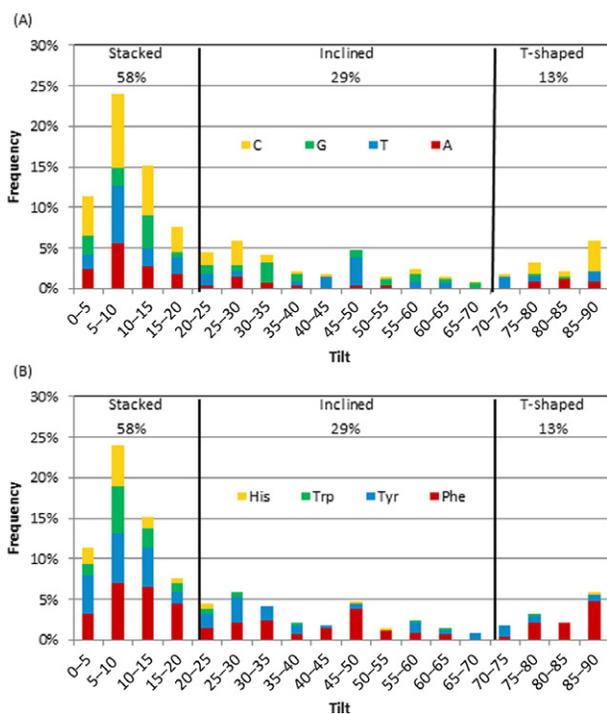


Figure 4. Frequency of tilt angle (degrees) between the ring planes for all interactions according to the (A) nucleobase or (B) amino acid.

shaped configuration (13%). The T-shaped interactions are also less frequent than the inclined structures ($\omega = 20\text{--}70^\circ$, 29%, Figure 4), but this is due to the large number of angles in the inclined category, while the frequency for a given angle in the T-shaped and inclined categories are nearly equal (approximately $<5\%$). Within the $\pi\text{--}\pi$ stacking interactions, the dimers more commonly adopt a tilt of $5\text{--}10^\circ$ rather than a perfectly parallel orientation ($\omega = 0$). Conversely, the perfectly perpendicular arrangement ($\omega = 90^\circ$) is the preferred T-shaped configuration. The most common inclined structures ($\omega = 20\text{--}70^\circ$) involve either a $\omega = 25\text{--}30^\circ$ or a maximum tilt of $45\text{--}50^\circ$ (Figure 4).

Dependence of $\pi\text{--}\pi$ binding arrangement on the nucleobase. A correlation exists between the nucleobase in the dimer and the tilt angle adopted (Figure 4A). Specifically, although all nucleobases prefer a stacked orientation, the largest frequency occurs with $\omega = 5\text{--}10^\circ$ for T, C and A, but with $\omega = 10\text{--}15^\circ$ for G. Among the inclined orientations, C and G prefer only slight deviations from stacking ($\omega = 25\text{--}35^\circ$), T prefers the maximum degree of tilt ($\omega = 45\text{--}50^\circ$) and A rarely adopts an inclined orientation ($<5\%$ frequency for $\omega = 30\text{--}70^\circ$). Cytosine is the most likely nucleobase to adopt a T-shaped structure (15% frequency for $\omega = 85\text{--}90^\circ$). Although A and T also adopt T-shaped orientations with $>10\%$ frequency, G rarely forms a T-shaped dimer ($<5\%$ frequency). Interestingly, A is only found in a T-shaped orientation with Phe. Furthermore, 74% of the identified T-shaped interactions and 21% of the inclined interactions involve a nucleobase edge and an amino acid face.

Dependence of $\pi\text{--}\pi$ binding arrangement on the amino acid. As discussed for the nucleobases, all amino acids show a

preference for the $\omega = 5\text{--}10^\circ$ stacked orientation, except His which equally prefers a $0\text{--}5^\circ$ tilt (Figure 4B). In fact, His and Trp are rarely found in any orientation besides a stacked structure (5 and 8% frequency for $\omega = 20\text{--}90^\circ$, respectively). Although Tyr adopts almost the full range of tilt angles, a stacked or slightly tilted orientation is most frequent adopted. Unlike the other amino acids, Phe exhibits a substantial occupancy of both inclined ($\omega = 45\text{--}50^\circ$) and T-shaped ($\omega = 85\text{--}90^\circ$) orientations (32 and 20%, respectively).

Trends in the distances between monomers. In addition to the varied tilt angles adopted by the nucleobase–amino acid dimers, many different separation distances are observed (Supplementary Figure S4A). Overall, the closest heavy atom distances fall between 3.0 and 4.2 Å in the nucleobase–amino acid $\pi\text{--}\pi$ dimers, with nearly a quarter of all interactions adopting a 3.5 Å separation. Interestingly, there is no clear correlation between the separation distance and tilt angle (Supplementary Figure S4B). Furthermore, unlike the stacking angle, which preferentially adopts a different value for each nucleobase, all bases have the same trend in the preferred separation distance (Supplementary Figure S4C). Conversely, the amino acids do not follow a particular trend in the separation distance. Specifically, Tyr adopts a large range of distances and His general adopts shorter distances ($<5\%$ occupancy of distances greater than 3.7 Å; Supplementary Figure S4D), while Phe and Trp display the same overall trend as across all $\pi\text{--}\pi$ contacts.

Quantum chemical calculations of nucleobase–aromatic amino acid interaction energies

The discussion above shows that nucleobase–amino acid dimers adopt a wide range of $\pi\text{--}\pi$ structures and therefore it is not surprising that the dimers also span a significant range of binding strengths (Figure 5). The magnitude of the nucleobase–amino acid stacking or T-shaped $\pi\text{--}\pi$ interaction depends on several factors such as the relative monomer orientation (including tilt angle), and the identity of the nucleobase and amino acid. For all DNA–protein pairs, the largest (most negative) binding energy occurs when the amino acid and nucleobase adopt a stacked ($\omega = 0\text{--}20^\circ$), not T-shaped ($\omega = 70\text{--}90^\circ$), orientation. With the exception of the fact that the maximum interaction energies generally occur for T and G, the most dominant trends depend on the amino acid. Therefore, interesting features of the binding energies will be discussed below as a function of the amino acid.

Phenylalanine. Phe interactions are up to -26.3 kJ mol^{-1} . In the stacked orientation, G or T generally leads to stronger contacts than A or C, while G or C interactions are generally stronger than T or A T-shaped interactions (Figure 5A). This leads to, for example, an 18.8 kJ mol^{-1} energy difference between the strongest T:Phe stacking and T-shaped dimers (Figure 5A).

Tryptophan. Similarly, the Trp interactions are up to -31.3 kJ mol^{-1} , with the strongest stacking interactions occurring with T or G (Figure 5B). However, no general conclu-

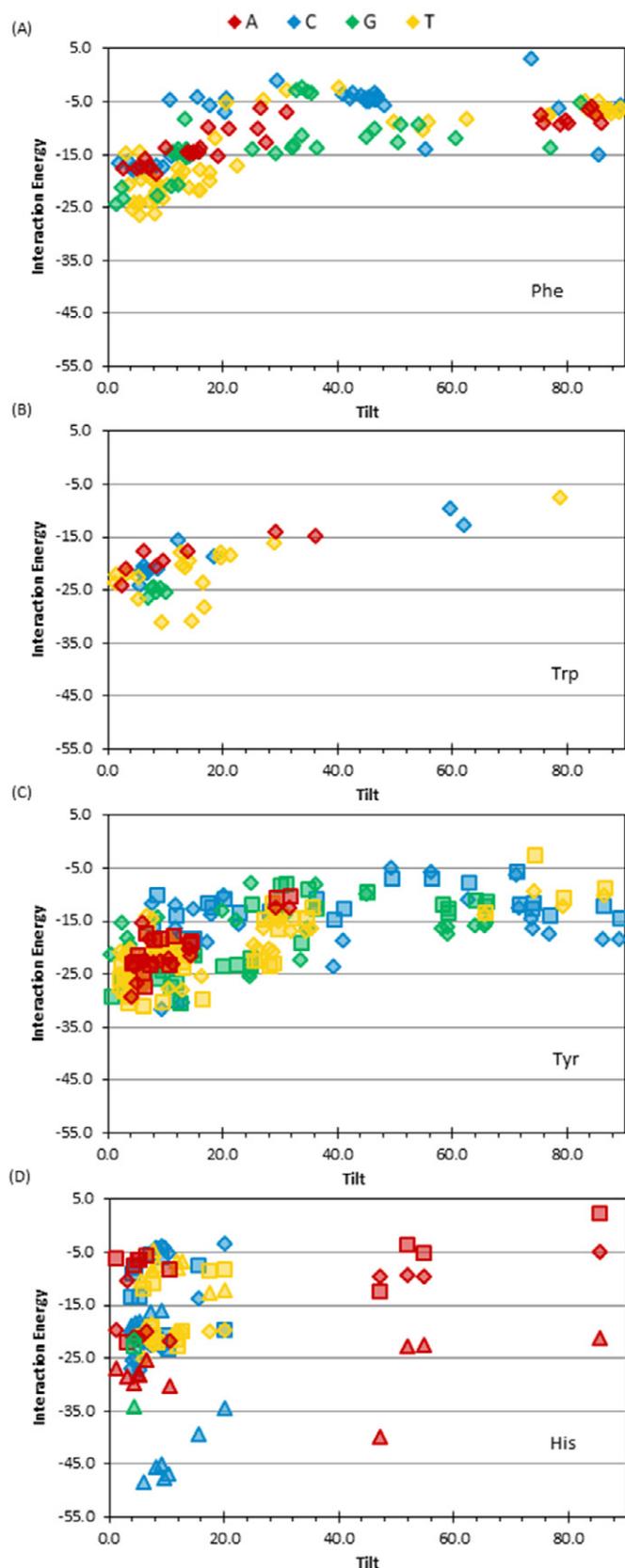


Figure 5. Binding energy of nucleobase–amino acid π – π interactions with respect to the tilt angle (degrees) for dimers involving (A) Phe, (B) Trp, (C) Tyr (for Tyr^{CW} (diamonds) and Tyr^{CCW} (squares)) and (D) His (for His^δ (diamonds), His^ε (squares) and His⁺ (triangles) (see Supplementary Figure S1, S1, for the definition of different Tyr and His conformations).

sions about the strength of Trp T-shaped interactions can be drawn since only one such contact was identified (Figure 5B).

Tyrosine. Unlike Trp and Phe, Tyr can adopt multiple conformations when stacked with the nucleobases, which differ in the orientation of the hydroxyl moiety (Supplementary Figure S1). However, the hydroxyl orientation has a negligible effect on the binding energy, with less than a 5 kJ mol⁻¹ energy difference between the two conformations for 74% of the interactions considered (Figure 5C). As discussed for Phe and Trp, Tyr interactions are stronger in the stacked rather than T-shaped orientation, with the largest deviation (up to 28.7 kJ mol⁻¹) occurring for T dimers (Figure 5C). The overall strongest Tyr interaction occurs with C (–31.6 kJ mol⁻¹, Figure 5C). Tyr nucleobase interactions are similar in strength to the corresponding Phe contact. Furthermore, although Tyr, Phe and Trp bind strongest to the pyrimidines, there is only a 5 kJ mol⁻¹ difference in the corresponding strongest interaction energies for these three amino acids.

Histidine. Similar to Tyr, (neutral) His can adopt two orientations (protonation states) with respect to the nucleobase (Supplementary Figure S1). However, unlike Tyr interactions, His contacts are highly dependent on the amino acid orientation, with 60% of the structures considered displaying a greater than 10 kJ mol⁻¹ energy difference with a change in His orientation and the largest difference (18 kJ mol⁻¹) occurring in a C dimer (Figure 5D). The greatest number of contacts and strongest interactions (–27.1 kJ mol⁻¹) with (neutral) His occur when stacked with C, which contrasts the greatest number and strongest interactions found with T for all other amino acids. As previously mentioned, very few His contacts were found to adopt a T-shaped orientation in nature (Figure 5D), where the only T-shaped interaction is –5.0 kJ mol⁻¹ and occurs with A. Interactions with cationic His are up to –48.7 kJ mol⁻¹, which is 21.6 kJ mol⁻¹ stronger than the neutral dimer. As for neutral His, the strongest interaction for cationic His occurs when stacked with C. Interestingly, although the interaction strengths between His and A, G or C always increase, and the interaction strengths with T decrease upon protonation. The different behaviour of T:His dimers upon protonation has been previously noted in the literature (49) and is attributed to the more positive π -system of T compared to the other nucleobases.

Crystal structure analysis of deoxyribose sugar–aromatic amino acid contacts in nature

Overall distribution of sugar– π contacts in DNA–protein complexes. Among the 428 crystal structures searched in the present study, 230 sugar– π contacts were identified in 137 structures. Although crystal structures containing sugar– π contacts typically have only one such interaction, up to six sugar– π contacts can be observed in a single structure (Figure 6A). The sugar– π contacts occur in a wide variety of DNA–binding proteins (Figure 6B). Interestingly, 68% of the structures do not contain a sugar– π interaction (Figure 6A), which is more than the 59% that do not contain a nucleobase–amino acid contact (Figure 2A), while

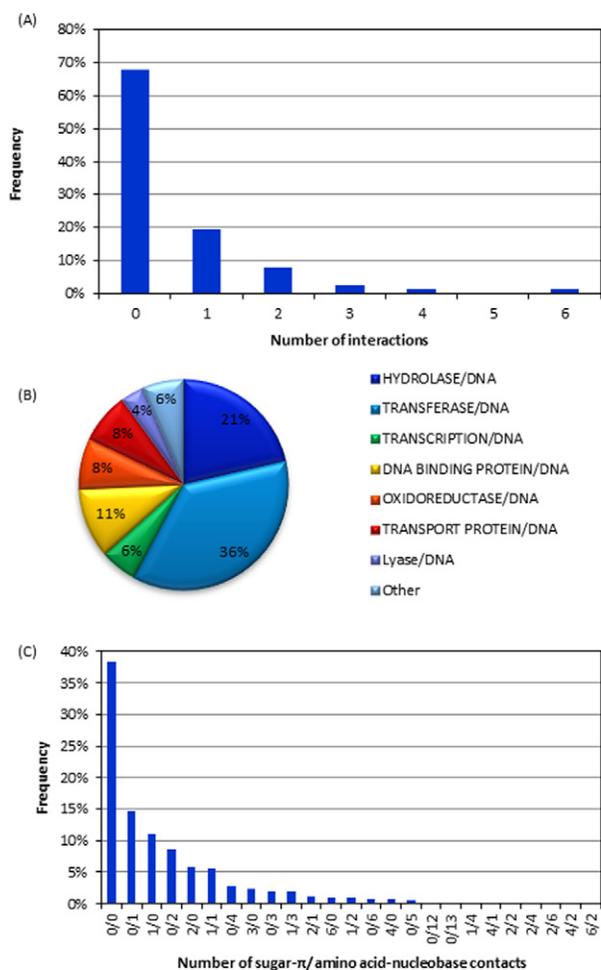


Figure 6. (A) The number of sugar- π contacts found in each structure. (B) Types of proteins in which sugar- π interactions were found. (C) The number of sugar- π and nucleobase-amino acid interactions observed in crystal structures considered in the present work.

38% of the structures do not contain any nucleobase π - π or sugar- π interactions (Figure 6C). Nevertheless, both types of amino acid interactions can be found in 11% of the structures, with these DNA-protein complexes typically possessing one of each type, but can contain up to six of one and two of the other class (Figure 6C).

Occurrence of aromatic amino acids in sugar- π contacts. Sugar- π interactions occur with all four aromatic amino acids (Figure 7A). However, most sugar- π contacts involve Tyr (45%), which is closely followed by Phe (36%). In contrast, few sugar- π interactions are found with His (4%) despite a similar natural abundance as Phe and Tyr (3–4%). Trp interactions make up 14% of all sugar- π interactions, which is consistent with the relative natural abundance of Trp (1%) in comparison to Tyr and Phe.

Classification of sugar- π contacts in DNA-protein complexes. A variety of contacts occur between the π -systems (faces) of the aromatic amino acids and deoxyribose in nature, which can be classified according to the sugar “edge” (Figure 8). The sugar edge that interacts with the π -system

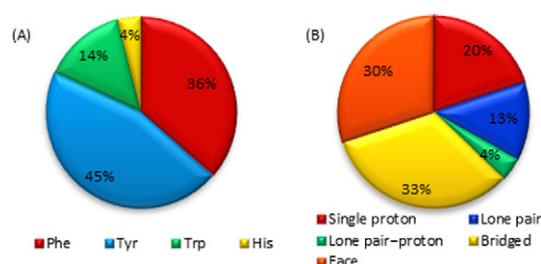


Figure 7. (A) Composition of sugar- π interactions found in nature as a function of amino acid. (B) Frequency of sugar- π interactions found in nature with respect to the class of contact.

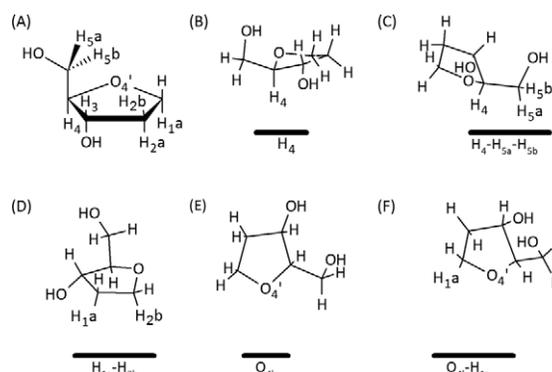


Figure 8. (A) Numbering scheme of the sugar moiety. Representative sugar- π interactions identified in crystal structures for (B) single proton, (C) face, (D) bridged, (E) lone pair and (F) lone pair-proton interactions (the amino acid is represented by a solid black line below the sugar).

can involve a single proton, two protons (a bridge), three protons (a face), a lone pair, or both a lone pair and a proton (lone pair-proton). Furthermore, these contacts can involve any of the hydrogen atoms in the sugar ring. The bridged and face interactions are the most common in the structures searched, with overall abundances of 33 and 30%, respectively (Figure 7B). While lone pair-proton interactions are fairly uncommon (4%), distinction between lone pair-proton and lone pair interactions is difficult, which collectively account for 17% of the contacts and is similar to the proportion of single proton interactions (20%, Figure 7B). Example orientations of the four most common interactions from select crystal structures are provided in Figure 9, which further clarifies the geometry of these contacts in nature.

Relative monomer orientations in sugar- π contacts. Figure 9 displays overlays of all contacts identified for each of the four most common sugar- π contacts, which were obtained using RMS fitting of the sugar atoms involved in the interaction. From these representative examples, it can be seen that the sugar- π interactions display significant variation in the amino acid position, which covers nearly all relative monomer orientations for a given sugar-edge type and leads to a continuum between the edges. Variations in the sugar are also evident from the overlays, which mainly arise due to different puckering in the crystal structures.

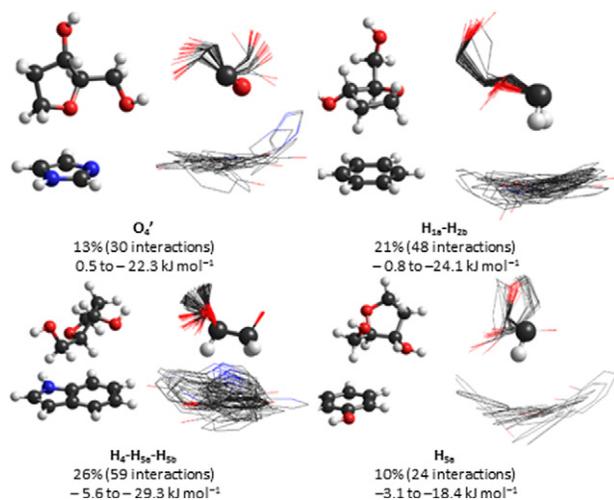


Figure 9. Example dimer and overlay of all dimers for the four most common sugar- π contacts identified in crystal structures, including calculated binding strengths.

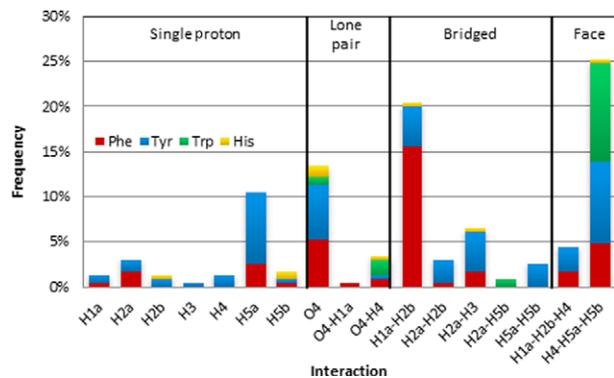


Figure 10. Frequency of sugar- π interactions found in nature with respect to the type of contact and the amino acid.

Dependence of binding arrangement on the sugar atoms involved. Within each category of sugar- π interactions, there is a clear preference for contacts with certain atoms (Figure 10). For example, single proton interactions occur with H_{5a} more than twice as frequently as any other proton. Similarly, the H_{1a}-H_{2b} bridged contact occurs more than three times as often as any other contact in this category and the H₄-H_{5a}-H_{5b} contact dominates the face class, which is in fact the overall most frequent sugar- π interaction (25% frequency). All lone pair interactions identified involve O_{4'} (rather than O_{5'} or O_{3'} phosphate backbone atoms) and more frequently do not involve a proton. When O_{4'} lone pair-proton interactions occur, contacts involving H₄ are twice as likely as those involving H_{1a}.

Dependence of binding arrangement on the amino acid. Within a given type of interaction, certain amino acids are more prevalent (Figure 10). Specifically, the single proton interactions are most common with Tyr. On the other hand, lone pair and bridged interactions involving each of the four aromatic amino acids can be identified, with Tyr or Phe involved in the majority of the contacts. Conversely, Trp and

Tyr compose approximately two-third of all face interactions. When the trend is instead considered as a function of amino acid and interaction adopted (Supplementary Figure S5), substantial variation in the types of contacts identified for each amino acid is noted. Trp only forms four types of sugar- π interactions in the crystal structures searched, which is fewer than for any other amino acid and does not include a single proton contact. The H₄-H_{5a}-H_{5b} face interaction makes up 76% of all sugar-Trp interactions, while the other three Trp interactions include two O_{4'} interactions and the H_{2a}-H_{5b} bridged interaction. Unlike Trp, His forms seven different sugar- π interactions that span all four categories of sugar- π contacts, with the O_{4'} interaction being the most common (30%) and the H_{5b} interaction also prevalent (20%, Supplementary Figure S5). In addition to being significantly more common, interactions with Phe and Tyr are markedly more varied, with more than 8 and 15 types of contacts found, respectively (Supplementary Figure S5). The most prevalent sugar- π Phe interaction is the H_{1a}-H_{2b} bridged interaction (43%), where Phe bridged interactions are in general considerably more common (59%) than face, lone pair and single proton contacts (19%, 16% and 13%, respectively). Unlike the other amino acids, Tyr does not substantially prefer one specific interaction. However, Tyr has some similarities to the other amino acids, where three of the four most common Tyr interactions include H₄-H_{5a}-H_{5b} (most common for Trp), O_{4'} (most common for His) and H_{1a}-H_{2b} (most common for Phe).

Quantum chemical calculations of deoxyribose sugar-aromatic amino acid interaction energies

The previous section shows that sugar- π interactions with the aromatic amino acids can adopt many different orientations in DNA-protein complexes. This structural variation leads to binding strengths for (neutral) sugar- π interactions between approximately 0 and -30 kJ mol⁻¹ (Figure 1180). Interactions with Trp are particularly strong, with magnitudes of up to -29.3 kJ mol⁻¹ and generally more stable than -20 kJ mol⁻¹. Interactions with Tyr can also be strong (up to -31.6 kJ mol⁻¹), but cover the full range of binding energies (i.e. from 0 to -30 kJ mol⁻¹). In general, the Tyr interactions do not greatly depend on the orientation of the hydroxyl moiety, with 86% of all sugar-Tyr interactions displaying a less than 5 kJ mol⁻¹ difference between the two orientations, but the dependence can be up to 22.1 kJ mol⁻¹ when a hydrogen bond forms in addition to the sugar- π interaction. Conversely, although Phe and (neutral) His contacts are generally weaker, they exhibit a significant range (from 0 to -20 kJ mol⁻¹, Figure 11). Similar to Tyr, the His binding strength depends on the amino acid orientation by 0.1-20 kJ mol⁻¹. The overall strongest sugar- π contacts typically occur when His is cationic (especially when interacting with O_{4'}), with binding strengths up to -68.2 kJ mol⁻¹.

Dependence on sugar edge. Among all sugar edge-aromatic amino acid combinations, only interactions with H_{2a}, H_{2b}, O_{4'}-H₄, H_{2a}-H₃, H_{1a}-H_{2b} and H₄-H_{5a}-H_{5b} have (neutral) interaction energies stronger than -20 kJ mol⁻¹ and only occur with Trp and Tyr. The strongest interac-

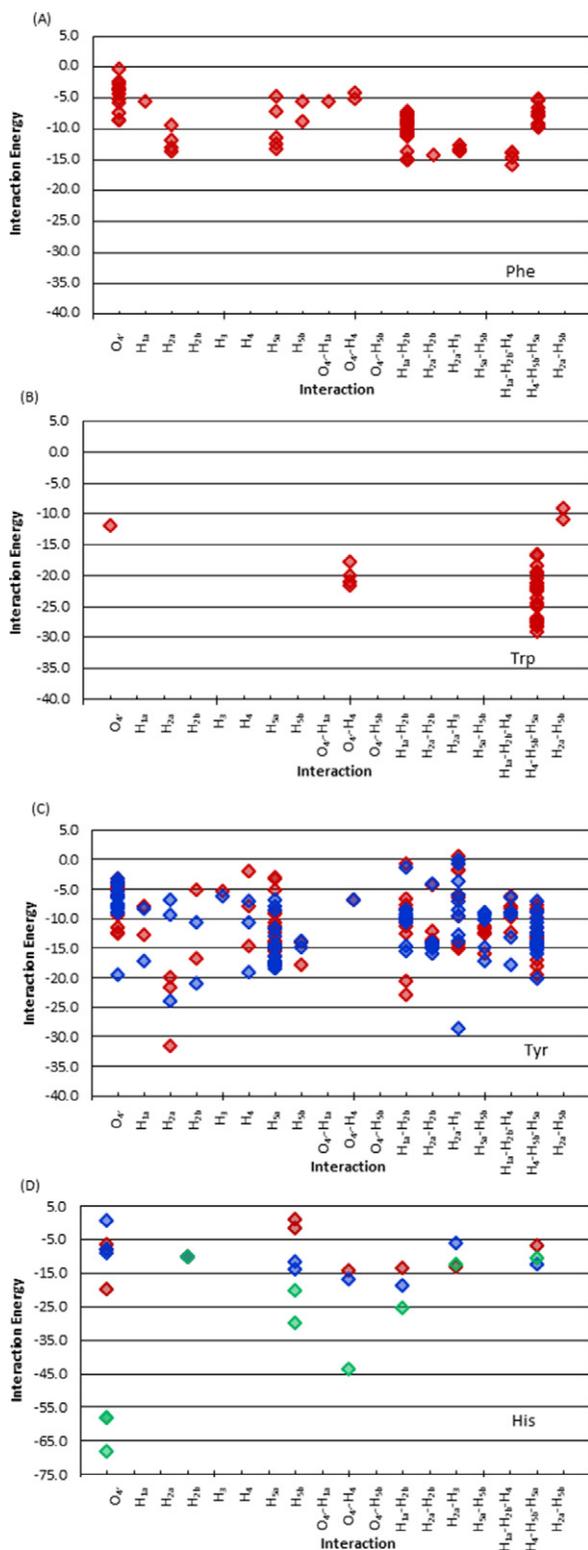


Figure 11. Binding energies of sugar- π interactions with respect to the type of contact for dimers involving (A) Phe, (B) Trp, (C) Tyr (for Tyr^{CW} (red) and Tyr^{CCW} (blue)), and (D) His (for His ^{δ} (red), His ^{ϵ} (blue) and His ^{σ} (green)), (see Supplementary Figure S1, for the definition of different Tyr and His conformations).

tions with Trp, Tyr, Phe and (neutral) His occur for H₄-H_{5a}-H_{5b} (-29.3 kJ mol⁻¹), H_{2a} (-31.6 kJ mol⁻¹), H_{1a}-H_{2b}-H₄ (-16.2 kJ mol⁻¹) and H_{1a}-H_{2b} (-18.9 kJ mol⁻¹), respectively. The overall four strongest interactions are the H₄-H_{5a}-H_{5b} dimer (-29.3 kJ mol⁻¹), followed by the H_{1a}-H_{2b} (-24.1 kJ mol⁻¹), O_{4'} (-22.3 kJ mol⁻¹) and H_{5a} (-18.4 kJ mol⁻¹) contacts (Figure 11). Furthermore, the binding strength of these four structures can vary by up to approximately 25 kJ mol⁻¹ due to differences in the relative orientation of the amino acid residue (Figure 11).

DISCUSSION

Abundance of nucleobase-aromatic amino acid π - π interactions

In the 428 crystal structures containing DNA-protein π -interactions (see Supplementary Data), 344 nucleobase-aromatic amino acid π - π contacts were identified and, for the first time in the literature, unambiguously confirmed through visual inspection. These contacts were found in all types of proteins (Figure 2B). However, the protein distribution directly correlates with the protein composition of the DNA complexes investigated (Figure 2C), which suggests that the observed distribution is a consequence of the structures searched rather than one protein class being more likely to rely on nucleobase-amino acid π - π interactions.

Structure of nucleobase-aromatic amino acid π - π interactions

Among the nucleobase interactions identified, stacked orientations (with a 5–10° angle (tilt) between ring planes) are more prevalent than T-shaped arrangements in a 3:2 ratio (Figure 4). Nevertheless, structures ranging from perfectly parallel to perfectly perpendicular relative monomer orientations appear in nature. Interestingly, the typical closest heavy atom-heavy atom distance between the two monomers (3.5 Å; Supplementary Figure S4) matches the preferred distance previously identified in computational studies of isolated monomers (45,46), and therefore some features of the relative monomer orientations in crystal structures may arise due to the inherent nature of the interactions.

Composition of nucleobase-aromatic amino acid π - π interactions

The pyrimidines are more likely to be involved in π - π interactions with aromatic amino acids than the purines (Figure 3A), which contrasts expectations that a larger ring size may lead to more π -interactions in nature due to greater possible overlap. In terms of the amino acids, more interactions occur with Phe and Tyr than with Trp and His in nature (Figure 3B), which does not directly relate to the relative natural abundances of these amino acids. This finding also contrasts previous literature that reports His to be the most likely aromatic amino acid to be involved in DNA-protein π - π interactions (8). Furthermore, our observation that Phe, Tyr and Trp contacts decrease in abundance with respect to the nucleobase as T > C > A ~ G. His was found

to form the most contacts with C. No contacts between His and G were identified (Figure 3C). These findings contrast previous reports that His selectively binds to T and G, while Phe selectively binds to T and A (7,8). Discrepancies between the present study and previous work may arise due to the careful visual inspection implemented herein as additional verification prior to classifying the π - π interactions.

Strength of nucleobase–aromatic amino acid π - π interactions

Since there is a large variation in the geometry of nucleobase–amino acid π - π interactions in nature (Figure 4), it is not surprising that there is also significant variation in the calculated binding strengths (Figure 5), as reported previously in computational studies of isolated dimers (40–47) or select crystal structure geometries (8,34,40–42). The magnitude of the nucleobase–amino acid π - π interactions are up to approximately -30 kJ mol⁻¹ and vary with the monomers involved and their relative orientation (with stacked structures being more stable than T-shaped). However, the trends in the binding strengths are not always the same as those found by considering two monomers in the absence of geometrical constraints imposed by an enzyme (45–47,49–57). Interestingly, most interactions identified in nature are on average 4.9 kJ mol⁻¹ weaker than the corresponding optimal interaction previously reported between two monomers in the absence of geometrical constraints imposed by the enzyme (Supplementary Table S2) (45–47,49–57). This difference arises due to deviations in the geometries (Supplementary Table S2), including greater separation distances and tilt in the crystal structures, which likely arise due to constraints imposed by the protein versus the perfectly parallel (stacked) or perpendicular (T-shaped) monomer arrangements implemented in the potential energy surface searches. The perfectly stacked or T-shaped orientations, as well as the step size implemented, in previous calculations also explain why three of the interaction energies calculated in the natural orientations are slightly stronger than the “optimal” values identified by searching the potential energy surface. These features underscore the influence of the relative monomer orientations on the binding strengths. In agreement with previous studies of charged DNA–protein interactions (41,49,50,53) and reports that π - π and $\pi_{\text{cation}}-\pi$ interactions are distinct (111), cationic His has significantly stronger interactions than the neutral amino acids, with interaction energies up to approximately -50 kJ mol⁻¹.

Biological relevance of nucleobase–aromatic amino acid π - π interactions

Nucleobase–aromatic amino acid π - π interactions have been implicated in the discriminatory and catalytic removal of damaged bases from the human genetic code by the DNA repair enzyme alkyladenine DNA glycosylase (AAG) (4,112). Specifically, unlike other DNA repair enzymes in the same glycosylase family, the active site of AAG is lined with three aromatic amino acids and there is limited hydrogen bonding to the substrate (Figure 12A). Although the resolution of the associated crystal structure (PDB ID:

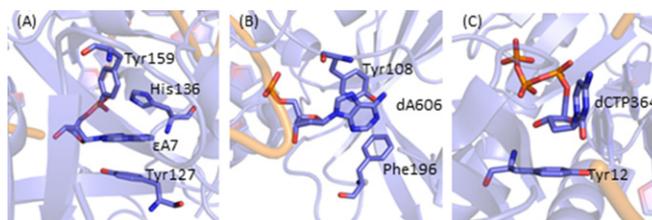


Figure 12. (A) The damaged nucleobase–amino acid π - π interactions in the AAG active site (PDB ID: 1EWN), (B) the natural nucleobase–amino acid π - π in the active site (PDB ID: 1G38) and (C) the sugar- π interaction in the Dpo4 active site (PDB ID: 3QZ8).

1EWN) is lower than the criteria used to select PDB structures in this study, and the interactions occur with a damaged nucleobase, the strengths of contacts between AAG and the bound substrate, ethenoadenine (ϵ A), were evaluated using the same methodology employed in the present work. Specifically, the interactions were determined to be -24.4 kJ mol⁻¹ for the ϵ A:Tyr127 stacking interaction, -6.9 kJ mol⁻¹ for the ϵ A:His136 tilted (inclined) contact and -1.0 kJ mol⁻¹ for the ϵ A:Tyr159 T-shaped (amino acid–edge) interaction. In particular, the strength of the ϵ A:Tyr127 contact suggests that such active site π - π interactions could be involved in substrate identification and/or binding.

The broader implications of the DNA–protein π - π contacts in the AAG active site were determined by a computational study of the associated catalytic mechanism using a full DNA–AAG model and different substrates (112). Specifically, the individual effects of sequentially removing each AAG active site amino acid suggest that the π -rings are catalytic (by approximately 30 kJ mol⁻¹) for the removal of neutral damaged nucleobases, but anti-catalytic for the removal of charged (cationic) alkylated nucleobases (by up to 35 kJ mol⁻¹). Coupled with previous work studying the strength of isolated dimers between a natural/damaged DNA base and an aromatic amino acid (47,51,52,57), a proposal was developed that AAG has evolved to take advantage of active site amino acid π -systems in several ways. First, the flexibility provided by the active composition (lack of discriminatory hydrogen bonding) explains why AAG can excise many different substrates. Second, the π - π interactions with the substrate maximize the catalytic power towards neutral lesions that are inherently difficult to excise. Finally, although the ability to remove neutral DNA lesions comes at the expense of the excision of cationic lesions, the inherent nature of $\pi_{\text{cation}}-\pi$ interactions (47,51,52,57) allows AAG to more strongly attract and bind cationic lesions.

Although AAG provides an exemplary example of the multiple roles π - π contacts can play in biology, interactions between damaged nucleobases and an aromatic amino acid residue may also be involved in the catalytic mechanism of other enzymes. Repair enzymes such as hUNG2 (113,114) and hOgg1 (115,116) are known to have π - π interactions in their active sites (involving Phe or Tyr), which may contribute towards the catalytic function of these enzymes. Notably, although AAG, hUNG2 and hOgg1 all involve damaged DNA nucleobase active site π - π interactions, π - π interactions are also known to contribute to

the binding and catalytic function of proteins that process natural DNA. For example, the extrahelical target A of N⁶-adenine DNA methyltransferase (PDB ID: 1G38; Figure 12B) forms an active site stacking interaction with Tyr108 (−21.6 kJ mol^{−1}) and a T-shaped interaction with Phe196 (−7.7 kJ mol^{−1}). Furthermore, as discussed for the DNA repair enzymes, the π – π interactions in the active site of N-DNA methyltransferases (including N⁶-adenine DNA methyltransferase) have been proposed to contribute to catalysis (117).

Abundance of deoxyribose–aromatic amino acid sugar– π interactions

Among the 428 crystal structures searched in the present work, 230 sugar– π contacts between the deoxyribose moiety and an aromatic amino acid were identified. Although a considerable number of nucleobase π – π interactions were expected based on previous literature (7,8,21–23,34), this is the first time that the significance of sugar– π contacts has been highlighted. Indeed, sugar– π contacts represent approximately 40% of all DNA–protein π –contacts found in the present work, and therefore occur with nearly the same frequency as nucleobase–amino acid π – π interactions. As discussed for the nucleobase–aromatic amino acid interactions, the sugar– π contacts are found in a variety of different proteins, with the relative abundances equal to the types of proteins searched (Figures 2C and 6B).

Structure of deoxyribose–aromatic amino acid sugar– π interactions

Although only π –interactions between the entire sugar face of pyranose and the aromatic amino acid were considered in previous work (61,62,67,76), a range of sugar– π contacts were identified for deoxyribose in the present study, which can involve a single proton, two protons (a bridge), three protons (a face), a lone pair, or both a lone pair and a proton (lone pair–proton; Figures 7B, 8 and 10). As a result, we introduce a classification system for DNA–protein sugar– π interactions based on the sugar edge participating in the contact, which can yield C–H... π and/or lone–pair... π interactions. In the literature, pyranoses involved in stacking interactions simultaneously participated in hydrogen bonding via a hydroxyl group and/or other van der Waals contact(s) (82–84). Although this preference was not explicitly examined in the present work, such hydrogen-bonding contacts are likely less important in the case of deoxyribose due to the lack of hydroxyl substituents on the sugar in DNA helices (except at the terminal positions). Interestingly, for each class of sugar– π interactions, the amino acid adopts a continuum of positions with respect to the sugar moiety (Figure 9).

Composition of deoxyribose–aromatic amino acid sugar– π interactions

Across the deoxyribose contacts identified in nature, each hydrogen atom in the sugar ring is involved in an interaction with the π –system of an aromatic amino acid (Figure

10). Nevertheless, certain atoms are more prone to participate in particular types of contacts (H_{5a} dominates the single proton, H_{1a}–H_{2b} the bridged and H₄–H_{5a}–H_{5b} the face interactions). Furthermore, although the bridged and face interactions are the most common overall relative monomer arrangements (Figure 10), interactions with the ring oxygen (rather than the O_{3'} or O_{5'} phosphate atoms) are also prevalent and are sometimes accompanied by a C–H... π contact.

The abundance of interactions with respect to the amino acid involved (Figure 7A) is similar to that discussed for the amino acid–nucleobase contacts (Figure 3B), with most interactions involving Tyr and Phe. The preferred binding arrangement is different for each amino acid, which likely occurs due to differences in the relative size of the π –systems. Specifically, Trp displays a preference for face interactions, Phe prefers bridged contacts, and His adopts the most lone pair– π contacts (Figure 10). Although Tyr assumes a wide variety of conformations with respect to the sugar moiety, most single proton interactions occur with Tyr (Figure 10).

Strength of deoxyribose–aromatic amino acid sugar– π interactions

The variation in the sugar– π conformations leads to a significant range in the binding energies (Figure 11), which are as strong as, or even stronger than, nucleobase–amino acid interactions (Figure 5). Indeed, the magnitude of sugar– π contacts found in nature can be up to approximately −70 kJ mol^{−1}. Among the neutral dimers, the sugar interactions with Trp are the strongest (most negative), which is consistent with the highly stable nucleobase–Trp interactions found in the present work and reported previously (45,46,50), as well as carbohydrate–Trp contacts (83). Nevertheless, the strongest interactions overall occur with cationic His, as discussed for the nucleobase π –contacts, which typically represent lone pair binding arrangements.

Interestingly, although the strongest interactions occur when a pyranose C–H is directed at the center of the aromatic face (76), the amino acid displays a wide range of locations with respect to the sugar in DNA sugar– π contacts. This implies that the sugar composition plays a large role in determining the preferred geometry of the interaction. To gain further fundamental information about sugar– π contacts, calculations as previously conducted for nucleobase–amino acid pairs (45,46,49) that consider the preferred relative orientation of isolated dimers in the absence of an enzyme, as well as the associated inherent interaction energy, should be considered for sugars of varying composition.

Biological relevance of deoxyribose–aromatic amino acid sugar– π interactions

Despite the fact that DNA sugar– π contacts with aromatic amino acid residues are rarely discussed in the literature, the importance of analogous carbohydrate– π interactions in many fields (62–74) coupled with the number of contacts found in nature in the present study suggests that these interactions may also be important for biological processes, either by providing stability to DNA–protein complexes, facilitating DNA binding/recognition, or possibly even having a greater (catalytic) role. As an example, the DNA polymerases in the RT, Y, X and B-families that are involved

in crucial cell replication have a conserved Tyr/Phe in their active sites. It has been proposed that the conserved π -containing amino acid uses stacking with the deoxyribose sugar through the R-group and hydrogen bonding with the 3'-OH through the backbone to select DNA deoxyribose nucleotide triphosphates (dNTPs) over RNA ribose nucleotide triphosphates (rNTPs) in a 1 000 000 (118) to 100 ratio (119). Indeed, the conserved Tyr/Phe has been referred to as a 'steric gate' since steric clashes may prevent incorporation of rNTP (enhance dNTP incorporation) (120). Nevertheless, the only support for this proposal comes from crystal structures (119,121) or mutational studies (120,122–125) that replace Tyr/Phe by Gly/Ala/Val, which significantly reduces the size of the R-group and removes the π -system.

In the present work, the sugar- π interactions in crystal structures with a nucleoside triphosphate bound in the active site were re-evaluated and determined to almost exclusively represent either H_{1a}-H_{2b} or H_{1a}-H_{2b}-H₄ contacts with Tyr or Phe depending on the dNTP orientation. A representative example is the H_{1a}-H_{2b} sugar- π interaction between Tyr12 and the incoming dCTP in the Dpo4 active site (a Y-family polymerase; PDB ID: 3QZ8; Figure 12C), which has a corresponding calculated binding energy of -15.6 kJ/mol (Tyr^{CCW} or -12.6 kJ/mol Tyr^{CCW}; see Supplementary Figure S1 for definition of Tyr orientations). This is a significant magnitude and indicates that the sugar- π contact with Tyr12 may be more than simply a steric constraint and, for example, may contribute to the selection of dNTP over rNTP. Indeed, modification of the sugar to the corresponding ribose analogue severely impacts this interaction in the polymerase active site, decreasing the closest heavy atom contact distance between the sugar and Tyr planes to 2.126 Å (3.397 Å with deoxyribose present) and is repulsive by approximately 95 kJ mol⁻¹ (with same hydroxyl orientation, which makes the sugar- π interaction highly repulsive). Although the RNA sugar- π interaction is repulsive compared to the stabilizing interaction with the DNA analogue in the Dpo4 example discussed above, this calculation was performed on a structure obtained by replacing the sugar without geometry relaxation. Therefore, it is possible that different relative monomer orientations in RNA-protein complexes allow sugar- π contacts to be capitalized for cellular RNA processing. Nevertheless, this example illustrates the potential importance of DNA sugar- π contacts in human biology.

CONCLUSIONS

In summary, our calculations yield important insight into the abundance and strength of over 500 DNA-protein interactions in nature. This in turn can be used to estimate the magnitude of similar contacts identified in lower resolution or newly released crystal structures. Most importantly, the present contribution suggests that nucleobase-amino acid contacts are wider spread than perhaps originally believed and highlights the role of novel interactions between the deoxyribose moiety and the aromatic amino acids, which parallel the carbohydrate- π contacts identified in glycobiology (62–68). Furthermore, we confirm for the first time that both broad classes of DNA-protein π -contacts are varied

in structure and can provide significant stability to DNA-protein complexes. We therefore propose that the critical role of nucleobase-aromatic amino acids π - π interactions and deoxyribose-aromatic amino acid sugar- π contacts in many biological processes may yet to be uncovered. Indeed, examples can be found of both types of DNA-protein contacts in the active sites of enzymes crucial for human survival. Understanding the DNA-protein π -interactions in such systems may lead to advances in nanotechnology (69–74) and (anticancer (4,126,127) or antiviral (128–130)) drug development.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENT

Computational resources from the Upscale and Robust Abacus for Chemistry in Lethbridge (URACIL) and those provided by Westgrid and Compute/Calcul Canada are greatly appreciated.

FUNDING

Natural Sciences and Engineering Research Council of Canada (249598-07); Canada Research Chair Program (950-228175); Canadian Foundation for Innovation (22770); Natural Sciences and Engineering Research Council of Canada Undergraduate Summer Research Award (to K.A.W.); Natural Sciences and Engineering Research Council of Canada Alexander Graham Bell Canada Graduate Scholarship-Doctoral and the University of Lethbridge (to J.L.K.).

Conflict of interest statement. None declared.

REFERENCES

- Ciccia, A. and Elledge, S.J. (2010) The DNA damage response: making it safe to play with knives. *Mol. Cell*, **40**, 179–204.
- Hoeijmakers, J.H.J. (2001) Genome maintenance mechanisms for preventing cancer. *Nature*, **411**, 366–374.
- Jackson, S.P. and Bartek, J. (2009) The DNA-damage response in human biology and disease. *Nature*, **461**, 1071–1078.
- Stivers, J.T. and Jiang, Y.L. (2003) A mechanistic perspective on the chemistry of DNA repair glycosylases. *Chem. Rev.*, **103**, 2729–2759.
- von Hippel, P.H. and Berg, O.G. (1986) On the specificity of DNA-protein interactions. *Proc. Natl. Acad. Sci. U.S.A.*, **83**, 1608–1612.
- Rohs, R., Jin, X., West, S.M., Joshi, R., Honig, B. and Mann, R.S. (2010) Origins of specificity in protein-DNA recognition. *Ann. Rev. Biochem.*, **79**, 233–269.
- Luscombe, N.M., Laskowski, R.A. and Thornton, J.M. (2001) Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. *Nucleic Acids Res.*, **29**, 2860–2874.
- Baker, C.M. and Grant, G.H. (2007) Role of aromatic amino acids in protein-nucleic acid recognition. *Biopolymers*, **85**, 456–470.
- Rutledge, L.R. and Wetmore, S.D. (2010), *Quantum biochemistry*. Wiley-VCH Verlag GmbH & Co. KGaA, pp. 307–336.
- Mavromoustakos, T., Durdagi, S., Koukoulitsa, C., Simcic, M., G. Papadopoulos, M., Hodosek, M. and Golic Grdadolnik, S. (2011) Strategies in the rational drug design. *Curr. Med. Chem.*, **18**, 2517–2530.
- Speck-Planche, A., Kleandrova, V.V., Luan, F. and Cordeiro, M.N.D.S. (2012) Rational drug design for anti-cancer

- chemotherapy: multi-target QSAR models for the in silico discovery of anti-colorectal cancer agents. *Bioorg. Med. Chem.*, **20**, 4848–4855.
12. Bultinck, J., Lievens, S. and Tavernier, J. (2012) Protein-protein interactions: network analysis and applications in drug discovery. *Curr. Pharm. Des.*, **18**, 4619–4629.
 13. Dermitzakis, E.T. (2008) From gene expression to disease risk. *Nat. Genet.*, **40**, 492–493.
 14. Cuccato, G., Gatta, G.D. and di Bernardo, D. (2009) Systems and synthetic biology: tackling genetic networks and complex diseases. *Heredity*, **102**, 527–532.
 15. Stranger, B.E. and Dermitzakis, E.T. (2006) From DNA to RNA to disease and back: the 'central dogma' of regulatory disease variation. *Coron. Artery Dis.*, **2**, 383–390.
 16. Stranger, B.E. and Raj, T. (2013) Genetics of human gene expression. *Curr. Opin. Genet. Dev.*, **23**, 627–634.
 17. Matthews, B.W. (1988) No code for recognition. *Nature*, **335**, 294–295.
 18. Pabo, C.O. and Sauer, R.T. (1992) Transcription factors: structural families and principles of DNA recognition. *Annu. Rev. Biochem.*, **61**, 1053–1095.
 19. Suzuki, M. (1994) A framework for the DNA-protein recognition code of the probe helix in transcription factors: the chemical and stereochemical rules. *Structure*, **2**, 317–326.
 20. Mandel-Gutfreund, Y., Schueler, O. and Margalit, H. (1995) Comprehensive analysis of hydrogen bonds in regulatory protein DNA-complexes: in search of common principles. *J. Mol. Biol.*, **253**, 370–382.
 21. Luscombe, N.M. and Thornton, J.M. (2002) Protein-DNA interactions: amino acid conservation and the effects of mutations on binding specificity. *J. Mol. Biol.*, **320**, 991–1009.
 22. Lejeune, D., Delsaux, N., Charlotteaux, B., Thomas, A. and Brasseur, R. (2005) Protein-nucleic acid recognition: statistical analysis of atomic interactions and influence of DNA structure. *Proteins: Struct. Funct. Bioinform.*, **61**, 258–271.
 23. Sathyapriya, R., Vijayabaskar, M. and Vishveshwara, S. (2008) Insights into protein-DNA interactions through structure network analysis. *PLoS Comput. Biol.*, **4**, e1000170.
 24. Meyer, E.A., Castellano, R.K. and Diederich, F. (2003) Interactions with aromatic rings in chemical and biological recognition. *Angew. Chem. Int. Ed. Engl.*, **42**, 1210–1250.
 25. Salonen, L.M., Ellermann, M. and Diederich, F. (2011) Aromatic rings in chemical and biological recognition: energetics and structures. *Angew. Chem. Int. Ed.*, **50**, 4808–4842.
 26. Brandl, M., Weiss, M.S., Jabs, A., Sühnel, J. and Hilgenfeld, R. (2001) C-H... π interactions in proteins. *J. Mol. Biol.*, **307**, 357–377.
 27. Thomas, A., Meurisse, R., Charlotteaux, B. and Brasseur, R. (2002) Aromatic side-chain interactions in proteins. I. Main structural features. *Proteins: Struct. Funct. Bioinform.*, **48**, 628–634.
 28. Thomas, A., Meurisse, R. and Brasseur, R. (2002) Aromatic side-chain interactions in proteins. II. Near- and far-sequence phe-X pairs. *Proteins: Struct. Funct. Bioinform.*, **48**, 635–644.
 29. Meurisse, R., Brasseur, R. and Thomas, A. (2003) Aromatic side-chain interactions in proteins. Near- and far-sequence his-X pairs. *Biochim. Biophys. Acta*, **1649**, 85–96.
 30. Chourasia, M., Sastry, G.M. and Sastry, G.N. (2011) Aromatic-aromatic interactions database, A2ID: an analysis of aromatic π -networks in proteins. *Int. J. Biol. Macromol.*, **48**, 540–552.
 31. Jenkins, D.D., Harris, J.B., Howell, E.E., Hinde, R.J. and Baudry, J. (2013) STAAR: statistical analysis of aromatic rings. *J. Comput. Chem.*, **34**, 518–522.
 32. Yan, C., Wu, F., Jernigan, R.L., Dobbs, D. and Honavar, V. (2008) Characterization of protein-protein interfaces. *The Protein Journal*, **27**, 59–70.
 33. Ellis, J.J., Broom, M. and Jones, S. (2007) Protein-RNA interactions: structural analysis and functional classes. *Proteins: Struct. Funct. Bioinform.*, **66**, 903–911.
 34. Mao, L., Wang, Y., Liu, Y. and Hu, X. (2004) Molecular determinants for ATP-binding in proteins: a data mining and quantum chemical analysis. *J. Mol. Biol.*, **336**, 787–807.
 35. Řezáč, J., Riley, K.E. and Hobza, P. (2012) Evaluation of the performance of post-hartree-fock methods in terms of intermolecular distance in noncovalent complexes. *J. Comput. Chem.*, **33**, 691–694.
 36. Sherrill, C.D. (2012) Energy component analysis of π interactions. *Acc. Chem. Res.*, **46**, 1020–1028.
 37. Riley, K.E. and Hobza, P. (2011) Noncovalent interactions in biochemistry. *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, **1**, 3–17.
 38. Grimme, S. and Steinmetz, M. (2013) Effects of London dispersion correction in density functional theory on the structures of organic molecules in the gas phase. *Phys. Chem. Chem. Phys.*, **15**, 16031–16042.
 39. Johnson, E.R., Otero-de-la-Roza, A., Dale, S.G. and DiLabio, G.A. (2013) Efficient basis sets for non-covalent interactions in XDM-corrected density-functional theory. *J. Chem. Phys.*, **139**, 214109.
 40. Copeland, K.L., Anderson, J.A., Farley, A.R., Cox, J.R. and Tschumper, G.S. (2008) Probing phenylalanine/adenine π -stacking interactions in protein complexes with explicitly correlated and CCSD(T) computations. *J. Phys. Chem. B*, **112**, 14291–14295.
 41. Cauet, E., Rومان, M., Wintjens, R., Lievin, J. and Biot, C. (2005) Histidine-aromatic interactions in proteins and protein-ligand complexes: quantum chemical study of X-ray and model structures. *J. Chem. Theory Comput.*, **1**, 472–483.
 42. Copeland, K.L., Pellock, S.J., Cox, J.R., Cafiero, M.L. and Tschumper, G.S. (2013) Examination of tyrosine/adenine stacking interactions in protein complexes. *J. Phys. Chem. B*, **117**, 14001–14008.
 43. Cysewski, P. (2008) A Post-SCF complete basis set study on the recognition patterns of uracil and cytosine by aromatic and π -aromatic stacking interactions with amino acid residues. *Phys. Chem. Chem. Phys.*, **10**, 2636–2645.
 44. Riley, K.E., Pitoňák, M., Černý, J.I. and Hobza, P. (2010) On the structure and geometry of biomolecular binding motifs (hydrogen-bonding, stacking, X-H... π): WFT and DFT calculations. *J. Chem. Theory Comput.*, **6**, 66–80.
 45. Rutledge, L.R., Campbell-Verduyn, L.S. and Wetmore, S.D. (2007) Characterization of the stacking interactions between DNA or RNA nucleobases and the aromatic amino acids. *Chem. Phys. Lett.*, **444**, 167–175.
 46. Rutledge, L.R., Durst, H.F. and Wetmore, S.D. (2009) Evidence for stabilization of DNA/RNA-protein complexes arising from nucleobase-amino acid stacking and T-shaped interactions. *J. Chem. Theory Comput.*, **5**, 1400–1410.
 47. Rutledge, L.R., Campbell-Verduyn, L.S., Hunter, K.C. and Wetmore, S.D. (2006) Characterization of nucleobase-amino acid stacking interactions utilized by a DNA repair enzyme. *J. Phys. Chem. B*, **110**, 19652–19663.
 48. Wells, R.A., Kellie, J.L. and Wetmore, S.D. (2013) Significant strength of charged DNA-protein π - π interactions: a preliminary study of cytosine. *J. Phys. Chem. B*, **117**, 10462–10474.
 49. Churchill, C.D.M. and Wetmore, S.D. (2009) Noncovalent interactions involving histidine: the effect of charge on π - π stacking and T-shaped interactions with the DNA nucleobases. *J. Phys. Chem. B*, **113**, 16046–16058.
 50. Rutledge, L.R. and Wetmore, S.D. (2010) The Assessment of density functionals for DNA-protein stacked and T-shaped complexes. *Can. J. Chem.*, **88**, 815–830.
 51. Rutledge, L.R., Durst, H.F. and Wetmore, S.D. (2008) Computational comparison of the stacking interactions between the aromatic amino acids and the natural or (Cationic) methylated nucleobases. *Phys. Chem. Chem. Phys.*, **10**, 2801–2812.
 52. Rutledge, L.R. and Wetmore, S.D. (2008) Remarkably strong T-shaped interactions between aromatic amino acids and adenine: their increase upon nucleobase methylation and a comparison to stacking. *J. Chem. Theory Comput.*, **4**, 1768–1780.
 53. Leavens, F.M.V., Churchill, C.D.M., Wang, S. and Wetmore, S.D. (2011) Evaluating how discrete water molecules affect protein-DNA π - π and π + π stacking and T-shaped interactions: the case of histidine-adenine dimers. *J. Phys. Chem. B*, **115**, 10990–11003.
 54. Churchill, C.D.M., Navarro-Whyte, L., Rutledge, L.R. and Wetmore, S.D. (2009) Effects of the biological backbone on DNA-protein stacking interactions. *Phys. Chem. Chem. Phys.*, **11**, 10657–10670.
 55. Churchill, C.D.M., Rutledge, L.R. and Wetmore, S.D. (2010) Effects of the biological backbone on stacking interactions at DNA-protein interfaces: the interplay between the backbone- π and π - π components. *Phys. Chem. Chem. Phys.*, **12**, 14515–14526.

56. Rutledge, L.R., Navarro-Whyte, L., Peterson, T.L. and Wetmore, S.D. (2011) Effects of extending the computational model on DNA-protein T-shaped interactions: the case of adenine-histidine dimers. *J. Phys. Chem. A*, **115**, 12646–12658.
57. Rutledge, L.R., Churchill, C.D.M. and Wetmore, S.D. (2010) A preliminary investigation of the additivity of π - π or π + $-\pi$ stacking and T-shaped interactions between natural or damaged DNA nucleobases and histidine. *J. Phys. Chem. B*, **114**, 3355–3367.
58. Spiwok, V., Lipovová, P., Skálová, T., Buchtelová, E., Hašek, J. and Králová, B. (2004) Role of CH/ π interactions in substrate binding by *Escherichia coli* β -galactosidase. *Carbohydr. Res.*, **339**, 2275–2280.
59. Spiwok, V., Lipovová, P., Skálová, T., Vondráčková, E., Dohnálek, J., Hašek, J. and Králová, B. (2006) Modelling of carbohydrate-aromatic interactions: ab initio energetics and force field performance. *J. Comput. Aided Mol. Des.*, **19**, 887–901.
60. Sujatha, M.S., Sasidhar, Y.U. and Balaji, P.V. (2004) Energetics of galactose- and glucose-aromatic amino acid interactions: implications for binding in galactose-specific proteins. *Protein Science*, **13**, 2502–2514.
61. Wimmerová, M., Kozmon, S., Nečasová, I., Mishra, S.K., Komárek, J. and Koča, J. (2012) Stacking interactions between carbohydrate and protein quantified by combination of theoretical and experimental methods. *PLoS ONE*, **7**, e46032.
62. Asensio, J.L., Ardá, A., Cañada, F.J. and Jiménez-Barbero, J. (2012) Carbohydrate-aromatic interactions. *Acc. Chem. Res.*, **46**, 946–954.
63. Tatko, C. (2008) Sugars stack up. *Nat. Chem. Biol.*, **4**, 586–587.
64. Laughrey, Z.R., Kiehn, S.E., Riemen, A.J. and Waters, M.L. (2008) Carbohydrate- π interactions: what are they worth? *J. Am. Chem. Soc.*, **130**, 14625–14633.
65. Lee, Y.C. and Lee, R.T. (1995) Carbohydrate-protein interactions: basis of glycobiology. *Acc. Chem. Res.*, **28**, 321–327.
66. Stanca-Kaposta, E.C., Çarçabal, P., Cocinero, E.J., Hurtado, P. and Simons, J.P. (2013) Carbohydrate-aromatic interactions: vibrational spectroscopy and structural assignment of isolated monosaccharide complexes with p-hydroxy toluene and N-acetyl l-tyrosine methylamide. *J. Phys. Chem. B*, **117**, 8135–8142.
67. Kumari, M., Sunoj, R.B. and Balaji, P.V. (2012) Exploration of CH... π mediated stacking interactions in saccharide: aromatic residue complexes through conformational sampling. *Carbohydr. Res.*, **361**, 133–140.
68. V. Balaji, P. (2011) Contribution of C-H... π ; interactions to the affinity and specificity of carbohydrate binding sites. *Mini-Rev. Org. Chem.*, **8**, 222–228.
69. Cao, M., Fu, A., Wang, Z., Liu, J., Kong, N., Zong, X., Liu, H. and Gooding, J.J. (2014) Electrochemical and theoretical study of π - π stacking interactions between graphitic surfaces and pyrene derivatives. *J. Phys. Chem. C*, **118**, 2650–2659.
70. Cho, Y., Min, S.K., Yun, J., Kim, W.Y., Tkatchenko, A. and Kim, K.S. (2013) Noncovalent interactions of DNA bases with naphthalene and graphene. *J. Chem. Theory Comput.*, **9**, 2090–2096.
71. Gazit, E. (2007) Self-assembled peptide nanostructures: the design of molecular building blocks and their technological utilization. *Chem. Soc. Rev.*, **36**, 1263–1269.
72. Kumar, R.M., Elango, M. and Subramanian, V. (2010) Carbohydrate-aromatic interactions: the role of curvature on XH... π interactions. *J. Phys. Chem. A*, **114**, 4313–4324.
73. Sharma, R., McNamara, J.P., Raju, R.K., Vincent, M.A., Hillier, I.H. and Morgado, C.A. (2008) The interaction of carbohydrates and amino acids with aromatic systems studied by density functional and semi-empirical molecular orbital calculations with dispersion corrections. *Phys. Chem. Chem. Phys.*, **10**, 2767–2774.
74. Cozens, C., Pinheiro, V.B., Vaisman, A., Woodgate, R. and Holliger, P. (2012) A short adaptive path from DNA to RNA polymerases. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 8067–8072.
75. Tsuzuki, S., Uchimaru, T. and Mikami, M. (2009) Magnitude and nature of carbohydrate-aromatic interactions: ab initio calculations of fucose-benzene complex. *J. Phys. Chem. B*, **113**, 5617–5621.
76. Kozmon, S., Matuška, R., Spiwok, V. and Koča, J. (2011) Three-dimensional potential energy surface of selected carbohydrates' CH/ π dispersion interactions calculated by high-level quantum mechanical methods. *Chem. Eur. J.*, **17**, 5680–5690.
77. Ramirez-Gualito, K., Alonso-Rios, R., Quiroz-Garcia, B., Rojas-Aguilar, A., Diaz, D., Jimenez-Barbero, J. and Cuevas, G. (2009) enthalpic nature of the CH/ π interaction involved in the recognition of carbohydrates by aromatic compounds, confirmed by a novel interplay of NMR, calorimetry, and theoretical calculations. *J. Am. Chem. Soc.*, **131**, 18129–18138.
78. Raju, R.K., Hillier, I.H., Burton, N.A., Vincent, M.A., Doudou, S. and Bryce, R.A. (2010) The effects of perfluorination on carbohydrate- π interactions: computational studies of the interaction of benzene and hexafluorobenzene with fucose and cyclodextrin. *Phys. Chem. Chem. Phys.*, **12**, 7959–7967.
79. Bautista-Ibanez, L., Ramirez-Gualito, K., Quiroz-Garcia, B., Rojas-Aguilar, A. and Cuevas, G. (2008) Calorimetric measurement of the CH/ π interaction involved in the molecular recognition of saccharides by aromatic compounds. *J. Org. Chem.*, **73**, 849–857.
80. Raju, R.K., Ramraj, A., Vincent, M.A., Hillier, I.H. and Burton, N.A. (2008) Carbohydrate-protein recognition probed by density functional theory and Ab initio calculations including dispersive interactions. *Phys. Chem. Chem. Phys.*, **10**, 6500–6508.
81. Su, Z., Cocinero, E.J., Stanca-Kaposta, E.C., Davis, B.G. and Simons, J.P. (2009) Carbohydrate-aromatic interactions: a computational and IR spectroscopic investigation of the complex, methyl alpha-L-fucopyranoside-toluene, isolated in the gas phase. *Chem. Phys. Lett.*, **471**, 17–21.
82. Stanca-Kaposta, E.C., Gamblin, D.P., Screen, J., Liu, B., Snoek, L.C., Davis, B.G. and Simons, J.P. (2007) Carbohydrate molecular recognition: a spectroscopic investigation of carbohydrate-aromatic interactions. *Phys. Chem. Chem. Phys.*, **9**, 4444–4451.
83. Tsuzuki, S., Uchimaru, T. and Mikami, M. (2011) Magnitude and nature of carbohydrate-aromatic interactions in fucose-phenol and fucose-indole complexes: CCSD(T) level interaction energy calculations. *J. Phys. Chem. A*, **115**, 11256–11262.
84. Mohamed, M.N.A., Watts, H.D., Guo, J., Catchmark, J.M. and Kubicki, J.D. (2010) MP2, density functional theory, and molecular mechanical calculations of C-H... π and hydrogen bond interactions in a cellulose-binding Module-Cellulose Model System. *Carbohydr. Res.*, **345**, 1741–1751.
85. Sujatha, M.S., Sasidhar, Y.U. and Balaji, P.V. (2005) Insights into the role of the aromatic residue in galactose-binding sites: MP2/6-311G+++ study on galactose- and glucose-aromatic residue analogue complexes. *Biochemistry*, **44**, 8554–8562.
86. Kozmon, S., Matuška, R., Spiwok, V. and Koča, J. (2011) Dispersion interactions of carbohydrates with condensate aromatic moieties: theoretical study on the CH- π interaction additive properties. *Phys. Chem. Chem. Phys.*, **13**, 14215–14222.
87. Lucas, R., Gomez-Pinto, I., Avino, A., Reina, J.J., Eritja, R., Gonzalez, C. and Morales, J.C. (2011) Highly polar carbohydrates stack onto DNA duplexes via CH/ π interactions. *J. Am. Chem. Soc.*, **133**, 1909–1916.
88. Maresca, M., Derghal, A., Carravagna, C., Dudin, S. and Fantini, J. (2008) Controlled aggregation of adenine by sugars: physicochemical studies, molecular modelling simulations of sugar-aromatic CH- π stacking interactions, and biological significance. *Phys. Chem. Chem. Phys.*, **10**, 2792–2800.
89. Gómez-Pinto, I., Vengut-Climent, E., Lucas, R., Aviñó, A., Eritja, R., González, C. and Morales, J.C. (2013) Carbohydrate-DNA interactions at G-quadruplexes: folding and stability changes by attaching sugars at the 5'-End. *Chem. Eur. J.*, **19**, 1920–1927.
90. Lucas, R., Vengut-Climent, E., Gomez-Pinto, I., Avino, A., Eritja, R., Gonzalez, C. and Morales, J.C. (2012) Apolar carbohydrates as DNA capping agents. *Chem. Commun.*, **48**, 2991–2993.
91. Schrodinger, L. (2010) The PyMOL molecular graphics system, Version 1.3r1.
92. Macrae, C.F., Bruno, I.J., Chisholm, J.A., Edgington, P.R., McCabe, P., Pidcock, E., Rodriguez-Monge, L., Taylor, R., van de Streek, J. and Wood, P.A. (2008) Mercury CSD 2.0 - new features for the visualization and investigation of crystal structures. *J. Appl. Crystallogr.*, **41**, 466–470.
93. Edgecomb, S.P. and Murphy, K.P. (2002) Variability in the pKa of histidine side-chains correlates with burial within proteins. *Proteins: Struct. Funct. Bioinform.*, **49**, 1–6.
94. HyperChem(TM). Professional 7.5 ed. Hypercube, Inc., Gainesville, Florida.
95. Cerny, J. and Hobza, P. (2007) Non-covalent interactions in biomacromolecules. *Phys. Chem. Chem. Phys.*, **9**, 5291–5303.

96. Raju, R.K., Ramraj, A., Hillier, I.H., Vincent, M.A. and Burton, N.A. (2009) Carbohydrate-aromatic π interactions: a test of density functionals and the DFT-D method. *Phys. Chem. Chem. Phys.*, **11**, 3411–3416.
97. Hujo, W. and Grimme, S. (2012) Performance of non-local and atom-pairwise dispersion corrections to DFT for structural parameters of molecules with noncovalent interactions. *J. Chem. Theory Comput.*, **9**, 308–315.
98. Burns, L.A., Vazquez-Mayagoitia, A., Sumpter, B.G. and Sherrill, C.D. (2011) Density-functional approaches to noncovalent interactions: a comparison of dispersion corrections (DFT-D), exchange-hole dipole moment (XDM) theory, and specialized functionals. *J. Chem. Phys.*, **134**, 084107.
99. Halkier, A., Helgaker, T., Jorgensen, P., Klopper, W., Koch, H., Olsen, J. and Wilson, A.K. (1998) Basis-set convergence in correlated calculations on Ne, N₂, and H₂O. *Chem. Phys. Lett.*, **286**, 243–252.
100. Halkier, A., Helgaker, T., Jorgensen, P., Klopper, W. and Olsen, J. (1999) Basis-set convergence of the energy in molecular hartree-fock calculations. *Chem. Phys. Lett.*, **302**, 437–446.
101. Jurecka, P., Spöner, J. and Hobza, P. (2004) Potential energy surface of the cytosine dimer: MP2 complete basis set limit interaction energies, CCSD(T) correction term, and comparison with the AMBER force field. *J. Phys. Chem. B*, **108**, 5466–5471.
102. Spöner, J., Jurecka, P., Marchan, I., Luque, F.J., Orozco, M. and Hobza, P. (2006) Nature of base stacking: reference quantum-chemical stacking energies in ten unique B-DNA base-pair steps. *Chem.-Eur. J.*, **12**, 2854–2865.
103. Spöner, J., Riley, K.E. and Hobza, P. (2008) Nature and magnitude of aromatic stacking of nucleic acid bases. *Phys. Chem. Chem. Phys.*, **10**, 2595–2610.
104. Morgado, C.A., Jurecka, P., Svozil, D., Hobza, P. and Spöner, J. (2010) Reference MP2/CBS and CCSD(T) quantum-chemical calculations on stacked adenine dimers. Comparison with DFT-D, MP2.5, SCS(MI)-MP2, M06–2X, CBS(SCS-D) and force field descriptions. *Phys. Chem. Chem. Phys.*, **12**, 3522–3534.
105. Mládek, A., Banáš, P., Jurečka, P., Otyepka, M., Zgarbová, M. and Špöner, J. (2013) Energies and 2'-Hydroxyl group orientations of RNA backbone conformations. Benchmark CCSD(T)/CBS database, electronic analysis, and assessment of DFT methods and MD simulations. *J. Chem. Theory Comput.*, **10**, 463–480.
106. Mladek, A., Krepl, M., Svozil, D., Cech, P., Otyepka, M., Banas, P., Zgarbova, M., Jurecka, P. and Spöner, J. (2013) Benchmark quantum-chemical calculations on a complete set of rotameric families of the DNA sugar-phosphate backbone and their comparison with modern density functional theory. *Phys. Chem. Chem. Phys.*, **15**, 7295–7310.
107. Jurecka, P., Spöner, J., Cerny, J. and Hobza, P. (2006) Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs. *Phys. Chem. Chem. Phys.*, **8**, 1985–1993.
108. Riley, K.E., Platts, J.A., Řezáč, J., Hobza, P. and Hill, J.G. (2012) Assessment of the performance of MP2 and MP2 variants for the treatment of noncovalent interactions. *J. Phys. Chem. A*, **116**, 4159–4169.
109. Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G.A. et al. (2009). *Revision A.02 ed. Gaussian, Inc.*, Wallingford CT.
110. Shao, Y., Molnar, L.F., Jung, Y., Kussmann, J., Ochsenfeld, C., Brown, S.T., Gilbert, A.T.B., Slipchenko, L.V., Levchenko, S.V., O'Neill, D.P. et al. (2006) Advances in methods and algorithms in a modern quantum chemistry program package. *Phys. Chem. Chem. Phys.*, **8**, 3172–3191.
111. Singh, N.J., Min, S.K., Kim, D.Y. and Kim, K.S. (2009) Comprehensive energy analysis for various types of π -interaction. *J. Chem. Theory Comput.*, **5**, 515–529.
112. Rutledge, L.R. and Wetmore, S.D. (2011) Modeling the chemical step utilized by human alkyladenine DNA glycosylase: a concerted mechanism AIDS in selectively excising damaged purines. *J. Am. Chem. Soc.*, **133**, 16258–16269.
113. Kellie, J.L., Navarro-Whyte, L., Carvey, M.T. and Wetmore, S.D. (2012) Combined effects of π - π stacking and hydrogen bonding on the (N1) acidity of uracil and hydrolysis of 2'-deoxyuridine. *J. Phys. Chem. B*, **116**, 2622–2632.
114. Przybylski, J.L. and Wetmore, S.D. (2011) A QM/QM investigation of the hUNG2 reaction surface: the untold tale of a catalytic residue. *Biochemistry*, **50**, 4218–4227.
115. Calvaresi, M., Bottoni, A. and Garavelli, M. (2007) Computational clues for a new mechanism in the glycosylase activity of the human DNA repair protein hOGG1. A generalized paradigm for purine-repairing systems? *J. Phys. Chem. B*, **111**, 6557–6570.
116. Bruner, S.D., Norman, D.P.G. and Verdine, G.L. (2000) Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA. *Nature*, **403**, 859–866.
117. Goedecke, K., Pignot, M., Goody, R.S., Scheidig, A.J. and Weinhold, E. (2001) Structure of the N6-adenine DNA methyltransferase MtaqI in complex with DNA and a cofactor analog. *Nat. Struct. Mol. Biol.*, **8**, 121–125.
118. Wang, W., Wu, E.Y., Helling, H.W. and Beese, L.S. (2012) Structural factors that determine selectivity of a high fidelity DNA polymerase for deoxy-, dideoxy-, and ribonucleotides. *J. Biol. Chem.*, **287**, 28215–28226.
119. Brown, J.A. and Suo, Z. (2011) Unlocking the sugar “steric gate” of DNA polymerases. *Biochemistry*, **50**, 1135–1142.
120. DeLucia, A.M., Grindley, N.D.F. and Joyce, C.M. (2003) An error-prone family Y DNA polymerase (DinB Homolog from *Sulfolobus solfataricus*) uses a ‘steric gate’ residue for discrimination against ribonucleotides. *Nucleic Acids Res.*, **31**, 4129–4137.
121. Kirouac, K.N., Suo, Z. and Ling, H. (2011) Structural mechanism of ribonucleotide discrimination by a Y-family DNA polymerase. *J. Mol. Biol.*, **407**, 382–390.
122. Yang, G., Franklin, M., Li, J., Lin, T.C. and Konigsberg, W. (2002) A conserved Tyr residue is required for sugar selectivity in a Pol α DNA polymerase. *Biochemistry*, **41**, 10256–10261.
123. Astatke, M., Ng, K., Grindley, N.D. and Joyce, C.M. (1998) A single side chain prevents *Escherichia coli* DNA polymerase I (klenow fragment) from incorporating ribonucleotides. *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 3402–3407.
124. Bonnin, A., Lázaro, J.M., Blanco, L. and Salas, M. (1999) A single tyrosine prevents insertion of ribonucleotides in the eukaryotic-type ϕ 29 DNA polymerase. *J. Mol. Biol.*, **290**, 241–251.
125. Gao, G., Orlova, M., Georgiadis, M.M., Hendrickson, W.A. and Goff, S.P. (1997) Conferring RNA polymerase activity to a DNA polymerase: a single residue in reverse transcriptase controls substrate selection. *Proc. Natl. Acad. Sci. U.S.A.*, **94**, 407–411.
126. Lindahl, T. and Wood, R.D. (1999) Quality control by DNA repair. *Science*, **286**, 1897–1905.
127. Sale, J.E., Lehmann, A.R. and Woodgate, R. (2012) Y-family DNA polymerases and their role in tolerance of cellular DNA damage. *Nat. Rev. Mol. Cell. Biol.*, **13**, 141–152.
128. Lorigian, A., Marsden, H.S. and Palù, G. (2002) Protein-protein interactions as targets for antiviral chemotherapy. *Rev. Med. Virol.*, **12**, 239–262.
129. Park, S.-H. and Raines, R.T. (2000) Genetic selection for dissociative inhibitors of designated protein-protein interactions. *Nat. Biotech.*, **18**, 847–851.
130. De Clercq, E. (2002) Strategies in the design of antiviral drugs. *Nat. Rev. Drug Discov.*, **1**, 13–25.