Article

The Importance of Charge Transfer and Solvent Screening in the Interactions of Backbones and Functional Groups in Amino Acid Residues and Nucleotides

Vladimir Sladek [1],* and Dmitri G. Fedorov [2],*

1   Institute of Chemistry, Slovak Academy of Sciences, Dubravska cesta 9, 845 38 Bratislava, Slovakia
2   Research Center for Computational Design of Advanced Functional Materials (CD-FMat) National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba 305-8568, Ibaraki, Japan
*   Correspondence: sladek.vladimir@savba.sk (V.S.); d.g.fedorov@aist.go.jp (D.G.F.)

Abstract: Quantum mechanical (QM) calculations at the level of density-functional tight-binding are applied to a protein–DNA complex (PDB: 2o8b) consisting of 3763 atoms, averaging 100 snapshots from molecular dynamics simulations. A detailed comparison of QM and force field (Amber) results is presented. It is shown that, when solvent screening is taken into account, the contributions of the backbones are small, and the binding of nucleotides in the double helix is governed by the base–base interactions. On the other hand, the backbones can make a substantial contribution to the binding of amino acid residues to nucleotides and other residues. The effect of charge transfer on the interactions is also analyzed, revealing that the actual charge of nucleotides and amino acid residues can differ by as much as 6 and 8% from the formal integer charge, respectively. The effect of interactions on topological models (protein -residue networks) is elucidated.

Keywords: fragment molecular orbital; molecular dynamics; molecular mechanics; non-covalent interactions; residue interaction networks; dielectric constant

1. Introduction

Proteins and nucleic acids are of paramount importance to life. Interactions between amino acid residues and nucleotides are key factors driving many processes in life sciences. Experimental evidence of the role of individual residues is typically indirect, based on mutations and differential properties of mutated species.

Computational studies can reveal contributions of residues to the binding in a given complex. Force fields have been very successful in describing molecular binding in biochemical systems [1]. However, they have their limitations, and it is desirable to compute interactions at a higher level using a quantum mechanical (QM) approach.

It is challenging to apply QM methods to biochemical systems because the cost of calculations scales steeply with their size. Regular QM methods provide total energies, but not interactions between parts of a system. There is a variety of low-scale fragment-based QM methods [2], which not only reduce the cost, but can also deliver the properties of fragments, including interactions between them.

The fragment molecular orbital (FMO) method [3–5] is one such approach. It has been applied to a variety of proteins and nucleic acids [6–12]. In this work, FMO is employed in combination with density-functional tight-binding (DFTB) [13] and the polarizable continuum model (PCM) [14].

Interaction energies [15,16] have been extensively used to analyze protein–ligand [17] and protein–protein binding, with the goal of designing better, more potent drugs in structure and interaction-based drug design (SIBDD) [18]. Interactions can be used as descriptors in structure–activity relationship (SAR) studies [19], for building models of protein residue networks (PRN) [20], and for coarse-grained force fields [21]. For a reliable analysis, interaction energies can be averaged for a selection of snapshots from MD

simulations [9,22–24]. Interaction energies differ from binding energies in the treatment of polarization and desolvation [25].

An important goal of this work is to develop a methodology for analyzing interaction energies between any combination of amino acid residues and nucleotides, appearing in proteins, and their complexes with other proteins and DNA, including DNA aptamers. Aptamers are promising candidates for pharmaceutical applications [26,27].

Secondly, it is of interest to compare interactions in molecular mechanics (MM) and QM and discuss the role of polarization [28] and charge transfer. The polarization of proteins can be critical to ligand binding [29]. In MM, atomic charges are fixed, and the charges of residues are integer numbers. In reality, because of charge transfer, the actual charge of a residue should be fractional. It is a goal of this work to compute interactions for residues with fractional charges, obtained from QM calculations. Although there are force fields that are polarizable [30,31] and some with charge transfer models [32], most MM applications continue to use traditional non-polarizable force fields. It is important to discuss the role of solvent screening, especially for charged nucleotides and amino acid residues. In QM calculations, a robust definition of solvent screening can be obtained based on induced solvent charges [33].

Thirdly, many applications of the FMO rely on interactions between fragments, which are slightly different from conventional residues. In this work, interactions for both typical FMO fragments and conventional residues are computed on equal footing.

Finally, it is of interest to understand the role of backbones. Are they mere scaffoldings, or do they also participate in the binding? If yes, how much do they contribute? Their role has a direct implication for many pharmacological problems, because mutations affect the contributions of side chains, but not backbones (although a mutation can also cause a structural deformation). In particular, understanding the role of backbones can be helpful to rationalizing site-directed mutagenesis and alanine scanning [34].

Biochemical processes occur at physiological temperature and reflect the flexibility of macromolecules. Therefore, in this study, all interactions are averaged over a set of 100 snapshots obtained from molecular dynamics (MD) simulations of a protein–DNA complex.

Interactions between base pairs in nucleic acids have attracted considerable interest with a variety of theoretical [35–40] and experimental [41–43] results available. Interactions between residues in proteins and nucleotides have been analyzed with the FMO [27,44–48].

2. Methods

2.1. Modeling Protein–DNA Complex

The MutS enzyme (PDB ID: 2o8b [49]) was used as a case study for protein–DNA interactions. MutS detects errors induced in the DNA replication by DNA polymerases. The complex of a part of this enzyme with a DNA helix was studied by Bouchal et al. [50] using MD simulations, and this trajectory was used in this work to extract relevant snapshots for conducting QM calculations.

The truncated protein model contains residues from 362 to 532 of the full protein, whereas the DNA has 30 nucleic acid residues. There are 3763 atoms in total, of which 2814 and 949 atoms are in the protein and DNA, respectively. The forward sequence of the DNA double-helix is 5′-GAACCGCXCGCTAGG-3′ and the reverse sequence 5′-CCTAGCGYGCGGTTC-3′, with X = G, Y = T, i.e., a mismatch.

From the 50 ns MD trajectory [50], one snapshot per 10 ps was extracted (5000 snapshots in total). This set of snapshots was used to analyze structural changes in the trajectory. The RMSD of all heavy (non-hydrogen) atoms was computed relative to the first snapshot (Figure 1a). It can be seen that the structure stabilized after 20 ns, with minor RMSD oscillations between 3 and 4 Å. The RMSD histogram in Figure 1b indicates that the structure does not undergo any significant conformational change after the RMSD reaches the value of 3–4 Å, as the corresponding distribution of the RMSD is unimodal, indicating an absence of multiple distinct conformers. Therefore, for conducting QM calculations, the part of the

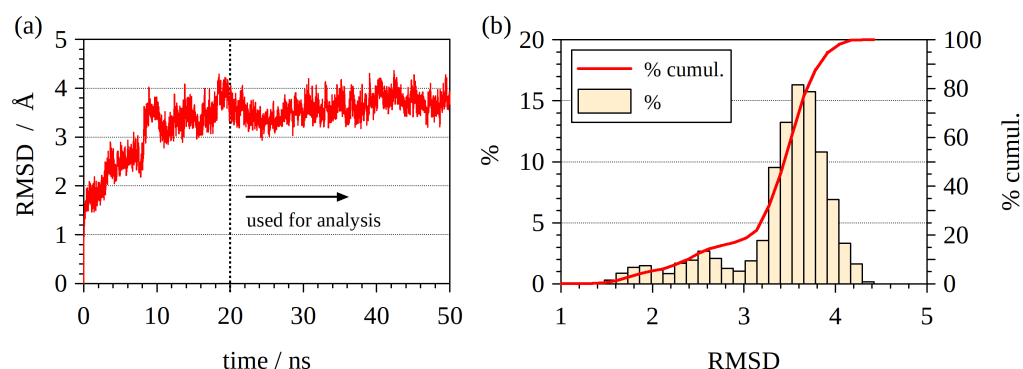MD trajectory between 20 and 50 ns was used, and 1 snapshot per 300 ps was extracted (100 snapshots in total).

Figure 1. (**a**) RMSD (root mean square displacement) of heavy atoms in the trajectory with respect to the first snapshot. (**b**) Histogram of RMSD values. The distribution has a single major peak.

2.2. Evaluation of Interactions

In the FMO method, fragments are defined in such a way to ensure the accuracy in the total properties such as energies or dipole moments. As a result, fragments may differ from the desirable units suitable for a discussion of their properties. FMO fragments in proteins differ from amino acid residues by a CO group, and nucleotides differ by a PO_3 group, namely a residue fragment has the same atomic composition as a conventional residue, but each CO group comes from an adjacent residue, and likewise, phosphate groups are shifted in nucleotides (see Figure S1).

In the FMO, the pair interaction energy (PIE) ΔE_{IJ} between fragments quantifies the binding between two units (nucleotides, etc). In the pair interaction energy decomposition analysis (PIEDA) [23], PIEs are decomposed into components, whose nature depends on the QM method. For DFTB/PCM, a PIE is decomposed into electrostatic (ES), 0-body (electron exchange, correlation, and repulsion), charge transfer (CT) coupling to ES (CT·ES), dispersion (DI), and solvent screening (solv) terms.

$$\Delta E_{IJ} = \Delta E_{IJ}^{ES} + \Delta E_{IJ}^{0} + \Delta E_{IJ}^{CT \cdot ES} + \Delta E_{IJ}^{DI} + \Delta E_{IJ}^{solv} \tag{1}$$

One reason for choosing DFTB in this work is the availability of partition analysis (PA) [51], which has been applied to identify solvent molecules strongly bound to residues in protein crystals [52] and to clarify the transition state stabilization in solid state catalysis [53]. Conceptually, PA resembles partial methods [54].

In PA, the basic units for decomposing DFTB energies are called segments. Segments can be defined exactly as conventional residues, both for amino acids and for nucleotides. Technically, a segment definition is accomplished by using a PDB file, the same one in MM and QM, where atoms are assigned to residues (see Figure S1 for details).

Fragments in the FMO are calculated with a QM method, so that there is an integer number of electrons and protons in each fragment, and hence, fragments themselves bear integer charges. This is similar to MM, where atomic charges in each residue sum up to an integer number. However, in the FMO, fragments are polarized by the electrostatic embedding, so that each fragment has a unique distribution of atomic charges. Charge transfer between fragments (residues), missing in MM, is accounted for at the two-body level in the FMO. On the other hand, in MM, each residue has a fixed set of atomic charges typically determined by atom types [55]. Carbon atoms with different atomic types (for example, $C\alpha$, and carboxyl C) have different atomic charges, albeit in each residue, a given atom type bears a fixed charge, unlike atoms in the FMO. Polarized protein-specific charges from QM can be used in MM simulations [29,56,57].

Segment charges in PA reflect both polarization and charge transfer, incorporated in segment energies (referred to as one-body terms), whereas interaction energies (two-body terms) have no explicit CT contribution and its coupling to the 0-order Hamiltonian, ΔE_{IJ}^0.

To distinguish segments from fragments, fragments are numbered with capital letters such as I and J, whereas segments are numbered with small letters i and j. In PA, pair interaction energies are decomposed into

$$\Delta E_{ij} = \Delta E_{ij}^{\text{ES}} + \Delta E_{ij}^{\text{DI}} + \Delta E_{ij}^{\text{solv}} \tag{2}$$

To discuss solvent effects, the values of the PIEs for the solute are defined as

$$\Delta E_{ij}^{\text{solute}} = \Delta E_{ij}^{\text{ES}} + \Delta E_{ij}^{\text{DI}} \tag{3}$$

There is another difference between fragments and segments. Due to the treatment of covalent boundaries between fragments, a PIE for two fragments connected by a covalent bond is large (for DFTB, on the order of -300 kcal mol^{-1}), but for segments, the values of such PIEs are on par with non-covalent interactions.

DFTB calculations were performed at the level of FMO2-DFTB3/C-PCM$\langle 1 \rangle$ with D3(BJ) dispersion [58], 3ob [59] parameters, and partial screening [33], as implemented [60] in GAMESS [61].

In MM, pair interactions are decomposed into electrostatic (elec) and van der Waals (vdW) terms as

$$\Delta E_{ij} = \Delta E_{ij}^{\text{elec}} + \Delta E_{ij}^{\text{vdW}} \tag{4}$$

$$\Delta E_{ij}^{\text{elec}} = \sum_{A \in i} \sum_{B \in j} \frac{Q_A Q_B}{\varepsilon R_{AB}} \tag{5}$$

$$\Delta E_{ij}^{\text{vdW}} = \sum_{A \in i} \sum_{B \in j} \frac{C_{AB}^{12}}{R_{AB}^{12}} - \frac{C_{AB}^6}{R_{AB}^6} \tag{6}$$

where A and B are the number of atoms, R_{AB} is the distance between A and B, Q_A is the charge of atom A, ε is the dielectric constant, and C_{AB}^{12} and C_{AB}^6 are parameters. The AMBER force field imposes some restrictions on the sums (not explicitly shown in Equation (5)), excluding some close pairs of atoms A, B and weighing 1–4 interactions of terminal atoms of dihedrals [62].

In this work, the ff14SB [63] and Parmbsc1 [64] parameters were used for the protein and DNA, respectively, the same parameters as in the MD trajectory [50].

In Equation (5), ε is the relative permittivity of the medium, also known as the dielectric constant. It describes the solvent screening. In AMBER, it is conventional to use $\varepsilon = 1$ [62]. Other values of ε are used in computing binding energies in Poisson–Boltzmann models [65]. The inability of force fields to take into account local variations of the electric field is often criticized as a drawback when evaluating accurate binding free energies [66,67]. On the other hand, in QM/PCM calculations, the locality of solvent screening is fully taken into account. Fragmentwise effective values of ε representing the local features of the solute/solvent interface can be defined in QM/PCM, although they are not directly used for computing the screening [68].

The actual dielectric constant of "dry" proteins was estimated to be $\varepsilon = 3.23 \pm 0.04$ [69]; however, this global value does not describe local effects. The dielectric constant inside a protein may differ from that on its surface at the solute/solvent interface. It was assessed [70] that ε is $6 \ldots 7$ inside a protein and $20 \ldots 30$ on the surface. Elaborate schemes for dealing with solvent screening in MM have been proposed [71,72]; however, the bulk of common MD simulations continues to use the traditional vacuum form of the Coulomb interaction in the solute with $\varepsilon = 1$, whereby the solute is in a mechanical embedding of the solvent, and both the solute and solvent are not polarizable.

In DFTB, the ES term is computed with a different functional form than in force fields, although $\varepsilon = 1$ is likewise used, whereas the solvent screening is described in the separate solv term in Equation (2). In addition to the use of polarizable atomic charges in DFTB compared to fixed charges in MM, damping functions γ and Γ describe the effect of charge penetration for close interatomic contacts, as well as charge transfer via the chemical hardness in the third-order Γ terms [73].

$$\Delta E_{ij}^{ES} = \sum_{A \in i} \sum_{B \in j} \left\{ \left(\gamma_{AB} + \frac{1}{3} \Gamma_{AB} Q_A + \frac{1}{3} \Gamma_{BA} Q_B \right) Q_A Q_B \right\}. \tag{7}$$

On the other hand, the dispersion in DFTB is conceptually very similar to the vdW term in force fields, a non-polarizable contribution, computed, however, with different parameters so that the values of dispersion in QM and MM differ.

2.3. Definition of Fragments and Segments

In this work, the structural units used for the analyses are as follows. In MM and PA, conventional nucleotides and amino acid residues were used (segments). FMO fragments have a shifted phosphate (nucleotides) or carbonyl (amino acids); see Figure S1.

Side chains in amino acid residues and bases in nucleotides are referred to as functional units (f). The remaining part in each unit is called the backbone (b). This splitting was only performed for segments in PA and MM, but not for fragments in the FMO.

A detailed atomic structure of all nucleotide pairs is shown in Figures 2 and S3, where the atom names as used in the PDB are also indicated.
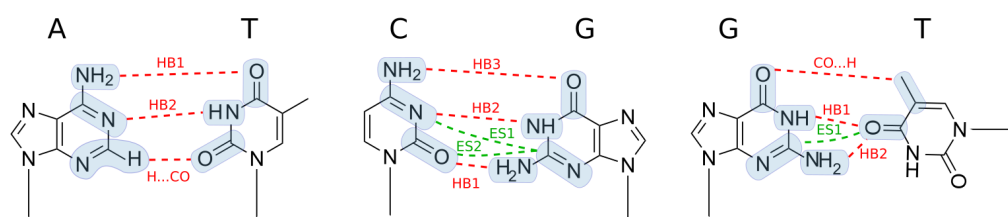

Figure 2. Three types of base pairs, adenine–thymine (A13:T18, regular), cytosine–guanine (C7:G24, regular), and guanine–thymine (G8:T23, mismatch). Hydrogen bonds (HB) and electrostatic (ES) interactions between functional groups (shadowed in grey) are labeled. See also Figure S3 for more details.

3. Results and Discussion

3.1. Interactions between Residues

Pair interaction energies for selected pairs (Figure S2), averaged over all 100 sampled structures, are shown in Table 1. For all interactions, the standard deviations are quite small, typically a few percent, indicating that the structure does not undergo major changes during the production part of the MD trajectory.

The most direct comparison of MM with QM can be performed for MM(solute,res) with QM(PA,solute,res), because MM interactions are between solute atoms without solvent screening, similar to the solute QM results. The correlation between MM and QM solute interactions is shown in Figure 3a. The correlation coefficient R^2 is high, 0.993. As can be seen from Table 1, solute pair interaction energies can be very large (up to -105.1 kcal mol^{-1}), because they are computed in a vacuum (although QM solute results reflect the polarization of the solute by the solvent).

The last pair in Table 1, Met107:Asp108, is excluded from Figure 3a. This pair is special because there is a covalent bond between the two residues. The MM and QM interactions are -27.6 and -4.9 kcal mol^{-1}, respectively. In MM, near-neighbor interactions between some atoms are excluded, but in QM, all pairs of atoms contribute. In QM(DFTB), the electrostatic interaction is not of a simple Coulomb form, but includes a damping in the form of the γ and Γ functions (Equation (7)) modeling charge penetration at short distances.

To discern the origin of the QM and MM discrepancy, the ES term was calculated using MM charges for all atom pairs, which decreased the absolute value by only 2 kcal mol^{-1}. Therefore, the bulk of the difference between −27.6 (MM) and −4.9 kcal mol^{-1} (QM) is due to other reasons, that is the damping, the three-body ES term in DFTB, and the distribution of atomic charges in the residues.

Table 1. Selected pair interaction energies (kcal mol^{-1}) in the protein–DNA complex (PDB: 2o8b), averaged over 100 snapshots of the MD trajectory [a].

Pair (State,Unit) (Unit Charge) (Definition)	MM(Amber) Solute,Res Integer ΔE_{ij} in Equation (4)	QM(PA) Solute,Res Fractional ΔE_{ij} in Equation (3)	QM(PA) Solution,Res Fractional ΔE_{ij} in Equation (2)	QM(PA) Solution,Frag Fractional ΔE_{ij} in Equation (2)	QM(FMO) Solution,Frag Integer ΔE_{IJ} in Equation (1)
C7:G24	−8.3 ± 1.8	−13.7 ± 2.7	−21.3 ± 2.3	−21.3 ± 2.3	−21.7 ± 2.1
A13:T18	8.8 ± 1.4	10.1 ± 1.8	−8.8 ± 1.4	−8.8 ± 1.4	−8.9 ± 1.1
G6:C25	−9.0 ± 1.6	−13.7 ± 2.8	−22.8 ± 2.1	−22.7 ± 2.1	−23.4 ± 1.6
A3:T28	8.8 ± 1.3	9.9 ± 1.9	−8.8 ± 1.2	−8.7 ± 1.2	−8.9 ± 1.2
A2:T29	9.1 ± 1.3	10.4 ± 1.7	−8.9 ± 1.2	−8.8 ± 1.2	−8.9 ± 1.1
G8:T23	7.8 ± 1.2	5.3 ± 1.8	−13.4 ± 1.6	−13.5 ± 1.6	−13.0 ± 1.5
G10:Arg180	−70.6 ± 2.6	−64.7 ± 4.0	−20.2 ± 3.0	1.9 ± 0.5	−1.5 ± 0.5
C9:Arg180	−53.6 ± 2.7	−50.5 ± 3.2	−12.6 ± 2.0	−33.3 ± 2.9	−32.9 ± 2.8
G24:Phe101	−5.3 ± 0.6	−5.8 ± 1.1	−4.2 ± 0.6	−3.9 ± 0.6	−4.1 ± 0.5
C7:Phe101	−0.6 ± 0.4	−1.0 ± 1.1	−0.6 ± 0.3	−0.6 ± 0.3	0.1 ± 0.5
Gly78:Tyr96	−4.9 ± 1.6	−7.0 ± 1.7	−6.7 ± 1.8	−0.6 ± 0.2	−0.6 ± 0.3
Trp82:Lys86	−12.4 ± 1.2	−15.7 ± 2.8	−12.3 ± 1.5	−6.1 ± 0.9	−4.8 ± 1.0
Trp82:Met107	−0.7 ± 0.6	−1.4 ± 0.3	−1.3 ± 0.3	−1.4 ± 0.3	−1.2 ± 0.2
Lys86:Asp108	−105.1 ± 3.2	−104.9 ± 4.0	−40.2 ± 3.7	−40.2 ± 3.7	−43.1 ± 3.9
Met107:Asp108	−27.6 ± 1.6	−4.9 ± 2.3	−5.4 ± 0.8	−5.9 ± 0.9	−394.0 ± 14.5

[a] QM at the level of FMO2-DFTB3/PCM/D3(BJ)/3ob; res refers to conventional residues, whereas frag denotes conventional fragments in the FMO. Integer or fractional refer to the total charge on residues. The interaction energies in solution include solvent screening.
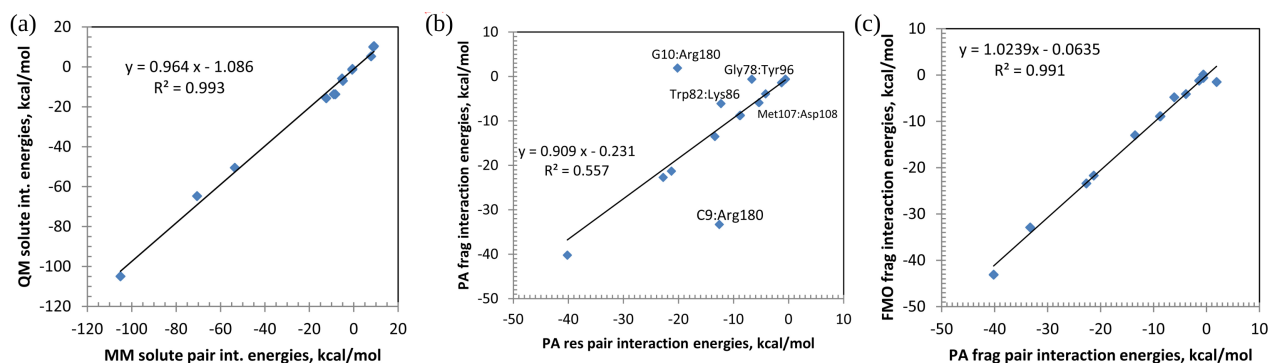

Figure 3. Correlation between pair interactions for conventional residues (res) and fragments (frag): (**a**) MM(solute,res) vs. QM(PA,solute,res), (**b**) PA(solution,res) vs. PA(solution,frag), and (**c**) PA(solution,frag) vs. FMO(solution,frag).

The largest PIEs are between charged residues (the formal charges of residues are shown in Figure S2). All nucleotides have the formal charges of −1. The atoms in the phosphate group bear most of that charge. Two negatively charged nucleotides may still have an overall attractive interaction, and the strong Coulomb repulsion of the phosphate groups is somewhat compensated by the attraction of the nucleoside bases.

The electrostatic interaction and solvent screening are large for ionic pairs, e.g., for the pair G10:Arg180, the value of −64.7 in QM(PA,solute,res) is reduced to −20.2 (kcal mol^{-1}) in QM(PA,solution,res) by adding the solvent screening. A similarly large weakening of the interactions in solution is seen for the pair Lys86:Asp108, due to the addition of solvent screening $\Delta E_{ij}^{\text{solv}}$. The inclusion of the screening can even change the sign of the PIEs, as seen for all A:T pairs and the mismatched G8:T23 pair. The slope in Figure 3a indicates that MM interactions are overestimated by a few % vs. QM.

Next, it is interesting to compare the interactions between conventional residues and shifted fragments. The results are shown in Figure 3b. Although there is a nice correlation for most pairs, there are two noticeable outliers, G10:Arg180 and C9:Arg180. The PO_3 group is shifted in the FMO from C9 to G10, as shown in Figure S1. It can be seen in Figure S2 that Arg180 interacts via its positively charged side chain with the negatively charged phosphate group of G10. The PO_3 group of residue G10 is, however, assigned to fragment C9. The PIE of C9:Arg180 contains the interaction of NH_2^+ from Arg180 and the negatively charged phosphate group from G10 and O3′ of C9. This is the reason for a mismatch between interactions as defined for fragments and segments in this case. If the pair interaction energies of C9:Arg180 and G10:Arg180 are added, then the res and frag results are very similar, -32.8 and -31.4 kcal mol^{-1}, respectively. Thus, the difference between the results is due to the shifted definitions of units, fragments vs. residues.

A similar mismatch occurs in Gly78:Tyr96 and Trp82:Lys86. In the former, one H-bond between the phenolic OH group of Tyr96 and the CO group of the Gly78 backbone is assigned to fragment 79, not Gly78 fragment. For the latter pair, it is the CO backbone group of Trp82 that forms a H-bond with the NH backbone group of Lys86. The CO group of Trp82 is assigned to another fragment.

The fact that the point for the connected pair Met107:Asp108 does not stand out in Figure 3b (the regression line almost passes through it) can be seen as the success of PA in defining pair interaction energies for residues connected by a covalent bond.

Figure 3c shows the correlation between fragments in PA and the FMO. The point of this comparison is to probe the effect of charge transfer: in the FMO, fragments have integer charges reflecting polarization; in PA, charges are fractional, computed at a higher level. They reflect both polarization and charge transfer (the fractional and integer residue charges differ by about 2%). There is a high correlation between them with the correlation coefficient of $R^2 = 0.991$. The last pair Met107:Asp108 in Table 1 is excluded from this plot, also because of the difference of defining pair interactions related to a covalent bond. In the FMO, due to the boundary definition, the interaction energy is large, -394.0 kcal mol^{-1}, but in PA, it is -5.9 kcal mol^{-1}.

3.2. Charges of Residues

The charges of residues are shown in Table 2. The charges in the non-polarizable force field are fixed and do not depend on the conformation (no deviations). On the other hand, the charges in QM reflect conformation-specific polarization and charge transfer. It can be seen that the QM(res) from the MM(res) charges differs by at most about 8% (for Asp108), whereas for QM(frag), the deviations are slightly more, because of the difference in the definition of residues. In MM, charges are not specific, so that, for example, all A nucleotides have the same charge. In contrast, QM charges show variation depending on the environment.

The charges of functional units (side chains for amino acids and bases for DNA) in MM(res) and QM(res) are similar. The charge transfer amount of a few percent agrees with other studies of the many-body charge transfer [4,74].

It is interesting to observe that the division of the negative charge of -1 between the phosphate+pentose and base is roughly the same for all four types (the values for QM(res) are discussed next): the phosphate+pentose part obtains $-0.75 \ldots -0.80$, and the rest of the charge goes to the base. The bases in A, C, G, and T have the charges of -0.15, -0.14, -0.19, and -0.20, respectively. For amino acids, it is the side chain that bears almost all of the formal charge. Actual charges for side chains deviate within 9% from the total formal charges.

Other QM studies report a similar behavior. At the level of FMO2-MP2/6-31G*, natural population analysis yielded [46] the base charges of -0.17 for A, -0.18 for C, -0.16 for G, and -0.19 for T, which are in very good agreement with the DFTB Mulliken charges in this work.

Table 2. Charges (a.u.) of residues averaged in MD [a].

Residue	MM(total) [b]	QM(res,total)	QM(frag,total)	MM(funct) [b]	QM(res,funct)
A2	−1.0	−0.990 ± 0.035	−0.988 ± 0.023	−0.105	−0.155 ± 0.017
A3	−1.0	−0.986 ± 0.033	−0.987 ± 0.022	−0.105	−0.153 ± 0.016
G6	−1.0	−1.025 ± 0.034	−1.024 ± 0.021	−0.089	−0.195 ± 0.018
C7	−1.0	−0.967 ± 0.036	−0.956 ± 0.027	−0.063	−0.141 ± 0.018
G8	−1.0	−1.012 ± 0.034	−1.014 ± 0.022	−0.089	−0.191 ± 0.016
C9	−1.0	−0.957 ± 0.035	−0.918 ± 0.024	−0.063	−0.148 ± 0.016
G10	−1.0	−0.955 ± 0.033	−0.970 ± 0.029	−0.089	−0.189 ± 0.017
A13	−1.0	−0.990 ± 0.033	−0.983 ± 0.023	−0.105	−0.154 ± 0.016
T18	−1.0	−1.018 ± 0.027	−1.018 ± 0.018	−0.127	−0.205 ± 0.014
T23	−1.0	−0.986 ± 0.032	−0.985 ± 0.022	−0.127	−0.193 ± 0.016
G24	−1.0	−0.994 ± 0.037	−0.967 ± 0.028	−0.089	−0.195 ± 0.020
C25	−1.0	−0.941 ± 0.030	−0.964 ± 0.022	−0.063	−0.144 ± 0.018
T28	−1.0	−1.012 ± 0.034	−1.015 ± 0.021	−0.127	−0.204 ± 0.017
T29	−1.0	−1.017 ± 0.033	−1.009 ± 0.022	−0.127	−0.205 ± 0.017
Gly78 [c]	0.0	−0.014 ± 0.038	−0.015 ± 0.019		
Trp82	0.0	0.016 ± 0.040	0.009 ± 0.018	0.029	0.038 ± 0.010
Lys86	1.0	0.956 ± 0.037	0.929 ± 0.027	1.026	0.981 ± 0.018
Tyr96	0.0	−0.009 ± 0.037	−0.020 ± 0.022	0.028	0.010 ± 0.013
Phe101	0.0	−0.003 ± 0.033	−0.012 ± 0.019	0.019	0.014 ± 0.009
Met107	0.0	−0.003 ± 0.034	−0.007 ± 0.023	0.050	0.041 ± 0.010
Asp108	−1.0	−0.921 ± 0.035	−0.917 ± 0.022	−0.858	−0.911 ± 0.015
Arg180	1.0	0.924 ± 0.043	0.914 ± 0.025	1.036	0.943 ± 0.015

[a] QM at the level of PA combined with FMO2-DFTB3/PCM/D3(BJ)/3ob. Total refers to the total charge; funct is the charge of the functional part (side chain for amino acids and base for DNA). Here, res refers to conventional residues, whereas frag denotes fragments, which are shifted by a few atoms. [b] All residues of the same type in MM have the same charges (computed from the .prmtop file), identical to published RESP values [55]. [c] Glycines have no side chains.

3.3. Components of Interactions

For a more detailed comparison of QM and MM, individual components for pair interactions are shown in Table 3 and Figure 4 (the covalently bound pair of Met107:Asp108 is excluded). There is a good correlation of the electrostatic term, but no correlation for the dispersion. The solvent screening is an important contribution to the interaction energies.

Table 3. Components of pair interaction energies (kcal mol^{-1}) in the protein–DNA complex, averaged over 100 snapshots of the MD trajectory.

| Pair | MM(res) [a] | | QM(PA,res) [b] | | |
	ΔE_{ij}^{elec}	ΔE_{ij}^{vdW}	ΔE_{ij}^{ES}	ΔE_{ij}^{DI}	ΔE_{ij}^{solv}
C7:G24	−8.6 ± 2.8	0.2 ± 1.9	−10.1 ± 2.7	−3.7 ± 0.2	−7.6 ± 1.7
A13:T18	9.5 ± 1.9	−0.8 ± 1.2	13.1 ± 1.7	−2.9 ± 0.3	−18.9 ± 1.2
G6:C25	−9.3 ± 2.8	0.2 ± 1.8	−10.0 ± 2.8	−3.7 ± 0.2	−9.0 ± 1.5
A3:T28	9.4 ± 1.7	−0.6 ± 1.2	12.8 ± 1.8	−2.9 ± 0.2	−18.7 ± 1.4
A2:T29	9.9 ± 1.9	−0.9 ± 1.3	13.3 ± 1.7	−2.9 ± 0.3	−19.3 ± 1.4
G8:T23	8.5 ± 1.5	−0.7 ± 1.1	8.0 ± 1.8	−2.7 ± 0.3	−18.7 ± 1.5
G10:Arg180	−70.7 ± 3.2	0.1 ± 1.3	−61.4 ± 3.9	−3.2 ± 0.3	44.4 ± 2.3
C9:Arg180	−52.6 ± 2.7	−1.0 ± 0.3	−49.3 ± 3.2	−1.2 ± 0.2	37.9 ± 2.2
G24:Phe101	−1.9 ± 0.3	−3.4 ± 0.5	−1.9 ± 0.9	−3.8 ± 0.5	1.6 ± 0.9
C7:Phe101	0.0 ± 0.3	−0.6 ± 0.3	−0.5 ± 1.0	−0.5 ± 0.2	0.4 ± 1.0
Gly78:Tyr96	−4.7 ± 1.9	−0.2 ± 0.9	−5.8 ± 1.7	−1.2 ± 0.2	0.2 ± 0.5
Trp82:Lys86	−9.5 ± 1.2	−2.9 ± 0.8	−11.1 ± 2.7	−4.6 ± 0.6	3.4 ± 2.0
Trp82:Met107	−0.2 ± 0.1	−0.5 ± 0.6	−0.2 ± 0.1	−1.2 ± 0.2	0.1 ± 0.1
Lys86:Asp108	−106.5 ± 4.4	1.5 ± 1.7	−102.9 ± 3.9	−1.9 ± 0.2	64.6 ± 3.3
Met107:Asp108	−26.8 ± 1.5	−0.8 ± 0.8	−0.4 ± 2.3	−4.5 ± 0.3	−0.5 ± 1.8

[a] The sum of MM terms (Equation (4)) is given in Table 1 as MM(Amber). [b] The sum of QM terms (Equation (2)) is given in Table 1 as QM(PA,total,res), whereas the sum (Equation (3)) is QM(PA,solute,res). QM calculations are at the FMO2-DFTB3/PCM/D3(BJ)/3ob level.
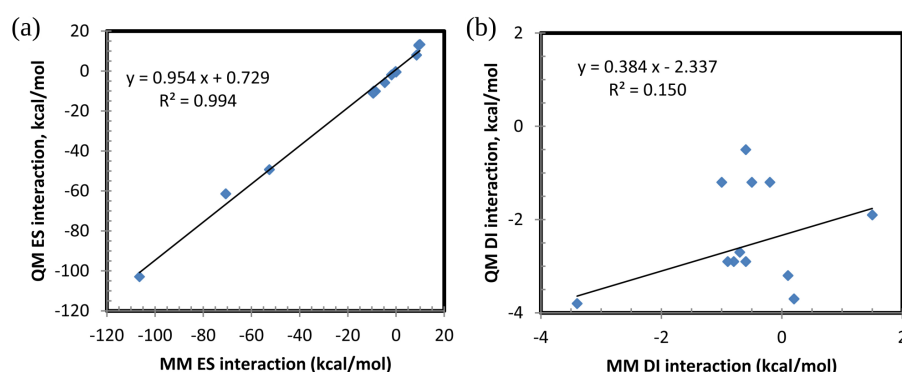
Figure 4. Correlation between components of pair interactions for conventional residues (res), MM vs. QM: (**a**) electrostatics (ES vs. elec) and (**b**) dispersion (DI vs. vdW).

3.4. Contributions of Backbones and Functional Units

To gain more insight, amino acid residues were split into backbones and side chains and nucleotides were split into the phosphate+pentose backbones and bases. For the uniformity of the notation, side chains and bases are called functional units. The contributions of backbones and functional units are shown in Table 4.

Table 4. Contributions of backbones (b) and functional units (f) to the total interaction energies (kcal mol^{-1}) ^{a}, averaged in MD.

Pair	MM(res,vacuum)				QM(res,solute)				QM(res,solution)			
	bb	bf	fb	ff	bb	bf	fb	ff	bb	bf	fb	ff
C7:G24	14.9	0.4	2.9	−26.4	10.6	2.6	3.4	−30.4	0.1	0.2	0.0	−21.7
A13:T18	12.6	3.8	2.4	−9.9	10.4	4.9	2.7	−7.9	0.1	0.0	−0.2	−8.8
G6:C25	14.2	3.2	0.3	−26.7	10.2	3.9	2.3	−30.2	0.1	−0.1	0.2	−23.0
A3:T28	12.6	3.8	2.4	−10.0	10.4	4.8	2.6	−7.9	0.1	0.0	−0.2	−8.8
A2:T29	12.8	3.8	2.5	−10.0	10.6	4.9	2.8	−7.9	0.1	0.0	−0.2	−8.9
G8:T23	14.4	3.3	3.4	−13.3	10.9	4.1	5.5	−15.4	0.1	0.1	0.1	−13.8
G10:Arg180	1.8	−69.9	0.1	−2.6	−0.1	−57.8	−0.0	−6.8	−0.1	−20.1	0.0	−0.1
C9:Arg180	5.1	−54.9	0.4	−4.2	−0.4	−41.9	−0.0	−8.1	0.0	−11.8	0.0	−0.8
G24:Phe101	−0.8	−2.3	−0.2	−1.9	−0.8	−1.4	−0.2	−3.3	−0.1	−1.0	−0.1	−2.9
C7:Phe101	0.7	0.0	−0.6	−0.7	0.0	−0.2	−0.5	−0.3	0.0	−0.1	−0.1	−0.4
Gly78:Tyr96	−0.2	−4.5			−0.3	−6.7			0.0	−6.7		
Trp82:Lys86	−3.9	−4.8	−0.7	−3.0	−6.9	−5.5	−0.8	−2.4	−6.2	−1.5	−0.6	−3.9
Trp82:Met107	0.0	0.0	−0.1	−0.6	0.0	0.0	−0.3	−1.1	0.0	0.0	−0.2	−1.1
Lys86:Asp108	0.1	1.3	−4.7	−101.8	0.0	0.5	4.3	−109.7	0.0	0.0	1.3	−41.5
Met107:Asp108	−23.0	−1.6	−1.5	−1.5	−0.9	−0.1	−1.6	−2.2	−1.7	−1.8	−1.7	−0.1

^{a} Standard deviations are not shown. As an example to explain the notation, bf for C7:G24 means the backbone (b) of C7 and the functional (f) unit of G24. QM calculations are at the level of PA with FMO2-DFTB3/PCM/D3(BJ)/3ob. Glycines have no side chains.

The most direct correspondence is between MM(res,vacuum) and QM(res,solute). MM calculations were performed in explicit solvent, which, however, did not affect pair interactions because MM is not polarizable and there is no solvent screening. On the other hand, QM(res,solute) calculations reflect solute polarization, to which solvent screening is added in QM(res,solution).

The correlation between the MM(res,vacuum) and QM(res, solute) results is shown in Figure 5. The correlation for backbone–backbone (bb) interactions is high, $R^2 = 0.945$, but the slope of 0.852 indicates that interactions in QM are systematically smaller than MM by about 15%. It is worth noting that bb interactions are relatively small, ranging between −7 and 15 kcal mol^{-1}, but nevertheless, the interactions are strongly correlated. This is attributed to the electrostatic nature of these interactions.
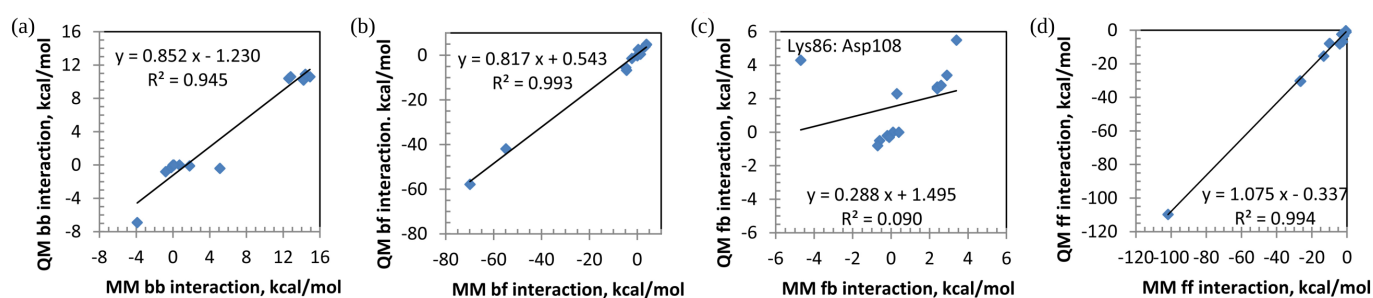
Figure 5. Correlation between pair interactions for conventional residues (res) divided into backbone (b) and functional (f) parts, MM vs. QM (solute): (**a**) bb, (**b**) bf, (**c**) fb, and (**d**) ff. For the distinction between bf and fb, see Table 4.

For bf interactions, the correlation is even higher than for bb, with $R^2 = 0.993$, and a similar systematical difference of about 18% is observed between MM and QM. It may be noted that in the bf pairs, b is always negatively charged (as shown above, the backbone charge is about -0.8), whereas f may be both charged (cationic side chains in Arg180 and Lys86 and anionic side chain in Asp108) and neutral (the rest, although the side chains of nucleotides do have a charge of about -0.2).

On the other hand, fb interactions show low correlation when all points are considered, which is caused by a single outlier, Lys86:Asp108. The reason why Lys86:Asp108 is rather different in QM and MM may be that, in QM, there is a strong polarization of these residues, not described in MM.

For ff interactions, the correlation is high, $R^2 = 0.994$, and the slope indicates a systematic deviation of about 8% in the absolute values.

The effect of the solvent screening is included in the QM(res,solution) set of results in Table 4. These results reveal many interesting details about the interaction in solution.

First, backbone–backbone interactions for nucleotides are essentially zero (0.1 kcal mol^{-1}). This is because nucleotides bind via the base and the anionic phosphate+pentose faces the solvent (or protein). Ignoring solvent screening would result in the opposite conclusion, that backbones are important and repel each other (10 kcal mol^{-1} at the QM level).

Second, backbone(amino acid)–backbone(nucleotide) interactions are likewise essentially zero (-0.1 kcal mol^{-1}), for the same reason as for nucleotidic pairs. However, backbone–backbone interactions for two amino acid residues can be substantial. This is because peptide backbones have CO and NH groups that can form hydrogen bonds. These backbone–backbone interactions are essential for secondary structure stabilization and are not directly affected by mutations. It is very important for drug design to identify if an interaction is for the side chain or the functional group [34].

Third, both bf and fb interactions for nucleotides are totally screened (the largest is 0.2 kcal mol^{-1}), attributed to the phosphates facing the solvent. As the backbone of amino acid residues is neutral, functional(nucleotide)–backbone(amino acid) interactions are also nearly zero (-0.1 kcal mol^{-1}). However, for backbone(nucleotide)–functional(amino acid), there is some residual interaction for charged amino acids, which is only partially screened. Both bf and fb interactions for two amino acids may be substantial, because backbones can form hydrogen bonds.

Fourth, many (but not all) residue–residue interactions are dominated by the functional unit pairs (ff). For nucleotides, it is the only significant interaction. When one residue is an amino acid, then the functional unit in the amino acid is the principle contributor to the interaction as a rule, although for some pairs, the backbone can contribute as much or even more, as in the Trp82:Met107 pair.

3.5. Comparison of Base Pairs

In order to analyze the stability of natural and mismatched base pairs, the contributions of individual hydrogen bonds were computed by defining segments appropriate for the purpose. The details are shown in Figures 2 and S3.

The results are shown in Table 5, where the total ff contribution in Table 4 is split among functional groups. Each functional unit (that is, a Watson–Crick base) is split into functional groups, as shown in Figure 2. For example, to obtain the interaction energy for HB1 in A13:T18, two small segments are defined, one for NH_2 of A13 and another for CO of T18. There are many pairs of such small segments, and only the most important of them is explicitly listed in Table 5. The contribution of the rest of the segment pairs is summed up and listed as the "rest" term. The segment pairs with a substantial electrostatic interaction are listed as weak ES "bonds".

Table 5. Interactions ΔE_{ij} (kcal mol^{-1}) between functional groups forming hydrogen bonds (HBs), $CO\cdots H$ bonds, and electrostatic (ES) bonds in selected base pairs in the protein–DNA complex, averaged over 100 snapshots of the MD trajectory [a].

Bond	A13:T18	C7:G24	G8:T23
HB1	-3.7 ± 1.2	-4.8 ± 1.1	-3.5 ± 1.1
HB2	-4.3 ± 0.6	-4.4 ± 0.6	-4.3 ± 0.9
HB3		-4.9 ± 0.9	
$CO\cdots H$	-0.8 ± 0.4		-1.1 ± 0.4
ES1		-2.2 ± 0.4	-2.7 ± 0.4
ES2		-1.0 ± 0.5	
rest	0.0 ± 0.5	-4.4 ± 1.2	-2.2 ± 0.9
total	-8.8 ± 1.4	-21.7 ± 2.3	-13.8 ± 1.6

[a] Computed at the QM(res,solution) level with FMO2-DFTB3/PCM/D3(BJ)/3ob. The total value here is equal to the ff value in Table 4. See Figure 2 for a picture of the bonds.

As can be seen from Table 4, the three nucleotide pairs chosen for Table 5 are representative as there is little variation of interactions for a given pair of nucleotides due to the local environment.

To complement the energetics, bond distances were computed between the closest pair of heavy atoms in each bond. The main results are in Table 6, and the details of the statistical averaging are shown in Figure S4.

Table 6. Theoretical and experimental bond lengths R_{AB} (Å) in Watson–Crick base pairs [a].

Pair	Bond	A	B	Amber [b]	MP2/TRVPP [c]	BP86/TZ2P [d]	Expt. [e]
A13:T18	HB1	N6	O4	3.04 ± 0.22	2.86	2.85	2.95/2.93
	HB2	N1	N3	2.95 ± 0.11	2.83	2.81	2.82/2.85
	$CO\cdots H$	C2	O2	3.59 ± 0.26			
C7:G24	HB1	N1	N3	2.96 ± 0.10	2.90	2.88	2.95
	HB2	O6	N4	2.90 ± 0.13	2.75	2.73	2.91
	HB3	O2	N2	2.88 ± 0.13	2.89	2.87	2.86
G8:T23	HB1	O4	N1	3.03 ± 0.18			
	HB2	O4	N2	2.87 ± 0.11			
	$CO\cdots H$	C7	O6	3.47 ± 0.29			

[a] Atoms $A \in$ first base and $B \in$ second base are the atom names used in the PDB file. [b] Base pairs are embedded in the solvated protein–DNA complex, from the MD trajectory [50] processed in this work. [c] RI-MP2/TRVPP geometries optimized in a vacuum [35]. [d] BP86/TZ2P geometries optimized in a vacuum [75]. [e] X-ray crystallographic measurements in sodium adenylyl-3′,5′-uridine hexahydrate [76] as a model of a A:T pair (two values for different environments) and in sodium guanylyl-3′,5′-cytidine nonahydrate [77] as a model of a C:G pair.

There is a clear correlation between R_{AB} in Table 6 and the interaction ΔE_{ij} in Table 5: the shorter the bond, the stronger the interaction. Hydrogen bonds have a much shorter length of about 2.95 Å, compared to 3.53 Å for $CO\cdots H$ bonds [78].

3.6. Comparison of Energies

A comparison of computational and experimental results is summarized in Table 7. Due to variations in both the chemical composition (bases vs. nucleotides) and environment (artificially capped vs. naturally embedded), it is difficult to compare various values directly.

Table 7. Association energies [a] (kcal mol^{-1}) for neutral Watson–Crick base pairs and anionic nucleotides.

Embedding	Pair	Solvent	Method	Reference	C:G	A:T	G:T
none [b]	nucleotides	vacuum [b]	AMBER	this work	−8.7	8.9	7.8
embedded [c]	nucleotides	water	M06-2X/6-31G(d,p)/PCM	[38]	−16.8	−10.1	
embedded	nucleotides	water	FMO2-DFTB3/PCM	this work	−22	−8.8	−13.4
none	bases	vacuum	AMBER (parameters of Ref. [55])	[36]	−28.0	−12.8	
none [b]	bases	vacuum [b]	AMBER	this work	−26.6	−10.0	−13.3
none	bases	vacuum	CNDO-CI	[41]	−16.79	−7.00	−9.86
none	bases	vacuum	B3LYP/6-31G **	[35]	−25.5	−12.3	
none	bases	vacuum	B3LYP/D95 **	[37]	−26.65	−12.56	
none	bases	vacuum	RI-MP2/a[TQ]Z [d]	[35,39]	−28.2	−15.4	
none	bases	vacuum	MP2/D95 **	[37]	−24.50	−11.69	
none	bases	vacuum	IQA/M06-2X [e]	[40]	−29.1	−14.4	−14.6
embedded	bases	water [f]	FMO2-MP2/6-31G *	[46]	−51 [g]	−20	
embedded	bases	vacuum	FMO2-MP3/CP/MCP/dzp	[47]	−38	−13	
embedded	bases	vacuum	FMO2-CCSD(T)/MCP/dzp	[47]	−45	−20	
embedded	bases	water	FMO2-DFTB3/PCM	this work	−22	−8.8	−13.8
none	bases [h]	none	Experiment(enthalpy)	[42]	−21.0	−13.0	
none	bases	none	Experiment(enthalpy)	[43]	−10.75		

[a] Interaction energies for FMO methods; otherwise, binding energies. Pairs are hydrogen capped in standalone model systems (no embedding) or computed inside actual DNA (embedded). Watson–Crick bases are conceptually comparable to the ff values in Table 4. [b] In mechanical embedding. [c] In a A-DNA trimer duplex. [d] aTZ → aQZ BSL extrapolated [79]. [e] Bond dissociation energies. [f] With solvent polarization, but no explicit screening. [g] There is some variation for different pairs, and the average is given with the number of significant figures determined by the variation. [h] Methylated monomers, measured in vapor.

Many theoretical methods in a vacuum predict the binding between C and G bases to be between −25 and −28 kcal mol^{-1}, both QM and MM, which is close to −22 kcal mol^{-1} for DFTB in this work and to one experimental enthalpy of −21.0 kcal mol^{-1}.

For A:T bases, many theoretical estimates are between −12 and −15 kcal mol^{-1}, whereas the DFTB result in this work is −9 kcal mol^{-1}, and the experimental value is −13 kcal mol^{-1}. For G:T bases, theoretical predictions are −10... −15, and the DFTB result in this work is −14 kcal mol^{-1}. The general trend is that the binding decreases in the order of C:G, G:T, and A:T, consistent with the present work. Some part of the difference between DFTB and other QM methods may be due to the small basis set superposition error in DFTB, owing to not treating core electrons as particles, which leads to smaller values of interactions [23].

The MM interaction energies in this work have no explicit screening and the repulsion between the anionic basis overwhelms the attraction for A:T and G:T pairs, so they are repulsive. It is shown above that a similar trend was observed for the solute-only part of the interactions at the DFTB level.

3.7. Statistical Analysis of All Interactions in Topological Models

While it is useful to discuss individual interactions in detail, an integral measure can be defined for a complete set of all interactions. This can be conveniently performed with protein residue network (PRN) models.

These are topological representations of the interaction network in proteins (or other molecular systems) as graphs (Figure 6), where residues are represented as vertices (nodes) and the interactions between them as edges connecting the vertices [80–83]. Accord-

ing to the convention [20], only attractive interactions stronger than a threshold (usually -1 kcal mol^{-1}) were used. The PRN can help to elucidate the role of individual residues [84] in static or dynamic models [85–90]. Macromolecular interfaces can be studied using singular-value decomposition (SVD) analysis [91] and network differential analysis (NDA) [92]. The mutual information (MI) [93,94] is useful to compare PRNs generated with different methods; see the SI for details.
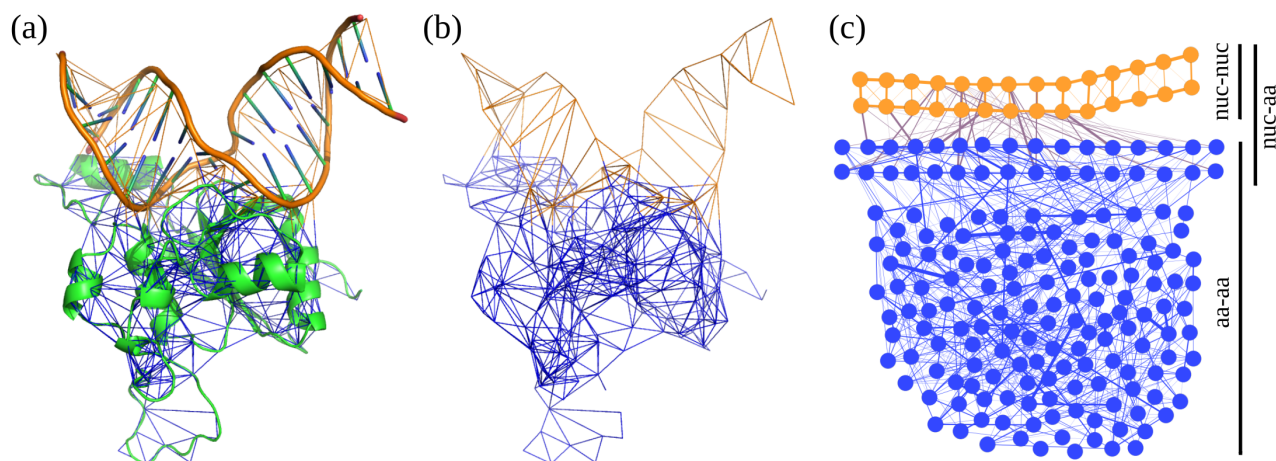

Figure 6. PRN models of the protein–DNA complex: (**a**) the PRN representation overlaid with the structure; (**b**) only the PRN without the structure; (**c**) a flat layout with the line thickness proportional to the interaction energy. Parts a-b are depicted with pyProGA [92], and part c is drawn with Gephi [95]. Orange and blue color denotes nucleotides (nuc) and amino acid (aa) residues, respectively.

The differences in the PRN based on the FMO fragments and conventional residues are caused by the shift of the CO group in amino acids and the PO$_3$ group in nucleotides; see Figure S1. In about 40% of cases (the coefficients of constraint defined in the SI are $C^{(1|2)} \approx C^{(2|1)} \approx 0.4$), the two methods could predict the existence of an edge with a weight within $\sim$1 kcal mol^{-1} accuracy. Much of the measure quantifying the agreement comes from pairs for which ΔE_{ij} is small (between 0 and -1 kcal mol^{-1}) without an edge in the graph (the region with small energies in Figure S7, the most populated). The nuc–nuc interactions are not much affected by the misassignment of the PO$_3$ group as the dominant interacting groups are the nucleotide bases (Table 4), as indicated by the population of states near the diagonal in Figure S7b. On the other hand, nuc–aa interactions in Figure S7c are more affected by the shifted PO$_3$, as these groups are orientated towards the protein. For the aa–aa interactions in Figure S7d, the shift of the CO group can cause discrepancies in the PIE values if the backbone is involved in interactions. However, stronger interactions with $\Delta E_{ij} \leq -15$ kcal mol^{-1} usually come from charged side chains, and the CO shift does not affect them much (states closer to the diagonal in Figure S7d).

To investigate the effect of fractional versus integer charges of interacting units, the FMO/solution/frag and PA/solution/frag interactions were compared, for which the atomic constitution is identical. The mutual information (MI) is larger, and the coefficients of constraint exceed 0.66. Figure S8 shows that the joint probability distribution is more diagonal in all residue type combinations. That is, off-diagonal elements are less frequent, indicating that the mutual prediction of equivalent edge weight is to be expected. The change of the treatment of charge transfer in the monomer (PA) vs. dimer (FMO) terms can shift the PIEs by a few kcal mol^{-1} only.

The final comparison is for PA/solution/res vs. MM/solute/res, for which atomic compositions are identical and the focus is on comparing QM and MM. In MM, pair interactions are overestimated (no screening), and thus, there are 4–5-times more edges in the graph (an edge is added if E_{lim} is stronger than -1 kcal mol^{-1}); see Table S1. Two

methods, using integer charges while neglecting the screening, FMO/solution/frag and MM/solute/res can be compared (Table 8). The coefficients of constraint are virtually unchanged for the nuc–nuc interactions. As discussed above, the effect of the shifted PO_3 assignment is negligible in nuc–nuc interactions. Hence, the difference in the PIEs between PA (or FMO) vs. MM is mainly due to the inclusion of solvent screening.

Table 8. Mutual information (MI) and coefficients of constraint C between two different interaction graphs (ΔE_{ij}), labeled by 1 and 2 [a].

Graph 1	Graph 2		All	nuc–nuc	nuc–aa	aa–aa
FMO solution frag	PA solution res	$MI^{(1;2)}$	0.1004	0.5269	0.0754	0.1040
		$C^{(1|2)}$	0.4009	0.9268	0.5794	0.3728
		$C^{(2|1)}$	0.4207	0.9432	0.6592	0.3872
FMO solution frag	PA solution frag	$MI^{(1;2)}$	0.1647	0.5249	0.0868	0.1813
		$C^{(1|2)}$	0.6666	0.9271	0.7873	0.6491
		$C^{(2|1)}$	0.6899	0.9397	0.7592	0.6748
PA solution res	MM solute res	$MI^{(1;2)}$	0.0684	0.1669	0.0604	0.0859
		$C^{(1|2)}$	0.0868	0.2233	0.0500	0.1315
		$C^{(2|1)}$	0.2732	0.2936	0.4643	0.3079
FMO solution frag	MM solute res	$MI^{(1;2)}$	0.0459	0.1672	0.0486	0.0548
		$C^{(1|2)}$	0.0582	0.2236	0.0402	0.0838
		$C^{(2|1)}$	0.1922	0.2993	0.4249	0.2039

[a] Averaged over 100 structures, using 40 bins in the histogram analysis with the parameter $E_{lim} = -1$ kcal mol^{-1}. Graphs are labeled by the computational method, as in Table 1. The values are computed either for all pairs or limited to pairs shown (nuc and aa stand for nucleotide and amino acid residues, respectively).

Because MM overestimates the interaction due to the lack of solvent screening (in the conventional AMBER model), one can ask whether MM-based PRNs are a reliable representation of residue interaction networks. Their choice as possible features in advanced analysis of MD trajectories remains to be tested [96]. Especially alarming is that some residue pairs (A:T) have a repulsive interaction in MM, about +8 kcal mol^{-1}, in contrast to about −8 kcal mol^{-1} in PA; see Table 1. Similarly, C:G interactions are −8 kcal mol^{-1} in MM against −21 kcal mol^{-1} in PA. The lack of screening may be a problem for nucleotide interactions beyond those of pairs in a double-helix [97], as occur in packed DNA, in non-canonical or transitional structures [98–101], or in complexes with ligands [102–104].

4. Conclusions

A methodology for analyzing interactions in proteins, DNA, and their complexes was developed, based on QM and MM calculations, decomposed into appropriate units: backbones, bases, and side chains. This methodology is facilitated by the convenient framework offered by the partition analysis developed for the fragment molecular orbital method combined with density-functional tight-binding.

As a case study of applying this methodology, a protein–DNA complex was computed and analyzed in detail. For a realistic treatment of temperature, 100 snapshots from MD simulations were averaged.

It was shown that there is good agreement between the MM and solute QM results, when both methods do not include solvent screening. Both a very high correlation and agreement in absolute values were found for the studied variety of interactions between nucleotides and amino acid residues. However, this agreement was observed only for the electrostatic interaction, but not for dispersion.

Nevertheless, some deviations were observed, as much as 5 kcal mol^{-1}. These deviations were attributed to charge transfer facilitated by hydrogen bonds, which was as large as 6 and 8% of the formal charge for nucleotides and amino acid residues, respectively,

The addition of solvent screening in the DFTB/PCM calculations resulted in a pronounced change of the picture of the interactions. The repulsive A:T interactions (between anionic nucleotides) became attractive when the screening contribution was added. One can question if MD simulations of DNA with a non-polarizable force field without a solvent screening are reliable at all.

A separate point of interest is the comparison of conventional residues to standard fragments in the FMO, which differ in the backbone definition. It was found that, for most pairs, where the interaction was between base pairs or between a base and a side chain, there was very good agreement, but for some adjacent nucleotide–residue pairs (as G10:Arg180 and C9:Arg180), the difference in the backbone definition of amino acid residues resulted in a very different definition of the interactions, which may be a source of confusion in discussing the results. This misassignment affects particular interactions, but not their total sum.

To provide a composite measure of the difference in two sets of interactions, a statistical analysis of all attractive residue–residue interactions was conducted using PRN models. The inclusion of charge transfer (PA vs. FMO) had on average a small effect on the PIEs, and the two sets had a high correlation. The use of non-conventional residues (FMO fragments) can result in larger differences, as much as 20 kcal mol^{-1} in some pairs, attributed to the shift in the unit definition. QM and MM have large discrepancies because solvent screening is not accounted for in MM.

The definition of segments in PA makes it possible to examine the strength of individual hydrogen bonds between nucleotide bases. A:T pairs have two conventional H-bonds; C:G pairs have three. The mismatched G:T pair has two H-bonds, similar to A:T in strength, albeit its additional CO$\cdots$H and ES1 interactions place G:T between A:T and C:G pairs, in agreement with other reports.

The interaction energies obtained in this work were compared to the binding energies reported in other studies, and in general, good agreement was observed. The developed methodology can be used in future applications of the FMO to biochemical studies such as drug design, where its ability to quantify the contributions of individual functional groups can be very useful.

References

1. Geng, C.; Xue, L.C.; Roel-Touris, J.; Bonvin, A.M.J.J. Finding the $\Delta\Delta G$ Spot: Are Predictors of Binding Affinity Changes upon Mutations in Protein-Protein Interactions Ready for it? *WIREs Comput. Mol. Sci.* **2019**, *9*, e1410. [CrossRef]

2. Gordon, M.S.; Fedorov, D.G.; Pruitt, S.R.; Slipchenko, L.V. Fragmentation Methods: A Route to Accurate Calculations on Large Systems. *Chem. Rev.* **2012**, *112*, 632–672. [CrossRef] [PubMed]

3. Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. Fragment Molecular Orbital Method: an Approximate Computational Method for Large Molecules. *Chem. Phys. Lett.* **1999**, *313*, 701–706. [CrossRef]

4. Fedorov, D.G. The Fragment Molecular Orbital Method: Theoretical Development, Implementation in GAMESS, and Applications. *WIREs Comput. Mol. Sci.* **2017**, *7*, e1322. [CrossRef]

5. Fukuzawa, K.; Tanaka, S. Fragment Molecular Orbital Calculations for Biomolecules. *Curr. Opin. Struct. Biol.* **2022**, *72*, 127–134. [CrossRef]

6. Sawada, T.; Fedorov, D.G.; Kitaura, K. Role of the Key Mutation in the Selective Binding of Avian and Human Influenza Hemagglutinin to Sialosides Revealed by Quantum-Mechanical Calculations. *J. Am. Chem. Soc.* **2010**, *132*, 16862–16872.

7. Morao, I.; Fedorov, D.G.; Robinson, R.; Southey, M.; Townsend-Nicholson, A.; Bodkin, M.J.; Heifetz, A. Rapid and Accurate Assessment of GPCR-Ligand Interactions Using the Fragment Molecular Orbital-Based Density-Functional Tight-Binding Method. *J. Comput. Chem.* **2017**, *38*, 1987–1990. [CrossRef]

8. Kim, J.; Lim, H.; Moon, S.; Cho, S.Y.; Kim, M.; Park, J.H.; Park, H.W.; No, K.T. Hot Spot Analysis of YAP-TEAD Protein-Protein Interaction Using the Fragment Molecular Orbital Method and its Application for Inhibitor Discovery. *Cancers* **2021**, *2021*, 4246. [CrossRef]

9. Tanaka, S.; Tokutomi, S.; Hatada, R.; Okuwaki, K.; Akisawa, K.; Fukuzawa, K.; Komeiji, Y.; Okiyama, Y.; Mochizuki, Y. Dynamic Cooperativity of Ligand-Residue Interactions Evaluated with the Fragment Molecular Orbital Method. *J. Phys. Chem. B* **2021**, *125*, 6501–6512. [CrossRef]

10. Deb, I.; Wong, H.; Tacubao, C.; Frank, A.T. Quantum Mechanics Helps Uncover Atypical Recognition Features in the Flavin Mononucleotide Riboswitch. *J. Phys. Chem. B* **2021**, *125*, 8342–8350. [CrossRef]

11. Xu, F.; Tanaka, S.; Watanabe, H.; Shimane, Y.; Iwasawa, M.; Ohishi, K.; Maruyama, T. Computational Analysis of the Interaction Energies between Amino Acid Residues of the Measles Virus Hemagglutinin and Its Receptors. *Viruses* **2018**, *10*, 236. [CrossRef] [PubMed]

12. Fujii, M.; Tanaka, S. Interspecies Comparison of Interaction Energies between Photosynthetic Protein RuBisCO and 2CABP Ligand. *Int. J. Mol. Sc.* **2022**, *23*, 11347. [CrossRef] [PubMed]

13. Nishimoto, Y.; Fedorov, D.G.; Irle, S. Density-Functional Tight-Binding Combined with the Fragment Molecular Orbital Method. *J. Chem. Theory Comput.* **2014**, *10*, 4801–4812. [CrossRef] [PubMed]

14. Nishimoto, Y.; Fedorov, D.G. The Fragment Molecular Orbital Method Combined with Density-Functional Tight-Binding and the Polarizable Continuum Model. *Phys. Chem. Chem. Phys.* **2016**, *18*, 22047–22061. [CrossRef] [PubMed]

15. Raha, K.; van der Vaart, A.J.; Riley, K.E.; Peters, M.B.; Westerhoff, L.M.; Kim, H.; Merz, K.M. Pairwise Decomposition of Residue Interaction Energies Using Semiempirical Quantum Mechanical Methods in Studies of Protein-Ligand Interaction. *J. Am. Chem. Soc.* **2005**, *127*, 6583–6594. [CrossRef] [PubMed]

16. Phipps, M.J.S.; Fox, T.; Tautermann, C.S.; Skylaris, C.K. Energy Decomposition Analysis Approaches and their Evaluation on Prototypical Protein-Drug Interaction Patterns. *Chem. Soc. Rev.* **2015**, *44*, 3177–3211. [CrossRef]

17. Yamamoto, Y.; Takei, K.; Arulmozhiraja, S.; Sladek, V.; Matsuo, N.; Han, S.i.; Matsuzaka, T.; Sekiya, M.; Tokiwa, T.; Shoji, M.; et al. Molecular Association Model of PPARα and its New Specific and Efficient Ligand, Pemafibrate: Structural Basis for SPPARMα. *Biochem. Biophys. Res. Commun.* **2018**, *499*, 239–245. [CrossRef] [PubMed]

18. Mironov, V.; Shchugoreva, I.A.; Artyushenko, P.V.; Morozov, D.; Borbone, N.; Oliviero, G.; Zamay, T.N.; Moryachkov, R.V.; Kolovskaya, O.S.; Lukyanenko, K.A.; et al. Structure- and Interaction-Based Design of anti-SARS-CoV-2 Aptamers. *Chem. Eur. J.* **2022**, *28*, e202104481. [CrossRef]

19. Yoshida, T.; Hirono, S. A 3D-QSAR Analysis of CDK2 Inhibitors Using FMO Calculations and PLS. *Chem. Pharm. Bull.* **2019**, *67*, 546–555. [CrossRef]

20. Sladek, V.; Tokiwa, H.; Hitoshi, S.; Yasuteru, S. Protein Residue Networks from Energetic and Geometric Data: Are They Identical? *J. Chem. Theory Comput.* **2018**, *14*, 6623–6631. [CrossRef]

21. Okuwaki, K.; Doi, H.; Fukuzawa, K.; Mochizuki, Y. Folding Simulation of Small Proteins by Dissipative Particle Dynamics (DPD) with Non-empirical Interaction Parameters Based on Fragment Molecular Orbital Calculations. *Appl. Phys. Expr.* **2020**, *13*, 017002. [CrossRef]

22. Ishikawa, T.; Ishikura, T.; Kuwata, K. Theoretical Study of the Prion Protein Based on the Fragment Molecular Orbital Method. *J. Comput. Chem.* **2009**, *30*, 2594–2601. [CrossRef]

23. Fedorov, D.G.; Kitaura, K. Pair Interaction Energy Decomposition Analysis for Density Functional Theory and Density-Functional Tight-Binding with an Evaluation of Energy Fluctuations in Molecular Dynamics. *J. Phys. Chem. A* **2018**, *122*, 1781–1795. [CrossRef] [PubMed]

24. Takaba, K.; Watanabe, C.; Tokuhisa, A.; Akinaga, Y.; Ma, B.; Kanada, R.; Araki, M.; Okuno, Y.; Kawashima, Y.; Moriwaki, H.; et al. Protein-Ligand Binding Affinity Prediction of Cyclin-Dependent Kinase-2 Inhibitors by Dynamically Averaged Fragment Molecular Orbital-Based Interaction Energy. *J. Comput. Chem.* **2022**, *43*, 1362–1371. [CrossRef] [PubMed]

25. Fedorov, D.G.; Kitaura, K. Subsystem Analysis for the Fragment Molecular Orbital Method and its Application to Protein-Ligand Binding in Solution. *J. Phys. Chem. A* **2016**, *120*, 2218–2231. [CrossRef]

26. Bouchal, T.; Durník, I.; Illík, V.; Réblová, K.; Kulhánek, P. Importance of Base-Pair Opening for Mismatch Recognition. *Nucleic Acids Res.* **2020**, *48*, 11322–11334. [CrossRef]

27. Morozov, D.; Mironov, V.; Moryachkov, R.V.; Shchugoreva, I.A.; Artyushenko, P.V.; Zamay, G.S.; Kolovskaya, O.S.; Zamay, T.N.; Krat, A.V.; Molodenskiy, D.S.; et al. The role of SAXS and Molecular Simulations in 3D Structure Elucidation of a DNA Aptamer against Lung Cancer. *Mol. Ther. Nucl. Acids* **2021**, *25*, 316–327. [CrossRef]

28. Fedorov, D.G. Polarization Energies in the Fragment Molecular Orbital Method. *J. Comput. Chem.* **2022**, *43*, 1094–1103. [CrossRef]

29. Ji, C.G.; Zhang, J.Z.H. Protein Polarization Is Critical to Stabilizing AF-2 and Helix-2' Domains in Ligand Binding to PPAR-γ. *J. Am. Chem. Soc.* **2008**, *130*, 17129–17133. [CrossRef]

30. Simmonett, A.C.; Pickard, F.C.; Shao, Y.; Cheatham, T.E.; Brooks, B.R. Efficient Treatment of Induced Dipoles. *J. Chem. Phys.* **2015**, *143*, 074115. [CrossRef]

31. Corrigan, R.A.; Qi, G.; Thiel, A.C.; Lynn, J.R.; Walker, B.D.; Casavant, T.L.; Lagardere, L.; Piquemal, J.P.; Ponder, J.W.; Ren, P.; et al. Implicit Solvents for the Polarizable Atomic Multipole AMOEBA Force Field. *J. Chem. Theory Comput.* **2021**, *17*, 2323–2341. [CrossRef] [PubMed]

32. Kim, Y.; Bui, Y.; Tazhigulov, R.N.; Bravaya, K.B.; Slipchenko, L.V. Effective Fragment Potentials for Flexible Molecules: Transferability of Parameters and Amino Acid Database. *J. Chem. Theory Comput.* **2020**, *16*, 7735–7747. [CrossRef] [PubMed]

33. Fedorov, D.G. Solvent Screening in Zwitterions Analyzed with the Fragment Molecular Orbital Method. *J. Chem. Theory Comput.* **2019**, *15*, 5404–5416. [CrossRef] [PubMed]

34. Monteleone, S.; Fedorov, D.G.; Townsend-Nicholson, A.; Southey, M.; Bodkin, M.; Heifetz, A. Hotspot Identification and Drug Design of Protein-Protein Interaction Modulators using the Fragment Molecular Orbital Method. *J. Chem. Inf. Model.* **2022**, *62*, 3784–3799. [CrossRef] [PubMed]

35. Šponer, J.; Jurečka, P.; Hobza, P. Accurate Interaction Energies of Hydrogen-Bonded Nucleic Acid Base Pairs. *J. Am. Chem. Soc.* **2004**, *126*, 10142–10151. [CrossRef]

36. Hobza, P.; Kabelac, M.; Sponer, J.; Mejzlik, P.; Vondrasek, J. Performance of Empirical Potentials (AMBER, CFF95, CVFF, CHARMM, OPLS, POLTEV), Semiempirical Quantum Chemical Methods (AM1, MNDO/M, PM3), and ab Initio Hartree–Fock Method for Interaction of DNA Bases: Comparison with Nonempirical beyond Hartree–Fock results. *J. Comp. Chem.* **1997**, *18*, 1136–1150.

37. Asensio, A.; Kobko, N.; Dannenberg, J.J. Cooperative Hydrogen-Bonding in Adenine-Thymine and Guanine-Cytosine Base Pairs. Density Functional Theory and Møller-Plesset Molecular Orbital Study. *J. Phys. Chem. A* **2003**, *107*, 6441–6443. [CrossRef]

38. Zubatiuk, T.; Kukuev, M.A.; Korolyova, A.S.; Gorb, L.; Nyporko, A.; Hovorun, D.; Leszczynski, J. Structure and Binding Energy of Double-Stranded A-DNA Mini-helices: Quantum-Chemical Study. *J. Phys. Chem. B* **2015**, *119*, 12741–12749. [CrossRef]

39. Jurečka, P.; Hobza, P. True Stabilization Energies for the Optimal Planar Hydrogen-Bonded and Stacked Structures of Guanine···Cytosine, Adenine···Thymine, and Their 9- and 1-Methyl Derivatives: Complete Basis Set Calculations at the MP2 and CCSD(T) Levels and Comparison with Experiment. *J. Am. Chem. Soc.* **2003**, *125*, 15608–15613. [CrossRef]

40. Cuyacot, B.J.R.; Durník, I.; Foroutan-Nejad, C.; Marek, R. Anatomy of Base Pairing in DNA by Interacting Quantum Atoms. *J. Chem. Inf. Model.* **2020**, *61*, 211–222. [CrossRef]

41. Kudritskaya, Z.G.; Danilov, V.I. Quantum Mechanical Study of Bases Interactions in Various Associates in Atomic Dipole Approximation. *J. Theor. Biol.* **1976**, *59*, 303–318. [CrossRef]

42. Yanson, I.K.; Teplitsky, A.B.; Sukhodub, L.F. Experimental Studies of Molecular Interactions Between Nitrogen Bases of Nucleic Acids. *Biopolymers* **1979**, *18*, 1149–1170. [CrossRef] [PubMed]

43. Kyogoku, Y.; Lord, R.C.; Rich, A. An Infrared Study Of Thehydrogen - Bonding Specificity Of Hypoxanthine And Other Nucleic Acid Derivatives. *Biochim. Biophys. Acta* **1969**, *179*, 10–17. [CrossRef]

44. Watanabe, T.; Inadomi, Y.; Fukuzawa, K.; Nakano, T.; Tanaka, S.; Nilsson, L.; Nagashima, U. DNA and Estrogen Receptor Interaction Revealed by Fragment Molecular Orbital Calculations. *J. Phys. Chem. B* **2007**, *111*, 9621–9627. [CrossRef] [PubMed]

45. Watanabe, T.; Inadomi, Y.; Umeda, H.; Fukuzawa, K.; Tanaka, S.; Nakano, T.; Nagashima, U. Fragment Molecular Orbital (FMO) and FMO-MO Calculations of DNA: Accuracy Validation of Energy and Interfragment Interaction Energy. *J. Comp. Theor. Nanosc.* **2009**, *6*, 1328–1337. [CrossRef]

46. Fukuzawa, K.; Kurisaki, I.; Watanabe, C.; Okiyama, Y.; Mochizuki, Y.; Tanaka, S.; Komeiji, Y. Explicit Solvation Modulates Intra- and Inter-Molecular Interactions within DNA: Electronic Aspects Revealed by the ab Initio Fragment Molecular Orbital (FMO) Method. *Comp. Theor. Chem.* **2015**, *1054*, 29–37. [CrossRef]

47. Yamada, H.; Mochizuki, Y.; Fukuzawa, K.; Okiyama, Y.; Komeiji, Y. Fragment Molecular Orbital (FMO) Calculations on DNA by a Scaled Third-Order Møller-Plesset Perturbation (MP2.5) Scheme. *Comp. Theor. Chem.* **2017**, *1101*, 46–54. [CrossRef]

48. Okiyama, Y.; Nakano, T.; Watanabe, C.; Fukuzawa, K.; Mochizuki, Y.; Tanaka, S. Fragment Molecular Orbital Calculations with Implicit Solvent Based on the Poisson-Boltzmann Equation: Implementation and DNA Study. *J. Phys. Chem. B* **2018**, *122*, 4457–4471. [CrossRef]

49. Warren, J.J.; Pohlhaus, T.J.; Changela, A.; Iyer, R.R.; Modrich, P.L.; Beese, L.S. Structure of the Human MutSα DNA Lesion Recognition Complex. *Mol. Cell* **2007**, *26*, 579–592. [CrossRef]

50. Bouchal, T.; Durník, I.; Kulhánek, P. Bending of Canonical and G/T Mismatched DNAs. *J. Chem. Inf. Model.* **2021**, *61*, 6000–6011. [CrossRef]

51. Fedorov, D.G. Partition Analysis for Density-Functional Tight-Binding. *J. Phys. Chem. A* **2020**, *124*, 10346–10358. [CrossRef] [PubMed]

52. Nakamura, T.; Yokaichiya, T.; Fedorov, D.G. Quantum-Mechanical Structure Optimization of Protein Crystals and Analysis of Interactions in Periodic Systems. *J. Phys. Chem. Lett.* **2021**, *12*, 8757–8762. [CrossRef] [PubMed]

53. Nakamura, T.; Fedorov, D.G. The Catalytic Activity and Adsorption in Faujasite and ZSM-5 Zeolites: The Role of Differential Stabilization and Charge Delocalization. *Phys. Chem. Chem. Phys.* **2022**, *24*, 7739–7747. [CrossRef] [PubMed]

54. Jacob, C.R.; Neugebauer, J. Subsystem Density-Functional Theory. *WIREs Comput. Mol. Sci.* **2014**, *4*, 325–362. [CrossRef]

55. Cornell, W.D.; Cieplak, P.; Bayly, C.I.; Gould, I.R.; Merz, K.M.; Ferguson, D.M.; Spellmeyer, D.C.; Fox, T.; Caldwell, J.W.; Kollman, P.A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197. [CrossRef]

56. Ji, C.; Mei, Y.; Zhang, J.Z. Developing Polarized Protein-Specific Charges for Protein Dynamics: MD Free Energy Calculation of pKa Shifts for Asp26/Asp20 in Thioredoxin. *Biophys. J.* **2008**, *95*, 1080–1088. [CrossRef]

57. Kato, K.; Masuda, T.; Watanabe, C.; Miyagawa, N.; Mizouchi, H.; Nagase, S.; Kamisaka, K.; Oshima, K.; Ono, S.; Ueda, H.; et al. High-Precision Atomic Charge Prediction for Protein Systems Using Fragment Molecular Orbital Calculation and Machine Learning. *J. Chem. Inf. Model.* **2020**, *60*, 3361–3368. [CrossRef]

58. Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J. Comp. Chem.* **2011**, *32*, 1456–1465. [CrossRef]

59. Gaus, M.; Goez, A.; Elstner, M. Parametrization and Benchmark of DFTB3 for Organic Molecules. *J. Chem. Theory Comput.* **2013**, *9*, 338–354. [CrossRef] [PubMed]

60. Fedorov, D.G.; Kitaura, K. The Importance of Three-Body Terms in the Fragment Molecular Orbital Method. *J. Chem. Phys.* **2004**, *120*, 6832 6840. [CrossRef]

61. Barca, G.M.J.; Bertoni, C.; Carrington, L.; Datta, D.; De Silva, N.; Deustua, J.E.; Fedorov, D.G.; Gour, J.R.; Gunina, A.O.; Guidez, E.; et al. Recent Developments in the General Atomic and Molecular Electronic Structure System. *J. Chem. Phys.* **2020**, *152*, 154102. [CrossRef] [PubMed]

62. Case, D.A.; Babin, V.; Berryman, J.T.; Betz, R.M.; Cai, Q.; Cerutti, D.S.; Cheatham, T.E.; Darden, T.A.; Duke, R.E.; Gohlke, H.; et al. *AMBER 14*; Technical Report; University of California: San Francisco, CA, USA, 2022.

63. Maier, J.A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K.E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713. [CrossRef] [PubMed]

64. Ivani, I.; Dans, P.D.; Noy, A.; Pérez, A.; Faustino, I.; Hospital, A.; Walther, J.; Andrio, P.; Goñi, R.; Balaceanu, A.; et al. Parmbsc1: A Refined Force Field for DNA Simulations. *Nat. Meth.* **2015**, *13*, 55–58. [CrossRef] [PubMed]

65. Nguyen, H.; Roe, D.R.; Simmerling, C. Improved Generalized Born Solvent Model Parameters for Protein Simulations. *J. Chem. Theory Comput.* **2013**, *9*, 2020–2034. [CrossRef] [PubMed]

66. Cournia, Z.; Allen, B.; Sherman, W. Relative Binding Free Energy Calculations in Drug Discovery: Recent Advances and Practical Considerations. *J. Chem. Inf. Model.* **2017**, *57*, 2911–2937. [CrossRef] [PubMed]

67. King, E.; Qi, R.; Li, H.; Luo, R.; Aitchison, E. Estimating the Roles of Protonation and Electronic Polarization in Absolute Binding Affinity Simulations. *J. Chem. Theory Comput.* **2021**, *17*, 2541–2555. [CrossRef]

68. Fedorov, D.G.; Kitaura, K. Energy Decomposition Analysis in Solution Based on the Fragment Molecular Orbital Method. *J. Phys. Chem. A* **2012**, *116*, 704–719. [CrossRef]

69. Amin, M.; Küpper, J. Variations in Proteins Dielectric Constants. *ChemistryOpen* **2020**, *9*, 691–694. [CrossRef]

70. Li, L.; Li, C.; Zhang, Z.; Alexov, E. On the Dielectric "Constant" of Proteins: Smooth Dielectric Function for Macromolecular Modeling and Its Implementation in DelPhi. *J. Chem. Theory Comput.* **2013**, *9*, 2126–2136. [CrossRef]

71. Janeček, M.; Kührová, P.; Mlýnský, V.; Otyepka, M.; Šponer, J.; Banáš, P. W-RESP: Well-Restrained Electrostatic Potential-Derived Charges. Revisiting the Charge Derivation Model. *J. Chem. Theory Comput.* **2021**, *17*, 3495–3509. [CrossRef]

72. Liebl, K.; Zacharias, M. Tumuc1: A New Accurate DNA Force Field Consistent with High-Level Quantum Chemistry. *J. Chem. Theory Comput.* **2021**, *17*, 7096–7105. [CrossRef] [PubMed]

73. Gaus, M.; Cui, Q.; Elstner, M. DFTB3: Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method (SCC-DFTB). *J. Chem. Theory Comput.* **2011**, *7*, 931–948. [CrossRef] [PubMed]

74. Fedorov, D.G. Electron Density from the Fragment Molecular Orbital Method Combined with Density-Functional Tight-Binding. *Chem. Phys. Lett.* **2021**, *780*, 138900. [CrossRef]

75. Guerra, C.F.; Bickelhaupt, F.M.; Snijders, J.G.; Baerends, E.J. Hydrogen Bonding in DNA Base Pairs: Reconciliation of Theory and Experiment. *J. Am. Chem. Soc.* **2000**, *122*, 4117–4128. [CrossRef]

76. Seeman, N.C.; Rosenberg, J.M.; Suddath, F.; Kim, J.J.P.; Rich, A. RNA Double-Helical Fragments at Atomic Resolution: I. The Crystal and Molecular Structure of Sodium Adenylyl-3′,5′-Uridine Hexahydrate. *J. Mol. Biol.* **1976**, *104*, 109–144. [CrossRef]

77. Rosenberg, J.M.; Seeman, N.C.; Day, R.O.; Rich, A. RNA Double-Helical Fragments at Atomic Resolution: II. The Crystal Structure of Sodium Guanylyl-3′,5′-Cytidine Nonahydrate. *J. Mol. Biol.* **1976**, *104*, 145–167. [CrossRef]

78. Schwalbe, C.H. June Sutor and the C–H⋯O Hydrogen Bonding Controversy. *Crystallogr. Rev.* **2012**, *18*, 191–206. [CrossRef]

79. Helgaker, T.; Klopper, W.; Koch, H.; Noga, J. Basis-Set Convergence of Correlated Calculations on Water. *J. Chem. Phys.* **1997**, *106*, 9639–9646. [CrossRef]

80. Vijayabaskar, M.; Vishveshwara, S. Interaction Energy Based Protein Structure Networks. *Biophys. J.* **2010**, *99*, 3704–3715. [CrossRef]

81. Yao, X.Q.; Momin, M.; Hamelberg, D. Establishing a Framework of Using Residue-Residue Interactions in Protein Difference Network Analysis. *J. Chem. Inf. Model.* **2019**, *59*, 3222–3228. [CrossRef]

82. Viloria, J.S.; Allega, M.F.; Lambrughi, M.; Papaleo, E. An Optimal Distance Cutoff for Contact-Based Protein Structure Networks Using Side-Chain Centers of Mass. *Sci. Rep.* **2017**, *7*, 2838–2848. [CrossRef] [PubMed]

83. Tasdighian, S.; Paola, L.D.; Ruvo, M.D.; Paci, P.; Santoni, D.; Palumbo, P.; Mei, G.; Venere, A.D.; Giuliani, A. Modules Identification in Protein Structures: The Topological and Geometrical Solutions. *J. Chem. Inf. Model.* **2013**, *54*, 159–168. [CrossRef] [PubMed]

84. Lim, H.; Chun, J.; Jin, X.; Kim, J.; Yoon, J.H.; No, K.T. Investigation of Protein-Protein Interactions and Hot Spot Region between PD-1 and PD-L1 by Fragment Molecular Orbital Method. *Sci. Rep.* **2019**, *9*, 16727. [CrossRef]

85. Yao, X.Q.; Hamelberg, D. From Distinct to Differential Conformational Dynamics to Map Allosteric Communication Pathways in Proteins. *J. Phys. Chem. B* **2022**, *126*, 2612–2620. [CrossRef] [PubMed]

86. Yao, X.Q.; Hamelberg, D. Residue–Residue Contact Changes during Functional Processes Define Allosteric Communication Pathways. *J. Chem. Theory Comput.* **2022**, *18*, 1173–1187. [CrossRef] [PubMed]

87. Doshi, U.; Holliday, M.J.; Eisenmesser, E.Z.; Hamelberg, D. Dynamical Network of Residue-Residue Contacts Reveals Coupled Allosteric Effects in Recognition, Catalysis, and Mutation. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 4735–4740. [CrossRef] [PubMed]

88. Calvó-Tusell, C.; Maria-Solano, M.A.; Osuna, S.; Feixas, F. Time Evolution of the Millisecond Allosteric Activation of Imidazole Glycerol Phosphate Synthase. *J. Am. Chem. Soc.* **2022**, *144*, 7146–7159. [CrossRef]

89. Osuna, S. The Challenge of Predicting Distal Active Site Mutations in Computational Enzyme Design. *WIREs Comput. Mol. Sci.* **2020**, *11*, e1502. [CrossRef]

90. Guzzi, P.H.; Paola, L.D.; Giuliani, A.; Veltri, P. PCN-Miner: An open-source extensible tool for the analysis of Protein Contact Networks. *Bioinformatics* **2022**, *38*, 4235–4237. [CrossRef]

91. Tanaka, S.; Watanabe, C.; Honma, T.; Fukuzawa, K.; Ohishi, K.; Maruyama, T. Identification of Correlated Inter-Residue Interactions in Protein Complex Based on the Fragment Molecular Orbital Method. *J. Mol. Graph. Modell.* **2020**, *100*, 107650. [CrossRef]

92. Sladek, V.; Yamamoto, Y.; Harada, R.; Shoji, M.; Shigeta, Y.; Sladek, V. pyProGA-A PyMOL Plugin for Protein Residue Network Analysis. *PLoS ONE* **2021**, *16*, e0255167. [CrossRef] [PubMed]

93. Bossomaier, T.; Barnett, L.; Harré, M.; Lizier, J.T. *An Introduction to Transfer Entropy: Information Flow in Complex Systems*; Springer International Publishing: Berlin/Heidelberg, Germany, 2016. [CrossRef]

94. Ben-Naim, A. *A Farewell to Entropy: Statistical Thermodynamics Based on Information*; World Scientific: Singapore, 2008.

95. Bastian, M.; Heymann, S.; Jacomy, M. *Gephi: An Open Source Software for Exploring and Manipulating Networks*; Technical Report; Gephi, WebAtlas: London, UK, 2009.

96. Diez, G.; Nagel, D.; Stock, G. Correlation-Based Feature Selection to Identify Functional Dynamics in Proteins. *J. Chem. Theory Comput.* **2022**, *18*, 5079–5088. [CrossRef] [PubMed]

97. Cheatham, T.E.; Young, M.A. Molecular Dynamics Simulation of Nucleic Acids: Successes, Limitations, and Promise. *Biopolymers* **2000**, *56*, 232–256. [CrossRef]

98. Galindo-Murillo, R.; Cheatham, T.E.; Hopkins, R.C. Exploring Potentially Alternative Non-canonical DNA Duplex Structures through Simulation. *J. Biomol. Struct. Dyn.* **2018**, *37*, 2201–2210. [CrossRef]

99. Waters, J.T.; Lu, X.J.; Galindo-Murillo, R.; Gumbart, J.C.; Kim, H.D.; Cheatham, T.E.; Harvey, S.C. Transitions of Double-Stranded DNA between the A- and B-Forms. *J. Phys. Chem. B* **2016**, *120*, 8449–8456. [CrossRef]

100. Rajasekaran, T.; Freestone, G.C.; Galindo-Murillo, R.; Lugato, B.; Rico, L.; Salinas, J.C.; Gaus, H.; Migawa, M.T.; Swayze, E.E.; Cheatham, T.E.; et al. Backbone Hydrocarbon-Constrained Nucleic Acids Modulate Hybridization Kinetics for RNA. *J. Am. Chem. Soc.* **2022**, *144*, 1941–1950. [CrossRef]

101. Zgarbová, M.; Jurečka, P.; Banáš, P.; Otyepka, M.; Šponer, J.E.; Leontis, N.B.; Zirbel, C.L.; Šponer, J. Noncanonical Hydrogen Bonding in Nucleic Acids. Benchmark Evaluation of Key Base–Phosphate Interactions in Folded RNA Molecules Using Quantum-Chemical Calculations and Molecular Dynamics Simulations. *J. Phys. Chem. A* **2011**, *115*, 11277–11292. [CrossRef]

102. Šimunková, M.; Štekláč, M.; Malček, M. Spectroscopic, Computational and Molecular Docking Study of Cu(II) Complexes with Flavonoids: From Cupric Ion Binding to DNA Intercalation. *New J. Chem.* **2021**, *45*, 10810–10821. [CrossRef]

103. Galindo-Murillo, R.; Winkler, L.; Ma, J.; Hanelli, F.; Fleming, A.M.; Burrows, C.J.; Cheatham, T.E. Riboflavin Stabilizes Abasic, Oxidized G-Quadruplex Structures. *Biochemistry* **2022**, *61*, 265–275. [CrossRef]

104. Šimunková, M.; Biela, M.; Štekláč, M.; Hlinčík, A.; Klein, E.; Malček, M. Cu(II) Complexes of Flavonoids in Solution: Impact of the Cu(II) Ion on the Antioxidant and DNA-Intercalating Properties. *J. Mol. Liq.* **2022**, *359*, 119230. [CrossRef]