

Article

Not peer-reviewed version

Quantum Mechanics Characterization of Non-Covalent Interaction in Nucleotides Fragments

[Mayar Tarek Ibrahim](#) , Elizabeth Wait , [Pengyu Ren](#) *

Posted Date: 6 June 2024

doi: 10.20944/preprints202406.0349.v1

Keywords: Nucleotides; ALMO EDA2; SAPT2+(3)/dmp2; Dispersion; Induction



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Quantum Mechanics Characterization of Non-Covalent Interaction in Nucleotides Fragments

Mayar Tarek Ibrahim ¹, Elizabeth Wait ² and Pengyu Ren ^{1,2,*}

¹ Department of Biomedical Engineering, The University of Texas at Austin, Austin, TX 78712, USA

² Interdisciplinary Life Sciences Graduate Program, The University of Texas at Austin, Austin, TX 78712, USA

* Correspondence: pren@utexas.edu

Abstract: Accurate calculation of non-covalent interaction energies in nucleotides is crucial for understanding the driving forces governing nucleic acid structures and functions, as well as developing advanced molecular mechanics forcefields or machine learning potentials tailored to nucleic acids. Here, we dissect the nucleotides' structure into three main constituents: nucleobases (A, G, C, T, and U), sugar moieties (ribose and deoxyribose), and phosphate group. The interactions among these fragments and between fragment and water were analyzed. Different quantum mechanical methods were compared for their accuracy in capturing the interaction energy. The non-covalent interaction energy was decomposed into electrostatics, exchange-repulsion, dispersion, and induction using two ab-initio methods: Symmetry-Adapted Perturbation Theory (SAPT) and Absolutely Localized Molecular Orbitals (ALMO). These calculations provide the benchmark for different QM methods while providing valuable understanding of the roles of various intermolecular forces in hydrogen bonding and aromatic stacking.

Keywords: nucleotides; ALMO EDA2; SAPT2+(3)/dmp2; dispersion; induction

1. Introduction

Nucleotides are the main building blocks of nucleic acid, either DNA or RNA. Nucleotides are composed of neutral sugar moiety (deoxyribose in DNA or ribose in RNA), negatively charged phosphate group, and nucleobase. Nucleobases are either purine (Adenine (A) or Guanine (G) or pyrimidine (Cytosine (C), Thymine (T), or Uracil (U)) [1]. Nucleic acids are not only essential for the storage of genetic material, transcription, and translation but also for various cellular processes as ligand binding, enzymatic function, regulation of different metabolic activities, genomic instability, apoptosis, and stress response [2–5].

The most common configuration of DNA is the B-DNA duplex [6,7]. Transition from B-DNA geometry to A-DNA geometry is observed upon adding solvents or interacting with other proteins or drug molecules [8–10]. Z-DNA form was detected in crystal structures and environments with high-salt concentration [11–13]. A-form helix is the dominant configuration for RNA as well as hairpins loops and tertiary structures depending on the sequence and the number of strands [14]. Depending on the B-DNA or A-DNA geometry, the sugar moiety adopts either C2' endo conformation or C3' endo conformation, respectively [15]. On the other hand, RNAs typically exist in A form. These complex and different configurations of nucleic acids are maintained and stabilized by the non-covalent interactions among the three main components of the nucleotides. Two main types of non-covalent interactions are present in the nucleic acid hydrogen bond and stacking interactions.

Due to the complex nature of DNA and RNA molecules with the dominance of non-covalent interactions, molecular simulations using the current molecular mechanics (MM) forcefields have been challenging. Current nucleic acid forcefields treat the non-covalent interactions as a sum of electrostatics interactions between fixed atomic charges and van-der-Waals interactions accounted

for by Lennard-Jones potentials [16]. AMOEBA (atomic multipole optimized energetics for biomolecular applications) employs polarizable atomic multipoles to improve the representation of electrostatics [17–19]. Drude NA force fields account for polarization in NAs based on classical Drude oscillators [20–22]. These efforts significantly improved our ability to describe many-body polarizations. Nonetheless, challenges remain. For example, in the case of stacked aromatic molecules, any point charge or multipole models are unable to capture electrostatic interactions between the atomic rings [23]. This is attributed to the lack of charge penetration effect needed to account for the π -stacking interactions due to the overlapping of the π -electron clouds and atomic nuclei [24–26]. The improvement in the accuracy of atomic potentials and force field parameters depends on the quality of the underlying QM benchmarks used.

Here we applied five different high-quality QM methods to the non-covalent interactions of nucleic acid based on Symmetry-Adapted Perturbation Theory (SAPT) and Absolutely Localized Molecular Orbitals (ALMO): SAPT0 with jun-cc-pVDZ, SAPT2+ with aug-cc-pVDZ, SAPT2+(3)/dmp2 with aug-cc-pVDZ, SAPT2+(3)/dmp2 with aug-cc-pVTZ, and ALMO EDA2 (original energy decomposition) with B97M-V and aug-cc-pVTZ. Nucleotides were divided into three main components based on the difference in the chemical structure and charge (Figure 1). First, the three most probable configurations of the negatively charged dimethyl phosphate: DMP gg, DMP gt, and DMP tt. Second, the neutral C 2'endo deoxyribose sugar moiety, the most common in B-DNA structure, and the neutral C3'endo ribose sugar moiety, the most common structure in the RNA and the A-DNA. The main non-covalent interaction that governs the bonding of the last two moieties is the hydrogen bond. The third component of nucleotides are the nucleobases, which interact with both hydrogen bonds and stacking interactions. The interaction of the three main components of the nucleotides with the water molecule over a wide range of locations and distances was studied to probe the hydrogen bond interaction aspect of the nucleotides. Besides the interaction with the water molecule, the stacked nucleobases (intrastrand) and the hydrogen-bonding nucleobases (interstrand) were examined. ALMO EDA2 was found to be consistent among selected DFT methods and independent of the basis sets used. On the other hand, the three levels of SAPT investigated (SAPT0, SAPT2+, and SAPT2+(3)/dmp2) showed high dependence on the basis set used. Upon studying the interactions of the intrastrand stacked nucleobases, interstrand nucleobases, and the three main fragments with water and nucleobases, the two of the best ab-initio methods for the nucleotides are ALMO EDA2 (B97M-V) with aug-cc-pVTZ basis set with original energy decomposition scheme and SAPT2+(3)/dmp2 with aug-cc-pVTZ. By examining the energy components of the interactions among the various constituents of nucleotides, it was found that exchange-repulsion dominates within the neutral sugar moiety. Electrostatic attraction is also strong, largely offsetting the repulsion. Similar trends were observed for neutral nucleobases' interaction with water. As expected, electrostatics were more significant in the charged DMP interaction with water molecules. In all these cases, there is a large cancellation between the electrostatic attraction and exchange-repulsion, and as a result the attractive induction and dispersion are responsible for stabilizing the molecular complexes with water, highlighting the significance of explicit consideration of induction effect. For interstrand or intrastrand nucleobase dimers, however, dispersion was found to be the most important while electrostatics is slightly attractive, and induction is insignificant.

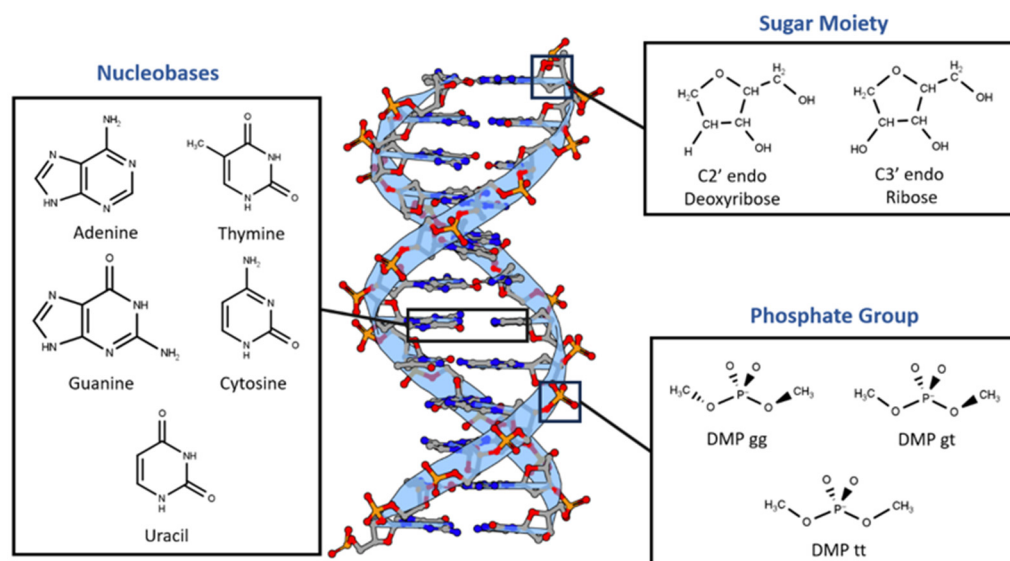


Figure 1. Nucleotides dissected into three main constituents based on the difference in the chemical nature: Dimethyl phosphate, Sugar Moiety, and Nucleobases.

2. Results

2.1. Benchmarking of Small Molecules

2.1.1. Various Theoretical Methods

Nucleotides mainly compromise two types of interactions: stacking interactions between the nucleobases and hydrogen bond interactions involving nucleobases, sugar moiety, phosphate group and water solvent. To determine the best theoretical method to study the interaction energies in such a complex system, benchmarking was done on 6 hydrogen-bond forming compounds, Ethylamine-water (CCN-Wat), Methanethiol-Water (CS-Wat), Ethylamine-Methanethiol (CCN-CS), Methanol-Water (CO-Wat), Methanol-Ethanethiol (CCS-CO), and Chlorobenzene-Ethylamine (BenCl-CCN), and 6 stacked molecular dimers, Pyrrole-Pyrrole, Benzene-Imidazole, Benzene-Benzene, Fluorobenzene-Fluorobenzene, Fluorobenzene-Imidazole, Chlorobenzene-Chlorobenzene. For these dimers, interactions energy at the MP2 (Møller-Plesset perturbation), CCSD(T) Coupled Cluster Theory through perturbative triples and Complete Basis Set CBS CCSD(T) were previously reported in the DE Shaw database [27]. Here we compared these methods with 1) DFT based Absolutely Localized Molecular Orbitals Energy Decomposition Analysis, ALMO EDA2, with the original and the new energy decomposition schemes [28–31] and 2) different level of theory of Symmetry-Adapted Perturbation Theory (SAPT): SAPT0, SAPT2+, and SAPT2+(3)/dmp2 [32–34]. The aug-cc-pVTZ basis set was used for both ALMO and SAPT calculations. These methods also offer value interaction energy components that allows us better understanding the driving force for intra- and intermolecular complexes and developing accurate atomic potentials for molecular simulations. The most accurate method here, CBS CCSD(T), is used as the benchmark. The interaction energies of the hydrogen-bond forming compounds from the different methods were comparably similar with most deviation observed in SAPT0 and MP2 methods (Figure 2a). MP2 method was found to underestimate the interaction energies compared to CBS CCSD(T), with most difference of 0.7 kcal/mol observed in the interaction energy of CCN-CS dimer. SAPT0 overestimated the interaction energies with the biggest deviation from CBS CCSD(T), -1.0 kcal/mol, in BenCl-CCN. SAPT2+ performed well for most of the hydrogen-bond forming dimers except for CCN-CS, where the difference from CBS CCSD(T) was -0.57 kcal/mol. In summary, all the tested methods showed relatively small deviation compared to CBS CCSD(T). The best performance was given by SAPT2+(3)/dmp2 and ALMO EDA2 with both old and new energy decomposition scheme – the maximum deviation from CBS CCSD(T) is 0.1 kcal/mol (**Figure 2a**).

The interaction energies obtained from the various theoretical methods were found to vary much more significantly for the stacked complexes (**Figure 2b**) compared to the hydrogen-bond forming complexes. CCSD(T), MP2, SAPT0, and SAPT2+ overestimated the interaction energies compared to CBS CCSD(T) with most deviation observed for SAPT0. The interaction energy of BenCl-BenCl dimer obtained from the various methods was found to be associated with the maximum difference compared to CBS CCSD(T) interaction energies. For example, SAPT0 interaction energy deviated from the CBS CCSD(T) value by 2.54 kcal/mol and for MP2 method the deviation was 1.77 kcal/mol. BenF-BenF and BenF-Imidazole interaction energies also showed large variations among different methods. This showcases the limitation of these methods with the halogen substituted aromatic systems in general. The best theoretical method with the least deviation from CBS CCSD(T) for the stacked molecules was found to be SAPT2+(3)/dmp2, with maximum of 0.2 kcal/mol observed in BenCl-BenCl, followed by ALMO EDA2 with the original energy decomposition with maximum deviation of 0.36 kcal/mol observed in BenCl-BenCl, and then ALMO EDA2 with the new energy decomposition with maximum difference of 0.54 kcal/mol observed in BenF-BenF (**Figure 2b**).

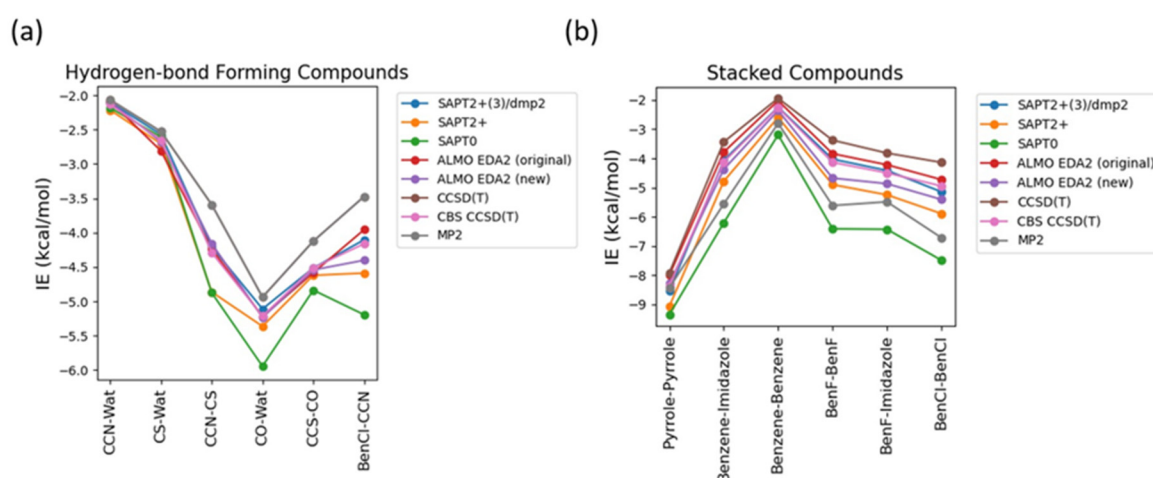


Figure 2. Benchmarking of the interaction energies of different small molecules using various theoretical methods, mainly different levels of SAPT (SAPT0, SAPT2+, and SAPT2+(3)/dmp2) and ALMO EDA2 (B97M-V) with original and new energy decomposition against CBS CCSD(T). All the calculations were performed using aug-cc-pVTZ basis set. (a) Interaction energies of hydrogen-bond forming compounds as obtained from different methods; (b) Interaction energies of stacked compounds.

2.1.2. Comparison of DFT Methods

As some of the tested theoretical methods were inconsistent in the case of the stacked complexes compared to the hydrogen-bond forming complexes, further analysis was performed on the stacked compounds using different DFT levels of theory. Four different DFT methods B97M-V, ω B97M-V, ω B97X-V, and ω B97X-D3 were tested and contrasted with CBS CCSD(T) and DLPNO/CCSD(T) an approximate CBS CCSD(T) method. This additional comparison is also performed due to the superior performance of both the original and the new energy decomposition schemes of ALMO EDA2, based on DFT calculation, for the hydrogen-bonding and stacked complexes. The interaction energies obtained from the four tested DFT level of theory were found to be very similar to each other and to CBS CCSD(T) interaction energies for the stacked molecule (**Figure 3**). B97M-V showed the most deviation from the CBS CCSD(T) method, and yet the maximum error (pyrrole-pyrrole) is only 0.36 kcal/mol. Interestingly, pyrrole-pyrrole was the dimer showing the largest variation among the different DFT methods, except for ω B97X-D3 where the largest error was associated with BenCl-BenCl molecule. ω B97X-D3 method was found to be the most consistent with CBS CCSD(T) with maximum difference of 0.21 kcal/mol among the test sets. Overall, the results indicate that all these DFT methods tested here produce satisfying interaction energies that closely match the CBS CCSD(T) results (**Figure 3**).

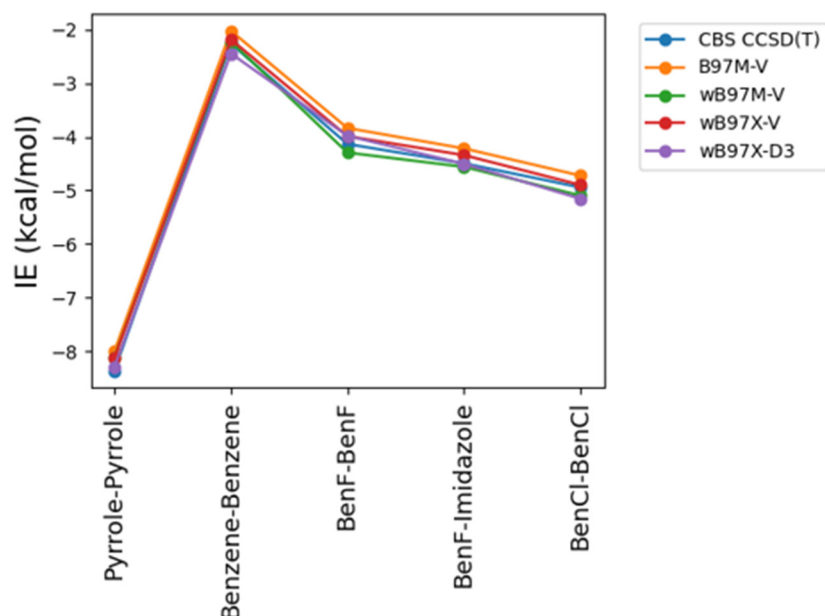


Figure 3. Benchmarking of different levels of DFT theory (B97M-V, ω B97M-V, ω B97X-V, and ω B97X-D3) for the stacked molecules against CBS CCSD(T) method.

2.1.3. Basis Set Dependence

The superior performance of ALMO EDA2 (B97M-V) with original energy decomposition scheme and SAPT2+(3)/dmp2 methods for both stacked and hydrogen-bond forming compounds was based on using aug-cc-pVTZ basis set. Benchmarking of the basis set dependence of both ALMO and SAPT methods is also crucial. Three different basis sets of aug-cc-pVTZ, aug-cc-PVDZ, and jun-cc-PVDZ were used for the five stacked molecules using different levels of theory: SAPT0, SAPT2+, SAPT2+(3)/dmp2, and ALMO EDA2 (**Figure 4**).

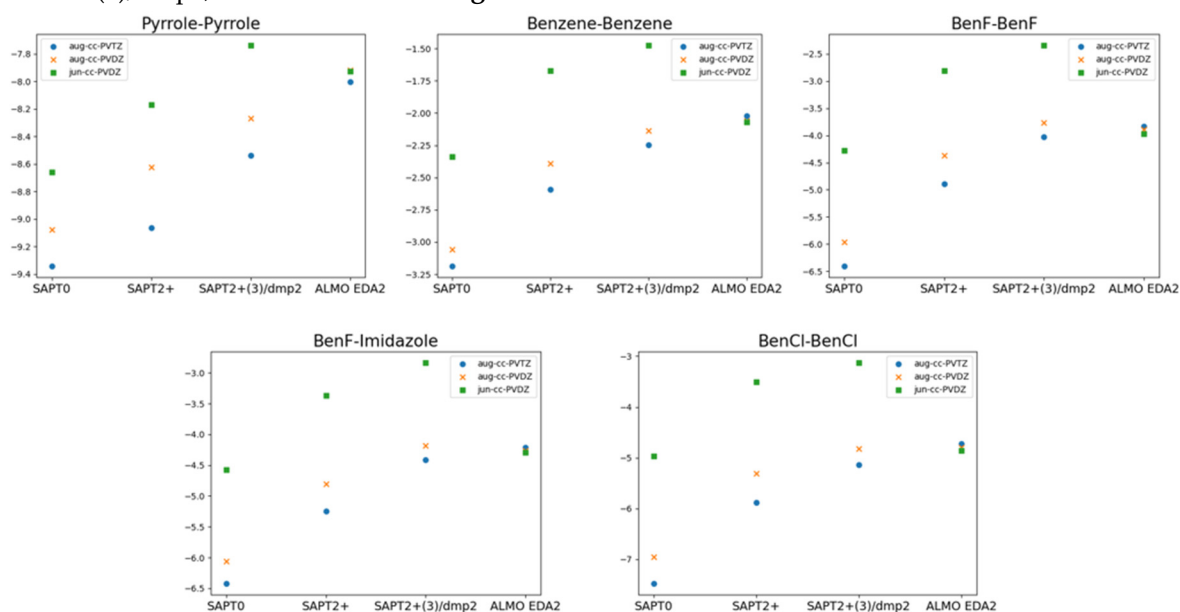


Figure 4. Basis set dependence of ALMO EDA2 generation with B97M-V, SAPT0, SAPT2+, and SAPT2+(3)/dmp2 in the stacked molecules using jun-cc-PVDZ, aug-cc-PVDZ, and aug-cc-pVTZ. Thus, using higher level of SAPT or large basis set do not guarantee better accuracy. Empirically, it was determined that SAPT0 with jun-cc-pVDZ basis and SAPT2+ and SAPT2+(3)/dmp2 with aug-cc-pVDZ are in reasonable agreement with ALMO EDA2, with an average deviation of 0.2 kcal/mol for

all stacking energies. The maximum difference among the three SAPT methods was 0.3 kcal/mol, observed for BenCl-BenCl.

ALMO EDA2 was found to be independent of the basis set used, where the maximum difference in the interaction values for all 5 stacked molecules obtained from different basis sets is around 0.1 kcal/mol. This reflects the robustness of the DFT methods such as B97M-V.

The situation with SAPT is complicated. Using the larger basis sets aug-cc-pVTZ and aug-cc-pVDZ with SAPT0 resulted in overestimation of the interaction energy by 3.0 kcal/mol on average, especially for halogen substituted benzene molecules, compared to SAPT0 with the smallest basis set jun-cc-pVDZ. Surprisingly, SAPT0 with jun-cc-pVDZ resulted in interaction energies that are more consistent with those obtained from ALMO EDA2. In higher levels of SAPT, SAPT2+ and SAPT2+(3)/dmp2, using jun-cc-pVDZ resulted in underestimation of the interaction energy for all the stacked molecules, by ~2.0 kcal/mol for the 5 stacked dimers. The interaction energies from using aug-cc-pVDZ with SAPT2+ were closer to those obtained from ALMO compared to using aug-cc-pVTZ basis set. The maximum difference of using the two basis sets is around 0.58 kcal/mol, observed in BenCl-BenCl molecule. The performance of SAPT2+(3)/dmp2 with aug-cc-pVDZ was slightly better, compared to ALMO energies, than with aug-cc-pVTZ. The difference between the two basis sets aug-cc-pVTZ and aug-cc-pVDZ in the highest level of theory SAPT2+(3)/dmp2 was found to be less compared to SAPT2+.

2.2. Total Interaction Energies of The Main Three Constituents of Nucleotides

Based on the benchmarking so far, we applied the following five protocols to examine the interaction energies of nucleotide fragments: ALMO EDA2 (B97M-V, original scheme), SAPT0 with jun-cc-pVDZ, SAPT2+ with aug-cc-pVDZ, and SAPT2+(3)/dmp2 with either aug-cc-pVTZ or aug-cc-pVDZ.

2.2.1. Sugar Moiety

To probe the hydrogen binding interaction around the sugar moiety, three different configurations of a water molecule interacting with C2' endo deoxyribose was selected based on the optimal docking scores obtained from aISS workflow [35] and on the uniqueness of the water position (**Figure 5**). In Wat1 and Wat5 configurations, the water molecules form hydrogen bond with the oxygen moiety of the 5-membered ring sugar moiety and with the terminal hydroxyl group. Both configurations interact with different angles with the sugar moiety but with the similar distances 1.8 or 1.9 Å. Wat13 configuration forms a single hydrogen bond with the hydroxyl group attached to C3' at 1.9 Å. Similarly, three different configurations of the water molecule interacting with C3' endo ribose sugar moiety were selected (**Figure 5**). Wat 1 configuration forms hydrogen bond with the oxygen of the 5-membered ring and the hydroxyl group attached to the C3', Wat 3 configuration forms hydrogen bond with the oxygen in the ring and the terminal hydroxyl group, and Wat 9 configuration forms hydrogen bond with the terminal hydroxyl group and the hydroxyl group attached to C2'. To probe the interaction energy surface, the water probe was moved away from the minimum energy distance along the vector connecting the two molecular centers of mass. For each distance, the single point interaction energy of sugar-water was calculated.

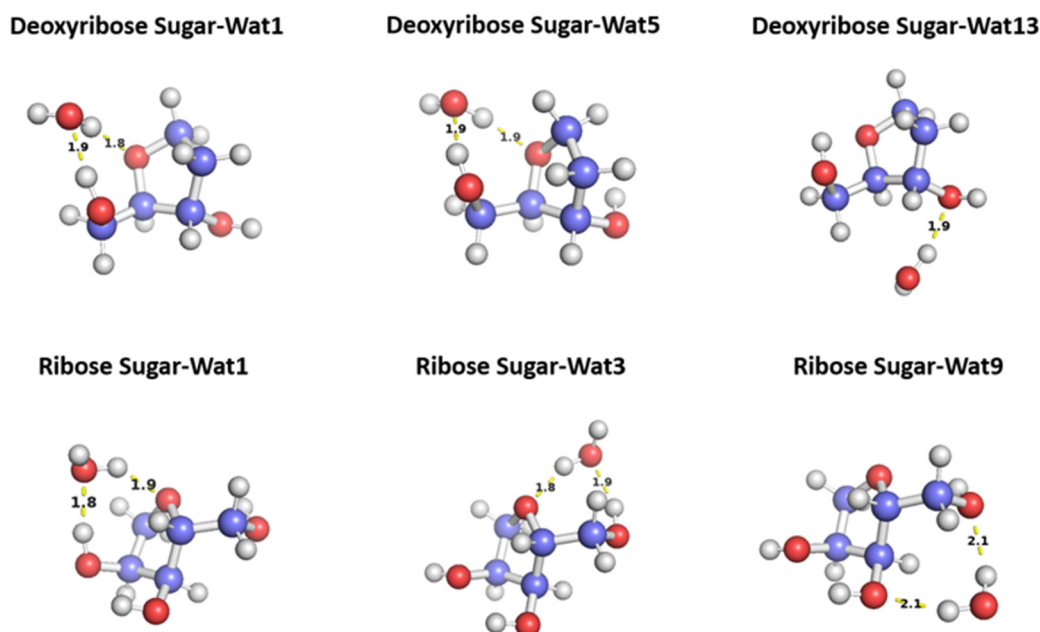


Figure 5. Representative structures for the C2' endo deoxyribose sugar and C3' endo ribose sugar interact with water probe at different positions. The distances correspond to the equilibrium distance obtained from MP2 optimization of the structures with 6-31G* basis.

Deoxyribose Wat1 and Wat5 configurations were found to have lower energy compared to Wat13 configuration due to the additional hydrogen bond formed. Wat5 configuration has slightly lower energy than Wat1 configuration by about 0.6 kcal/mol. For clarify, only Wat 1 and Wat13 energy profiles are shown in **Figure 6a**.

Ribose Wat1 configuration was associated with the lowest energy followed by Wat3 configuration then Wat9 configuration. This indicates that forming hydrogen bond with the oxygen of the ring is crucial to have the lowest energy configurations. SAPT2+(3)/dmp2 with aug-cc-pVDZ resulted in the most inconsistent interaction energies compared to the other methods with an average difference of 1.2 kcal/mol at the equilibrium distance for all three configurations. This suggests that SAPT2+(3)/dmp2 with aug-cc-pVDZ should not be recommended for QM benchmark for the nucleic acid, even though the stacking calculations above indicated this protocol was "reasonable." Similar to deoxyribose-water interaction curve, SAPT0 with jun-cc-pVDZ, SAPT2+(3)/dmp2 with aug-cc-pVTZ, and ALMO EDA2 generation were found to be very consistent in the predicted interaction energies, except in Wat9 configuration (**Figure 6b**). The stabilization energy is about -6 or -7 kcal/mol for a single hydrogen-bond and about -11 kcal/mol for double hydrogen-bonds. The only exception was observed in Wat9 configuration, SAPT2+ with aug-cc-pVDZ was found to be more consistent with the other protocols compared to SAPT0, with average difference of 0.2 kcal/mol between the two protocols for the different distances.

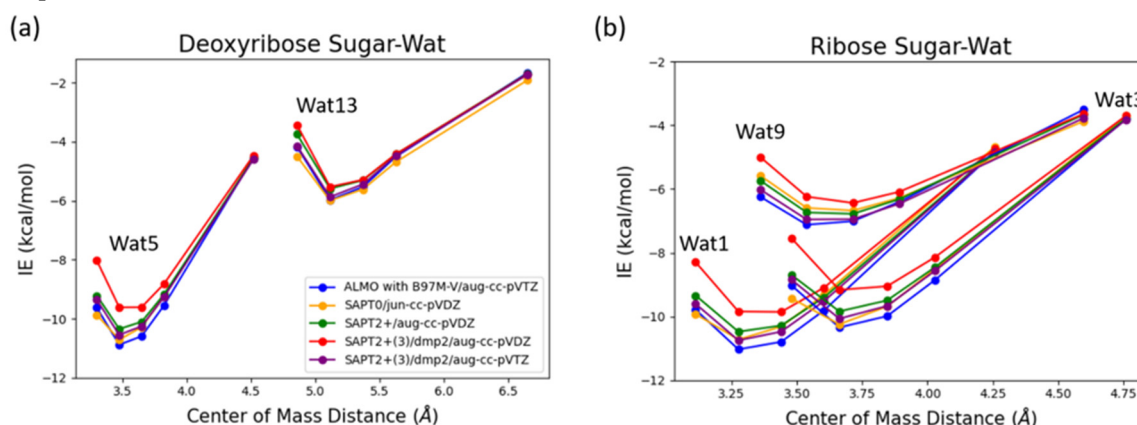


Figure 6. Interaction energy curve of the sugar moiety with different positions of water molecule at various percents of the equilibrium COM distance (0.95, 1.0, 1.0, 1.05, 1.1, and 1.3 times of equilibrium separation). (a) Interaction energy curve of C2' endo deoxyribose sugar with water; (b) Interaction energy curve of C3' endo ribose sugar with water.

The interaction energy components were obtained from SAPT2+(3)/dmp2 and ALMO EDA2 (original scheme) B97M-V aug-cc-pVTZ (ATZ) for the equilibrium structures of deoxyribose and ribose sugar moieties with the water probe (**Figure 7**). Overall, the EDA results from SAPT and ALMO are quite consistent with each other. The exchange repulsion is the most dominant positive component. Typically, the repulsion is ~25 kcal/mol for double h-bond complexes, ~10 kcal/mol for a single h-bond complex. The electrostatic, dispersion and induction (polarization + charge transfer in ALMO) make the attractive contributions. The electrostatic is clearly the major stabilizing force, contributing ~20 kcal/mol in double h-bond complexes, and ~ -9 kcal/mol in single h-bond complexes. The separation among dispersion and induction is slightly different between SAPT and ALMO. In SAPT, the dispersion is about -7 and -3 kcal/mol for double and single h-bond complexes, respectively. In ALMO, the dispersion is about 1-2 kcal/mol less negative than that of SAPT. Nonetheless the sum of polarization and charge transfer in ALMO is slightly more negative than the induction in SAPT but a similar amount. Thus, the total interaction energies are still consistent. Both EDA methods show that induction is significant, amounting to ~30-40% of the electrostatic energy. This again highlights the importance of explicit consideration of electronic polarization in classical force fields.

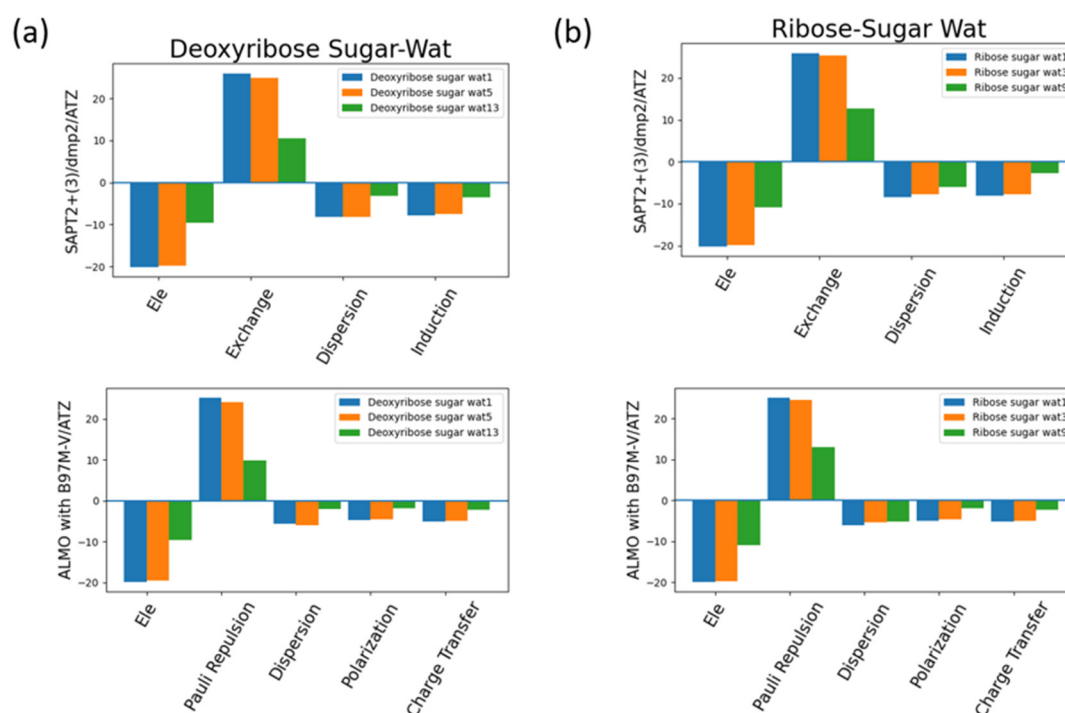


Figure 7. Analysis of total energy components using SAPT2+(3)/dmp2 and ALMO EDA2 generation scheme for sugar moiety-water configurations. (a) Energy components of C2' endo deoxyribose sugar-water configurations; (b) Energy components of C3' endo ribose sugar-water configurations.

2.2.2. DMP

Three different positions for the water molecules surrounding the low energy configurations of DMP [17] were selected to probe the key functional groups of DMP molecule. In Wat1 DMP tt configuration, the water molecule forms two hydrogen bonds with the phosphate group. While in Wat 9 and Wat10 configurations the water molecule forms a single hydrogen bond with different oxygen atoms of the phosphate group and with the oxygen of the -OCH3 groups attached to the

phosphate group (**Figure 8**). For DMP gt configuration, Wat7 forms a single hydrogen bond with the oxygen atom of the phosphate group, Wat10 configuration forms a hydrogen bond with the phosphate group and with the oxygen of the -OCH₃ group, and Wat15 configuration forms two hydrogen bonds with the oxygens of the two -OCH₃ groups attached to the phosphate group (**Figure 8**). In Wat10 of DMP gg configuration, water molecule forms hydrogen bond with the phosphate group and with the oxygen of -OCH₃ group attached to it, Wat 12 and Wat13 configurations form two hydrogen bonds with the phosphate group (**Figure 8**).

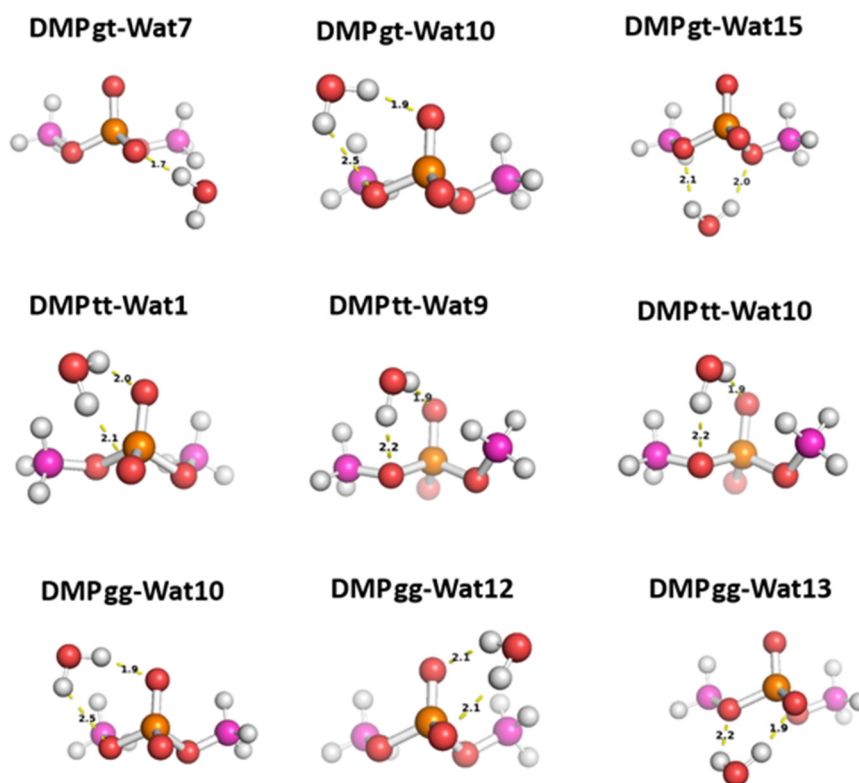


Figure 8. The optimized structures of the three dominant conformations of DMP [17]: DMP gg, DMP gt, and DMP tt interact with water molecules in various positions. The distances between DMP and water moieties shown on the structures correspond to the equilibrium distances retrieved from MP2/6-31G* optimization.

The interaction energy of DMP tt-Wat1 configuration was found the most favorable among the three configurations of DMP tt-water, followed by Wat9 and Wat10 configurations that have similar interaction energy (**Figure 9**). For clarity, Wat10 profiles are not shown here since it largely overlaps with those of Wat9. The two hydrogen bonds involving the two charged oxygen atoms in the phosphate group in Wat1 configuration resulted in higher interaction energy of ~ 2 kcal/mol near the equilibrium distances compared to those where a hydrogen bond is with the sp³ oxygen of the phosphate group. Wat10 and Wat 15 in DMP gt configuration followed a similar trend except that the interaction energy is slightly less favorable than those in the DMP tt conformation. DMP gt-Wat7 has only one hydrogen bond and the interaction energy is much weaker as expected. Water interaction energies with the DMP gg conformer are in general similar to those observed for DMP gt conformation (**Figure 9**). The results suggest that interaction of DMP with water has a nonnegligible conformational dependence, with the tt conformation being preferred.

Again, the different QM methods resulted in consistent interaction energies except SAPT0 with jun-cc-pVDZ basis set (**Figure 9**). Therefore, SAPT0 with jun-cc-pVDZ is not recommended as a general method for evaluating nonvalent interactions.

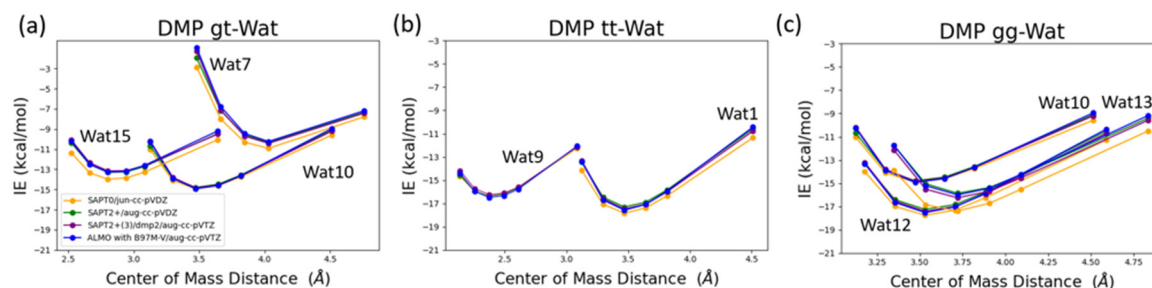


Figure 9. Interaction energy curve of the DMP with different positions of water molecule at various percents of the equilibrium COM distance (0.9, 0.95, 1.0, 1.0, 1.0, 1.05, 1.1, and 1.3). (a) DMP tt-Wat configurations; (b) DMP gt-Wat configurations; (c) DMP gg-Wat configurations.

The EDA of the interaction of the negatively charged DMP gg with water molecule at the equilibrium distance showed that the attractive electrostatics is counterbalanced with the positive exchange-repulsion, similar to what was observed for sugar-water (**Figure 10a**). The main difference here is electrostatic is A similar pattern was observed in ALMO energy decomposition where the dominant electrostatics component, with an average contribution of -22 kcal/mol for all three configurations, was counterbalanced with the Pauli repulsion, with average of 17 kcal/mol for the three configurations, (**Figure 10a**). Electrostatics components were found to be dominant in both DMP gt-Wat and DMP tt-Wat configurations with an average contribution of -23.5 kcal/mol and -23.7 kcal/mol for SAPT2+(3)/dmp2 and with average of -23.3 kcal/mol for both in ALMO energy decomposition, respectively (Figure 10b, 10c). DMP gt 7 was the only exception where the exchange component had the largest contribution to the total energy with 48.5 kcal/mol followed by the electrostatics component with a contribution of -32.1 kcal/mol (**Figure 10b**). Also, the contribution of the induction and dispersion in DMP gt-Wat7 in SAPT2+(3)/dmp2 scheme was found to be double of that of the remaining configurations. This was similar to what observed in the ALMO energy decomposition where polarization and charge transfer was also twice as high as in the remaining configurations (**Figure 10b**). This high contribution of the polarization and energy transfer may be contributed to the small distance of 1.7Å between the water molecule and the oxygen molecule of the phosphate group of the DMP gt molecule compared to 1.9Å in DMP gt-Wat10 (**Figure 8**). Besides DMP gt-Wat7 configuration, the contribution of the induction and dispersion components in the SAPT2+(3)/dmp2 decomposition were found to be less compared to the electrostatics and exchange. This was also observed in the ALMO energy decomposition, the dispersion, polarization, and charge transfer were found to be contributing less than the electrostatics and the Pauli repulsion for all DMP configurations (**Figure 10**).

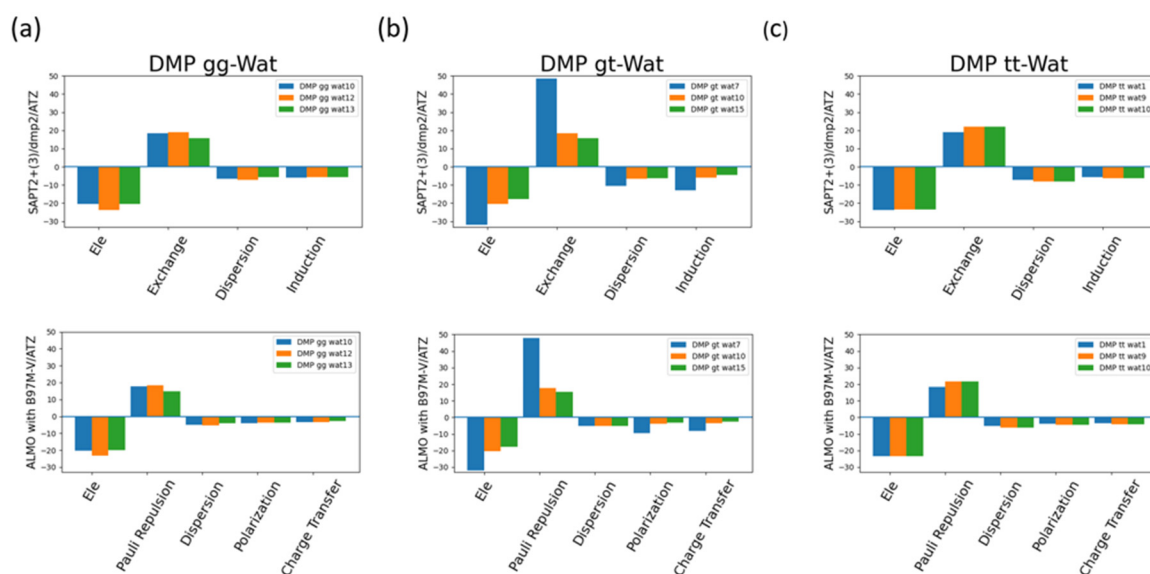


Figure 10. Analysis of total energy components using SAPT2+(3)/dmp2 and ALMO EDA2 generation scheme for DMP-water configurations. (a) Energy components of DMP gg-Wat; (b) Energy components of DMP gt-Wat; (c) Energy components of DMP tt-Wat.

2.2.3. Nucleobases

2.2.3.1. Nucleobase-Water

aISS workflow [35] was used to generate 15 different conformations of purine (A and G) and pyrimidine (U, T, and C) nucleobases interacting with water probes at different positions. Three different positions of the water molecule probe around A, U, T, and C nucleobases and four positions around G nucleobase were selected to cover the interactions of the water molecule with all the corresponding polarizable groups of the nucleobases (**Figure 11**). Optimization of the selected nucleobase-water structures was performed using MP2/6-31G* basis to obtain the equilibrium distances of the interaction of the nucleobases with the water probe at various positions.

The water molecules interacting with the Thymine nucleobase form two hydrogen bonds; one with the hydrogen atom of amine group and the other with the oxygen of the carbonyl group at different positions and at various distances. T-Wat1 interacts with N1 (1.9 Å) and oxygen at C2 (1.9 Å), T-Wat9 interacts with N3 (2.0 Å) and oxygen at C4 (1.9 Å), and T-Wat15 interacts with N3 (1.9 Å) and the oxygen atom at C2 (2.0 Å). Adenine-Wat1 (A-Wat1) interacts with the terminal NH₂ group at 1.9 Å and N7 of the purine ring, A-Wat9 interacts with N3 and N9 at equal distance of 2.0 Å, and A-Wat15 interacts with N1 of the purine ring at 2.0 Å and with the terminal NH₂ group. Uracil-Wat1 and Uracil-Wat7 show similar interaction patterns, where U-Wat1 interacts with N1 and with the carbonyl group at C2 at 1.9 Å and 1.7 Å, respectively. U-Wat7 interacts with N3 and the carbonyl of C4 at 1.9 Å each. On the other hand, U-Wat12 forms a single hydrogen bond with the carbonyl group of C4. All the water molecules interact with the Cytosine nucleobase share the same interaction pattern of forming two hydrogen bonds with the hydrogen of NH₂ and the oxygen of the carbonyl group at different positions. Cytosine-Wat1 interacts with N1 and the oxygen atom at of the carbonyl group at C2, C-Wat9 interacts with N3 and the oxygen of the carbonyl group at C4, and C-Wat15 interacts with N3 and the oxygen atom of the carbonyl group at C2. Four water probes were needed to cover all the polarizable groups of the Guanine nucleobase. G-Wat1 and G-Wat10 interact in a similar manner to each other and to the other water probes interacting with the other nucleobases. Both Wat1 (hydrogen atom of N1 and oxygen atom at carbonyl group of C6) and Wat10 (N7 oxygen atom at carbonyl group of C6) interact with an amine group and with the oxygen atom of the carbonyl group. G-Wat3 and G-Wat4 interact with amine groups. Wat3 and Wat4 interact with equal distances of 2 Å with N3 and NH₂ attached at C2 and with N3 and N9, respectively (**Figure11**).

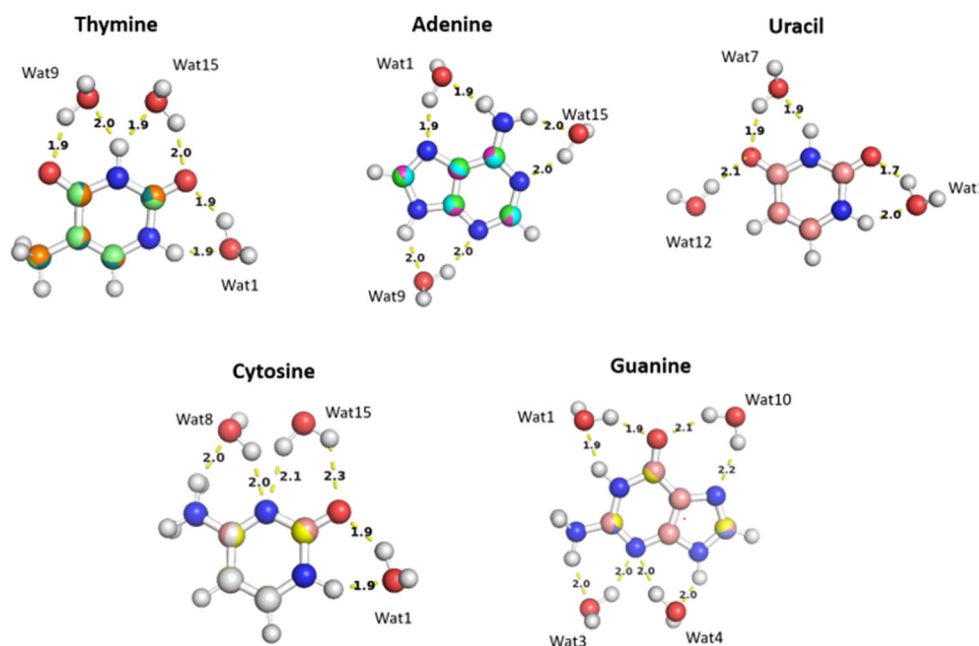


Figure 11. The optimized structures of the nucleobases: A, G, U, T, and C interact with water molecules in various positions. The distances between the purine and pyrimidine nucleobases and water moieties shown on the structures correspond to the equilibrium distances retrieved from MP2/6-31G* optimization.

The water probe interacting with the nucleobases at different positions was transitioned at 0.90, 0.95, 1.0, 1.1, and 1.3 percent of the equilibrium distance between the center of mass of the two molecules, water and nucleobase, obtained from the optimized structures to get an interaction energy curve of the nucleobases with the water probes. T-Wat1 was found to have the lowest energy among the three conformations of T-Wat interacting complexes (**Figure 12**). This may be attributed to the shorter distance of the two hydrogen bonds formed by Wat1 probe at 1.9Å of compared to the other two Wat probes at 1.9Å and 2.0Å (**Figure 11**). A-Wat9 was found to have the lowest energy conformation followed by A-Wat1. A-Wat15 was found to have the highest energy among the three conformations. This is attributed to the difference in the hydrogen bonding pattern of the different water probes. Wat12 interacting with Uracil nucleobase was found to be the least energetically stable as it forms a single hydrogen bond compared to Wat1 and Wat7 that form two hydrogen bonds (**Figure 11**). Wat1 probe has lower interaction energy with Uracil compared to Wat7 probe due to the shorter distances of the hydrogen bonds formed. C-Wat15 had the lowest interaction energies due to the large distances of the two formed hydrogen bonds at 2.1Å and 2.3Å compared to C-Wat1 (1.9Å) and C-Wat8 (2.0Å) interaction energies. G-Wat1 was found to have the lowest energy among the four G-Wat conformations due to the short hydrogen bond formed unlike Wat10 probe that was found to have the lowest interaction energy despite having nearly similar interaction of Wat1 probe. G-Wat3 and G-Wat4 complexes were found to have higher energy compared to G-Wat1 (**Figure 12**). This indicates that the interaction with the oxygen of the carbonyl group at short distance resulted in stronger interaction. Wat4, forms two bonds with the amine of the Guanine ring, was found to have higher interaction energy than Wat3 probe, one amine in the ring and one terminal amine, due to the resonance effect found in the purine ring.

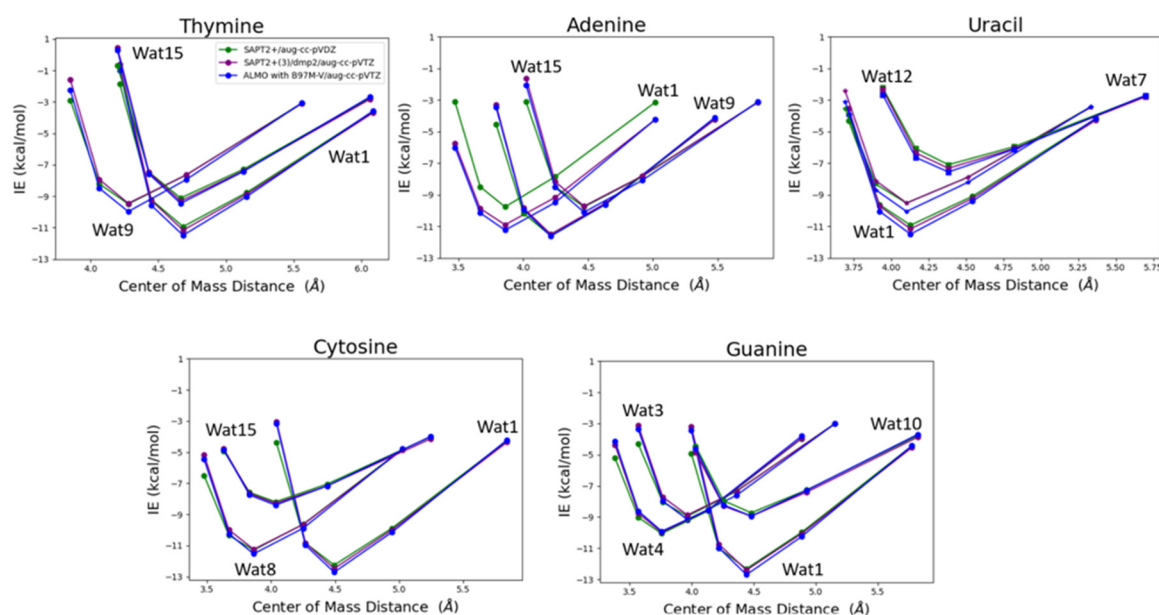


Figure 12. Interaction energy curve of the purine and pyrimidine nucleobases with a water probe molecule over various distances. The distance between center of mass of the two molecules (nucleobase and water in this case), correspond to various percents of the equilibrium distance (0.9, 0.95, 1.0, 1.1, and 1.3). The distance between the center of mass shown in the plot does not reflect the interaction distance between the water probe and the nucleobase.

SAPT2+(3)/dmp2 with aug-cc-pVTZ and ALMO EDA2 generation with B97M-V were found to be in good agreement with each other for studying the interaction energy of the water probe with the different nucleobases (**Figure 12**). A maximum difference of -1.2 kcal/mol between ALMO EDA2 generation and SAPT2+ with aug-cc-pVDZ was observed at the smallest interacting distance of Wat 15 probe with Adenine (A-Wat15), where RMSE was found to be 0.52. This reflects that SAPT2+ with aug-cc-pVDZ falls behind when the interacting molecules are so close to each other. That is why this protocol was dropped as a possible QM benchmark to study the interaction energy of the different components of nucleic acid.

The energy components of Thymine total interaction energy and of Adenine total interaction energy were investigated as an example of pyrimidine and purine interactions with water (**Figure 13**). Similar to the neutral sugar moiety (**Figure 7**), exchange in SAPT2+(3)/dmp2 scheme and Pauli repulsion in ALMO EDA2 generation scheme were found to be the most dominant component, which was counterbalanced with the attractive electrostatics' component. Dispersion, Induction, and Charge transfer were found to be contributing less with an average of -6.5 kcal/mol for each component in the different Nucleobase-Wat configurations. In other words, induction and dispersion are the main contributing forces to stabilize the nucleobase-water interactions. The difference in the purine and pyrimidine structure did not make a difference in the contribution of the different energy components to the total interaction energy.

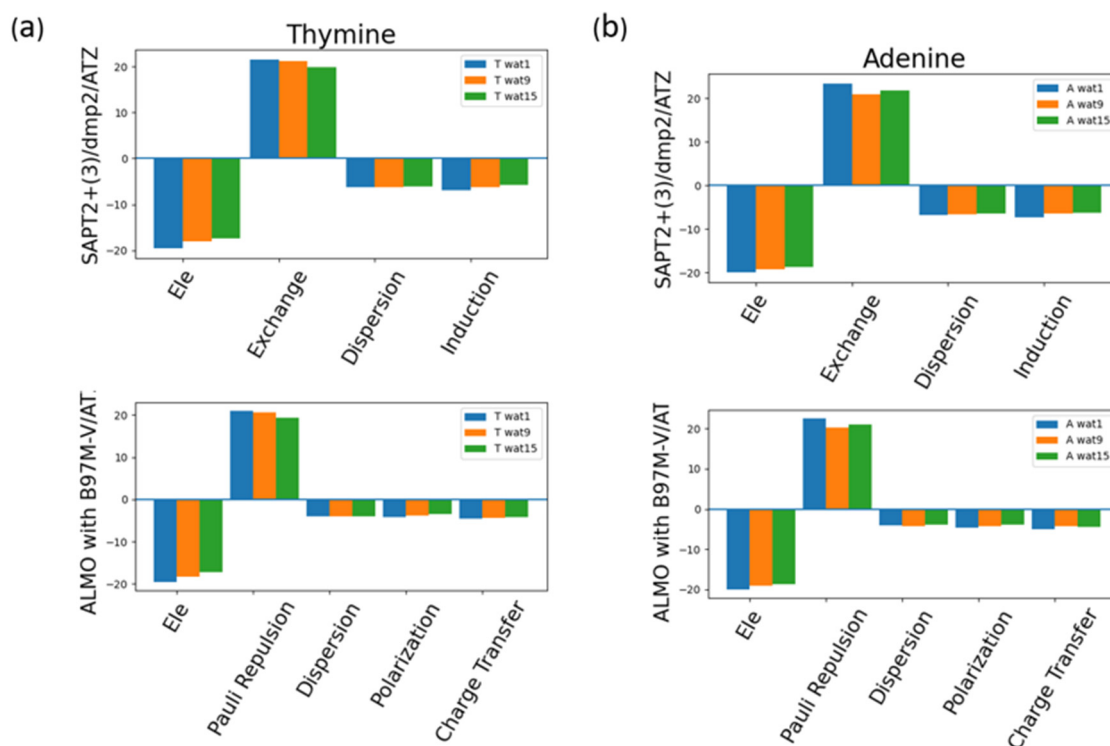


Figure 13. Analysis of total energy components using SAPT2+(3)/dmp2 and ALMO EDA2 generation scheme for Nucleobase-water configurations. (a) Energy components of Thymine-Wat; (b) Energy components of Adenine-Wat.

2.2.3.2. Intrastrand and Interstrand Nucleobases

Another important non-covalent interaction to consider in the nucleic acid structure is the π -stacking interaction. That is why the interaction energies of the intrastrand stacked nucleobases and the interstrand nucleobases were studied using SAPT2+(3)/dmp2 and ALMO EDA2 with B97M-V (**Figure 14**). The structures of the of B-DNA geometry nucleobases were obtained from Sherill paper following the same nomenclature for the interstrand and intrastrand nucleobases [26]. The two methods were found to result in different interaction energies for the stacked intrastand nucleobases (**Figure 14a**), however both methods were in good agreement with each other for the interstrand

nucleobases (Figure 14b). The average root-mean-square-error (RMSE) for both 11 and 22 intrastrand nucleobases was 0.68, while the average RMSE for both 12 and 21 interstrand nucleobases was found to be 0.25.

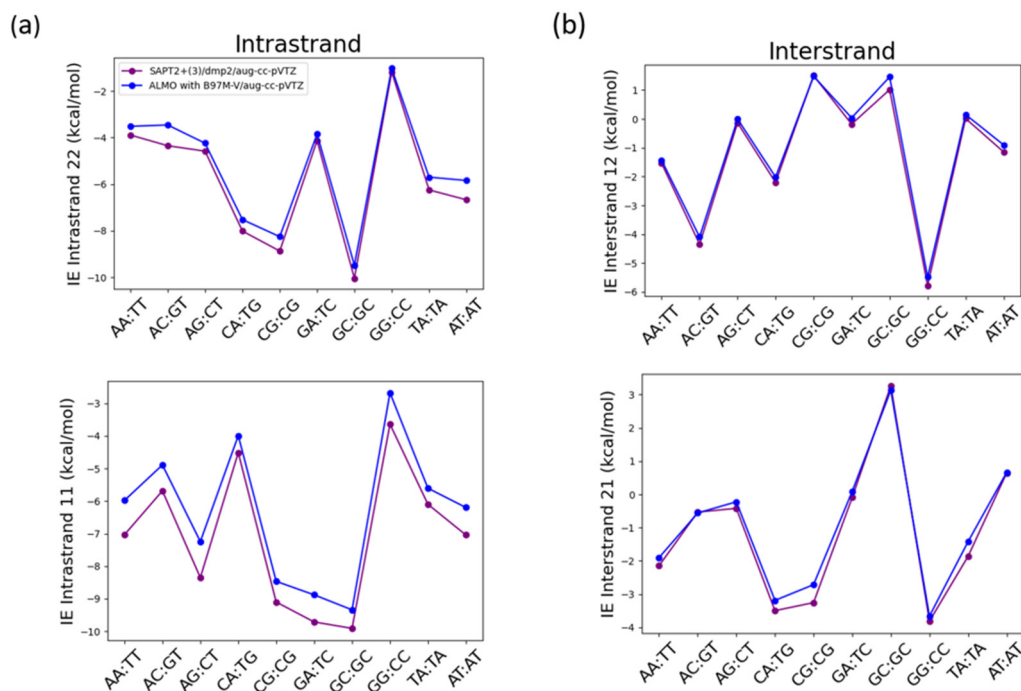


Figure 14. Interaction energies of the nucleobases with each other using SAPT2+(3)/dmp2 and ALMO with B97M-V generation scheme (a) Interaction energies of intrastrand nucleobases governed by π -stacking interactions; (b) Interaction energies of interstrand nucleobases governed by hydrogen bond.

The maximum difference in the intrastrand 11 nucleobases of -1.11 kcal/mol between SAPT2+(3)/dmp2 and ALMO with B97M-V was observed in the stacked AG base pair followed by stacked AA base pair with a difference of -1.06 kcal/mol (**Figure 14a**). The maximum difference in the intrastrand 22 nucleobases was observed in GT base pair (-0.89 kcal/mol) followed by AT base pair with a difference of -0.83 kcal/mol (**Figure 14a**). A much smaller difference was observed in the interstrand nucleobases, where the maximum difference observed in interstrand 12 nucleobases between the two methods was -0.25 kcal/mol in AC base pair followed by AT base pair at -0.24 kcal/mol (**Figure 14b**). The maximum difference of -0.54 kcal/mol observed in 21 interstrand nucleobases was observed in CG base pair followed by TA base pair with a difference of -0.44 kcal/mol.

Analyzing the energy components of the intrastrand AA and interstrand AA as an example for the π -stacking interactions in the nucleic acid (**Figure 15**). Dispersion was found to be the most contributing component in the intrastrand AA (**Figure 15a**) as expected due to the dominating π -stacking interaction. Exchange-Pauli Repulsion was found to be the second most dominant contributor to the total energy, followed by a much lesser extent attractive electrostatics components.

Figure 15b showed that dispersion was also the most dominating factor, to a much lesser extent compared to intrastrand AA, to the interaction of the interstrand AA. This was surprising as the dominating type of interaction in the intrastrand AA is the hydrogen bond. The second most dominating factor in the interstrand interaction was found to be the Exchange-Pauli Repulsion similar to what observed in the intrastrand AA. Unlike the intrastrand AA base pair, the electrostatics component was found to be repulsive, having a positive contribution, rather than attractive.

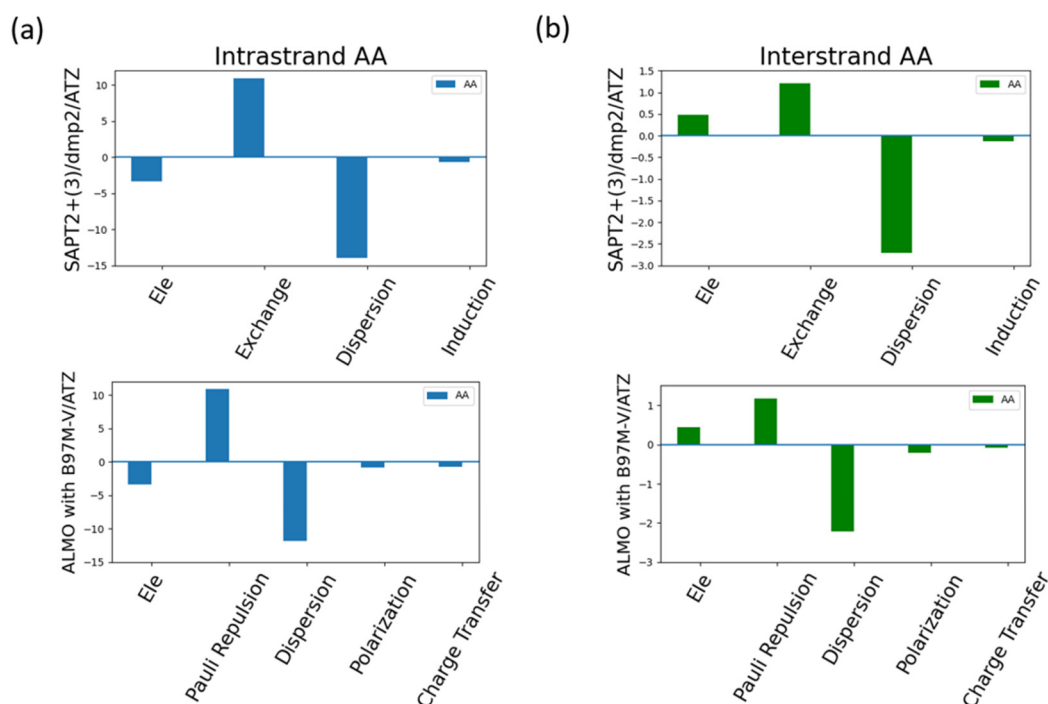


Figure 15. Analysis of total energy components using SAPT2+(3)/dmp2 and ALMO EDA2 with B97MV generation scheme for interstranded and intrastranded AA base pair. (a) Energy components of Intrastranded AA nucleobases; (b) Energy components of interstranded AA nucleobase.

3. Discussion

Nucleic acid plays a vital role in several biological processes from contributing to the central dogma of the cell to functioning as an enzyme or binding as a ligand. With the current limitations of the MM forcefields in dealing with the non-covalent interactions dominating the nucleic acid, especially the π -stacking interactions due to the lack of charge penetration factor, accurate description of the non-covalent interactions will help in developing better forcefields tailored for the nucleic acid.

Benchmarking of SAPT0, SAPT2+, SAPT2+(3)/dmp2, ALMO EDA2 with the original and the new energy decomposition, MP2, and CCSD(T) on a set of small molecules with available CBS CCSD(T) interaction energies was performed. The small molecules comprised the two types of interactions prevalent in the nucleic acid structure: hydrogen bond and π -stacking interactions. Based on the benchmarking results, ALMO EDA2 was found to be independent of the DFT method, or the basis set used. On the other hand, the level of SAPT was highly dependent on the basis set used. The approximate CCSD(T), DLPNO/CCSD(T), was found to be less reliable compared to the DFT methods due to the large deviation from the CBS CCSD(T) interaction energies, especially for the halogen substituted benzene molecules. The interaction energy of BenCl-BenCl complex was the most challenging to the various theoretical methods. Further analysis for this deviation is needed to tackle the inconsistent and inaccurate interaction energy for the halogen substituted benzene and heterocycles in general, especially BenCl.

After testing the different theoretical methods along with dependence on DFT method and basis set, five different protocols were selected to be tested on the three main components of the nucleotides' structure. SAPT0 with jun-cc-pVDZ, SAPT2+ with aug-cc-pVDZ, SAPT2+(3)/dmp2 with aug-cc-pVDZ, SAPT2+(3)/dmp2 with aug-cc-pVTZ, ALMO EDA2 (B97M-V) with the original energy decomposition and aug-cc-pVTZ. The five protocols were tested on the interaction of neutral sugar moiety (C2' endo deoxyribose sugar and C3' endo ribose sugar) with water, the negatively charged DMP (DMP gg, DMPgt, and DMP tt) with water, nucleobases with water, and the interstrand and intrastrand nucleobases to govern the two types of interactions found in the structure of the nucleic acid. SAPT2+(3)/dmp2 and ALMO EDA2 were found to be the most consistent with each other with

respect to the obtained interaction energies. Most of the difference observed between the two methods was in the stacked interaction energies of the nucleobases. Another important point to consider is the computational cost of the calculations, ALMO EDA2 is faster on average than SAPT2+(3)/dmp2 with similar accuracy.

Analyzing the energy components of the total interaction energies obtained from the ALMO EDA2 and SAPT2+(3)/dmp2 schemes showed that Exchange-Pauli repulsion, the dominating contributor to the total interaction energy of the neutral molecules (sugar moiety and nucleobases) with water was counterbalanced with the attractive electrostatics component, resulting in making the induction and dispersion the stabilizing factors for these interactions. In the negatively charged DMP interaction with water, electrostatics components were found to be associated with the largest contribution to the total energy. In the interstrand and intrastrand nucleobases, dispersion was found to be the dominating factor in this type of π -stacking and hydrogen bond interactions.

4. Materials and Methods

4.1. Preparation of Structures

The canonical structures of the nucleotides were constructed using Nucleic.exe implemented in Tinker8 [36,37]. C2'endo deoxyribose sugar was extracted from B-DNA geometry, while C3' endo ribose sugar was extracted from A-DNA geometry. The purine and pyrimidine nucleobases were obtained from B-DNA geometry of the nucleic acid as indicated by Tinker8 [36,37]. The three dominant conformations of DMP: DMP gg, DMP gt, and DMP tt were constructed using Pymol [17]. The structures of the B-DNA interstrand and intrastrand nucleobases were obtained from Sherill paper following the same nomenclature [26].

After obtaining the canonical structures of the nucleotides' components, water molecules were placed at various positions surrounding the structures following aISS workflow [35]. aISS workflow is based on searching for the lowest energy conformation of two fragments interacting non-covalently. The first step in the workflow is to calculate the electronic structure properties using semi-empirical GFNN-xTB [38,39] followed by grid-search based on xTB-IFF [40] interaction energies of the two fragments. In other words, several geometries are generated for the two interacting fragments by moving the first fragment around the second one. Then, the interaction energy is evaluated using xTB-IFF interaction energies. The generated structures from the grid-search are subjected to genetic optimization followed by final GFNN-xTB or GFN-FF optimization.

15 different conformations of the water molecule interacting with the different constituents were generated (nucleobase-water, sugar moiety-water, and DMP-water). Three different conformations of the water molecule probe were selected based on the docking score and the uniqueness of the water molecule position. In other words, the position of the water molecule was chosen to cover the entire conformational space around the counter interacting moieties and to interact with all the polarizable groups present in the structure.

The selected three complexes with the water molecule probe were subjected to energy minimization using MP2/6-31G* in GAUSSIAN16 [41]. This was followed by translation of the water molecule probe along the center of mass of the water molecule. The translation distances corresponded to percents of the equilibrium distance of the optimized structures. The water molecule was translated to five or six positions corresponding to 0.90, 0.95, 1.0, 1.05, 1.1, and 1.3 of the equilibrium distance between the center of mass of the water molecule and the center of mass of the counter interacting moiety. These distances do not reflect the interacting distances between the two molecules. This translation of the water probe helped us to generate an interaction energy curve of the nucleobase-water, sugar moiety-water, and DMP-water, which helped us in identifying the local minimum of each structure.

4.2. Ab-Initio Calculations

The interaction energies of the generated nucleobase-water, sugar moiety-water, DMP-water complexes, and interstrand and intrastrand base pairs were computed using the original and the new

decomposition energy scheme of ALMO (Absolutely Localized Molecular Orbital) ALMO EDA2 [29–31] implemented in QCHEM 6.1 package [28]. ALMO EDA decomposes the interaction energy into three main terms: frozen density component, polarization, and charge transfer. The frozen density core includes the electrostatics, pauli-repulsion, and dispersion. ALMO EDA2 resolved the shortcomings of ALMO EDA as the dependence of the polarization term on the basis set used. It was found by the larger basis set utilized the more contaminated the polarization term become by the charge transfer term. Two energy decomposition schemes are popular while using ALMO EDA2: the first is the original energy decomposition scheme which includes the Frozen energy decomposition with ALMO polarization (AO-block-based), the new energy decomposition scheme involves Frozen energy decomposition with nDQ-FERF polarization. Both were tested on a set of small molecules to determine which will be better for computing the interaction energies of the nucleotides.

Different levels of SAPT (Symmetry-Adapted Perturbation Theory): SAPT0, SAPT2+, and SAPT2+(3)/dmp2 [34,42–45] implemented in psi4 package [32] were also used to compute the interaction energies. SAPT determines the interaction energy between two fragments directly by decomposing the energy into four physically meaningful terms: electrostatics, exchange, induction, and dispersion. SAPT0 mainly depends on error cancellation. Higher-order SAPT depends on including several corrections, for example SAPT2+ involves second-order correction for the monomeric fluctuations and dispersion energy correction by adding the effect of the intramolecular correlation [33]. SAPT2+(3)/dmp2 includes higher-order perturbative terms and variational correction terms.

5. Conclusions

QM benchmarks for nucleotides, the monomeric structure of nucleic acid, is essential to develop better forcefields with improved parameters for the non-covalent interactions preserving the complex structure of the nucleic acid. Upon testing five different protocols, we recommend using SAPT2+(3)/dmp2 with aug-cc-pVTZ and ALMO EDA2 with B97M-V and aug-cc-pVTZ as high-quality ab-initio methods for studying the non-covalent interactions of the nucleic acid. The attractive induction and dispersion were found to be responsible for stabilizing the interactions of the neutral moieties (sugar moiety and nucleobase) with water due to the cancellation of the large Exchange-Pauli repulsion component with the attractive electrostatics' components. Dispersion was found to be the most important in the interstrand and intrastrand base pairs.

Author Contributions: Conceptualization, Mayar Tarek Ibrahim and Pengyu Ren; methodology, Mayar Tarek Ibrahim and Pengyu Ren; validation, Mayar Tarek Ibrahim, Elizabeth Wait and Pengyu Ren; formal analysis, Mayar Tarek Ibrahim and Pengyu Ren; investigation, Mayar Tarek Ibrahim, Elizabeth Wait, and Pengyu Ren; resources, Pengyu Ren.; data curation, Mayar Tarek Ibrahim.; writing—original draft preparation, Mayar Tarek Ibrahim, Elizabeth Wait, and Pengyu Ren; writing—review and editing, Mayar Tarek Ibrahim, Elizabeth Wait, and Pengyu Ren.; visualization, Mayar Tarek Ibrahim.; supervision, Pengyu Ren; funding acquisition, Pengyu Ren. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Institutes of Health, grant number R01GM106137 and R01GM114237, the Welch Foundation (F-2120), and the Cancer Prevention and Research Institute of Texas grant (RP210088).

Data Availability Statement: SAPT2+(3)/dmp2 and ALMO EDA2 with original energy decomposition scheme interaction energies for all the complexes are provided in separate excel sheet.

Acknowledgments: The authors are grateful for support from the National Institutes of Health (R01GM106137 and R01GM114237), the Welch Foundation (F-2120), and the Cancer Prevention and Research Institute of Texas grant (RP210088).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. F. B. Rudolph, "The Biochemistry and Physiology of Nucleotides¹²³," *J Nutr*, vol. 124, no. 13, pp. 124S–127S, 1994, doi: https://doi.org/10.1093/jn/124.suppl_1.124S.

2. R. R. Breaker and G. F. Joyce, "The expanding view of RNA and DNA function," *Chem Biol*, vol. 21, no. 9, pp. 1059–1065, 2014.
3. W. H. Hudson and E. A. Ortlund, "The structure, function and evolution of proteins that bind DNA and RNA," *Nat Rev Mol Cell Biol*, vol. 15, no. 11, pp. 749–760, 2014.
4. D. Varshney, J. Spiegel, K. Zyner, D. Tannahill, and S. Balasubramanian, "The regulation and functions of DNA and RNA G-quadruplexes," *Nat Rev Mol Cell Biol*, vol. 21, no. 8, pp. 459–474, 2020.
5. R. W. Carthew and E. J. Sontheimer, "Origins and mechanisms of miRNAs and siRNAs," *Cell*, vol. 136, no. 4, pp. 642–655, 2009.
6. F. A. Hays *et al.*, "How sequence defines structure: a crystallographic map of DNA structure and conformation," *Proceedings of the National Academy of Sciences*, vol. 102, no. 20, pp. 7157–7162, 2005.
7. D. Svozil, J. Kalina, M. Omelka, and B. Schneider, "DNA conformations and their sequence preferences," *Nucleic Acids Res*, vol. 36, no. 11, pp. 3690–3706, 2008.
8. L. E. Minchenkova, A. K. Schyolkina, B. K. Chernov, and V. I. Ivanov, "CC/GG contacts facilitate the B to A transition of DNA in solution," *J Biomol Struct Dyn*, vol. 4, no. 3, pp. 463–476, 1986.
9. D. Jose and D. Porschke, "The dynamics of the B– A transition of natural DNA double helices," *J Am Chem Soc*, vol. 127, no. 46, pp. 16120–16128, 2005.
10. D. R. Whelan, T. J. Hiscox, J. I. Rood, K. R. Bambery, D. McNaughton, and B. R. Wood, "Detection of an en masse and reversible B-to A-DNA conformational transition in prokaryotes in response to desiccation," *J R Soc Interface*, vol. 11, no. 97, p. 20140454, 2014.
11. C. Zhang *et al.*, "AMOEBA polarizable atomic multipole force field for nucleic acids," *J Chem Theory Comput*, vol. 14, no. 4, pp. 2084–2108, 2018.
12. A. Herbert and A. Rich, "The Biology of Left-handed Z-DNA (*)," *Journal of Biological Chemistry*, vol. 271, no. 20, pp. 11595–11598, 1996.
13. R. V Gessner, C. A. Frederick, G. J. Quigley, A. Rich, and A. H. J. Wang, "The molecular structure of the left-handed Z-DNA double helix at 1.0-Å atomic resolution: Geometry, conformation, and ionic interactions of d (CGCGCG)," *Journal of Biological chemistry*, vol. 264, no. 14, pp. 7921–7935, 1989.
14. N. Leontis and E. Westhof, *RNA 3D structure analysis and prediction*, vol. 27. Springer Science & Business Media, 2012.
15. S. N. Rao and P. Kollman, "On the role of uniform and mixed sugar puckers in DNA double-helical structures," *J Am Chem Soc*, vol. 107, no. 6, pp. 1611–1617, 1985.
16. R. Galindo-Murillo *et al.*, "Assessing the current state of amber force field modifications for DNA," *J Chem Theory Comput*, vol. 12, no. 8, pp. 4114–4127, 2016.
17. C. Zhang, C. Lu, Q. Wang, J. W. Ponder, and P. Ren, "Polarizable multipole-based force field for dimethyl and trimethyl phosphate," *J Chem Theory Comput*, vol. 11, no. 11, pp. 5326–5339, 2015.
18. C. Zhang, D. Bell, M. Harger, and P. Ren, "Polarizable multipole-based force field for aromatic molecules and nucleobases," *J Chem Theory Comput*, vol. 13, no. 2, pp. 666–678, 2017.
19. C. Zhang *et al.*, "AMOEBA polarizable atomic multipole force field for nucleic acids," *J Chem Theory Comput*, vol. 14, no. 4, pp. 2084–2108, 2018.
20. J. A. Lemkul and A. D. MacKerell Jr, "Polarizable force field for DNA based on the classical Drude oscillator: II. Microsecond molecular dynamics simulations of duplex DNA," *J Chem Theory Comput*, vol. 13, no. 5, pp. 2072–2085, 2017.
21. J. A. Lemkul and A. D. MacKerell Jr, "Polarizable force field for DNA based on the classical Drude oscillator: I. Refinement using quantum mechanical base stacking and conformational energetics," *J Chem Theory Comput*, vol. 13, no. 5, pp. 2053–2071, 2017.
22. J. A. Lemkul and A. D. MacKerell Jr, "Polarizable force field for RNA based on the classical drude oscillator," *J Comput Chem*, vol. 39, no. 32, pp. 2624–2646, 2018.
23. C. D. Sherrill *et al.*, "Assessment of standard force field models against high-quality ab initio potential curves for prototypes of π - π , CH/ π , and SH/ π interactions," *J Comput Chem*, vol. 30, no. 14, pp. 2187–2193, 2009.
24. E. G. Hohenstein, J. Duan, and C. D. Sherrill, "Origin of the surprising enhancement of electrostatic energies by electron-donating substituents in substituted sandwich benzene dimers," *J Am Chem Soc*, vol. 133, no. 34, pp. 13244–13247, 2011.
25. A. Stone, *The theory of intermolecular forces*. oUP oxford, 2013.
26. T. M. Parker and C. D. Sherrill, "Assessment of empirical models versus high-accuracy ab initio methods for nucleobase stacking: Evaluating the importance of charge penetration," *J Chem Theory Comput*, vol. 11, no. 9, pp. 4197–4204, 2015.
27. A. G. Donchev *et al.*, "Quantum chemical benchmark databases of gold-standard dimer interaction energies," *Sci Data*, vol. 8, no. 1, p. 55, 2021.
28. R. Z. Khaliullin, E. A. Cobar, R. C. Lochan, A. T. Bell, and M. Head-Gordon, "Unravelling the origin of intermolecular interactions using absolutely localized molecular orbitals," *J Phys Chem A*, vol. 111, no. 36, pp. 8753–8765, 2007.

29. P. R. Horn, Y. Mao, and M. Head-Gordon, "Defining the contributions of permanent electrostatics, Pauli repulsion, and dispersion in density functional theory calculations of intermolecular interaction energies," *J Chem Phys*, vol. 144, no. 11, 2016.
30. P. R. Horn, Y. Mao, and M. Head-Gordon, "Probing non-covalent interactions with a second generation energy decomposition analysis using absolutely localized molecular orbitals," *Physical Chemistry Chemical Physics*, vol. 18, no. 33, pp. 23067–23079, 2016.
31. P. R. Horn and M. Head-Gordon, "Polarization contributions to intermolecular interactions revisited with fragment electric-field response functions," *J Chem Phys*, vol. 143, no. 11, 2015.
32. D. G. A. Smith *et al.*, "PSI4 1.4: Open-source software for high-throughput quantum chemistry," *J Chem Phys*, vol. 152, no. 18, 2020.
33. J. G. McDaniel and J. R. Schmidt, "Next-generation force fields from symmetry-adapted perturbation theory," *Annu Rev Phys Chem*, vol. 67, pp. 467–488, 2016.
34. E. G. Hohenstein and C. D. Sherrill, "Density fitting and Cholesky decomposition approximations in symmetry-adapted perturbation theory: Implementation and application to probe the nature of π - π interactions in linear acenes," *J Chem Phys*, vol. 132, no. 18, 2010.
35. C. Plett and S. Grimme, "Automated and efficient generation of general molecular aggregate structures," *Angewandte Chemie International Edition*, vol. 62, no. 4, p. e202214477, 2023.
36. J. W. Ponder, "TINKER: Software tools for molecular design," *Washington University School of Medicine, Saint Louis, MO*, vol. 3, p. 116, 2004.
37. J. A. Rackers *et al.*, "Tinker 8: software tools for molecular design," *J Chem Theory Comput*, vol. 14, no. 10, pp. 5273–5289, 2018.
38. C. Bannwarth *et al.*, "Extended tight-binding quantum chemistry methods," *Wiley Interdiscip Rev Comput Mol Sci*, vol. 11, no. 2, p. e1493, 2021.
39. S. Grimme, "Semiempirical Extended Tight-Binding Program Package." 2019.
40. S. Grimme, C. Bannwarth, E. Caldeweyher, J. Pisarek, and A. Hansen, "A general intermolecular force field based on tight-binding quantum chemical calculations," *J Chem Phys*, vol. 147, no. 16, 2017.
41. M. J. Frisch *et al.*, "Gaussian 16." Gaussian, Inc. Wallingford, CT, 2016.
42. E. G. Hohenstein, R. M. Parrish, C. D. Sherrill, J. M. Turney, and H. F. Schaefer, "Large-scale symmetry-adapted perturbation theory computations via density fitting and Laplace transformation techniques: Investigating the fundamental forces of DNA-intercalator interactions," *J Chem Phys*, vol. 135, no. 17, 2011.
43. E. G. Hohenstein and C. D. Sherrill, "Density fitting of intramonomer correlation effects in symmetry-adapted perturbation theory," *J Chem Phys*, vol. 133, no. 1, 2010.
44. E. G. Hohenstein and C. D. Sherrill, "Efficient evaluation of triple excitations in symmetry-adapted perturbation theory via second-order Møller–Plesset perturbation theory natural orbitals," *J Chem Phys*, vol. 133, no. 10, 2010.
45. B. Jeziorski, R. Moszynski, and K. Szalewicz, "Perturbation theory approach to intermolecular potential energy surfaces of van der Waals complexes," *Chem Rev*, vol. 94, no. 7, pp. 1887–1930, 1994.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.