

Density functional theory-symmetry adapted perturbation treatment energy decomposition of nucleic acid base pairs taken from DNA crystal geometry

Robert Sedláč

Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Flemingovo náměstí 2, 166 10 Prague 6, Czech Republic and Center for Biomolecules and Complex Molecular Systems, Flemingovo náměstí 2, 166 10 Prague 6, Czech Republic

Petr Jurečka

Department of Physical Chemistry, Palacký University, třída Svobody 26, 771 46 Olomouc, Czech Republic

Pavel Hobza^{a)}

Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Flemingovo náměstí 2, 166 10 Prague 6, Czech Republic; Center for Biomolecules and Complex Molecular Systems, Flemingovo náměstí 2, 166 10 Prague 6, Czech Republic; and Department of Physical Chemistry, Palacký University, třída Svobody 26, 771 46 Olomouc, Czech Republic

(Received 23 April 2007; accepted 20 June 2007; published online 20 August 2007)

[DOI: [10.1063/1.2759207](https://doi.org/10.1063/1.2759207)]

Stabilization energies of DNA base pairs are of utmost importance as they make it possible to deduce the origin of DNA duplex stabilization. Nevertheless, it is essential that highly accurate values be used, which is rather easy for H-bonded pairs but much more complicated for stacked pairs. A fairly reliable technique is the CCSD(T) method in combination with the complete basis set (CBS) limit, and such CCSD(T) energies for both gas-phase optimized and crystal structures have appeared recently.¹ The stabilization energies for planar H-bonded structures in the gas phase and in crystal are similar, which only confirms the well-known fact that the geometries of these pairs (the Watson-Crick structure) are similar in both phases. This is, however, not the case with stacked pairs, and the optimized and crystal structures differ substantially.

Despite being accurate, total stabilization energies do not provide any information about their partitioning into physically defined energy components, which can be of particular consequence for interpreting the differences between planar H bonding and vertical stacking. Such partitioning can be achieved in different ways and one of the most rigorous is using the symmetry adapted perturbation treatment (SAPT).² Nonetheless, the SAPT calculations are time consuming, and their application in extended complexes was made possible only when the SAPT intersystem treatment was combined with the density functional theory (DFT) description of subsystems.^{3,4} The DFT-SAPT or SAPT(DFT) method combining the DFT treatment for monomers and SAPT treatment for intermonomer interactions was used for decomposition of total interaction energy for both the benzene dimer⁵ and DNA base pairs.⁶ The latter study used the gas-phase optimized geometries of adenine (A)··thymine (T) and guanine (G)··cytosine (C) H-bonded [both in the Watson-Crick (WC) arrangement] and stacked structures. Whereas the crystal and gas-phase optimized structures are similar in terms of their H-bonded motifs, they differ considerably in the stacked ones. Furthermore, the DNA contains other

stacked motifs (not only the AT and GC investigated in Ref. 6).

The aim of the present communication is to decompose the total interaction energy for various H-bonded and stacked DNA base pairs in crystal geometry for which the CCSD(T)/CBS energies are available.¹ In total, we have investigated 6 complementary and noncomplementary H-bonded pairs, 8 intrastrand stacked pairs, and 31 interstrand stacked pairs.

In the DFT-SAPT method,³ the interaction energy is given as the sum of first- and second-order energies $[E^{(1)}, E^{(2)}]$ and also the $\delta(\text{HF})$ term. The former energy contains electrostatic $[E_{\text{el}}^{(1)}]$ and exchange-repulsion $[E_{\text{ex}}^{(1)}]$ contributions and the latter the induction, exchange-induction, dispersion, and exchange-dispersion contributions. Accordingly, second-order exchange components will be added into the induction $[E_i^{(2)}]$ and dispersion $[E_D^{(2)}]$ terms. The charge-transfer energy is considered as a part of induction energy. The $\delta(\text{HF})$ term estimates higher order Hartree-Fock contributions (induction, exchange induction, and charge transfer). We used PBE0AC exchange-correlation functional with density fitting and the aug-cc-pVDZ basis set for the decomposition. The PBE0AC functional was shown to give accurate first-order as well as induction and dispersion contributions.³ The aug-cc-pVDZ set is large enough to provide a reliable estimate of the electrostatic, induction, and exchange components. The dispersion component is underestimated by about 10%–20% in this basis set (see Ref. 7) but should serve well enough for the purpose of comparison. We implemented a gradient-controlled shift procedure, which needs a difference (shift) between the vertical ionization potential (IP) and the highest occupied molecular orbital (HOMO) energy of the DFT method used as an input.³ The IPs were calculated at the PBE0/TZVP level while the HOMO values were taken from the aug-cc-pVDZ calculation.

It should be noted that we use the term “dispersion” herein in two different meanings. First, in the same sense as

TABLE I. First- and second-order perturbative DFT-SAPT energies and the reference CCSD(T)/CBS interaction energies (energies in kcal/mol) for planar H-bonded DNA base pairs [(1–6)], intrastrand stacked pairs [(7–14)], and interstrand pairs [(15–27)]. Second-order energies include the respective exchange terms; the exchange-dispersion energy is presented in parentheses.

Structure ^a	$E_{\text{el}}^{(1)}$	$E_{\text{ex}}^{(1)}$	$E^{(1)}$	$E_i^{(2)}$	$E_D^{(2)}$	$E^{(2)}$	$\delta(\text{HF})$	E	ΔE^b	E_{disp}^c
(1) <i>mAmT</i> WC	−26.71	33.29	6.58	−6.13	−9.45(2.43)	−15.58	−5.13	−14.13	−16.40	−9.01
(2) <i>mCmG</i> WC	−48.28	52.51	4.22	−12.39	−13.17(3.75)	−25.56	−9.42	−30.76	−35.80	−13.38
(3) <i>mAmT</i> WC	−27.36	32.29	4.92	−6.21	−9.30(2.40)	−15.51	−5.23	−15.82	−18.40	−8.74
(4) GA	−24.96	34.77	9.81	−5.74	−9.88(2.49)	−15.62	−5.03	−10.84	−11.30	−9.79
(5) GC WC	−48.77	56.73	7.96	−12.75	−13.60(3.92)	−26.35	−9.94	−28.33	−30.70	−14.18
(6) GC WC	−47.96	55.39	7.44	−12.55	−13.44(3.84)	−25.99	−9.70	−28.25	−31.40	−13.96
(7) AT S	−3.30	9.83	6.53	−0.62	−11.84(1.53)	−12.46	−0.46	−6.39	−8.10	−10.31
(8) <i>mCmG</i> S	−2.98	3.08	0.10	−1.10	−5.84(0.58)	−6.94	−0.20	−7.04	−7.90	−4.57
(9) <i>mAmC</i> S	−1.02	7.95	6.93	−1.12	−10.55(1.39)	−11.67	−0.31	−5.05	−6.70	−8.70
(10) <i>mTmG</i> S	−0.53	7.11	6.58	−0.77	−10.21(1.10)	−10.98	−0.28	−4.68	−6.20	−9.13
(11) CG S	−5.93	11.54	5.61	−0.91	−10.07(1.67)	−10.98	−0.69	−6.06	−7.70	−8.54
(12) AG S	−4.39	12.78	8.39	−0.80	−11.91(1.87)	−12.71	−0.60	−4.92	−6.50	−9.97
(13) GC S	−6.97	9.78	2.81	−1.39	−10.48(1.56)	−11.87	−0.54	−9.60	−12.40	−8.36
(14) GC S	−7.68	9.51	1.83	−1.10	−9.89(1.49)	−10.99	−0.50	−9.66	−11.60	−8.05
(15) GC IS	−1.98	0.95	−1.03	−0.57	−1.93(0.16)	−2.50	−0.08	−3.61	−3.68	−1.53
(16) GG IS	1.24	8.54	9.78	−0.93	−10.48(1.39)	−11.41	−0.19	−1.82	−4.82	−8.90
(17) AT IS	−1.99	2.77	0.78	−0.34	−2.21(0.32)	−2.55	−0.22	−1.99	−2.34	−1.67
(18) TT IS	−1.31	8.69	7.38	−0.59	−9.28(1.19)	−9.87	−0.33	−2.82	−2.16	−8.22
(19) GG IS	−2.04	7.27	5.23	−0.72	−7.07(1.03)	−7.79	−0.33	−2.89	1.24	−6.23
(19) GG IS	−2.04	7.27	5.23	−0.72	−7.07(1.03)	−7.79	−0.33	−2.89	1.24	−6.23
(20) AG IS	−2.75	5.70	2.95	−0.48	−5.76(0.84)	−6.24	−0.27	−3.56	−4.22	−4.82
(21) TC IS	−0.60	0.10	−0.50	−0.12	−0.55(0.02)	−0.67	−0.01	−1.18	−1.15	−0.45
(22) AG IS	−2.71	3.47	0.76	−0.76	−3.43(0.49)	−4.19	0.25	−3.18	−4.06	−2.56
(23) AT IS	−0.57	0.36	−0.21	−0.09	−1.34(0.06)	−1.43	−0.02	−1.66	−1.71	−1.09
(24) <i>mGmG</i> IS	−1.86	6.70	4.84	−0.50	−7.57(1.02)	−8.07	−0.30	−3.53	−4.50	−6.48
(25) <i>mAmG</i> IS	−2.81	2.51	−0.30	−0.88	−3.13(0.38)	−4.01	−0.19	−4.50	−4.80	−2.31
(26) CA IS	−1.93	4.32	2.39	−0.83	−3.15(0.54)	−3.98	−0.36	−1.95	−3.00	−2.35
(27) GG IS	−4.06	1.71	−2.35	−0.61	−2.10(0.32)	−2.71	−0.16	−5.22	−5.20	−1.62

^aReference 1.

^b $\Delta E_{\text{CBS}}^{\text{CCSD(T)}}$.

^cNondamped empirical dispersion energy as in Ref. 8.

it is used in the SAPT theory, i.e., as a quality defined also at the van der Waals distances with significant overlap, and, second, meaning the empirical asymptotic description valid at large separations only. The empirical dispersion is not damped at short distances here. In the following, the particular meaning of the term dispersion can be understood from the context (SAPT/empirical dispersion).

The energy components as well as CCSD(T)/CBS stabilization energies from Ref. 1 are summarized in Table I. The DFT-SAPT stabilization energies are systematically smaller than the reference values, and this finding is in full agreement with the conclusion drawn in Ref. 6. Table I includes six planar H-bonded complexes, among which are three GC WC and two AT WC pairs and one GA mispair. The GC WC structures are considerably more stable than the AT WC pairs, with the GA mispair being the least stable. Dominant attraction for all H-bonded structures originates in the $E_{\text{el}}^{(1)}$ electrostatic term. Due to the much larger dipole moments of G and C (in comparison with A and T), the electrostatic $E_{\text{el}}^{(1)}$ energy of the GC pairs is much more attractive than that of the AT pairs and also the GA pair. The attractive $E_{\text{el}}^{(1)}$ term is systematically overcompensated by the exchange-repulsion term; thus the first-order $E^{(1)}$ energy is systematically repulsive and this repulsion is significant (more than 4 kcal/mol).

The second-order induction energy is large and again it is much larger for the GC pairs than for the other pairs, the explanation for which should be again sought in the varying values of the electric dipole moments of the various bases. The exchange component is now smaller (not shown), and the resulting effective second-order induction energy is systematically attractive. In the case of GC pairs, this attraction is approximately twice as big as that of the other pairs.

The $\delta(\text{HF})$ terms are rather large (more than half of dispersion energy) and attractive for planar H-bonded pairs. For the GC pairs, they are about twice as large than that for other pairs. Evidently, the value of the $\delta(\text{HF})$ term correlates with the value of induction energy.

In conclusion, we ascertained that stabilization in induction and dispersion terms is comparable, whereas the $\delta(\text{HF})$ term is smaller but definitely not negligible. The electrostatic $E_{\text{el}}^{(1)}$ term is significantly more attractive than the previous terms, but the absolute values of the exchange-repulsion term are even higher. The $E^{(1)}$ energies are repulsive while the $E^{(2)}$ and $\delta(\text{HF})$ energies are attractive. All these findings basically agree with the conclusions drawn by Hesselmann *et al.* in Ref. 6.

The situation found for stacked complexes was different, where intrasystem stacking will be investigated first. As in

the case of H-bonded pairs, the DFT-SAPT stabilization energies are systematically underestimated when compared with the accurate values. The $E_{\text{el}}^{(1)}$ electrostatic energy is systematically attractive and is the largest for the GC pairs and also for the AG pair. The electrostatic attraction is roughly compensated by exchange-repulsion terms (as in the case of H bonding). The range of $E^{(1)}$ energy is larger here (from 0.1 to 8.4 kcal/mol) than in the case of H bonding (from 4.2 to 9.8 kcal/mol), and GC stacked pairs exhibit more favorable (less repulsive) $E^{(1)}$ energies. The $E^{(1)}$ energies are, however, surprisingly comparable for planar H-bonded and stacked base pairs. Attractive and repulsive components of induction energy are practically compensated (not shown), and the resulting effective induction energy is negligible. This is an important difference in comparison with H-bonded systems, where the effective induction energy is very large. Unambiguously, the dominant attraction originating in all the cases is effective dispersion energy, containing only a small (repulsion) exchange part. Dispersion energy is systematically more attractive than the electrostatic term, and the ratio ranges from 10 (*mAmC* and *mTmG*) to approximately 1.5 (GC pairs). It is to be mentioned that even with two GC stacked pairs, the dispersion energy is about twice larger than the electrostatic term. Yet having compared the dispersion energies for stacked and H-bonded complexes, we reached a similar conclusion as with $E^{(1)}$ energies, which indicates that they are more or less comparable. This finding is surprising as it had been expected that dispersion energy should be larger for stacked pairs and it deserves a more detailed discussion, which is presented in the following paragraph. The $\delta(\text{HF})$ term is negligible for all stacked pairs, which makes the stacked pairs significantly different from H-bonded systems.

Intuitively, the similarity of magnitudes of the dispersion energy in stacked and H-bonded complexes is indeed puzzling. In the stacked complexes the distance of centers of masses is considerably smaller than in the H-bonded ones, and, more important, both monomers seem to have larger geometrical overlap in the stacking arrangement. Consequently, one would anticipate larger dispersion interaction energy. An explanation can be found, at least in part, in comparing the SAPT dispersion energies with the empirical dispersion energies as calculated by the well known C_6/r^6 formula. Table I shows the SAPT values along with the calculated empirical dispersion energy contributions (for details on the method and C_6 coefficients used see Ref. 8). Interestingly, the nondamped values are in very good agreement with the SAPT values for both stacked and H-bonded complexes, being on average 0.6 kcal/mol weaker and with standard deviation of 0.6 kcal/mol. Obviously our intuitive perception of the dispersion interaction is not accurate and very close contacts in the H-bonded complexes can bring as much dispersion stabilization as is found in the stacked molecules. The remaining 0.6 kcal/mol missing in the empirical description (and likely more; note that our SAPT values are underestimated in the aug-cc-pVDZ basis) can be explained

by overlap effects, which result in small but exponentially growing dispersion contribution at very short distances. We would like to point out that the present comparison is illustrative only, because the C_6/r^6 term is unphysically divergent at short distances. However, very similar magnitudes of the reference SAPT calculations indicate that the short range effects (damping) are still not so profound even at the short H-bonding distances, so the results drawn are likely not affected.

Interstrand stacking is characterized by smaller total interaction energies, which are attractive and repulsive. A total of 32 interstrand pairs have been investigated, of which only such pairs that had DFT-SAPT stabilization energy larger than 1 kcal/mol are shown in Table I. Six pairs (of the 13 presented in Table I) possess stabilization energy larger than 3 kcal/mol, including only one GC pair. The largest interstrand stacking was found for the GG pair. Three pairs had attractive $E^{(1)}$ energy (due to attractive electrostatic energy and the rather small exchange-repulsion term). Notice that this energy was systematically repulsive for all H-bonded and stacked pairs. The $\delta(\text{HF})$ term is mostly negligible and never exceeds 0.5 kcal/mol.

Having collected all the data, we can state that, in the DFT-SAPT type of analysis, the $E^{(1)}$ energies for H-bonded and intrastrand stacked pairs (in crystal geometries) are similar. The $E^{(2)}$ energies are much larger for planar H-bonded pairs, which is caused by $E_i^{(2)}$ induction energy. The dispersion energy is surprisingly similar for both structural types. The $E^{(1)}$ and $E_D^{(2)}$ energies thus do not show preference for either of the two motifs.

This work has been part of Research Project No. Z4 055 905 and was supported by grants from the Grant Agency of the Czech Republic (203/05/0009) and Ministry of Education of the Czech Republic (Center for Biomolecules and Complex Molecular Systems, LC512). It was also supported by a grant (Grant No. MSM6198959216) to one of the authors (P.J.) from the MSMT of the Czech Republic.

^{a)} Author to whom correspondence should be addressed. Electronic mail: pavel.hobza@uochb.cas.cz

¹ P. Jurečka, J. Šponer, J. Černý, and P. Hobza, *Phys. Chem. Chem. Phys.* **8**, 1985 (2006).

² B. Jeziorski, R. Moszynski, and K. Szalewicz, *Chem. Rev. (Washington, D.C.)* **94**, 1887 (1994).

³ A. Haselmann and G. Jansen, *Chem. Phys. Lett.* **357**, 464 (2002); **362**, 319 (2002); **367**, 778 (2003); *Phys. Chem. Chem. Phys.* **5**, 5010 (2003).

⁴ A. J. Misquitta and K. Szalewicz, *Chem. Phys. Lett.* **357**, 301 (2002); A. J. Misquitta, B. Jeziorski, and K. Szalewicz, *Phys. Rev. Lett.* **91**, 033201 (2003); A. J. Misquitta, R. Podesszwa, B. Jeziorski, and K. Szalewicz, *J. Chem. Phys.* **123**, 214103 (2005).

⁵ R. Podesszwa, R. Bukowski, and K. Szalewicz, *J. Phys. Chem. A* **110**, 10345 (2006).

⁶ A. Hesselmann, G. Jansen, and M. Schutz, *J. Am. Chem. Soc.* **128**, 11730 (2006).

⁷ A. Haselmann, G. Jansen, and M. Schütz, *J. Chem. Phys.* **122**, 014103 (2005).

⁸ P. Jurečka, J. Černý, P. Hobza, and D. R. Salahub, *J. Comput. Chem.* **28**, 555 (2007).

The Journal of Chemical Physics is copyrighted by the American Institute of Physics (AIP). Redistribution of journal material is subject to the AIP online journal license and/or AIP copyright. For more information, see <http://ojps.aip.org/jcpo/jcpcr/jsp>