

Article

# A Density Functional Theory and Information-Theoretic Approach Study of Interaction Energy and Polarizability for Base Pairs and Peptides

Dongbo Zhao<sup>1</sup>, Shubin Liu<sup>2,3,\*</sup> and Dahua Chen<sup>1,\*</sup><sup>1</sup>Institute of Biomedical Research, Yunnan University, Kunming 650500, Yunnan, P.R. China<sup>2</sup>Research Computing Center, University of North Carolina, Chapel Hill, North Carolina 27599-3420, USA<sup>3</sup>Department of Chemistry, University of North Carolina, Chapel Hill, North Carolina 27599-3290, USA

\*Correspondence: chendh@yun.edu.cn; shubin@email.unc.edu

**Abstract:** Using density functional theory (DFT) and the information-theoretic approach (ITA) quantities to appreciate the energetics and properties of biopolymers is still an unaccomplished and ongoing task. To this end, we studied the building blocks of nucleic acid base pairs and small peptides. For base pairs, we have dissected the relative importance of energetic components by using two energy partition schemes in DFT. Our results convincingly show that the exchange-correlation effect predominantly governs the molecular stability of base pairs while the electrostatic potential plays a minor but indispensable role, and the steric effect is trivial. Furthermore, we have revealed that simple density-based ITA functions are in good relationships with molecular polarizabilities for a series of 30 hydrogen-bonded base pairs and all 20 natural  $\alpha$ -amino acids, 400 dipeptides, and 8000 tripeptides. Based on these lines, one can easily predict the molecular polarizabilities of larger peptides, even proteins as long as the total molecular wavefunction is available, rather than solving the computationally demanding coupled-perturbed Hartree-Fock (CPHF) equation or its DFT counterpart coupled-perturbed Kohn-Sham (CPKS) equation.

**Keywords:** density functional theory (DFT); information-theoretic approach (ITA); base pairs; peptides; molecular polarizability

## 1. Introduction

Nucleic acids and proteins are important biological polymers in the living organism. Proteins are responsible for the catalysis of biological processes while nucleic acids or base pairs play a role as carriers of the genetic information. However, *ab initio* calculations of entire nucleic acids or proteins are still a hard nut to crack or even computationally intractable. Rather, one can easily delve into their building blocks, base pairs of nucleic acids and amino acids of proteins or peptides, respectively.

For base pairs, there exists a type of important noncovalent interaction: hydrogen-bonding interactions, which could stabilize the two strands of nucleic acids (DNA and RNA). To figure out the origin and nature of hydrogen-bonding interactions in base pairs is still of heated discussion and an ongoing task in the literature.[1–5] Normally, the role of electrostatic potential is greatly emphasized[2] and in most cases, this is the dominantly important energetic component of the interaction energy. However, this is not the whole picture of H-bonded interactions! Parthasarathi et al.[4] shown that the presence of critical points between hydrogen-bonded DNA base pairs indicate the closed-shell kind of interactions. With Bader's quantum theory of atoms in molecules (QTAIM), Toczyłowski et al.[5] revealed that electrostatic and the exchange energies have been found to be the most important components of the overall interaction energy, although the dispersion and the induction energies also play important roles. Suffice to note that the exchange term used is not the same quantity as that in DFT.[6] In this work, we also resort to the supermolecule interaction model as that of Toczyłowski et al. But, we employ the energetic components

in DFT to dictate which is responsible for the molecular stability. From two energy decomposition schemes,[6,7] we have clearly shown that the exchange-correlation effect in DFT predominantly governs the molecular stability of base pairs while the electrostatic potential plays a minor but indispensable role, and the steric effect is trivial.

Besides energetics, we have also explored the molecular polarizabilities of 30 base pairs and 20 natural  $\alpha$ -amino acids, 400 dipeptides, and 8000 tripeptides. A few simple density-based ITA functions, such as Shannon entropy,[8] Fisher information,[9] Ghosh–Berkowitz–Parr entropy,[10] Onicescu information energy,[11] relative Rényi entropy,[11] information gain,[12] and relative Fisher information,[13,14] are utilized to correlate with molecular polarizabilities. In addition, molecular volumes are also employed. We have found that exist strong linear correlations between molecular polarizabilities and ITA quantities. Implication of these lines is straightforward that one can readily predict the molecular polarizabilities of larger peptides, or even an entire protein when the molecular wavefunction can be obtained. In other words, one can bypass the time-consuming CPHF equation[15,16] or its DFT counterpart CPKS equation[17] by numerically integrating some density-based functions. Moreover, our work on molecular polarizability is physics-based (locality of electron density) and is markedly different from other mathematics-based predictions, such as machine learning (ML) [18–20] and regression model. [21]

## 2. Results

### 2.1. Validation

Hait et al.[22] have verified that the accuracy of the approximate hybrid density functionals are satisfactory in calculating molecular polarizabilities. Here, M06-2X,[23] use chosen without further test. We only validated a series of atomic-centered Gaussian-type basis sets 6-311G(d,p),[24] Def2-SVP,[25] Def2-TZVP,[25] and aug-cc-pVTZ.[26] We calculated the molecular polarizabilities for a total 20 natural  $\alpha$ -amino acids at different theoretical levels and compared the results with the experimental data as shown in **Table 1**. Two statistical parameters, mean unsigned error (**MUE**) and mean signed error (**MSE**) are employed to gauge the quality of the method. It is clearly shown that the combination of 6-311G(d,p) and M06-2X from this work or MP2 from the literature[27] leads to the largest deviations when compared with the experiment, with **MSEs** up to around 15 au. When the small Def2-SVP basis set is changed to a relatively larger one Def2-TZVP, the **MSE** dramatically decreases from 18.4 to 8.9 au. If an even better basis set aug-cc-pVTZ is used, the **MSE** is only 3.8 au. With both accuracy and efficiency taken into consideration, we have clearly showcased that M06-2X/Def2-TZVP is a good candidate in predicting molecular polarizabilities. Unless otherwise stated, the following results are obtained on top of this theoretical approach.

### 2.2. Total Energy Decomposition of Base Pairs

To determine the nature of hydrogen bonding interactions in base pairs is of heated discussion in the literature. In this work, we aim to use two energy decomposition schemes as mentioned above to dictate which energetic component is responsible for the hydrogen bonding interactions. Using the supramolecular approach, the interaction energy ( $\Delta E$ ) and its components, including kinetic ( $\Delta T_s$ ), exchange-correlation ( $\Delta E_{xc}$ ), electrostatic potential ( $\Delta E_e$ ), steric hindrance ( $\Delta E_s$ ), and quantum effect ( $\Delta E_q$ ), are obtained at the M06-2X/Def2-TZVP level, as collected in **Table 2**. Here, we only consider the electronic energy with thermal contributions and BSSE (basis set superposition error)[29,30] corrections neglected. The total energy difference is always negative, indicating that hydrogen-bonding interactions are energetically favorable. In addition, one can see that exchange-correlation ( $\Delta E_{xc}$ ), electrostatic potential ( $\Delta E_e$ ), steric hindrance ( $\Delta E_s$ ) all make positive contributions to the total energy difference. However, the large negative values of steric hindrance are compensated by the positive Fermionic quantum effect in the new scheme. In the conventional scheme, the positive kinetic energy also

cancels out part of the exchange-correlation effect and electrostatic potential. Next, we will figure out the relative importance of these three components. Shown in **Figure 2** are the linear correlations between the total energy difference ( $\Delta E$ ) and  $\Delta E_s$ ,  $\Delta E_e$ , and  $\Delta E_{xc}$ . These lines indicate that single-variable linear regression cannot distinguish the relative importance of these three energetic components. To resolve this issue, two-variable fits are used to make that happen. Shown in **Figure 4** are the strong linear correlations between the fitted and predicted total energy difference. The fitted regression equations for **Figure 4(a)** and **4(b)** are

$$\Delta E = 0.040\Delta E_s + 0.467\Delta E_e$$

and

$$\Delta E = 1.017\Delta E_{xc} + 0.350\Delta E_e$$

It is lucidly shown that it is the exchange-correlation effect that dominates the hydrogen-bonding interactions while the electrostatic potential also plays an important role and the steric hindrance is trivial.

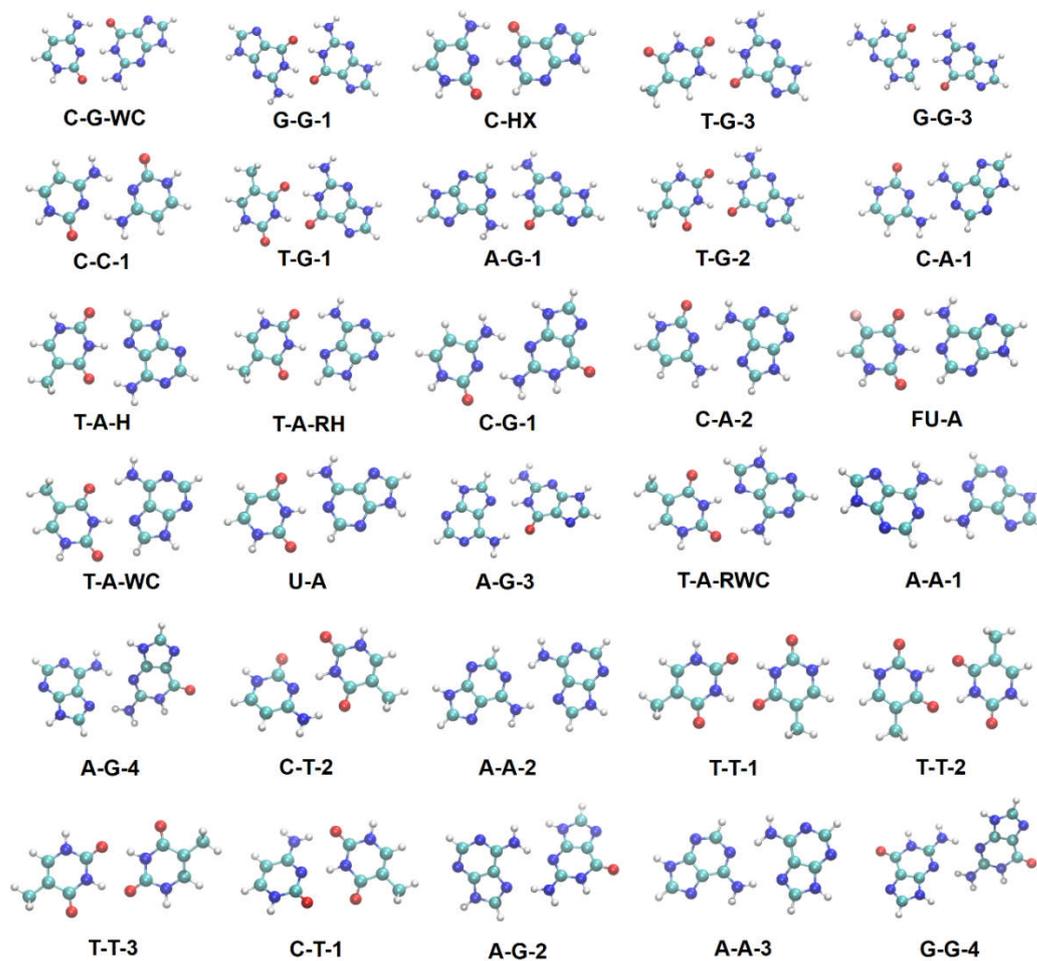
### 2.3. Molecular Polarizabilities of Base Pairs

In **Table 3**, we have collected the molecular polarizabilities, molecular volumes, and ITA quantities for 30 base pairs, which are obtained at the M06-2X/Def2-TZVP level. Also included in **Table 3** are the correlation coefficients ( $R^2$ ) between the molecular polarizabilities and molecular volumes and ITA quantities. One can observe that Shannon entropy ( $S_s$ ), Fisher information ( $I_F$ ), 2<sup>nd</sup> and 3<sup>rd</sup> relative Rényi entropy ( ${}^rR_2$  and  ${}^rR_3$ ), and  $G_3$  are in strong linear relationships with molecular polarizabilities, with  $R^2 > 0.8$ . It is intriguing to note that the absolute values of  $G_3$  are very close to molecular polarizabilities. However, the reason is unknown at present. In **Table 4**, we compare the polarizability results from conventional calculations,  $G_3$  data, and those predicted on top of the molecular wavefunctions obtained at the M06-2X/Def2-TZVP level. One can easily observe that the  $G_3$  data are very accurate in comparison to conventional polarizability data with **MSE**(%) to 1.5. However, the two sets of predicted data employing the original Tkatchenko-Scheffler (TS) approach[31] on top of Becke[32] or Hirshfeld[33] partitions are either strongly underestimated or overestimated, with **MUE**(%) up to -23.4 and 11.6, respectively. It is found that a mean value can greatly reduce the **MSE**(%) to 5.9. Still, a rational theoretical explanation is still lacking at the moment! Regarding the TS approach, a very recent publication has made some improvement[34] for small molecules. Yet, the performance of the variant TS method for macromolecules requires more extensive work.

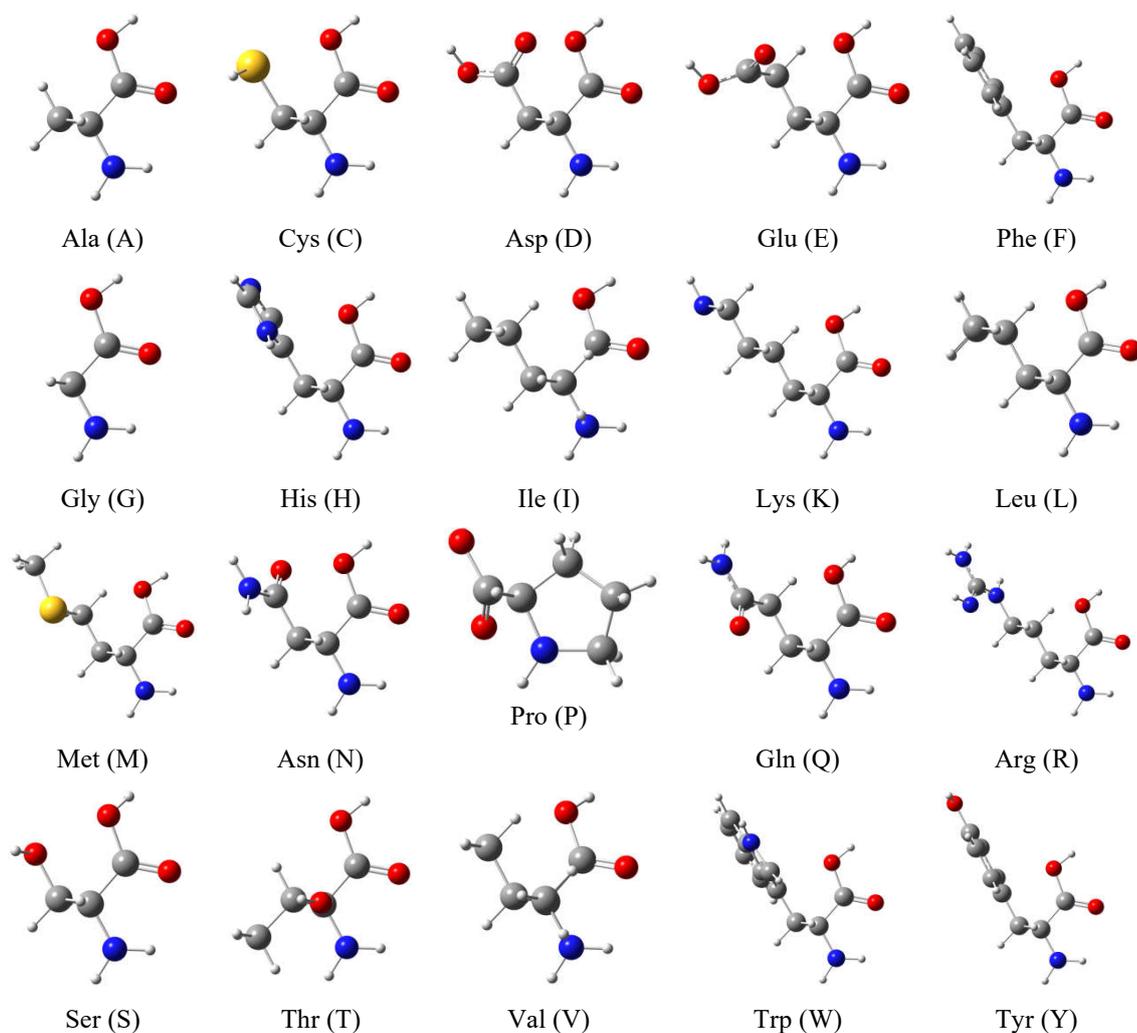
### 2.4. Molecular Polarizabilities of Amino Acids, Dipeptides and Tripeptides

We first cross-validate the accuracy of the molecular polarizabilities of 8000 tripeptides. The two theoretical levels are M06-2X/Def2-TZVP//M06-2X/Def2-SVP and B3LYP/6-31+G(d,p)//B3LYP/6-31G level[35,36], respectively. Shown in **Figure 5** is the strong correlation coefficient ( $R^2 = 0.992$ ) between the two sets of calculated molecular polarizabilities. The gap between the two methods is only 11.95 au, indicating that one can use a relatively cheaper theoretical method to obtain more accurate results. In addition, we have found some strong correlations for a total of 20 amino acids, 400 dipeptides, and 8000 tripeptides between molecular polarizabilities and molecular volumes, GBP entropy ( $S_{GBP}$ ), 2<sup>nd</sup> relative Rényi entropy ( ${}^rR_2$ ), and  $G_3$  as shown in **Figure 6**. The correlation coefficients are in the range of 0.87 to 0.95. We have to point out that molecular polarizabilities can be linearly correlated with molecular volumes.[37–39] Here, we have verified that ITA quantities can serve as good indicators of molecular properties, in our case, molecular polarizability. This means that one can directly predict molecular polarizabilities of larger peptides, or even proteins. It is well-documented that it is time-consuming to solve the CPHF equations[15,16]. When the molecular system becomes larger, it may be even intractable. For ITA quantities, as long as the total molecular wavefunction is obtained in a single-point calculation, numerical integration of ITA quantities normally requires much less time than iteratively solving the CPHF equations.

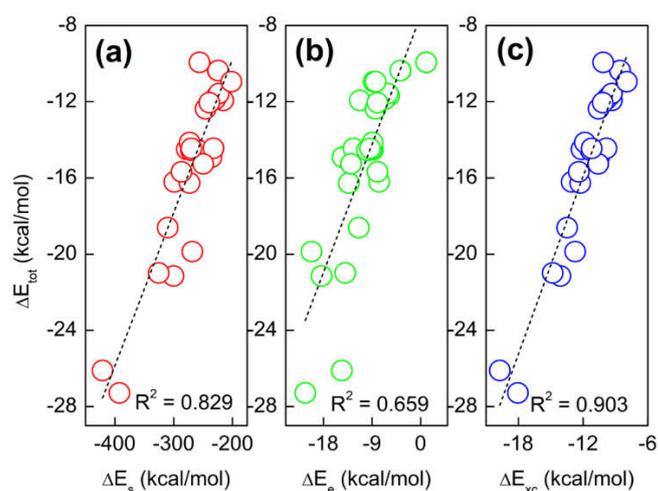
It is anticipated when it comes to real systems, like proteins, one can combine the linear-scaling fragment-based methods, such as GEBF (generalized-energy based fragmentation)[40–43], and ITA quantities to accurately and efficiently predict the molecular polarizabilities.



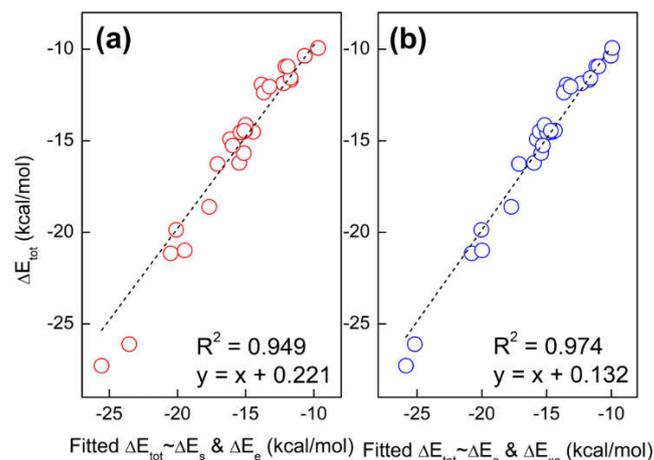
**Figure 1.** Optimized structures of the 30 base pairs. Color code: H in white, C in cyan, N in blue, O in red, and F in pink.



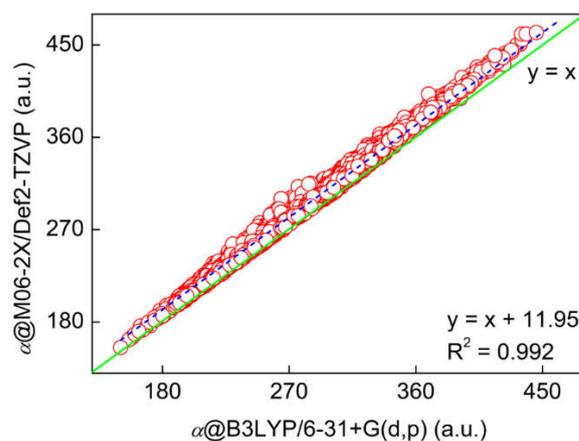
**Figure 2.** Schematic representation of 20 amino acids that form the basis of 400 dipeptides and 8000 tripeptides. Each molecule is named with a three-letter abbreviation and single-letter in the parenthesis. Color code: H in white, C in grey, N in blue, O in red, and S in yellow.



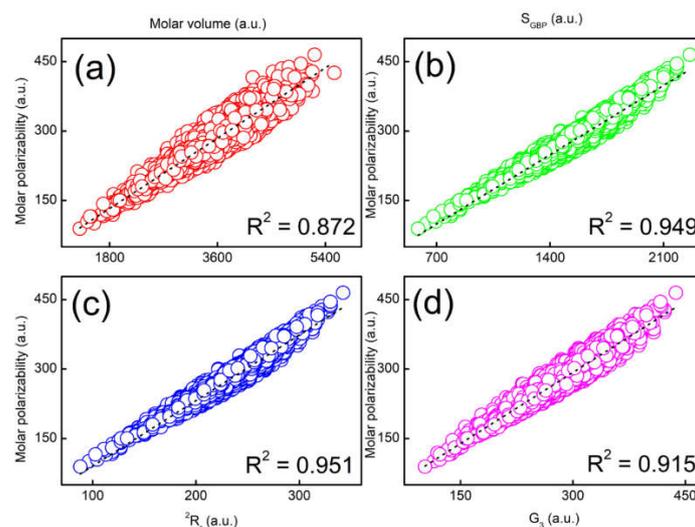
**Figure 3.** Strong correlations between the total energy difference ( $\Delta E_{\text{tot}}$ ) and (a) steric hindrance ( $\Delta E_s$ ), (b) electrostatic potential ( $\Delta E_e$ ), and (c) exchange-correlation effect ( $\Delta E_{\text{xc}}$ ). The  $y$ -axis of (a) spans to (b) and (c).



**Figure 4.** Two-variable fit of the total energy difference ( $\Delta E_{\text{tot}}$ ) with (a) steric hindrance ( $\Delta E_s$ ) and electrostatic potential ( $\Delta E_e$ ) and (b) electrostatic potential ( $\Delta E_s$ ) and exchange-correlation effect ( $\Delta E_{\text{xc}}$ ).



**Figure 5.** Cross-validation of molecular polarizabilities of 8000 tripeptides carried out at the M06-2X/Def2-TZVP//M06-2X/Def2-SVP and B3LYP/6-31+G(d,p)//B3LYP/6-31G level, respectively.



**Figure 6.** Strong correlations between molecular polarizability and (a) molecular volume, (b)  $S_{\text{GBP}}$ , (c)  ${}^2R_r$ , and (d)  $G_3$  for a total of 20 amino acids, 400 dipeptides, and 8000 tripeptides.

**Table 1.** Benchmark results of several basis sets combined with the M06-2X functional for 20 amino acids. Atomic units.

| Name                   | M06-2X/<br>6-311G(d,p) | M06-2X/<br>Def2-SVP | M06-2X/<br>Def2-TZVP | M06-2X/<br>aug-cc-pVTZ | MP2/<br>6-311G(d,p) <sup>a</sup> | exp. <sup>b</sup> |
|------------------------|------------------------|---------------------|----------------------|------------------------|----------------------------------|-------------------|
| Ala                    | 45.8                   | 43.6                | 50.1                 | 53.8                   | 45.5                             | 55.9              |
| Cys                    | 62.2                   | 57.8                | 67.6                 | 74.4                   | 62.0                             | --                |
| Asp                    | 59.6                   | 56.6                | 65.8                 | 70.6                   | 59.8                             | --                |
| Glu                    | 71.4                   | 68.1                | 78.1                 | 83.2                   | 70.4                             | 90.4              |
| Phe                    | 104.6                  | 102.0               | 113.6                | 119.6                  | 104.6                            | 122.9             |
| Gly                    | 33.8                   | 32.2                | 37.7                 | 42.2                   | 34.5                             | 44.3              |
| His                    | 85.2                   | 81.9                | 93.3                 | 99.0                   | 84.8                             | 102.6             |
| Ile                    | 78.9                   | 75.8                | 84.4                 | 88.4                   | 78.7                             | 95.2              |
| Lys                    | 87.2                   | 83.4                | 93.4                 | 99.3                   | 86.5                             | 101.2             |
| Leu                    | 79.4                   | 76.3                | 85.0                 | 89.1                   | 78.9                             | 94.5              |
| Met                    | 86.0                   | 81.5                | 93.0                 | 99.9                   | 84.0                             | 102.1             |
| Asn                    | 63.6                   | 60.4                | 70.1                 | 75.5                   | 64.1                             | 79.8              |
| Pro                    | 63.4                   | 61.0                | 68.4                 | 71.9                   | 63.3                             | 73.5              |
| Gln                    | 75.4                   | 71.8                | 82.3                 | 88.2                   | 75.3                             | 91.2              |
| Arg                    | 100.2                  | 95.4                | 108.6                | 116.3                  | 100.1                            | 115.6             |
| Ser                    | 48.7                   | 46.2                | 53.7                 | 58.0                   | 48.2                             | 61.2              |
| Thr                    | 59.9                   | 57.1                | 65.3                 | 69.7                   | 59.3                             | 73.7              |
| Val                    | 67.7                   | 64.9                | 72.8                 | 76.6                   | 67.6                             | 81.5              |
| Trp                    | 131.7                  | 128.8               | 142.7                | 149.8                  | 132.9                            | 157.8             |
| Tyr                    | 108.3                  | 105.4               | 117.6                | 123.7                  | 109.8                            | --                |
| <b>MSE<sup>c</sup></b> | -15.3                  | -18.4               | -8.9                 | -3.7                   | -15.6                            |                   |
| <b>MUE<sup>d</sup></b> | 15.3                   | 18.4                | 8.9                  | 3.8                    | 15.6                             |                   |

<sup>a</sup>Data taken from ref 27.<sup>b</sup>Data taken from ref 28.<sup>c</sup>MUE: mean unsigned error.<sup>d</sup>MSE: mean signed error.

**Table 2.** Total energy difference ( $\Delta E$ ) and its components obtained at the M06-2X/Def2-TZVP level for 30 base pairs, including kinetic ( $\Delta T_s$ ), exchange-correlation ( $\Delta E_{xc}$ ), electrostatic potential ( $\Delta E_e$ ), steric hindrance ( $\Delta E_s$ ), and quantum effect ( $\Delta E_q$ ). Units are in kcal/mol.

| Base pair | $\Delta T_s$ | $\Delta E_x$ | $\Delta E_c$ | $\Delta E_{xc}$ | $\Delta E_e$ | $\Delta E_s$ | $\Delta E_q$ | $\Delta E$ |
|-----------|--------------|--------------|--------------|-----------------|--------------|--------------|--------------|------------|
| C-G-WC    | 12.1         | -1.5         | -16.6        | -18.0           | -21.4        | -392.3       | 386.4        | -27.3      |
| G-G-1     | 8.2          | -2.4         | -17.4        | -19.7           | -14.6        | -421.0       | 409.4        | -26.1      |
| C-HX      | 11.3         | -0.4         | -13.7        | -14.1           | -18.4        | -300.3       | 297.5        | -21.1      |
| T-G-3     | 7.8          | -1.6         | -13.2        | -14.8           | -14.0        | -325.5       | 318.5        | -21.0      |
| G-G-3     | 6.3          | -0.9         | -12.5        | -13.5           | -11.5        | -310.0       | 302.9        | -18.6      |
| C-C-1     | 13.1         | 0.3          | -13.0        | -12.7           | -20.2        | -268.2       | 268.6        | -19.9      |
| T-G-1     | 4.5          | -0.8         | -12.3        | -13.1           | -7.7         | -298.9       | 290.4        | -16.2      |
| A-G-1     | 9.3          | 1.2          | -13.5        | -12.3           | -13.3        | -273.0       | 270.0        | -16.3      |
| T-G-2     | 4.7          | -0.6         | -11.8        | -12.4           | -8.0         | -286.7       | 279.1        | -15.7      |
| C-A-1     | 10.1         | 0.8          | -11.3        | -10.5           | -14.5        | -235.8       | 235.4        | -14.9      |
| T-A-H     | 6.3          | 0.8          | -12.0        | -11.2           | -9.6         | -270.3       | 265.4        | -14.5      |
| T-A-RH    | 6.2          | 0.8          | -12.0        | -11.2           | -9.4         | -270.0       | 264.9        | -14.4      |
| C-G-1     | 5.6          | 0.2          | -11.4        | -11.2           | -8.8         | -259.9       | 254.2        | -14.5      |
| C-A-2     | 7.8          | 1.5          | -11.3        | -9.8            | -12.4        | -232.2       | 230.2        | -14.4      |
| FU-A      | 6.7          | 0.6          | -12.7        | -12.1           | -9.1         | -277.6       | 272.2        | -14.5      |
| T-A-WC    | 6.7          | 0.8          | -12.1        | -11.3           | -9.9         | -270.7       | 266.0        | -14.5      |
| U-A       | 6.6          | 0.7          | -12.5        | -11.8           | -8.9         | -272.9       | 267.6        | -14.1      |
| A-G-3     | 8.2          | 2.0          | -12.6        | -10.6           | -12.9        | -249.8       | 247.5        | -15.3      |
| T-A-RWC   | 6.1          | 0.8          | -12.0        | -11.2           | -9.4         | -269.9       | 264.9        | -14.4      |
| A-A-1     | 8.7          | 0.9          | -10.2        | -9.3            | -11.3        | -215.8       | 215.1        | -11.9      |
| A-G-4     | 1.9          | 1.5          | -10.1        | -8.6            | -3.7         | -224.8       | 218.1        | -10.4      |
| C-T-2     | 6.5          | 1.3          | -11.9        | -10.6           | -8.4         | -245.6       | 241.6        | -12.4      |
| A-A-2     | 5.7          | 1.4          | -9.3         | -8.0            | -8.7         | -202.1       | 199.8        | -10.9      |
| T-T-1     | 3.6          | 0.1          | -9.7         | -9.5            | -5.8         | -226.1       | 220.3        | -11.7      |
| T-T-2     | 4.6          | 0.0          | -9.8         | -9.8            | -6.7         | -229.4       | 224.1        | -11.9      |
| T-T-3     | 3.8          | 0.2          | -9.5         | -9.3            | -6.1         | -223.2       | 217.7        | -11.6      |
| C-T-1     | 6.1          | 1.5          | -11.7        | -10.2           | -8.0         | -239.4       | 235.3        | -12.0      |
| A-G-2     | 1.9          | 1.5          | -10.1        | -8.6            | -3.7         | -224.4       | 217.8        | -10.4      |
| A-A-3     | 5.3          | 1.4          | -9.3         | -7.9            | -8.3         | -201.8       | 199.2        | -10.9      |
| G-G-4     | -0.9         | 0.8          | -11.0        | -10.1           | 1.1          | -256.2       | 245.2        | -9.9       |

**Table 3.** Molecular polarizabilities, molecular volumes, and ITA quantities for 30 base pairs.

| Base pair      | polar | Vol    | S <sub>S</sub> | I <sub>F</sub> | S <sub>GBP</sub> | E <sub>2</sub> | E <sub>3</sub> | rR <sub>2</sub> | rR <sub>3</sub> | G <sub>1</sub> | G <sub>2</sub> | G <sub>3</sub> |
|----------------|-------|--------|----------------|----------------|------------------|----------------|----------------|-----------------|-----------------|----------------|----------------|----------------|
| C-G-WC         | 170.2 | 1984.1 | 99.2           | 5854.3         | 924.3            | 866.2          | 48121.2        | 138.8           | 143.9           | -22.5          | 9.3            | 169.0          |
| G-G-1          | 197.2 | 1954.2 | 109.9          | 6775.8         | 1060.1           | 1002.5         | 55241.1        | 159.1           | 164.8           | -23.4          | 8.7            | 194.8          |
| C-HX           | 159.1 | 1863.3 | 93.6           | 5514.6         | 870.1            | 814.0          | 45149.3        | 130.6           | 135.6           | -23.6          | 11.2           | 157.8          |
| T-G-3          | 180.1 | 2060.0 | 105.0          | 6215.6         | 978.7            | 927.0          | 53287.8        | 147.0           | 152.7           | -26.6          | 12.1           | 177.2          |
| G-G-3          | 194.4 | 2184.6 | 110.2          | 6777.3         | 1060.3           | 1002.5         | 55250.6        | 159.1           | 164.9           | -23.6          | 8.7            | 194.0          |
| C-C-1          | 144.9 | 1731.0 | 88.8           | 4934.0         | 788.7            | 730.0          | 41006.7        | 118.5           | 123.3           | -21.5          | 9.9            | 142.1          |
| T-G-1          | 177.3 | 2058.5 | 105.1          | 6216.0         | 978.7            | 926.9          | 53282.0        | 147.0           | 152.7           | -26.7          | 12.5           | 177.1          |
| A-G-1          | 191.7 | 2146.7 | 108.8          | 6327.7         | 1005.6           | 921.3          | 48236.1        | 151.0           | 156.6           | -24.7          | 10.6           | 186.1          |
| T-G-2          | 177.2 | 2016.1 | 105.1          | 6216.1         | 978.8            | 926.9          | 53283.8        | 147.0           | 152.7           | -26.7          | 12.3           | 177.0          |
| C-A-1          | 165.7 | 1707.8 | 98.2           | 5406.3         | 869.9            | 785.1          | 41120.1        | 130.7           | 135.9           | -23.5          | 10.5           | 159.8          |
| T-A-H          | 169.8 | 2060.1 | 103.7          | 5766.3         | 924.1            | 845.6          | 46267.5        | 138.9           | 144.5           | -27.5          | 13.9           | 169.1          |
| T-A-RH         | 169.6 | 1997.3 | 103.8          | 5766.3         | 924.1            | 845.7          | 46272.0        | 138.9           | 144.5           | -27.5          | 13.8           | 169.0          |
| C-G-1          | 147.8 | 1899.5 | 99.6           | 5856.5         | 924.6            | 866.3          | 48137.5        | 138.8           | 144.0           | -22.8          | 9.7            | 167.6          |
| C-A-2          | 164.4 | 1810.0 | 98.2           | 5406.4         | 869.9            | 785.1          | 41118.5        | 130.7           | 135.9           | -23.6          | 10.8           | 159.9          |
| FU-A           | 158.2 | 1839.0 | 90.1           | 6089.2         | 925.1            | 933.8          | 59941.1        | 138.6           | 143.4           | -23.2          | 10.5           | 162.2          |
| T-A-WC         | 169.7 | 1950.4 | 103.7          | 5766.3         | 924.1            | 845.6          | 46264.0        | 138.9           | 144.5           | -27.5          | 14.0           | 169.1          |
| U-A            | 157.9 | 1892.1 | 93.6           | 5514.7         | 870.1            | 814.0          | 45157.7        | 130.6           | 135.6           | -23.7          | 11.0           | 157.6          |
| A-G-3          | 190.2 | 2205.6 | 108.9          | 6328.0         | 1005.6           | 921.4          | 48248.3        | 151.0           | 156.6           | -24.8          | 10.1           | 186.1          |
| T-A-RWC        | 169.7 | 2093.3 | 103.8          | 5766.3         | 924.1            | 845.7          | 46279.4        | 138.9           | 144.5           | -27.5          | 13.4           | 169.0          |
| A-A-1          | 186.9 | 2020.8 | 107.6          | 5878.4         | 951.0            | 840.0          | 41215.1        | 142.9           | 148.5           | -25.5          | 12.0           | 177.5          |
| A-G-4          | 189.6 | 2119.7 | 109.0          | 6328.3         | 1005.7           | 921.3          | 48241.8        | 151.0           | 156.6           | -24.7          | 10.4           | 185.3          |
| C-T-2          | 150.7 | 1701.9 | 94.4           | 5294.8         | 843.0            | 790.8          | 46189.2        | 126.7           | 131.9           | -25.7          | 12.1           | 150.9          |
| A-A-2          | 185.3 | 2157.6 | 107.7          | 5878.6         | 951.0            | 840.0          | 41219.9        | 142.9           | 148.5           | -25.6          | 11.8           | 177.6          |
| T-T-1          | 158.4 | 1822.9 | 100.1          | 5655.5         | 897.3            | 851.3          | 51315.6        | 135.0           | 140.6           | -29.8          | 16.3           | 159.7          |
| T-T-2          | 158.2 | 1918.8 | 100.1          | 5655.5         | 897.3            | 851.2          | 51313.7        | 134.9           | 140.6           | -29.8          | 16.5           | 159.7          |
| T-T-3          | 158.7 | 1961.9 | 100.1          | 5655.5         | 897.3            | 851.3          | 51318.3        | 135.0           | 140.6           | -29.8          | 16.0           | 159.8          |
| C-T-1          | 150.3 | 1874.4 | 94.4           | 5294.9         | 843.0            | 790.7          | 46172.2        | 126.7           | 131.9           | -25.7          | 12.9           | 150.8          |
| A-G-2          | 189.6 | 2023.7 | 109.0          | 6328.3         | 1005.7           | 921.2          | 48235.2        | 151.0           | 156.6           | -24.7          | 10.7           | 185.4          |
| A-A-3          | 185.3 | 2114.5 | 107.7          | 5878.6         | 951.0            | 840.0          | 41217.5        | 142.9           | 148.5           | -25.6          | 11.9           | 177.4          |
| G-G-4          | 195.4 | 2115.8 | 110.3          | 6777.9         | 1060.4           | 1002.5         | 55256.4        | 159.0           | 164.7           | -23.7          | 9.2            | 193.4          |
| R <sup>2</sup> | 1.000 | 0.632  | 0.834          | 0.722          | 0.824            | 0.544          | 0.048          | 0.826           | 0.828           | 0.007          | 0.141          | 0.904          |

**Table 4.** Comparison of molecular polarizabilities of 30 base pairs with conventional data as reference. Atomic units.

| Base pair                 | $\alpha_{\text{iso}}$ | Other work <sup>a</sup> |           |       | This work      |
|---------------------------|-----------------------|-------------------------|-----------|-------|----------------|
|                           |                       | Becke                   | Hirshfeld | avg.  | G <sub>3</sub> |
| C-G-WC                    | 170.2                 | 128.7                   | 188.1     | 158.4 | 169.0          |
| G-G-1                     | 197.2                 | 144.9                   | 213.1     | 179.0 | 194.8          |
| C-HX                      | 159.1                 | 122.0                   | 177.4     | 149.7 | 157.8          |
| T-G-3                     | 180.1                 | 135.3                   | 198.3     | 166.8 | 177.2          |
| G-G-3                     | 194.4                 | 145.2                   | 212.6     | 178.9 | 194.0          |
| C-C-1                     | 144.9                 | 112.9                   | 163.1     | 138.0 | 142.1          |
| T-G-1                     | 177.3                 | 135.5                   | 198.2     | 166.9 | 177.1          |
| A-G-1                     | 191.7                 | 142.3                   | 207.9     | 175.1 | 186.1          |
| T-G-2                     | 177.2                 | 135.5                   | 198.3     | 166.9 | 177.0          |
| C-A-1                     | 165.7                 | 126.2                   | 182.9     | 154.5 | 159.8          |
| T-A-H                     | 169.8                 | 132.3                   | 193.4     | 162.8 | 169.1          |
| T-A-RH                    | 169.6                 | 132.3                   | 193.4     | 162.9 | 169.0          |
| C-G-1                     | 169.9                 | 129.2                   | 187.5     | 158.4 | 167.7          |
| C-A-2                     | 164.4                 | 126.3                   | 183.2     | 154.7 | 159.9          |
| FU-A                      | 158.2                 | 122.1                   | 177.8     | 150.0 | 162.2          |
| T-A-WC                    | 169.7                 | 132.3                   | 193.4     | 162.8 | 169.1          |
| U-A                       | 157.9                 | 122.0                   | 177.4     | 149.7 | 157.6          |
| A-G-3                     | 190.2                 | 142.4                   | 208.0     | 175.2 | 186.1          |
| T-A-RWC                   | 169.7                 | 132.3                   | 193.4     | 162.9 | 169.0          |
| A-A-1                     | 186.9                 | 139.4                   | 202.6     | 171.0 | 177.5          |
| A-G-4                     | 189.6                 | 142.5                   | 207.5     | 175.0 | 185.3          |
| C-T-2                     | 150.7                 | 119.3                   | 173.4     | 146.4 | 150.9          |
| A-A-2                     | 185.3                 | 139.5                   | 202.8     | 171.2 | 177.6          |
| T-T-1                     | 158.4                 | 125.8                   | 183.6     | 154.7 | 159.7          |
| T-T-2                     | 158.2                 | 125.8                   | 183.6     | 154.7 | 159.7          |
| T-T-3                     | 158.7                 | 125.8                   | 183.7     | 154.7 | 159.8          |
| C-T-1                     | 150.3                 | 119.4                   | 173.4     | 146.4 | 150.8          |
| A-G-2                     | 189.6                 | 142.5                   | 207.5     | 175.0 | 185.4          |
| A-A-3                     | 185.3                 | 139.5                   | 202.8     | 171.2 | 177.4          |
| G-G-4                     | 195.4                 | 145.5                   | 212.1     | 178.8 | 193.4          |
| <b>MUE(%)<sup>b</sup></b> |                       | -23.4                   | 11.6      | -5.9  | -1.2           |
| <b>MSE(%)<sup>c</sup></b> |                       | 23.4                    | 11.6      | 5.9   | 1.5            |

<sup>a</sup>Tkatchenko-Scheffler approach on top of Becke or Hirshfeld partitions.<sup>b</sup>MUE: mean unsigned error.<sup>c</sup>MSE: mean signed error.

### 3. Discussion

We have systematically investigated the energetics of 30 base pairs and molecular polarizabilities of 30 base pairs and 20 amino acids, 400 dipeptides, and 8000 tripeptides. It is the first time that ITA quantities have been applied to correlate with molecular polarizabilities. Though at present, in theory we have not verified the existence of such a linear correlation between ITA quantities and molecular polarizabilities, they are both related to electron density and its derivatives. Also, our results presented here demonstrate that one can determine the molecular polarizability by numerically integrating the simple density-based functions when molecular wavefunction a given system is obtained from single-point calculations, rather than resorting to iteratively obtain the molecular orbital derivatives. Implication of our work is straightforward that one can use the linear relationships

as shown in the text to predict molecular polarizabilities of relatively larger peptides, or even larger proteins.

We have to point out that we are not the first to predict the molecular polarizabilities just as employing the electron density as the input. Jayatilaka et al.[44] has cast the molecular polarizability in terms of moments of the ground-state electron density matrix and the results are reasonably good against CPHF results. The corresponding performance for hyperpolarizability is far from satisfactory. Since hyperpolarizability is not considered in this work, we will figure out if abovementioned density-based ITA functions can predict the molecular hyperpolarizability well.

#### 4. Materials and Methods

##### 4.1. Energy Decomposition Schemes in DFT

In Kohn-Shan DFT, the total energy difference ( $\Delta E$ ) can be decomposed into its components via[3,4]

$$\Delta E[\rho] = \Delta T_s[\rho] + \Delta E_e[\rho] + \Delta E_{xc}[\rho] \quad (1)$$

and

$$\Delta E[\rho] = \Delta E_s[\rho] + \Delta E_e[\rho] + \Delta E_q[\rho] \quad (2)$$

where  $T_s$ ,  $E_e$ , and  $E_{xc}$  are the noninteracting kinetic, electrostatic, and exchange–correlation terms, respectively. The electrostatic potential  $E_e$  has three components: the nuclear–electron attraction, ( $V_{ne}$ ), the classical Coulombic repulsion, ( $J$ ), and the nuclear–nuclear repulsion ( $V_{nn}$ ). The last term  $E_{xc}$  consists of exchange ( $E_x$ ) and correlation ( $E_c$ ) components. In eq 2,  $E_s$  stands for the steric hindrance, and  $E_q$  signifies the contribution from Fermionic quantum effect. The steric effect  $E_s$  has been proved to be simply the Weizsäcker kinetic energy,

$$\tau_w(\mathbf{r}) = \frac{1}{8} \frac{|\nabla\rho(\mathbf{r})|^2}{\rho(\mathbf{r})} \quad (3)$$

Combining eqs 1, 2 and 3, one can simply defines  $E_q$ , which reads

$$\Delta E_q[\rho] = \Delta E_{xc}[\rho] + \Delta T_s[\rho] - \Delta E_s[\rho] \quad (4)$$

This new formulation has its own distinct physical meaning with a corresponding physical state. It has been applied to a number of molecular systems and phenomena, whose results are consistent with our chemical intuition and conventional wisdom.[45]

##### 4.2. Information-Theoretic Approach Quantities

Shannon entropy  $S_S$ [5] is a measure of the spatial delocalization of the electron density, and Fisher information  $I_F$ [6] measures the sharpness or localization of the same. They are defined as eqs 5 and 6, respectively

$$S_S = - \int \rho(\mathbf{r}) \ln \rho(\mathbf{r}) d\mathbf{r} \quad (5)$$

$$I_F = - \int \frac{|\nabla\rho(\mathbf{r})|^2}{\rho(\mathbf{r})} d\mathbf{r} \quad (6)$$

Additionally, for atoms and molecules,  $I_F$  has an equivalent expression[46] in terms of the Laplacian of the electron density,  $\nabla^2\rho(\mathbf{r})$

$$I'_F = - \int [\nabla^2\rho(\mathbf{r})] \ln \rho(\mathbf{r}) d\mathbf{r} \quad (7)$$

Ghosh–Berkowitz–Parr (GBP) entropy  $S_{GBP}$ [7]

$$S_{GBP} = - \int \frac{3}{2} k\rho(\mathbf{r}) \left[ c + \ln \frac{t(\mathbf{r}; \rho)}{t_{TF}(\mathbf{r}; \rho)} \right] d\mathbf{r} \quad (8)$$

where  $t(\mathbf{r}; \rho)$  is the kinetic energy density, which is related to the total kinetic energy  $T_s$  via

$$\int t(\mathbf{r}; \rho) d\mathbf{r} = T_s \quad (9)$$

$t_{TF}(\mathbf{r}; \rho)$  is the Thomas–Fermi (TF) kinetic energy density given by

$$t_{TF}(\mathbf{r}; \rho) = c_K \rho^{5/3}(\mathbf{r}) \quad (10)$$

with  $K$  as the Boltzmann constant (set to be unity for convenience in this work),  $c = (5/3) + \ln(4\pi c_K/3)$ , and  $c_K = (3/10)(3\pi^2)^{2/3}$ , the specific form of the local kinetic energy

$$t(\mathbf{r}; \rho) = \sum_i \frac{1}{8} \frac{\nabla \rho_i \cdot \nabla \rho_i}{\rho_i} - \frac{1}{8} \nabla^2 \rho \quad (11)$$

More recently, additional ITA quantities have been introduced as new reactivity descriptors in conceptual density functional theory (CDFT).[8] One example is Onicescu information energy[8] of order  $n$

$$E_n = \frac{1}{n-1} \int \rho^n(\mathbf{r}) d\mathbf{r} \quad (12)$$

relative Rényi entropy[8] of order  $n$

$$R_n^r = \frac{1}{n-1} \ln \left[ \int \frac{\rho^n(\mathbf{r})}{\rho_0^{n-1}(\mathbf{r})} d\mathbf{r} \right] \quad (13)$$

and information gain[9] (also called Kullback–Leibler divergence)  $I_G$  is given in eq 14

$$I_G = \int \rho(\mathbf{r}) \ln \frac{\rho(\mathbf{r})}{\rho_0(\mathbf{r})} d\mathbf{r} \quad (14)$$

where  $\rho_0(\mathbf{r})$  is the reference-state density satisfying the same normalization condition as  $\rho(\mathbf{r})$ .

Very recently,[10] we have proposed another three functions  $G_1$ ,  $G_2$ , and  $G_3$ , whose analytical forms as shown below:

$$G_3 = \sum_A \int \rho_A(\mathbf{r}) \left[ \nabla \ln \frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})} \right]^2 d\mathbf{r} \quad (15)$$

$$G_1 = \sum_A \int \nabla^2 \rho_A(\mathbf{r}) \frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})} d\mathbf{r} \quad (16)$$

$$G_2 = \sum_A \int \rho_A(\mathbf{r}) \left[ \frac{\nabla^2 \rho_A(\mathbf{r})}{\rho_A(\mathbf{r})} - \frac{\nabla^2 \rho_A^0(\mathbf{r})}{\rho_A^0(\mathbf{r})} \right] d\mathbf{r} \quad (17)$$

The  $G_2$  function involves Laplacian contribution and can be regarded as the relative Laplacian contribution to the steric potential. The quantifications and applications of eqs 15–17 can be found in ref 11. Suffice to note during the past decade we have attempted to seamlessly glue the density functional theory and information theory together as electron density can be used as a linker of these two theories. The progress and applications can be found in our recent review.[45] For example, very recently we have applied the information-theoretic approach to appreciate homochirality,[47,48] which is another very fundamental problem in biology and we will look into it in the near future.

### 4.3. Computational Details

#### 4.3.1. Base Pairs

Following the notation of ref 2, we build a total of 30 hydrogen-bonded nucleic acid base pairs as shown in **Figure 1**. For each complex, the nucleic acid bases are denoted by one or two letter abbreviations, among which are the four DNA bases: adenine (A), cytosine (C), guanine (G), and thymine (T), as well as fluorouracil (FU), hypoxanthine (HX), and uracil (U). For the most common configurations we use the usual abbreviations: Watson-Crick (WC), reversed Watson-Crick (RWC), Hoogsteen (H), and reversed Hoogsteen (RH). We also use numbers to distinguish between different configurations of the same complex and not to introduce any kind of ordering.

#### 3.3.2. Amino acids, Dipeptides, and Tripeptides

Shown in **Figure 2** are the 20 natural amino acids, directly taken from the template of the GaussView program. We used the tleap module in the AmberTools package to generate a series of capped (ACE and NME) dipeptides (400) and tripeptides (8000). For each of the peptide, a simulation box was built with water molecules and  $\text{Na}^+$  and  $\text{Cl}^-$  ions. We performed a total of 1 ns MD simulations with the ff19SB[49] force field for the peptide and TIP3P[50] for the water molecules. An NPT (1 atm, 300 K) ensemble was used with a time step of 2 fs. The temperature was controlled with underdamped Langevin simulations of the “virtual” solvent with the damping coefficient  $\gamma = 5 \text{ ps}^{-1}$ . [51] The pressure was held constant by applying the Langevin piston method.[52,53] Periodic long-range electrostatic interactions were computed using the particle mesh Ewald summation.[54] Covalent bonds associated with hydrogen atoms were constrained by the SHAKE

algorithm.[55] All classical MD simulations were carried out with the Amber20[56,57] CUDA version. A total of 10 MD snapshot structures of each peptide were evenly extracted for further quantum chemical calculations. Since we were not going to locate the global minimum conformers, we first performed single-point calculations of the 10 structures at the semi-empirical PM7[58] level, and selected the structure with the lowest energy for subsequent analysis.

All density functional theory (DFT) calculations were conducted with the Gaussian 16[59] package with ultrafine integration grids and tight self-consistent field convergence. A full geometric optimization at the M06-2X/Def2-SVP level with no symmetry constraint was conducted for 30 base pairs and 8420 peptides. The optimized Cartesian coordinates are supplied in the Supplementary Information. Since it is too computationally demanding to perform harmonic vibrational frequency analysis for all systems in this study, we chose 30 base pairs, 20 amino acids, 400 dipeptides, and 10 randomly selected tripeptides and no imaginary frequencies were observed. Molecular polarizabilities, volumes, and wavefunctions were obtained at the M06-2X/Def2-TZVP level. Total energy components were obtained via the keyword `iop(5/33=1)`. The Multiwfn 3.8[60] program was utilized to calculate all ITA quantities introduced above by using the checkpoint or wavefunction file from the Gaussian calculations as the input. Molecular volumes were obtained at 0.001 au contour surface of electronic density. The stockholder Hirshfeld partition scheme of atoms in molecules was employed when atomic contributions were concerned. The reference density was the neutral atom calculated at the same level of theory as molecules.

## 5. Conclusions

In this work, we have explored the interaction energy of base pairs and molecular properties of both base pairs and amino acids, dipeptides, and tripeptides. Using the total energy decomposition schemes, one can find that the exchange-correlation effect makes the predominant contribution to the molecular stability, followed by the electrostatic potential, and steric hindrance plays a trivial role. We further revealed that molecular polarizabilities can be linearly correlated with ITA quantities. These strong linear correlations can be used to predict the molecular polarizabilities of larger peptides and even proteins. We mention in passing that when a fragment-based method, such as the linear-scaling generalized energy-based fragmentation (GEBF) approach, is taken into consideration, one can directly predict the subsystem (fragment of a large system, usually with a few atoms within a distance threshold) polarizabilities rather than solving the derivatives of molecular orbitals and obtain the molecular polarizability of macromolecular systems via a linear combination of subsystem polarizabilities. More work along this line is ongoing and results will be presented elsewhere in the very near future.

**Supplementary Materials:** The optimized Cartesian coordinates of 30 base pairs and 20 amino acids, 400 peptides, and 8000 tripeptides.

**Author Contributions:** Conceptualization, Shubin Liu and Dahua Chen; Data curation, Dongbo Zhao; Formal analysis, Dongbo Zhao; Funding acquisition, Dongbo Zhao; Project administration, Shubin Liu and Dahua Chen; Supervision, Shubin Liu and Dahua Chen; Writing – original draft, Dongbo Zhao; Writing – review & editing, Shubin Liu and Dahua Chen.

**Funding:** This work was supported by the Yunnan Fundamental Research Projects (grant NO. 202101AU070012).

**Data Availability Statement:** Data is contained within the article.

**Acknowledgments:** Professor Pratim K. Chattaraj is acknowledged for this invitation.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Šponer, J.; Leszczynski, J.; Hobza, P. Structures and Energies of Hydrogen-Bonded DNA Base Pairs. A Nonempirical Study with Inclusion of Electron Correlation. *J. Phys. Chem.* **1996**, *100*, 1965–1974.

2. Hobza, P.; Šponer, J. Structure, Energetics, and Dynamics of the Nucleic Acid Base Pairs: Nonempirical *Ab Initio* Calculations. *Chem. Rev.* **1999**, *99*, 3247–3276.
3. 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.
4. Parthasarathi, R.; Amutha, R.; Subramanian, V.; Nair, B.U.; Ramasami, T. Bader's and Reactivity Descriptors' Analysis of DNA Base Pairs. *J. Phys. Chem. A* **2004**, *108*, 3817–3828.
5. Toczyłowski, R.R.; Cybulski, S.M. An Analysis of the Interactions between Nucleic Acid Bases: Hydrogen-Bonded Base Pairs. *J. Phys. Chem. A* **2003**, *107*, 418–426.
6. Parr, R.G.; Yang, W.T. *Density Functional Theory of Atoms and Molecules*; Oxford University Press: Oxford, U.K., 1989.
7. Liu, S.B. Steric effect: a quantitative description from density functional theory. *J. Chem. Phys.* **2007**, *126*, 244103.
8. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423.
9. Fisher, R.A. Theory of statistical estimation. *Math. Proc. Cambridge Philos. Soc.* **1925**, *22*, 700–725.
10. Ghosh, S. K.; Berkowitz, M.; Parr, R.G. Transcription of ground-state density-functional theory into a local thermodynamics. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 8028–8031.
11. Liu, S.B.; Rong, C.Y.; Wu, Z.M.; Lu, T. Rényi entropy, Tsallis entropy and Onicescu information energy in density functional reactivity theory. *Acta Phys.-Chim. Sin.* **2015**, *31*, 2057–2063.
12. Kullback, S. *Information Theory and Statistics*; Dover Publications: Mineola, NY, 1997.
13. Liu, S.B. Identity for Kullback-Leibler divergence in density functional reactivity theory. *J. Chem. Phys.* **2019**, *151*, 141103.
14. Wang, B.; Zhao, D.B.; Lu, T.; Liu, S.B.; Rong, C.Y. Quantifications and Applications of Relative Fisher Information in Density Functional Theory. *J. Phys. Chem. A* **2021**, *125*, 3802–3811.
15. McWeeny, R. Some recent advances in density matrix theory. *Rev. Mod. Phys.* **1960**, *32*, 335–369.
16. Langhoff, P.W.; Karplus, M.; Hurst, R.P. Approximations to Hartree–Fock Perturbation Theory. *J. Chem. Phys.* **1966**, *44*, 505–514.
17. Colwell, S.M.; Murray, C.W.; Handy, N.C.; Amos, R.D. The determination of hyperpolarisabilities using density functional theory. *Chem. Phys. Lett.* **1993**, *210*, 261–268.
18. Grisafi, A.; Wilkins, D.M.; Csányi, G.; Ceriotti, M. Symmetry-Adapted Machine Learning for Tensorial Properties of Atomistic Systems. *Phys. Rev. Lett.* **2018**, *120*, 036002.
19. Wilkins, D. M.; Grisafi, A.; Yang, Y.; Lao, K. U.; DiStasio, R. A. Jr.; Ceriotti, M. Accurate molecular polarizabilities with coupled cluster theory and machine learning. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116*, 3401–3406.
20. Nguyen, V.H.A.; Lunghi, A. Predicting tensorial molecular properties with equivariant machine learning models. *Phys. Rev. B* **2022**, *105*, 165131.
21. Amin, M.; Samy, H.; Küpper, J. Robust and Accurate Computational Estimation of the Polarizability Tensors of Macromolecules. *J. Phys. Chem. Lett.* **2019**, *10*, 2938–2943.
22. Hait, D.; Head-Gordon M. How accurate are static polarizability predictions from density functional theory? An assessment over 132 species at equilibrium geometry. *Phys. Chem. Chem. Phys.* **2018**, *20*, 19800–19810.
23. Zhao, Y.; Truhlar, D.G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, non-covalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* **2008**, *120*, 215–241.
24. McLean, A.D.; Chandler, G.S. Contracted Gaussian-basis sets for molecular calculations. 1. Second row atoms,  $Z=11-18$ . *J. Chem. Phys.* **1980**, *72*, 5639–5648.
25. Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.
26. Dunning, T. H. Jr., Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007–1023.
27. Krishtal, A.; Senet, P.; Alsenoy, C.V. Local softness, softness dipole, and polarizabilities of functional groups: Application to the side chains of the 20 amino acids. *J. Chem. Phys.* **2009**, *131*, 044312.
28. McMeekin, T.L.; Groves, M.L.; Wilensky, M. Refractive indices of proteins in relation to amino acid composition and specific volume. *Biochem. Biophys. Res. Commun.* **1962**, *7*, 151–156.
29. Boys, S. F.; Bernardi, F. The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Mol. Phys.* **1970**, *19*, 553–566.
30. Simon, S.; Duran, M.; Dannenberg, J. J. How does basis set superposition error change the potential surfaces for hydrogen-bonded dimers? *J. Chem. Phys.* **1996**, *105*, 11024–11031.
31. Tkatchenko, A.; Scheffler, M. Accurate Molecular Van Der Waals Interactions from Ground-State Electron Density and Free-Atom Reference Data. *Phys. Rev. Lett.* **2009**, *102*, 073005.
32. Hirshfeld, F.L. Bonded-Atom Fragments for Describing Molecular Charge Densities. *Theoret. Chim. Acta* **1977**, *44*, 129–138.
33. Szabó, P.; Góger, S.; Charry, J.; Karimpour, M.R.; Fedorov, D.V.; Tkatchenko, A. Four-Dimensional Scaling of Dipole Polarizability in Quantum Systems. *Phys. Rev. Lett.* **2022**, *128*, 070602.
34. Becke, A.D. A multicenter numerical integration scheme for polyatomic molecules. *J. Chem. Phys.* **1988**, *88*, 2547–2553.
35. Becke, A.D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
36. Lee, C.; Yang, W.; Parr, R.G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B*, **1988**, *37*, 785–789.

37. Brinck, T.; Murray, J.S.; Politzer, P. Polarizability and volume. *J. Chem. Phys.* **1993**, *98*, 4305–4306.
38. Politzer, P.; Jin, P.; Murray, J.S. Atomic polarizability, volume and ionization energy. *J. Chem. Phys.* **2002**, *117*, 8197–8202.
39. Blair, S.A.; Thakkar, A.J. Relating polarizability to volume, ionization energy, electronegativity, hardness, moments of momentum, and other molecular properties. *J. Chem. Phys.* **2014**, *141*, 074306.
40. Li, S.; Li, W.; Fang, T. An Efficient Fragment-Based Approach for Predicting the Ground-State Energies and Structures of Large Molecules. *J. Am. Chem. Soc.* **2005**, *127*, 7215–7226.
41. Li, W.; Li, S.; Jiang, Y. Generalized Energy-Based Fragmentation Approach for Computing the Ground-State Energies and Properties of Large Molecules. *J. Phys. Chem. A* **2007**, *111*, 2193–2199.
42. Li, S.; Li, W.; Ma, J. Generalized Energy-Based Fragmentation Approach and Its Applications to Macromolecules and Molecular Aggregates. *Acc. Chem. Res.* **2014**, *47*, 2712–2720.
43. Li, W.; Dong, H.; Ma, J.; Li, S. Structures and Spectroscopic Properties of Large Molecules and Condensed-Phase Systems Predicted by Generalized Energy-Based Fragmentation Approach. *Acc. Chem. Res.* **2021**, *54*, 169–181.
44. Jayatilaka, D.; Jha, K.K.; Munshi, P. Is it Reasonable to Obtain Information on the Polarizability and Hyperpolarizability Only from the Electron Density? *Aust. J. Chem.* **2018**, *71*, 295–306.
45. Rong, C.Y.; Wang, B.; Zhao, D.B.; Liu, S.B. Information-theoretic approach in density functional theory and its recent applications to chemical problems. *WIREs Comput. Mol. Sci.* **2020**, *10*, e1461.
46. Liu, S.B. On the relationship between densities of Shannon entropy and Fisher information for atoms and molecules. *J. Chem. Phys.* **2007**, *126*, 191107.
47. Liu, S.B. Homochirality Originates from Handedness of Helices. *J. Phys. Chem. Lett.* **2020**, *11*, 8690–8696.
48. Liu, S.B. Principle of Chirality Hierarchy in Three-Blade Propeller Systems. *J. Phys. Chem. Lett.* **2021**, *12*, 8720–8725.
49. Tian, C.; Kasavajhala, K.; Belfon, K.A.A.; Raguette, L.; Huang, H.; Miguez, A.N.; Bickel, J.; Wang, Y.; Pincay, J.; Wu, Q.; Simmerling, C. ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *J. Chem. Theory Comput.* **2020**, *16*, 528–552.
50. Jorgensen, W.L.; Chandrasekhar, J.; Madura, J.D.; Impey, R.W.; Klein, M.L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.
51. Adelman, S.; Doll, J. Generalized Langevin equation approach for atom/solid-surface scattering: General formulation for classical scattering off harmonic solids. *J. Chem. Phys.* **1976**, *64*, 2375–2388.
52. Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. Constant pressure molecular dynamics simulation: The Langevin piston method. *J. Chem. Phys.* **1995**, *103*, 4613–4621.
53. Martyna, G. J.; Tobias, D. J.; Klein, M. L. Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* **1994**, *101*, 4177–4189.
54. Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An  $N \cdot \log(N)$  method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
55. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of  $n$ -alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.
56. Goetz, A.W.; Williamson, M.J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R.C. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory Comput.* **2012**, *8*, 1542–1555.
57. Salomon-Ferrer, R.; Goetz, A.W.; Poole, D.; Le Grand, S.; Walker, R.C. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9*, 3878–3888.
58. Stewart, J.J.P. Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters. *J. Mol. Model.* **2013**, *19*, 1–32.
59. Frisch, M.J.; Trucks, G.W.; Schlegel, H.B.; Scuseria, G.E.; Robb, M.A.; Cheeseman, J.R.; Scalmani, G.; Barone, V.; Petersson, G.A.; Nakatsuji, H.; et al. *Gaussian 16 Revision C.01*; Gaussian Inc.: Wallingford, CT, USA, 2016.
60. Lu, T.; Chen, F.W. Multiwfn: a multifunctional wavefunction analyzer. *J. Comput. Chem.* **2012**, *33*, 580–592.