Ab initio studies on the H-bonding of hypoxanthine and DNA bases

Gábor Paragi,*[a] István Pálinkó,[b] Cris Van Alsenoy,[c] Iván K. Gyémánt,[d] Botond Penke[a] and Zoltán Timár[a]

[a] *Department of Medicinal Chemistry, University of Szeged, Dóm tér 8., H-6720, Szeged, Hungary. E-mail: paragi@sol.cc.u-szeged.hu; Fax: +36-62-545971; Tel: +36-62-545136*
[b] *Department of Organic Chemistry, University of Szeged, Dóm tér 8., H-6720, Szeged, Hungary*
[c] *Department of Chemistry, University Insstelling Antwerp, B-2610, Wilrijk, Belgium*
[d] *Department of Theoretical Physics, University of Szeged, Tisza Lajos krt. 84-86., H-6720, Szeged, Hungary*

Received (in London, UK) 14th May 2002, Accepted 12th August 2002
First published as an Advance Article on the web 9th September 2002

Novel and interesting points are underlined by *ab initio* (QM) calculations of H-bonding of all the six nucleobase dimers of hypoxanthine and DNA bases, namely Hyp^{anti}···A^{syn}, Hyp^{anti}···A^{anti}, Hyp^{syn}···G^{anti}, Hyp^{anti}···G^{syn}, Hyp^{anti}···T^{anti}, Hyp^{anti}···C^{anti}. HF, DFT (Becke) and B3LYP (hybrid-DFT) methods were applied with and without polarisation functions. The H-bonding preference of hypoxanthine to natural DNA bases is G^{syn} > C^{anti} > A^{anti} > A^{syn} > T^{anti} ≈ G^{anti}. Becke and B3LYP give buckle and propeller arrangements for Hoogsteen dimers, but the use of an auxillary basis set overestimates the angles. Hartree–Fock optimised structures are almost planar and this method does not find the local minima of Hoogsteen pairs. Among the methods giving reliable geometries the Becke calculation needs the least computational resources.

1. Introduction

Hypoxanthine (Hyp), the nucleobase of inosine, is a very important substance in biology and in biotechnology, since it has low selectivity in binding to nucleic acid bases (wobble or universal base).[1] The importance of Hyp is shown by recent publications.[2] Early investigations on the hybridisation of Hyp with the DNA bases are available but ambiguous.[3,4] It was, however, impossible to decide how the H-bonding of Hyp and DNA bases influences the hybridisation of nucleic acid chains, since *ab initio* level calculations on Hyp and DNA bases were not available until now. The most effective tool in the study of the H-bonding of nucleobase pairs is the *ab initio* calculation,[5] but which of the methods (HF or DFT[6] or hybrid-DFT) to be used is of prime importance. The aim of this contribution is to clear up the above points and provide unambiguous answers to the open questions.

Several mutual effects influence the strength and specificity of the hybridisation of nucleic acid chains, such as H-bonding, van der Waals forces, Coulomb repulsion of the backbones, *etc*. H-bonding is mainly responsible for attractive forces and thus the selectivity in hybridisation. Furthermore, H-bonding energy values between two uncharged nucleobases are high, therefore this interaction is determinative in the agglomeration of nucleic acid chains. Calculations provide accurate data for the H-bonding energies and geometries of nucleobases. However, we need to simplify the nucleic acid chain interactions to make calculations tractable. The simplest case for molecular modelling is the nucleic acid base dimer in the gas phase. The binding energy values calculated in this model cannot be easily measured. However, these calculations give useful information on the binding energies, geometries, mechanisms or the selectivity order of Hyp with the 4 natural nucleobases. The H-bonding energies of nucleic acid–base dimers are known for all kind of base pairs. However, *ab initio* level data on Hyp selectivity order, H-bonded to all the four DNA bases, adenine (A), guanine (G), thymine (T) and cytosine (C), are missing. The main purpose of this work was to determine the relative bonding preferences for Hyp–DNA base pairs by *ab initio* calculations and to compare the selectivity order to experimental data. First, we performed geometrical optimisation on all the six possible Hyp–DNA base pairs where two H-bonds are formed. These dimers are Hyp^{anti}···A^{syn}, Hyp^{anti}···A^{anti}, Hyp^{syn}···G^{anti}, Hyp^{anti}···G^{syn}, Hyp^{anti}···T^{anti}, Hyp^{anti}···C^{anti}, following the nucleotide nomenclature. The nucleotide bases were methylated in the calculations on the N^9 (purines) or on the N^3 (pyrimidines) in order to simulate the glycosidic bond, and to avoid H-bonding to N^9 or N^3. It is important to explore which theoretical level or basis set should be used to save time and still get reliable results. Our calculations provide the possibility to compare different methods (HF, DFT, hybrid-DFT),[7] and the effects of the polarisation functions.

2. Computational methods

Preoptimisation of the six base pairs was performed by a semi-empirical method (PM3) included in the MOPAC package. *Ab initio* calculations were done with HF, DFT (Becke) and B3LYP (hybrid-DFT) methods included in the Gaussian software package.[8] Spin polarised calculations were chosen for all methods and for the optimisation the full gradient Berny method was utilised with the default optimisation criteria. The 6-311 and 6-311** basis sets were applied and the Basis Set Superposition Error (BSSE) was taken into account in all cases. The Hartree–Fock and the density functional calculation schemes were compared at exchange-only level and the correlation was applied only in the hybrid-DFT case. The interaction energies were calculated by the $E_{\text{base-pair}}{}^{\text{total}} - (E_{\text{Hyp}}{}^{\text{total}} + E_{\text{DNA–base}}{}^{\text{total}})$ equation. The largest part of this energy is originated from H-bonding, which is determinative in these cases.

This journal is © The Royal Society of Chemistry and the Centre National de la Recherche Scientifique 2002

3. Results and discussion

General consideration

A comparison of the Watson-Crick (W–C) and the Hoogsteen type H-bonding energies reveals that in the case of Hyp, the W–C type bonding is always preferred. For the Hyp–G pair, the W–C H-bonding energy is double the H-bonding energy of the Hoogsteen bonding. The Hartree–Fock method did not find the local energy minimum for the Hoogsteen type H-bonding, even if it was started from the Hoogsteen bonding geometry optimised by the Becke method. Furthermore, it was found that the Hoogsteen type H-bonding preferred the buckle and propeller conformations for Hyp–G pair and energetically was less favourable than the W–C type bonds that are almost planar in most cases. The optimised Hoogsteen type geometries of Hyp–G are shown in Fig. 1 for the Becke and B3LYP calculations with different basis sets. These were the only examples when the use of the polarisation functions provided essential differences in the geometries. In the other cases the optimum geometry had planar conformation, and the polarisation functions had negligible effects.

It is to be noted that in Fig. 1 the hydrogens at the N^2 atom of guanine show the pyramidal effect as was pointed out by Sponer and Hobza.[5] However, this effect has never been significant ($< 1°$), when the planar conformation is preferred. The optimum geometries for the investigated DNA-pairs are presented in Fig. 2, and except for the above-mentioned situations, planar conformations were found without any pyramidal effect.

For Hyp···A and Hyp···G there was the possibility to bind either in the W–C or the triplex method. The Hyp···A H-bonding in both cases is of the NH···O and the N···HN type, one H donor in each nucleobase. They are similar in energies, but the W–C contact is always stronger by 3–7%. The Hyp···G W–C H-bonding consists of two NH···O H-bonds, the Hoogsteen bonding contains one NH···N and one NH···O contact (both H donors on the G) and is weaker by around 50% than the W–C type. The weakest Hyp···G energy value is close to the Hyp···A H-bonding energies, and these all are of two NH···N and NH···O H-bonds. The other difference in the Hoogsteen type H-bondings of Hyp is that the Hoogsteen attack site of Hyp is of two H acceptors (N^7 and O^6). It follows that the NH···N H-bonding is weaker than the NH···O type and, in the case of close and parallel H-bondings, alternation in H donation is preferred. Consequently, *triplex formation prefers energetically and geometrically the NH···O type and alternating H-bonds*. This statement is a corner point in the research of triple-helical nucleic acid chemistry.

In our hands the H–F calculations were less time consuming, however they were not appropriate for Hoogsteen H-bonding optimisation. It means that they did not reproduce the buckle and propeller conformations, as other calculations and crystallographic experiments did. This feature was checked by geometrical optimisation starting from the preferred propeller and buckle conformations of the $Hyp^{anti}\cdots G^{syn}$ pair of the other calculations. HF optimisation, in both cases and with each basis set, transformed the propeller and buckle conformations to the planar W–C arrangement.

Geometry

Both the BECKE and the B3LYP calculation reproduced the buckle (α) and propeller (β) geometries and are appropriate in calculating the energy and geometry for Hoogsteen type H-bonding. These correlated methods give slightly buckled (α) and twisted (β) conformations ($\alpha \approx 10°$ and $\beta \approx 15°$) without the polarisation functions, however the use of the polarisation functions gives too large values for these angles ($\alpha^{Becke} \approx 30°$, $\alpha^{B3LYP} \approx 40°$ and $\beta \approx 50°$). When checking an optimised B3LYP/6-311** conformation by an optimisation without the polarisation functions, the correct geometry was reproduced, *i.e.*, small degree of buckle and propeller torsions. The W–C conformations show little of the above conformational torsions (α and $\beta < 1°$) in any of our calculations. In respect of the polarisation function we can state that it did not have any effect on the order of the H-bonds. The energy values are smaller using the correlation and the complementary basis sets. As concerns H-bonding lengths, data in Table 1 reveal that the correlated calculation provided the smallest value in all cases. Furthermore, the polarisation function

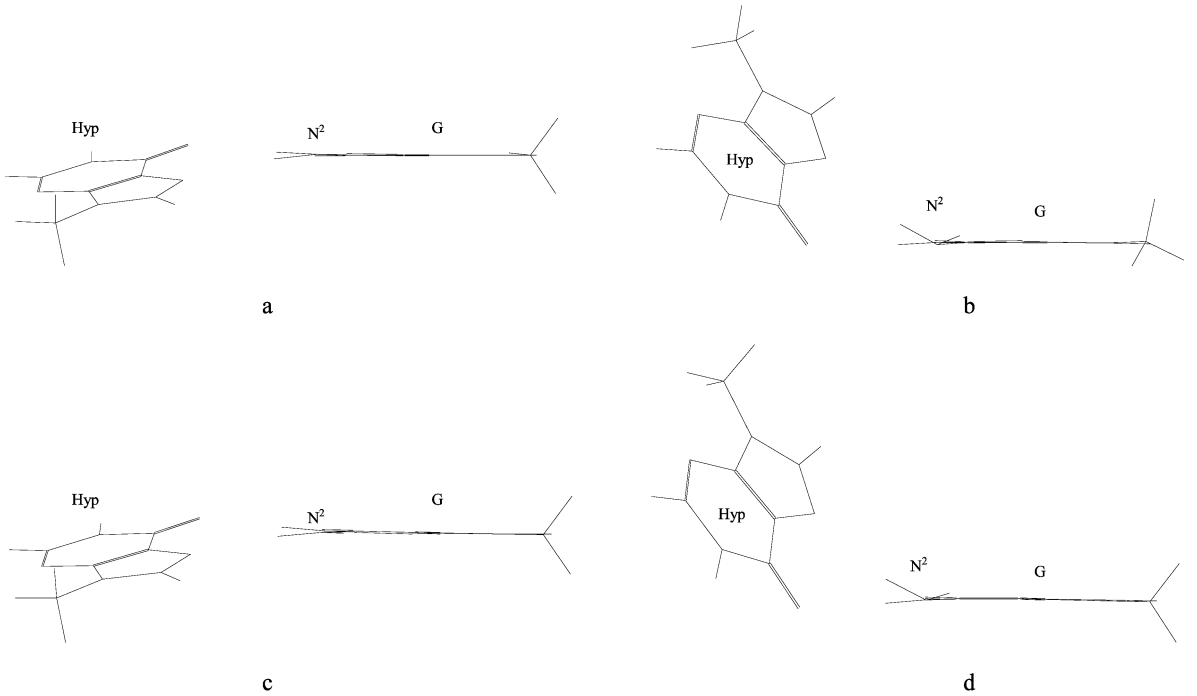

Fig. 1 The buckle and propeller conformations of $Hyp^{syn}\cdots G^{anti}$ pair at Becke (a, b) and B3LYP (c, d) optimum. Polarisation function was used in the cases of (b) and (d).

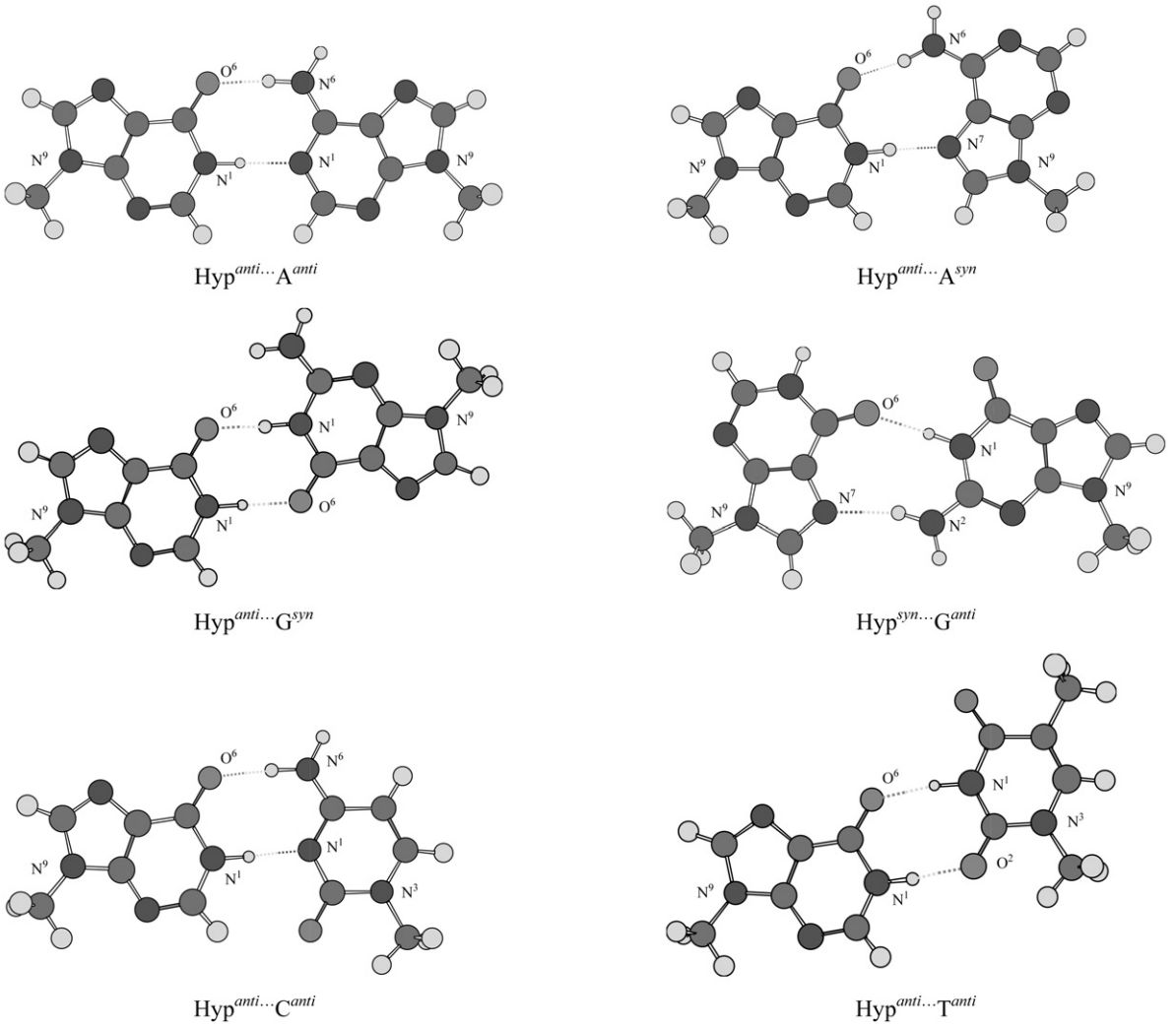

Fig. 2 Optimised conformations of hypoxanthine with A,G,C,T bases.

always increased the optimum distance, and thus, it weakened H-bonding. Comparing the two exchange-only calculations (Becke and HF), the DFT method always gave longer H-bonds than the HF calculation.

Energy

One of the main aims of this work was to gain reliable data on the binding preference order of Hyp to all the four DNA nucleobases. Surprisingly, it was found that the strongest binding is between $Hyp^{anti}\cdots G^{syn}$ as is shown in Table 2. Unfortunately, however, it may not be easily checked by synthetic experiments, since the DNA backbone would not support such a big distortion for guanine in a mixed DNA sequence. The real possibilities for H-bonding of Hyp in any mixed DNA sequence are the other 5 conformations (A^{syn}, A^{anti}, G^{anti}, T^{anti}, C^{anti}). From these, C was found to be the strongest partner of Hyp. A^{anti} then A^{syn} followed except in the H–F calculations, where T had somewhat higher H-bonding energy values than A. G^{anti} and T closed the binding preference sequence. G^{anti} is preferred in the Becke and T in the B3LYP calculations. Summarising: the binding preference in H-bonding to Hyp is

Table 1 H-bond distances (in Å) with hypoxanthine base in Hartree–Fock, Becke and B3LYP calculations with polarisation functions (**) and without them. The conformation of Hyp was always *anti*, except with G^{anti} (see Fig. 2)

H-bond length	Hartree–Fock	Hartree–Fock**	Becke	Becke**	B3LYP	B3LYP**
(Hyp)$O^6\cdots HN^1$(T)	1.885	1.995	2.023	2.111	1.789	1.843
(Hyp)$N^1H\cdots O^2$(T)	1.821	1.923	1.944	2.025	1.746	1.805
(Hyp)$O^6\cdots HN^1$(G)	1.772	1.864	1.959	1.959	1.700	1.745
(Hyp)$N^1H\cdots O^6$(G)	1.798	1.895	1.962	1.962	1.706	1.751
(Hyp)$O^6\cdots HN^6$(C)	1.898	1.993	1.967	2.040	1.782	1.835
(Hyp)$N^1H\cdots N^1$(C)	1.855	1.951	1.952	2.031	1.746	1.816
(Hyp)$O^6\cdots HN^6$(A)	1.912	1.995	1.982	2.052	1.797	1.844
(Hyp)$N^1H\cdots N^1$(A)	1.962	2.067	2.033	2.121	1.814	1.894
(Hyp)$O^6\cdots HN^6$(A)	1.908	2.000	1.987	2.064	1.800	1.857
(Hyp)$N^1H\cdots N^7$(A)	1.984	2.080	2.042	2.129	1.834	1.909
(Hyp)$O^6\cdots HN^1$(G)	—	—	2.157	2.255	1.891	1.952
(Hyp)$N^7\cdots HN^2$(G)	—	—	2.408	2.459	2.141	2.186

Table 2 Binding preferences with hypoxanthine base in Hartree–Fock, Becke and B3LYP calculations with polarisation functions (**) and without them. The conformation of Hyp was always *anti*, except with G^{anti} (see Fig. 2). In the subscripts the interaction energy is in SI units (kJ mol^{-1})

Hartree–Fock		DFT with Becke		B3LYP	
6-311	6-311**	6-311	6-311**	6-311	6-311**
$G^{syn}_{-26.09}$	$G^{syn}_{-20.32}$	$G^{syn}_{-16.36}$	$G^{syn}_{-12.70}$	$G^{syn}_{-27.29}$	$G^{syn}_{-23.00}$
$C^{anti}_{-23.95}$	$C^{anti}_{-18.64}$	$C^{anti}_{-15.19}$	$C^{anti}_{-11.92}$	$C^{anti}_{-25.70}$	$C^{anti}_{-21.02}$
$T^{anti}_{-15.12}$	$A^{anti}_{-11.79}$	$A^{anti}_{-10.48}$	$A^{anti}_{-8.17}$	$A^{anti}_{-18.63}$	$A^{anti}_{-15.29}$
$A^{anti}_{-14.86}$	$T^{anti}_{-11.12}$	$A^{syn}_{-10.16}$	$A^{syn}_{-7.75}$	$A^{syn}_{-17.95}$	$A^{syn}_{-14.49}$
$A^{syn}_{-14.14}$	$A^{syn}_{-11.02}$	$G^{anti}_{-8.04}$	$G^{anti}_{-6.32}$	$T^{anti}_{-16.38}$	$T^{anti}_{-13.21}$
—	—	$T^{anti}_{-8.02}$	$T^{anti}_{-5.99}$	$G^{anti}_{-14.42}$	$G^{anti}_{-11.61}$

$G^{syn} > C^{anti} > A^{anti} > A^{syn} > T^{anti} \approx G^{anti}$. In the followings these calculations are compared to earlier experimental results. Two publications are available on the thermodynamic values of the hybridisation of Hyp to the DNA bases.[3,4] In these we found a binding preference referring to the ΔG: C > A > T > G or C > A > G > T[3] and C > A > T > G or A > G > C > T.[4] However the measured ΔG values of the experimental works contain the entropy changes during hybridisation of DNA–DNA or PNA–DNA strands, they cannot be compared directly to our computed results. Nevertheless, we can conclude that the binding preference of Hyp to the DNA bases is C > A > T ≈ G as most of the available results suggests.

The expense of calculations

We present a summary of the computation time in Table 3. The least time consuming method is the Hartree–Fock method. Using the polarisation function the computation time increased significantly. Among the DFT methods, the Becke calculation needs less time without the polarisation function, however, with the polarisation function the computation time increased significantly. For instance, the Becke calculation was twice as long as the standard calculation, in the case of B3LYP the polarisation function increased the runtime need by a factor of 2.4. The use of polarisation function did not influence significantly the energy values but resulted in very different geometries for Hoogsteen dimers. Since the Becke method without the polarisation function resulted in Hoogsteen geometries (closest to experimental values) and it was the least time consuming, we prefer its use for nucleobase–pair calculations.

Conclusions

Calculations revealed that the binding preference of DNA bases in H-bonding to Hyp is $G^{syn} > C^{anti} > A^{anti} > A^{syn} > T^{anti} \approx G^{anti}$. It was also shown that triplex formation prefers energetically and geometrically the NH···O type and alternating H-bonds. The Hartree–Fock method proved to be unable to find the local minimum for the Hoogsteen-type H-bondings. The Becke and B3LYP calculations overestimate the buckle and propeller angles when using the polarisation functions. Becke calculations gave such geometries, that are closest to experimental values and still do not need much computational means.

Table 3 The calculation time (in s) of an SCF run is presented for the calculation methods used on the Hyp···G Hoogsteen pair. Number of iterations to reach the convergency limit is shown in the iteration row

	Hartree–Fock		DFT with Becke		B3LYP	
	6-311	**	6-311	**	6-311	**
Iteration	16	16	15	13	17	16
Time	3281	12 433	9444	18 665	11 569	27 536

Acknowledgements

This work was supported by OTKA D 38492 and the Bilateral Scientific and Technological Cooperation between Hungary and Flanders (project No. B-2/01)

References

1 (*a*) R. Green, J. W. Szostak, S. A. Benner, A. Rich and N. Usman, *Nucleic Acids Res.*, 1991, **19**, 4161–4166; (*b*) E. Cubero, R. Guimil-Garcia, F. J. Luque, R. Eritja and M. Orozco, *Nucleic Acids Res.*, 2001, **29**, 2522–2534.

2 (*a*) M. K. Shukla and J. Leszczynski, *J. Mol. Struct.: THEOCHEM*, 2000, **529**, 99–112; (*b*) M. E. Costas and R. Acevedo-Chávez, *J. Mol. Struct.: THEOCHEM*, 2000, **532**, 143–156; (*c*) R. Ramaekers, G. Maes, L. Adamowicz and A. Dkhissi, *J. Mol. Struct.: THEOCHEM*, 2001, **560**, 205–221.

3 F. H. Martin and M. M. Castro, *Nucleic Acids Res.*, 1985, **13**, 8927–8938.

4 Z. Timár, S. Bottka, L. Kovács and B. Penke, *Nucleosides Nucleotides*, 1999, **18**, 1131–1133.

5 J. Sponer and P. Hobza, *Encyclopedia of Computational Chemistry*, ed. P. v. R. Schleyer, Wiley, Chichester, 1998, pp. 777–789.

6 C. Tuma, A. D. Boese and N. C. Handy, *Phys. Chem. Chem. Phys.*, 1999, **1**, 3939–3947.

7 (*a*) C. C. J. Roothan, *Rev. Mod. Phys.*, 1951, **23**, 69; (*b*) A. D. Becke, *Phys. Rev. A*, 1988, **38**, 3098; (*c*) A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648.

8 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, Jr., R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, A. G. Baboul, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P. M. W. Gill, B. G. Johnson, W. Chen, M. W. Wong, J. L. Andres, M. Head-Gordon, E. S. Replogle and J. A. Pople, Gaussian 98 (Revision A.x), Gaussian, Inc., Pittsburgh PA, 1998.