J|A|C|S

COMMUNICATIONS

Interaction Energy Contributions of H-Bonded and Stacked Structures of the AT and GC DNA Base Pairs from the Combined Density Functional Theory and Intermolecular Perturbation Theory Approach

Andreas Hesselmann,*,[†] Georg Jansen,*,[‡] and Martin Schütz*,[§]

Lehrstuhl für Theoretische Chemie, Universität Erlangen-Nürnberg, D-91058 Erlangen, Germany, Fachbereich Chemie, Universität Duisburg-Essen, Campus Essen, D-45117 Essen, Germany, and Institut für Physikalische und Theoretische Chemie, Universität Regensburg, D-93040 Regensburg, Germany

Received May 12, 2006; E-mail: andreas.hesselmann@chemie.uni-erlangen.de

In the past decade, we have seen a strong increase in the number of quantum chemical studies of the interaction between the purine and pyrimidine constituents of DNA.[1-5] One of the goals of these studies is to provide as accurate as possible interaction energies for which, with a notable exception,[6] there is hardly any experimental information available. Another important goal is to provide insight into the nature of the interactions between the nucleobases. The quantum chemical investigations carried out so far usually employed the supermolecular approach, often followed by an a posteriori rationalization of the results using notions of intermolecular perturbation theory. The interplay between the electrostatic, E_{el}, and the dispersion, E_{disp}, interaction energies has received particular attention. E_{el} often is calculated from point charges or multipoles[7] distributed over the monomers, thus neglecting its important penetration part.[8] In supermolecular calculations, E_{disp} usually is estimated from the electron correlation contribution to the interaction energy which, in fact, contains further interaction terms. A hybrid class of methods combines approximate treatments of the long-range part of E_{disp} with supermolecular density functional theory (DFT)[4] so as to correct for corresponding failures of the present exchange-correlation (xc) functionals.[5]

To put the discussion of the interplay of the various interaction energy contributions on firm quantitative grounds, it appears most appropriate to consistently use a well-defined variant of intermolecular perturbation theory, such as symmetry-adapted perturbation theory (SAPT).[9] In SAPT, the total interaction energy, E_{int}, is obtained as a sum of first-order electrostatic, $E_{el}^{(1)}$, and second-order induction and dispersion contributions, $E_{ind}^{(2)}$ and $E_{disp}^{(2)}$, respectively. All of these terms do contain the corresponding penetration contributions; the (second-order) charge-transfer contribution is included in $E_{ind}^{(2)}$. Furthermore, these terms are accompanied by corresponding first-, $E_{exch}^{(1)}$, and second-order exchange-corrections, $E_{exch-ind}^{(2)}$ and $E_{exch-disp}^{(2)}$, respectively, describing the repulsive effects of electron exchange between the overlapping molecular charge distributions. In particular, for hydrogen bridges induction, exchange-induction and charge-transfer effects of higher than second order in the intermolecular perturbation operator become non-negligible, their combined effect can be estimated from supermolecular Hartree–Fock calculations and is denoted as δ(HF).[9b]

Thanks to the combination of a DFT treatment of monomer properties with a SAPT treatment of intermonomer interactions (DFT–SAPT)[10,11] and the introduction of the density-fitting approximation, SAPT calculations with extended basis sets on medium-sized systems, such as the benzene dimer, recently have become possible.[12] Here we investigate the Watson–Crick (WC) and

stacked (S) structures of the adenine–thymine (AT) and guanine–cytosine (GC) base pairs with DFT–SAPT.

Their gas-phase-optimized structures were taken from the work of Jurečka and Hobza.[3a] The aug-cc-pVXZ (X = D, T, Q) sets[13] were used as atomic basis sets, and the cc-pV(X+1)Z JK-fitting and aug-cc-pVXZ MP2-fitting sets[14] were employed for the density-fitting approximation. Complete basis set (cbs) limits of DFT–SAPT interaction energies were obtained by extrapolating the augmented triple- and quadruple-ζ results for $E_{disp}^{(2)} + E_{exch-disp}^{(2)}$ with the scheme of Bak et al.,[15] adding all other contributions as calculated with the aug-cc-pVQZ basis set. Monomer DFT calculations were done with the LPBE0AC xc potential.[10b,12,16,17] All core electrons were kept frozen in the response calculations using the ALDA xc kernel. The calculations have been done with the Molpro quantum chemistry package.[18]

Table 1 displays the total interaction energies along with second-order Møller–Plesset (MP2) and coupled-cluster (CCSD(T)) results.[3] Note that in order to obtain overall *stabilization* energies the repulsive contributions of deformation of the monomers from their gas-phase equilibrium geometries to their structure within the dimer should be added. Clearly, large basis sets are required to obtain converged results. This is most obvious for the GC(WC) pair, where the difference between the augmented double-ζ result and the cbs limit amounts to −2.4 kcal/mol for DFT–SAPT and −2.9 kcal/mol for MP2. At present, a direct cbs extrapolation is not feasible for CCSD(T); here it is assumed that the differences between MP2 and CCSD(T) as found in a double-ζ basis set are constant upon increase of basis set quality so that they can be added to the cbs extrapolated MP2 energy.[3a] While these CCSD(T) corrections are relatively small for the WC structures (−0.5−0.0 kcal/mol), they are much more important and become repulsive for the stacked structures (1.8−2.8 kcal/mol). Similar trends were observed for the benzene dimer and can be traced back to a too attractive dispersion contribution contained in the supermolecular MP2 energy.[12] DFT–SAPT, on the other hand, is known to describe the dispersion contribution in good agreement with elaborate many-body SAPT calculations.[10d] With DFT–SAPT, the stacked structures of AT and GC are found to be even somewhat less bound than with CCSD(T) (by 1.4 and 1.2 kcal/mol, respectively). Since this is also true for the hydrogen-bonded structures, which deviate by 1.2 (AT) and 1.6 (GC) kcal/mol from the estimated CCSD(T) interaction energy, one finds a good agreement for the relative energies between both methods: while with MP2 the WC structure of AT is only 1.8 kcal/mol more stable than the stacked structure, CCSD(T) and DFT–SAPT yield 4.6 and 4.8 kcal/mol, respectively. For the GC pair, the WC structure is favored by 10.8 kcal/mol with MP2, while CCSD(T) and DFT–SAPT give 13.1 and 12.7 kcal/mol, respectively. In agreement with earlier observations,[3a] all methods describe GC(S) as even more stable than AT(WC).

[†] Universität Erlangen-Nürnberg.
[‡] Universität Duisburg-Essen.
[§] Universität Regensburg.

Table 1. Interaction Energies (in kcal/mol) of the AT and GC Base Pairs for the aug-cc-pVXZ Basis Sets

base pair	X	def[a,b]	MP2[b]	CCSD(T)[b]	DFT-SAPT
AT (WC)	D		−14.8		−14.1
	T		−16.0		−15.2
	Q		−16.5		−15.5
	cbs	1.4	−16.9	−16.9	−15.7
AT (S)	D		−13.1		−9.2
	T		−14.4		−10.3
	Q		−14.8		−10.7
	cbs	0.7	−15.1	−12.3	−10.9
GC (WC)	D		−28.7		−28.1
	T		−30.4		−29.8
	Q		−31.1		−30.2
	cbs	3.4	−31.6	−32.1	−30.5
GC (S)	D		−18.5		−15.7
	T		−20.0		−17.1
	Q		−20.5		−17.5
	cbs	2.0	−20.8	−19.0	−17.8

[a] MP2 deformation energy. [b] Taken from Jurečka et al.[3]
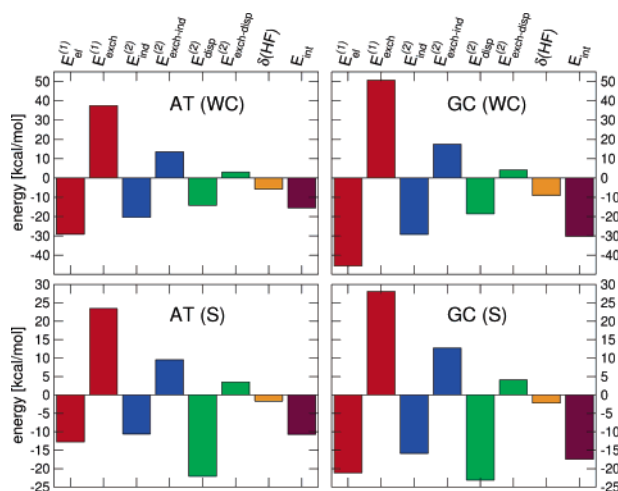
Figure 1. Interaction energy contributions for the four structures (see text) of the nucleic acid base pairs with the aug-cc-pVQZ basis set.

The prevailing point of view that the WC structures are mainly stabilized by $E_{el}^{(1)}$ is supported by Figure 1 in the sense that it is the dominating attractive contribution. It is, however, overcompensated through $E_{exch}^{(1)}$, leading to overall repulsive first-order contributions of 8.3 and 5.1 kcal/mol for AT and GC, respectively. A similar observation has been made by Fonseca Guerra et al.[2] via an energy decomposition scheme of the supermolecular DFT energy. As is clear from the figure, the interaction contributions of second and higher order in the intermolecular perturbation play a most important role for the stabilization of the WC structures.

For the stacked structures, Figure 1 shows that $E_{disp}^{(1)}$ is the most important contribution to E_{int}, as expected. Yet, a closer inspection reveals also unexpected findings: while for AT $E_{disp}^{(2)}$ is roughly 2-fold larger than $E_{el}^{(1)}$ and $E_{ind}^{(2)}$, for GC, this is not the case. Here $E_{el}^{(1)}$ is nearly as large as $E_{disp}^{(2)}$, and $E_{ind}^{(2)}$ amounts to about 70% of $E_{disp}^{(2)}$. So, for the GC(S) pair, all three types of interaction (electrostatic, induction, and dispersion) play an equally important role. One should note, however, that this applies to the gas-phase-optimized stacked dimer structure showing two fairly close distances of 2.28 and 2.45 Å, respectively, between hydrogen atoms of strongly nonplanar amino groups of one monomer and the carbonyl group oxygen atoms of the other. Such close contacts are not found for the intrastrand stacked structures of GC in DNA. A similar proviso can be made for the AT dimer, but here the gas-phase-optimized stacked structure displays only one comparable H···O contact with a much larger distance of 2.77 Å.

Furthermore, we note that (i) the dispersion energies for the WC structures still amount to 80% (GC) and 64% (AT) of that of the stacked structures; (ii) $E_{ind}^{(2)}$ is quenched to a large extent by $E_{exch-ind}^{(2)}$ for the stacked structures (by 90% for AT and 80% for GC), and much less for the WC structures (by 66% for AT and 60% for GC); and (iii) δ(HF) makes a quite substantial contribution to the total interaction energy of 30% (GC) to 38% (AC) for the WC and 13% (GC) to 16% (AC) for the stacked structures. Since intramonomer and higher-order intermonomer electron correlation effects are not included in this estimate, we suspect it to be responsible for most of the remaining errors in the method.

In summary, for the stacked and WC structures of AT and GC, cbs-extrapolated DFT−SAPT yields total interaction energies which deviate by an upward shift of 1.2−1.6 kcal/mol from estimated cbs CCSD(T) results. This is a very satisfactory agreement in view of the wildly differing theoretical foundations of both approaches. The rigorous determination of individual energy contributions to the total interaction energy available through DFT−SAPT furthermore allows for the detailed understanding of their interplay and should be useful for future theoretical descriptions and parametrizations of the fundamental base pair interactions in DNA.

Supporting Information Available: Tables with interaction energy contributions. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) (a) Hobza, P.; Šponer, J. *Chem. Rev.* **1999**, *99*, 3247. (b) Šponer, J.; Hobza, P.; Leszczynski, J. In *Computational Chemistry: Reviews of Current Trends*; Leszczynski, J., Ed.; World Scientific: Singapore, 2000; Vol. 5, p 171. (c) Šponer, J.; Leszczynski, J.; Hobza, P. *Biopolymers* **2002**, *61*, 3.

(2) (a) Fonseca Guerra, C.; Bickelhaupt, F. M.; Baerends, E. J.; Snijders, J. G. In *Computational Chemistry: Reviews of Current Trends*; Leszczynski, J., Ed.; World Scientific: Singapore, 2002; Vol. 7, p 17. (b) Fonseca Guerra, C.; Bickelhaupt, F. M. *Angew. Chem., Int. Ed.* **2002**, *41*, 2092.

(3) (a) Jurečka, P.; Hobza, P. *J. Am. Chem. Soc.* **2003**, *125*, 15608. (b) Jurečka, P.; Šponer, J.; Černý, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985.

(4) (a) Elstner, M.; Hobza, P.; Frauenheim, T.; Kaxiras, E. *J. Chem. Phys.* **2001**, *114*, 5149. (b) Wu, Q.; Yang, W. *J. Chem. Phys.* **2002**, *116*, 515. (c) Grimme, S. *J. Comput. Chem.* **2004**, *25*, 1463.

(5) (a) Cybulski, S. M.; Seversen, C. E. *J. Chem. Phys.* **2005**, *122*, 014117. (b) Černý, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2005**, *7*, 1624. (c) Zhao, Y.; Truhlar, D. G. *Phys. Chem. Chem. Phys.* **2005**, *7*, 2701.

(6) Yanson, I. K.; Teplitsky, A. B.; Sukhodub, L. F. *Biopolymers* **1979**, *18*, 1149.

(7) Jaubert, L.; Popelier, P. L. A. *Phys. Chem. Chem. Phys.* **2002**, *4*, 4353.

(8) Toczyłowski, R. R.; Cybulski, S. M. *J. Chem. Phys.* **2005**, *123*, 154312.

(9) (a) Jeziorski, B.; Moszynski, R.; Szalewicz, K. *Chem. Rev.* **1994**, *94*, 1887. (b) Szalewicz, K.; Patkowski, K.; Jeziorski, B. *Struct. Bonding* **2005**, *116*, 43.

(10) (a) Jansen, G.; Hesselmann, A. *J. Phys. Chem. A* **2001**, *105*, 11156. (b) Hesselmann, A.; Jansen, G. *Chem. Phys. Lett.* **2002**, *357*, 464. (c) Hesselmann, A.; Jansen, G. *Chem. Phys. Lett.* **2002**, *362*, 319. (d) Hesselmann, A.; Jansen, G. *Chem. Phys. Lett.* **2003**, *367*, 778. (e) Hesselmann, A.; Jansen, G. *Phys. Chem. Chem. Phys.* **2003**, *5*, 5010.

(11) For a similar approach called SAPT(DFT), see: Misquitta, A. J.; Podeszwa, R.; Jeziorski, B.; Szalewicz, K. *J. Chem. Phys.* **2005**, *123*, 214103 and references therein.

(12) Hesselmann, A.; Jansen, G.; Schütz, M. *J. Chem. Phys.* **2005**, *122*, 014103.

(13) Kendall, R. A.; Dunning, T. H., Jr.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796.

(14) (a) Weigend, F. *Phys. Chem. Chem. Phys.* **2002**, *4*, 4285. (b) Weigend, F.; Köhn, A.; Hättig, C. *J. Chem. Phys.* **2002**, *116*, 3175.

(15) Bak, K. L.; Jørgensen, P.; Olsen, J.; Helgaker, T.; Klopper, W. *J. Chem. Phys.* **2000**, *112*, 9229.

(16) (a) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865. (b) Adamo, C.; Barone, V. *J. Chem. Phys.* **1999**, *110*, 6158. (c) Grüning, M.; Gritsenko, O. V.; van Gisbergen, S. J. A.; Baerends, E. J. *J. Chem. Phys.* **2001**, *114*, 652.

(17) (a) Della Sala, F.; Görling, A. *J. Chem. Phys.* **2001**, *115*, 5718. (b) Hesselmann, A.; Manby, F. *J. Chem. Phys.* **2005**, *123*, 164116.

(18) Werner, H.-J.; Knowles, P. J.; Lindh, R.; Schütz, M. *MOLPRO*, version 2002.10; a package of ab inito programs; see http://www.molpro.net.

JA0633363