Computational Study of Substituent Effects on the Interaction Energies of Hydrogen-Bonded Watson–Crick Cytosine:Guanine Base Pairs

ARTICLE *in* THE JOURNAL OF PHYSICAL CHEMISTRY B · MAY 2008

Impact Factor: 3.3 · DOI: 10.1021/jp7108913 · Source: PubMed

Computational Study of Substituent Effects on the Interaction Energies of Hydrogen-Bonded Watson−Crick Cytosine:Guanine Base Pairs

Chunxia Xue and Paul L. A. Popelier*

Manchester Interdisciplinary Biocentre (MIB), 131 Princess Street, Manchester, M1 7DN, Great Britain

Received: November 14, 2007; In Final Form: January 31, 2008

The substituent effects on interaction energies of hydrogen-bonded DNA Watson−Crick base pairs in the gas phase were captured in a model using ab initio descriptors (at the B3LYP/6-311+G(2d,p) level). While forming a noncovalently bonded complex with unsubstituted guanine (G), cytosine (C) carried 42 possible substituents both at the C6 position (C^{6X}:G) and at the C5 position (C^{5X}:G). We rationalize why complexes possessing a more strongly electron-withdrawing group in C^X form less stable base pairs. Multivariate linear regression constructed the quantitative relationships between the interaction energies of the complexes and the descriptors, which were drawn from quantum chemical topology (QCT). For the C^{6X} dataset, the best model yielded $r^2 = 0.93$ and a root-mean-square (rms) energy of 0.53 kJ/mol for the 28 complexes in the training set. This model was evaluated by an external test set (14 complexes), yielding an r^2 value of 0.96 and an rms error of 0.42 kJ/mol. For the C^{5X} dataset, the QCT descriptors generated a linear model, with r^2 values of 0.92 and 0.97 and rms values of 1.69 and 1.24 kJ/mol for the training set (31 compounds) and the external test set (11 compounds), respectively. The models built here could therefore be useful for the assessment of the interaction energy of C^{6X}:G and C^{5X}:G purely from monomeric data.

1. Introduction

The hydrogen bond formation of a Watson−Crick type base pair is fundamental for molecular recognition in the duplex formation of nucleic acids.[1] It is also essential for the transmission of genetic information.[2] Because of their biological activity, most substituted (or modified) nucleic bases have been frequently studied[3] experimentally. For example, halogenated pyrimidines have been synthesized as potential antitumor, antibacterial, and antiviral agents.[4]

Rapid but accurate prediction of the interaction energies between base pairs is a valuable goal. Hobza and co-workers[5,6] reviewed the theoretical studies of the interaction energies of the natural nucleic acid base pairs. Guerra et al.[7] investigated the nature of hydrogen bond in DNA base pairs, focusing on the role of charge transfer and resonance assistance. From their BP86/TZ2P analysis of adenine−thymine (AT) and guanine−cytosine (GC), they disproved that hydrogen bonding in DNA base pairs is a predominantly electrostatic phenomenon. Instead, it has a substantial charge-transfer character caused by donor−acceptor orbital interactions between O and N lone pairs and N−H σ^*−acceptor orbitals. Grunenberg[8] calculated the inter-residue compliance constants for all hydrogen bonds in Watson−Crick base pairs AT and GC, permitting a quantification of individual hydrogen bond strength. The central inter-residue N−H···N hydrogen bond between G and C is by far the strongest hydrogen bond in both Watson−Crick base pairs. Many theoretical studies have been carried out on the interaction energies of the base pairs between natural nucleic acid bases, but few systematic studies on the substituent base pairs have been reported. In a more recent study Guerra et al. also investigated[9] how the hydrogen-bond lengths, strength, and bonding mechanism were affected by replacing the hydrogen atom H8 in a purine (A, G) and/or the H6 in a pyrimidine (T, C) by a fluorine, chlorine, or bromine substituent. They also analyzed[10] Watson−Crick G:C base pairs in which purine-C8 and/or pyrimidine-C6 positions carry the substituent X = NH^-, NH_2, NH_3^+, O^-, OH, or OH_2^+, using the same level of theory. Meng et al. discussed the geometries of CH_3-, CH_3O-, F-, and NO_2-substituted G:C base pair derivatives and their cations (optimized at the B3LYP/6-31G* level), and the substituent effects on the neutral and cationic geometric structures and energies.[11] They also investigated the substituent effects on AT and GC base pairs caused by CH_3, CH_3O, F, and NO_2,[12] and in a separate study[13] that of large substituents such as 2-nitronaphthalene, 2-hydroxynaphthalene, 1-nitro-4-vinylbenzene, and 1-hydroxyl-4-vinylbenzene. Systematic computational investigations were carried out by Kawahara and Uchimaru,[14] in pursuit of improving the stability of the base pair formed between a chemically modified nucleic acid base derivative and an unmodified one.[15−23]

In this contribution we study the influence of base pair substitution from the perspective of a quantitative model inspired by quantum topological molecular similarity (QTMS). This approach was developed in a series of papers[24−28] and is ultimately based on quantum chemical topology (QCT).[29,30] This method is "quantum" since it draws its data from computational schemes that explicitly incorporate the quantum nature of molecules. It is "topological" since it uses the theory of "atoms in molecules" pioneered by Bader and co-workers to discretize the quantum information contained in a molecular system.[31] Because of its generality and feasibility, QTMS has been applied successfully in a variety of quantitative structure−activity relationships (QSAR).[32−35]

To the best of our knowledge, the number of substituents studied in the literature is limited. Moreover, no quantitative structure−interaction energy relationship has hitherto been developed for substituted base pairs based on descriptors

* To whom correspondence should be addressed.

10.1021/jp7108913 CCC: $40.75
Published on Web 03/29/2008

TABLE 1: Selected Geometric Parameters, Counterpoise Corrected Interaction Energies (ΔE^{HB}), BSSE Values, Substitution Effects ($\Delta\Delta E$), and Predicted $\Delta\Delta E$ Values of C^{6X}:G

no.	substituent	ΔE^{HB} (kJ/mol)	BSSE (kJ/mol)	$\Delta\Delta E$ (kJ/mol)	$O_6\cdots N_4$ (Å)	$N_1\cdots N_3$ (Å)	$N_2\cdots O_2$ (Å)	pred $\Delta\Delta E^a$ (kJ/mol)
1	$-n$-C_4H_9	−107.48	2.22	−3.40	2.81	2.94	2.94	−2.06
2	$-n$-C_3H_7	−107.46	2.19	−3.38	2.81	2.95	2.94	−2.11
3[b]	$-C_2H_5$	−106.10	2.22	−2.02	2.81	2.95	2.94	−1.92
4	$-i$-C_4H_9	−106.03	2.20	−1.94	2.82	2.94	2.94	−1.46
5	$-OC_2H_5$	−105.83	2.23	−1.75	2.81	2.95	2.94	−1.67
6[b]	$-NHCH_3$	−105.62	2.23	−1.54	2.83	2.94	2.93	−1.88
7	$-t$-C_4H_9	−105.51	2.23	−1.43	2.82	2.94	2.94	−2.34
8	$-N(CH_3)_2$	−105.48	2.28	−1.40	2.83	2.94	2.93	−1.79
9[b]	$-OCH_3$	−105.39	2.27	−1.30	2.81	2.95	2.94	−1.23
10	$-i$-C_3H_7	−105.20	2.22	−1.12	2.81	2.94	2.94	−1.54
11	$-COCH_3$	−105.07	2.24	−0.99	2.79	2.95	2.95	−0.98
12[b]	$-C_6H_5$	−104.99	2.24	−0.91	2.81	2.95	2.94	−1.60
13	$-COOC_3H_7$	−104.98	2.23	−0.90	2.80	2.95	2.95	−0.92
14	$-CH_3$	−104.98	2.23	−0.90	2.81	2.95	2.94	−1.54
15[b]	$-COOC_2H_5$	−104.95	2.21	−0.87	2.80	2.95	2.95	−0.88
16	$-CHCH_2$	−104.70	2.22	−0.62	2.81	2.95	2.94	−1.41
17	$-CONH_2$	−104.69	2.25	−0.61	2.79	2.95	2.95	−0.40
18[b]	$-COOCH_3$	−104.65	2.26	−0.57	2.80	2.95	2.95	−0.54
19	$-NH_2$	−104.47	2.25	−0.38	2.82	2.94	2.94	−0.35
20	−OH	−104.11	2.24	−0.03	2.81	2.95	2.95	−0.22
21[b]	−H	−104.08	2.25	0.00	2.80	2.95	2.95	−0.25
22	$-CH_2F$	−103.96	2.25	0.12	2.80	2.95	2.96	0.35
23	−COOH	−103.74	2.25	0.34	2.79	2.96	2.96	0.12
24[b]	$-CH_2Cl$	−103.68	2.31	0.40	2.80	2.95	2.96	0.68
25	$-CH_2Br$	−103.63	2.28	0.45	2.80	2.95	2.96	0.52
26	−COH	−103.54	2.24	0.55	2.78	2.96	2.97	0.81
27[b]	−CCH	−103.52	2.21	0.56	2.80	2.95	2.95	−0.44
28	$-CHBr_2$	−103.47	2.36	0.61	2.79	2.95	2.97	1.03
29	$-CHCl_2$	−103.32	2.35	0.76	2.78	2.95	2.96	1.19
30[b]	−SH	−102.77	2.33	1.32	2.80	2.95	2.96	1.42
31	$-CBr_3$	−102.41	2.45	1.67	2.79	2.95	2.97	1.25
32	$-CCl_3$	−102.03	2.45	2.06	2.79	2.95	2.98	1.72
33[b]	$-CHF_2$	−102.02	2.30	2.07	2.79	2.96	2.97	2.00
34	−COCl	−101.74	2.27	2.35	2.78	2.96	2.98	1.69
35	$-CF_3$	−101.36	2.36	2.73	2.78	2.96	2.98	2.24
36[b]	−Cl	−101.32	2.31	2.77	2.80	2.96	2.97	2.53
37	−Br	−101.10	2.29	2.99	2.79	2.95	2.97	3.21
38	−F	−101.07	2.27	3.02	2.79	2.96	2.97	3.10
39[b]	−NO	−100.97	2.28	3.11	2.78	2.97	2.98	2.34
40	−CN	−100.36	2.22	3.72	2.78	2.96	2.99	3.60
41	$-NO_2$	−100.01	2.26	4.08	2.77	2.97	3.00	4.57
42[b]	$-SO_3H$	−99.83	2.36	4.26	2.78	2.96	2.99	4.56

[a] Predicted results obtained by the MLR model using the electron density evaluated at the BCPs. [b] Substituent belonging to the test set (14 compounds).

calculated from structure alone. Therefore, in this work, we studied the effect on the interaction energy of the C:G base pair caused by a large variety of electron-withdrawing (EWG) and electron-donating groups (EDG) introduced at the cytosine C6 and C5 position. The purpose is to find the different effects due to different substituents, which can be instructive in modifying cytosine interacting with guanine in DNA. For the first time quantitative models of the substituent effects were built, based on QCT. Compared with previous work the data set used in our investigation is more diverse, leading to a more general model, which is also practical given the economy in number of descriptors.

2. Computational Details

2.1. Substituents. The 42 electron-withdrawing and -donating substituents introduced at the 5 and 6 positions of cytosine are listed in Table 1. This unparalleled diversity enabled an extensive consideration of the substituent effects. Figure 1 shows the frame structure of the base pairs. The geometries of each isolated monomer and supermolecule (C^X:G Watson−Crick) were optimized with the B3LYP functional[36,37] in conjunction with the 6-311+G(2d,p) basis set,[38] using the computer program

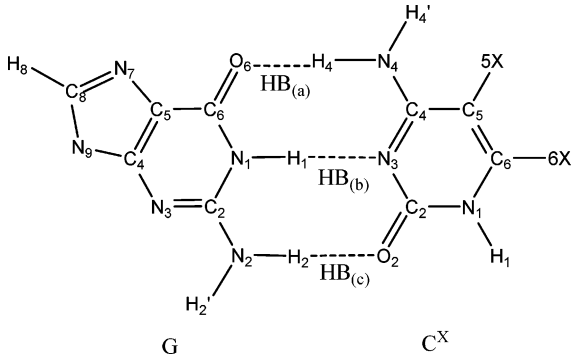

Figure 1. Structure of substituted Watson−Crick C^X:G base pairs.

GAUSSIAN03.[39] The corresponding wave functions of each monomer were also generated. According to previous work in our laboratory[40,41] this level is reliable as it correctly predicts intermolecular vibrational frequencies. The interaction energies of the complexes were corrected for basis set superposition error (BSSE) by the counterpoise method.[42] Hereafter we refer to the molecular interaction energy with BSSE correction as ΔE^{HB}, where HB refers to the association of the interaction energy with hydrogen bonding. The quantity ΔE^{HB} is defined as the

difference between the total energy of a base pair and the corresponding monomers (eq 1). A more negative ΔE^{HB} value means a more stable base pair complex. The quantity $\Delta\Delta E$ is defined as the substituent effect on ΔE^{HB} (eq 2). A more negative $\Delta\Delta E$ value means that the substituted base pair is more stable than the unsubstituted one

$$\Delta E^{HB} = E(C^X{:}G) - \{E(C^X) + E(G)\} \quad (1)$$

$$\Delta\Delta E = \Delta E^{HB}(C^X{:}G) - \Delta E^{HB}(C{:}G) \quad (2)$$

No symmetry was assumed in any of the geometry optimizations. Nonplanarity of the exocyclic amino moiety (i.e., pyramidalization) in isolated bases has been reported[5] before, and the differences in the energy derived from the planar and the nonplanar structure of cytosine are of the order of 0.6−1.6 kJ/mol depending on the level of theory.[43] This energy difference is not negligible within the context of the span of ΔE^{HB} energies and hence had to be taken into account.

2.2. QCT Descriptors. Here we briefly review the background of QCT descriptors.[29,30] For the purpose of this study we only focus on the so-called bond critical points (BCPs). Note that we were not using QCT integrated properties, which are known to consume much CPU time. BCPs are points in real 3D space where the gradient of the electron density, denoted by ρ, vanishes (or $\nabla\rho = \mathbf{0}$) and where the Hessian of ρ has two negative eigenvalues ($\lambda_1 < \lambda_2 < 0$) and one positive one ($\lambda_3 > 0$). BCPs appear roughly in between two bonded nuclei. The descriptors included in the list of QCT descriptors can be packed together in a vector, namely, $\rho,\lambda_1,\lambda_2,\lambda_3,\nabla^2\rho,\epsilon,K,R_e$, describing each bond.[24−28] The first seven vector components are quantities that are evaluated at a BCP, each characterizing a given bond. The Laplacian, denoted by $\nabla^2\rho$, is defined as the sum of the Hessian eigenvalues, the positive curvature (λ_3) being associated with an eigenvector that is tangent to the atomic interaction line. This line is a curve in real space linking two bonded nuclei, along which ρ is a maximum with respect to any neighboring line. The Laplacian measures the degree of charge concentration in space. When the Laplacian is negative at a BCP, charge is locally concentrated, which is typical for covalent and polar bonds. When the Laplacian is positive, we have charge depletion in the bond, which is usually associated with ionic or hydrogen bonds. The ellipticity ϵ is defined as $(\lambda_1/\lambda_2) - 1$ and is always positive since $\lambda_1 < \lambda_2 < 0$ at the BCP. The ellipticity measures the deviation of the electron density distribution from cylindrical symmetry. Hence, within certain limitations, the ellipticity can be used to measure the π character of a bond. However, not all bonds with large π contributions have large ellipticity values. The quantity K is a kinetic energy density. The equilibrium bond length, R_e, is not strictly a BCP property, but for sake of consistency, it can (somewhat artificially) be turned into one. For that purpose one regards R_e as a sum of two distances, i.e., the distance between the BCPs and one nucleus and the distance between the same BCPs and the other nucleus, neglecting any deviation from a straight line the bond path may exhibit. The electron density evaluated at a BCP, ρ, was considered as an important BCP property to serve as a quantum chemical signature for a bond. For example, it has been proposed to relate ρ to bond order via an exponential[44] or multiple linear relationship.[45] In that work,[45] ρ was interpreted as a measure of σ character of a bond. In the present work, only ρ and R_e were employed as the descriptors to explore the direct correlation with the substituent effects in C^{6X}:G. All QCT descriptors were employed for C^{5X}:G base pairs.

Once the wave function files of each monomer have been produced, they were then read by a local version of the program MORPHY98,[46] which extracts the required bond descriptors and exports them into a format that is convenient for subsequent statistical analysis.

2.3. Feature Selection and Regression Analysis. To have compounds for external validation, the data set was split into a training set and an external test set. The regression models were developed on the selected training set, and once the models were built, predictions were made for the test set. Once descriptors were generated, the correlation analysis of descriptors was performed first. In the process of correlation analysis, pairs of descriptors with an absolute correlation coefficient above 0.85 contain at least one descriptor that needs to be eliminated. To decide which descriptor, one focuses on the descriptor that causes correlation coefficients above 0.85 with another descriptor. For instance, the correlation coefficients of ($\rho_{C_5-C_6}$, $\rho_{C_4-C_5}$), ($\rho_{C_5-C_6}$,$\rho_{C_4-N_4}$), and ($\rho_{C_4-C_5}$, $\rho_{C_4-N_4}$) were −0.93, 0.77, and −0.87, respectively. Hence, $\rho_{C_4-C_5}$ was removed from the set of descriptors. After the correlation analysis of the descriptors, descriptor-screening methods were used to select the most relevant descriptor to establish the models for prediction of the molecular property. The selection of relevant descriptors is an important step to construct a predictive model. Here, the stepwise multiple linear regression (MLR) was used as the feature selection method to choose the subset of molecular descriptors.[47,48] This method combines the forward and backward procedures. Stepwise model-building techniques for regression designs with a single dependent variable involve identifying an initial model, repeatedly altering the model from the previous step by adding (forward stepwise) or removing (back stepwise) a predictor variable and terminating the search when stepping does not further improve the model. The best single predictor, which is the most significant variable, was used for the initial linear regression step. Next, descriptors were added one at a time, always adding the one that most improved the fit, until the fit was not significantly improved. Once all the significant variables were determined, the regression equation was constructed. The number of variables retained in the model is based on the levels of significance assumed for inclusion and exclusion of variables from the model. The package SPSS was used for this analysis, and the linear regression is described below.[49]

After the descriptor was selected, MLR was used to develop the linear model of the property of interest, which takes this form

$$Y = b_0 + b_1X_1 + b_2X_2 + ... + b_nX_n \quad (3)$$

In this equation, Y is the property, that is, the dependent variable, X_1, X_2, ..., X_n represent the descriptors, while b_1, b_2, ..., b_n represent the coefficients of those descriptors, and b_0 is the intercept. Regression coefficients represent the independent contributions of each calculated molecular descriptor. The model's performance was described by means of the parameters' fitting power (r^2), standard error of estimates (s.e.), and root-mean-square error (rms). Statistical significance is checked by an F-test of the overall fit at significance level 0.05, followed by t-tests (also at 0.05) of individual descriptors. Any developed model, based on a designed training set, was externally validated by evaluating the prediction errors and computing rms error and r^2. Good predictive properties are an additional indication that chance correlation has been avoided.

3. Results and Discussion

3.1. The C^{6X}:G Complexes. *3.1.1. The Substitution Effects Calculated by Gaussian.* A complete list of the substituents'

TABLE 2: Correlation Equations for the Linear Model Based on Electron Densities Evaluated at BCPs (Model a), Equilibrium Bond Lengths (Model b), and NPA Charges (Model c)

	descriptor	unstandardized coefficients	unstandardized error	standardized coefficients	*t* test	sig.
model a[a]	constant	1077.12	145.41		7.41	0.00
	$\rho_{C_4-N_4}$	625.54	48.57	0.73	12.88	0.00
	$\rho_{N_4-H_4}$	−3604.63	427.75	−0.48	−8.43	0.00
	$\rho_{C_2-N_1}$	−230.24	27.32	−0.48	−8.43	0.00
model b[b]	constant	−1897.27	213.12		−8.90	0.00
	R_{e,C_4-N_4}	−292.88	21.68	−1.19	−13.5 1	0.00
	R_{e,N_4-H_4}	1269.98	128.87	0.87	9.86	0.00
	R_{e,C_2-N_1}	85.948	11.694	0.450	7.35	0.00
model c[c]	constant	555.40	87.32		6.36	0.00
	C_{N_4}	566.71	65.42	1.67	8.66	0.00
	C_{C_5}	−20.04	4.25	−0.54	−4.71	0.00
	C_{H_6}	−326.50	104.65	−0.52	−3.12	0.01

[a] Training set: $r^2 = 0.93$; $F = 102.54$; s.e. = 0.57; rms = 0.53; $n = 28$; N=3. Test set: $r^2 = 0.96$; rms = 0.42. [b] Training set: $r^2 = 0.91$; $F = 81.98$; s.e. = 0.63; rms = 0.58; $n = 28$; $N = 3$. Test set: $r^2 = 0.94$; rms = 0.54. [c] Training set: $r^2 = 0.86$; $F = 48.16$; s.e. = 0.80; rms = 0.74; $n = 28$; $N = 3$. Test set: $r^2 = 0.61$; rms = 1.12.

names and the corresponding counterpoise corrected interaction energies (ΔE^{HB}), the substitution effects ($\Delta\Delta E$), and BSSE values of 42 C^{6X}:G base pairs are given in Table 1. The interaction energy of the C:G base pair (compound 21) was −104.08 kJ/mol. For compounds 1−20, the values of $\Delta\Delta E$ were all negative, which means that these modified base pairs were more stable than unmodified C:G. For compounds 22−42, the values of $\Delta\Delta E$ were all positive, and this indicated that these base pairs were less stable than C:G. The most stable and unstable complex was $C^{6-n-C_4H_9}$:G and C^{6-SO_3H}:G, with the interaction energies of −107.48 and −99.83 kJ/mol, respectively. Within this range of 7.7 kJ/mol, there is a fairly homogeneous distribution of data. Remarkably, the cytosine 6 position derivatives possessing stronger EWGs, such as NO_2, F, or CN, form a less stable base pair with guanine. This observation agrees with earlier work.[11,19] The data set was divided into two subsets, the training set and the test set, by ranking the compounds according to their interaction energy. Normally, the size of the test set equals to 20−50% of the whole data set. For the C^{6X}:G data set, the test set was constructed by selecting every third compound of the series. As a result the training set (1, 2, 4, 5, 7, 8, ...) contains twice as many elements (28) as the test set (3, 6, 9, ...) (14 compounds). The training set was used to build the MLR models, and the test set was used to evaluate its prediction ability.

The hydrogen bonds length of C^{6X}:G were also listed in Table 1. In our calculation, the lengths of the three hydrogen bonds $HB_{(a)}$, $HB_{(b)}$, and $HB_{(c)}$ in the C:G base pair are 2.80, 2.95, and 2.95 Å (short−long−long), respectively, while the corresponding lengths measured by experimental X-ray crystallography are 2.91, 2.95, and 2.86 Å (long−long−short), respectively.[3] This qualitative discrepancy has been noted before by Meng et al.[12,51] and also by Guerra and Bickelhaupt,[50] who showed that it was not due to neglecting ribose. Table 1 shows that, for C^{6X} derivatives possessing an EDG, most $HB_{(a)}$ lengths were elongated and $HB_{(b)}$ and $HB_{(c)}$ lengths contracted. On the contrary, for C^{6X} derivatives possessing an EWG, most $HB_{(a)}$ lengths were contracted, and $HB_{(b)}$ and $HB_{(c)}$ elongated.

As one would expect, simply from the hydrogen pattern shown in Figure 1, cytosine derivatives act as electron acceptors in the hydrogen bond $HB_{(a)}$ and as electron donors in $HB_{(b)}$ and $HB_{(c)}$. This assertion is confirmed by a charge transfer analysis[7] of the GC complex by Guerra et al. They used their Voronoi deformation density method to quantify the charge transfer from G and C through $HB_{(a)}$ and the reverse transfer through $HB_{(b)}$ and $HB_{(c)}$. The introduction of an EWG on C^{6X} is expected to aid the transfer through $HB_{(a)}$ and thereby strengthen this hydrogen bond. Similarly, this EWG would weaken $HB_{(b)}$ and $HB_{(c)}$. Now, from Table 1, a C^{6X} species possessing an increasingly powerful EWG weakens the GC complex progressively more. Thus one would conclude that the sum of the substituent effects on $HB_{(b)}$ and $HB_{(c)}$ overcomes the substituent effect on $HB_{(a)}$. Is this compatible with relative hydrogen bond strengths? The strength of the three hydrogen bonds in GC has been compared by Grunenberg using compliance constants.[8] Indeed, at the B3LYP/6-311++G(d,p) level, he found that $HB_{(b)}$ is twice as strong as $HB_{(c)}$ (while their bond lengths are virtually identical) and almost twice as strong as $HB_{(a)}$. This finding confirms the combined dominance of $HB_{(b)}$ and $HB_{(c)}$ over $HB_{(a)}$. This paragraph logically draws together computational findings to present a consistent picture of subsituent effects on base pair stability. However, this interpretation operates under the assumption that something as complex as an overall stability, which is dependent on many atom−atom interactions, is due to hydrogen bonding alone. The number of hydrogen bonds and their strength, as well as their generalization into the secondary interaction hypothesis, can be a misleading guide to complex stability.[52] If this guide is effective then it is for currently obscure reasons that should unravel when topological energy partitioning[53,54] becomes feasible for systems of the size of G:C.

3.1.2. The MLR Model Based on Electron Densities Evaluated at BCPs. The electron density was evaluated at 13 BCPs for each cytosine derivative. From Figure 1 it is clear that the BCPs correspond to the six bonds in the aromatic ring, the five bonds immediately attached to this ring, and the two extra bonds in the amino group (N_4-H_4 and N_4-H_4'). Note that all BCPs are sampled from the monomers. Including intermolecular BCPs would defeat the object of constructing a model purely from the monomeric properties. This objective stems from the fact that obtaining monomeric data represents a typical saving in CPU time from a given number of days to that same number in hours. The substitution effect $\Delta\Delta E$ was used as the dependent variable Y. After the correlation analysis of the descriptors, the pool of descriptors was reduced to 10 by correlation analysis. A stepwise regression routine (in SPSS) was used to develop the linear model for the prediction of the substitution effect of C^{6X}:G using the calculated densities of BCPs. The best linear model contains 3 descriptors. The regression coefficients of the descriptors are listed in Table 2. "Unstandardized coefficients" refers to the coefficients in the regression equation. Standardized coefficients have been standardized so that they have variances of 1. The values for "Sig." are all essentially zero, which indicates that the descriptors are significant predictors of

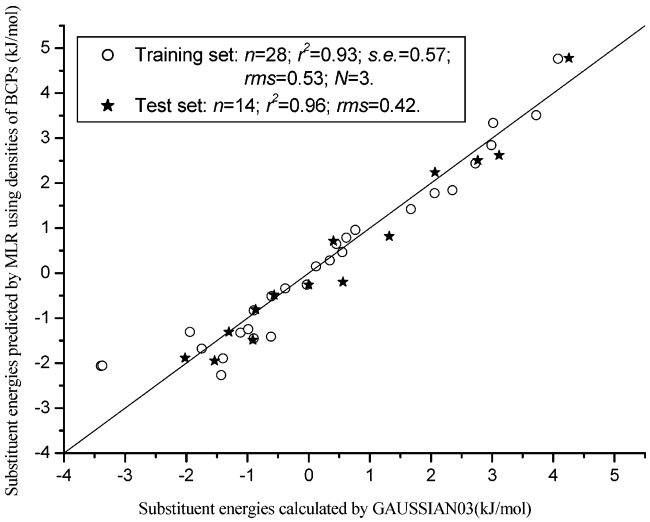

Figure 2. Correlation between C^{6X}:G substituent energies ($\Delta\Delta E$) predicted by an MLR model based on BCP electron densities vs energies calculated by Gaussian.

interaction energies. The linear correlation coefficient value of each of the two descriptors is less than 0.85, which means that the descriptors were independent in this MLR analysis. The three descriptors selected through stepwise regression (by SPSS) were the electron densities evaluated at the C_4-N_4, N_4-H_4, and C_2-N_1 BCPs. This model gave an rms error of 0.53 kJ/mol for the training set, 0.42 kJ/mol for the test set, and 0.49 kJ/mol for the whole set, and the corresponding correlation coefficients (r^2) were 0.93, 0.96, and 0.94, respectively. Figure 2 shows these predicted substitution effects obtained by a MLR based on electron densities evaluated at BCPs vs those calculated by Gaussian.

3.1.3. The MLR Model Based on Equilibrium Bond Lengths. Thirteen equilibrium bond lengths were also calculated for each cytosine derivative. After the correlation analysis, the descriptors were used to develop the second linear model for the prediction of the substitution effect of C^{6X}:G using a stepwise regression routine. The best linear model contained 3 descriptors. The regression coefficients of the descriptors are listed in Table 2. The three descriptors selected in the model were the equilibrium bond lengths of C_4-N_4, N_4-H_4, and C_2-N_1. This model gave an rms error of 0.58 kJ/mol for the training set, 0.54 kJ/mol for the test set, and 0.57 kJ/mol for the whole set, and the corresponding correlation coefficients (r^2) were 0.91, 0.94, and 0.91, respectively. Figure 3 shows these predicted substitution effects obtained the MLR model based on equilibrium bond length vs those calculated by Gaussian.

3.1.4. The MLR Model Based on NPA Charges. To test the suitability of the models that were proposed by QCT descriptors, we also built a third model based on NPA charges, obtained by natural population analysis. The separation of the training and test set was similar to that in the two previous models, and so was the process of building the MLR model. The results of the best linear regression model are also listed in Table 2. The three descriptors selected in this model were the NPA charges of N_4, C_5, and H_4 atoms, respectively. This model gave an rms error of 0.74 kJ/mol for the training set, 1.12 kJ/mol for the test set, and 1.40 kJ/mol for the whole set, and the corresponding correlation coefficients (r^2) were 0.80, 0.61, and 0.65, respectively. Figure 4 shows these predicted substitution effects vs those calculated by Gaussian.

3.1.5. Comparison of the Three MLR Models and Discussion of the Descriptors. Three MLR models for the prediction of

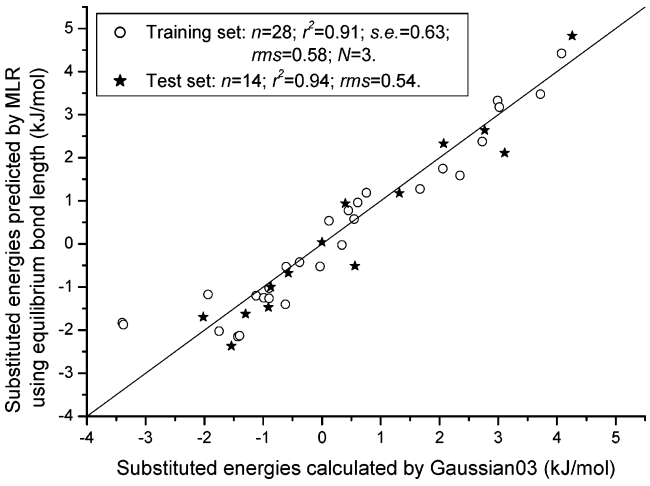

Figure 3. Correlation between C^{6X}:G substituent energies ($\Delta\Delta E$) predicted by an MLR model based on equilibrium bond length vs energies calculated by Gaussian.

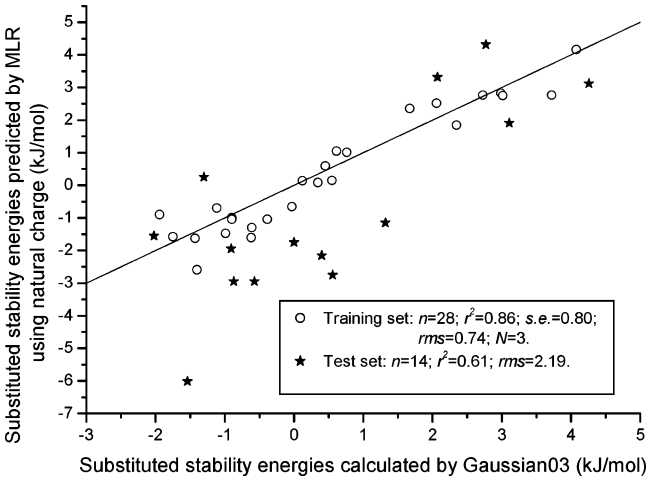

Figure 4. Correlation between C^{6X}:G substituent energies ($\Delta\Delta E$) predicted by an MLR model based on NPA charges vs energies calculated by Gaussian.

the substituent effect of the C^{6X}:G base pairs were generated, one based on (a) BCP electron densities, (b) equilibrium bond lengths, and (c) NPA charges. For these three models, the inputs of the external test set were presented to it, and the predicted results were obtained. The values (see Table 2) of r^2 of the training set and r^2 of the test perhaps lack the discriminatory power to clearly rank the models according to quality. However, the F test value of model a is highest, which means that this model is most significant. Second, the rms error of model a is lowest, and this indicates that model a has most predictive ability. Model a is the best, followed by model b, and then model c. We note that model c (NPA charges) is markedly worse than the two others. Consequently, the substituted effects of C^{6X}:G complexes were predicted from the model with electron densities at the BCPs, as listed in Table 1 for both training and test set.

By interpretation of the descriptors in the regression model, it is possible to gain some insight into factors that are likely to relate to the substituted effect of the C^{6X}:G base pair. Both the model based on ρ at BCPs and that based on R_e contained three descriptors. Table 2 reveals that the three descriptors in the former model belong to three bonds in C^{6X}, that is, C_4-N_4, N_4-H_4, and C_2-N_1. Interestingly, the three descriptors in the second model also belong to these three bonds. So, we conclude

TABLE 3: Selected Geometric Parameters, Counterpoise Corrected Interaction Energies (ΔE^{HB}), BSSE Values, Substitution Effects ($\Delta\Delta E$), and Predicted $\Delta\Delta E$ values of C^{5X}:G

no.	substituent	ΔE^{HB} (kJ/mol)	BSSE (kJ/mol)	$\Delta\Delta E$ (kJ/mol)	$O_6\cdots N_4$ (Å)	$N_1\cdots N_3$ (Å)	$N_2\cdots O_2$ (Å)	pred $\Delta\Delta E^a$ (kJ/mol)
1[b]	$-CH_2Br$	−117.32	2.37	−13.24	2.80	2.97	2.97	−12.13
2	$-CH_2Cl$	−114.55	2.36	−10.47	2.80	2.97	2.96	−6.17
3	$-CH_2F$	−113.87	2.36	−9.79	2.81	2.96	2.96	−6.19
4	$-OCH_3$	−110.12	2.32	−6.03	2.81	2.95	2.93	−3.34
5[b]	$-OC_2H_5$	−110.02	2.30	−5.93	2.81	2.95	2.93	−8.61
6	$-n$-C_4H_9	−106.45	2.26	−2.36	2.80	2.95	2.93	−2.59
7	$-n$-C_3H_7	−106.37	2.26	−2.28	2.80	2.95	2.93	−2.61
8	$-CH_3$	−106.02	2.31	−1.94	2.80	2.95	2.93	−2.35
9[b]	$-t$-C_4H_9	−105.99	2.33	−1.90	2.80	2.98	2.91	−3.31
10	$-C_2H_5$	−105.95	2.27	−1.86	2.80	2.95	2.93	−3.28
11	$-i$-C_3H_7	−105.82	2.29	−1.74	2.80	2.96	2.92	−2.64
12	$-i$-C_4H_9	−105.75	2.29	−1.66	2.81	2.96	2.92	−2.85
13[b]	−NO	−104.11	2.32	−0.03	2.79	2.98	3.01	1.34
14	−H	−104.08	2.25	0.00	2.80	2.95	2.95	−0.83
15	$-N(CH_3)_2$	−104.03	2.34	0.05	2.82	2.95	2.94	0.37
16	−OH	−104.00	2.35	0.08	2.80	2.95	2.94	−0.62
17[b]	$-NHCH_3$	−103.35	2.30	0.73	2.82	2.95	2.94	0.37
18	$-C_6H_5$	−103.27	2.33	0.82	2.81	2.96	2.94	−0.99
19	−F	−102.56	2.38	1.53	2.79	2.95	2.96	0.86
20	$-CHCH_2$	−102.47	2.33	1.62	2.80	2.96	2.95	−0.79
21[b]	$-NH_2$	−101.83	2.31	2.25	2.82	2.95	2.94	1.73
22	−Cl	−101.07	2.39	3.01	2.79	2.96	2.96	3.19
23	−Br	−100.84	2.42	3.25	2.84	3.00	2.98	3.88
24	−SH	−100.38	2.36	3.71	2.80	2.96	2.96	2.37
25[b]	−CCH	−100.08	2.29	4.01	2.80	2.96	2.96	4.56
26	$-CHCl_2$	−99.79	2.43	4.29	2.79	2.97	2.96	5.60
27	$-CHBr_2$	−99.56	2.60	4.52	2.79	2.97	2.96	5.78
28	$-CHF_2$	−98.64	2.43	5.44	2.79	2.97	2.97	4.96
29[b]	$-CCl_3$	−97.89	2.42	6.20	2.78	2.98	2.96	5.18
30	$-CF_3$	−97.62	2.41	6.47	2.78	2.97	2.98	3.91
31	$-CBr_3$	−97.33	2.54	6.76	2.78	2.99	2.96	6.52
32	−CN	−95.37	2.28	8.71	2.78	2.97	3.00	5.86
33[b]	$-COOC_3H_7$	−94.94	2.27	9.14	2.82	2.97	2.96	10.52
34	$-COOC_2H_5$	−94.81	2.32	9.27	2.82	2.97	2.96	10.54
35	$-COOCH_3$	−94.70	2.33	9.38	2.82	2.97	2.96	10.53
36	$-SO_3H$	−94.05	2.41	10.03	2.79	2.98	3.00	8.19
37[b]	−COOH	−93.77	2.30	10.32	2.81	2.97	2.97	11.05
38	$-CONH_2$	−93.75	2.33	10.33	2.83	2.97	2.97	10.70
39	$-COCH_3$	−92.93	2.30	11.15	2.82	2.98	2.97	13.10
40	−COH	−92.17	2.31	11.91	2.81	2.97	2.99	13.63
41[b]	−COCl	−91.78	2.31	12.30	2.79	2.98	3.00	11.59
42	$-NO_2$	−91.41	2.32	12.67	2.78	2.99	3.00	12.15

[a] Predicted results obtained by MLR model using the QCT method. [b] Subsituent belonging to the test set (11 compounds).

that these three bonds in C^{6X} play an important role in determining the stability of the base pair formed between C^{6X} and G.

In the following, we only discuss the BCP electron density model since this model was chosen as the best one for the prediction of the substituent effect. The quantity $\rho_{C_4-N_4}$ receives a positive coefficient in the regression; this indicates that the substituent effect $\Delta\Delta E$, increases and hence the substituted base pair becomes less stable with increasing ρ. The descriptors $\rho_{N_4-H_4}$ and $\rho_{C_2-N_1}$ receive negative regression coefficients, which indicates that increasing ρ leads to a base pair stabilization. The t test value of the$\rho_{C_4-N_4}$ descriptor is the largest (12.88, Table 2), which means that it has the largest influence on the interaction energy.

3.2. The C^{5X}:G Complexes. *3.2.1. The Substitution Effects Calculated by Gaussian.* Table 3 lists the counterpoise corrected interaction energies (ΔE^{HB}), the substitution effects ($\Delta\Delta E$), and BSSE values of 42 C^{5X}:G base pairs. The interaction energy of the unsubstituted C:G base pair (No. 14) was −104.08 kJ/ mol. Compounds 1−13 all have negative $\Delta\Delta E$ values, indicating that these modified base pairs are more stable than unmodified C:G. For compounds 15−42, the positive $\Delta\Delta E$ values means that these base pairs are less stable than C:G. The most stable and most unstable base pair was C^{5-CH_2Br}:G and C^{5-NO_2}:G, with the interaction energies of −117.32 and −91.41 kJ/mol, respectively. This range of 26 kJ/mol is 3.4 times larger than that of C^{5X}:G, with the same favorable (i.e., homogeneous) data distribution. From Table 3, the remarkable tendency of C^{5X} derivatives with stronger EDGs forming more stable base pairs again emerges, in agreement with earlier work.[11,19] The C^{5X}:G data set was divided into two subsets. The test set was constructed by selecting every fourth compound of the series. As a result the training set (2, 3, 4, 6, 7, 8, 10, ...) contains roughly three times as many elements (31) as the test set (1, 5, 9, ...) (11 compounds).

The hydrogen bond lengths of C^{5X}:G are also listed in Table 3. Table 3 shows that, in C^{5X} derivatives possessing an EDG, the lengths of most $HB_{(a)}$, $HB_{(b)}$, and $HB_{(c)}$ change slightly, compared with those in C:G. On the contrary, for C^{5X} derivatives possessing an EWG, the lengths of most $HB_{(a)}$ were contracted, and $HB_{(b)}$ and $HB_{(c)}$ were elongated. The effect of EWGs (or vice versa EDGs) on the stability of the complex can again be rationalized along the lines of the discussion (just before section 3.1.2) in connection with Table 1 (C^{6X} data). Note that the ranking in terms of stability of the substituents is different between C^{6X} and C^{5X}. However, here again and broadly speaking, a C^{5X} species possessing an increasingly powerful EWG weakens the G:C^{5X} complex progressively more. The

TABLE 4: Correlation Equations for the Linear Model Based on QCT Method for C^{5X}:G[a]

descriptor	unstandardized coefficients	unstandardized error	standardized coefficients	*t* test	sig.
constant	−296.45	19.31		−15.36	0.00
λ_{3,C_4-N_4}	859.94	51.67	0.91	16.64	0.00
$\epsilon_{C_4-C_5}$	−48.75	18.29	−0.15	−2.67	0.01

[a] Training set: $r^2 = 0.92$; $F = 102.54$; se = 1.77; rms = 1.69; $n = 31$; $N = 2$.Test set: $r^2 = 0.97$; rms = 1.24; $n = 11$.

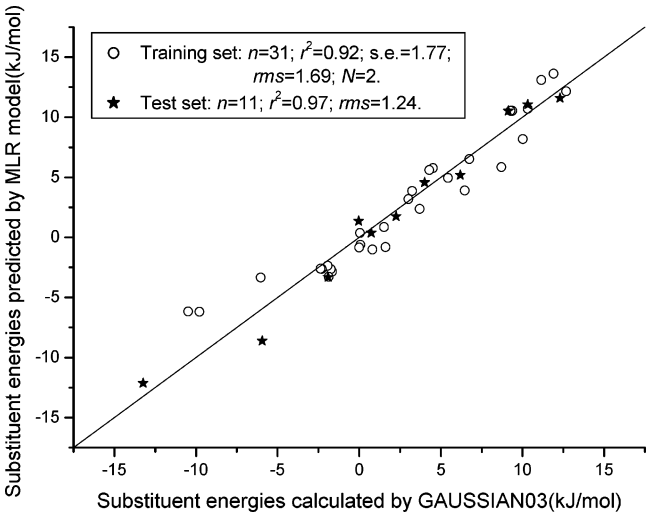

Figure 5. Correlation between C^{5X}:G substituent energies ($\Delta\Delta E$) predicted by an MLR model based on two QCT descriptors vs energies calculated by Gaussian.

arguments of specific charge transfer through the three hydrogen bonds and their relative strength, used for the C^{6X} case, can be invoked here as well. This is again done under the tacit assumption that these assertions remain valid for substituted C, since they were originally only quoted for unsubstituted C (i.e., in the "pure" G:C complex). As a result, $HB_{(b)}$ and $HB_{(c)}$ again dominate $HB_{(a)}$ when determining the stability of the complex. The former drown the influence of $HB_{(a)}$ in making the stronger EWGs stabilize the stability of the complex. In actual fact, the opposite is observed: stronger EWGs weaken the complexes. Again we retrieve a consistent picture.

3.2.2. Results of MLR Model Based on QCT. The BCP properties were calculated for each C^{5X} derivative. The substitution effect $\Delta\Delta E$ calculated by Gaussian was used as the dependent variable. For C^{5X}:G base pairs, unfortunately, the models only based on electron densities or equilibrium bond length are much worse than those based on all QCT descriptors. So, all BCP descriptors were used to develop a linear model for the prediction of $\Delta\Delta E$. The process of building the MLR model was similar with that for C^{6X}:G. The best linear model contains two descriptors. The correlation coefficient of the two descriptors is less than 0.85, which means they were independent in this MLR analysis. The two descriptors selected were λ_{3,C_4-N_4} and $\epsilon_{C_4-C_5}$. The quantity λ_{3,C_4-N_4} received a positive coefficient in the regression, indicating that $\Delta\Delta E$ increases and hence the base pairs become less stable upon an increase in λ_{3,C_4-N_4}. The descriptor $\epsilon_{C_4-C_5}$ received a negative coefficient in the regression, indicating that the base pairs become more stable upon an increase in $\epsilon_{C_4-C_5}$. By analysis of the descriptors in the regression model, we conclude that two bonds C_4-N_4 and C_4-C_5 in the C^{5X} derivative play an important role in determining the stability of the base pair formed with guanine. According to the *t* test values (Table 4), λ_{3,C_4-N_4} has most influence on the interaction energy.

This model gave an rms error of 1.77 kJ/mol and correlation coefficient r^2 of 0.92 for the training set. With the test set (Table 4), the prediction results were obtained, confirming the predictive capability of the model. The statistical parameters were $r^2 = 0.97$ and rms = 1.24 kJ/mol. Figure 5 shows these predicted substituent effects obtained by MLR model vs those calculated by Gaussian for all of 42 C^{5X}:G base pairs studied, the training set and the test set.

4. Conclusion

We performed quantum chemical calculations on 42 Watson−Crick type base pairs formed between unmodified guanine and modified cytosine monomers, substituted in the 5 and 6 positions. The presence of an EWG at the 5 or 6 positions of cytosine induces a less stable base pair with guanine. For the first time, quantitative models using the MLR method based were built with QCT descriptors to predict the interaction energy of C^{6X}:G and C^{5X}:G base pairs. For C^{6X}:G base pairs, another two models based on equilibrium bond lengths and NPA charges were generated to compare the performance among them. The MLR model based on electron densities evaluated at BCPs provided the best results with much better predictive ability than that of the other two models. We conclude that, for both C^{6X}:G and C^{5X}:G: (1) BCP properties represent the features of substituted cytosine responsible for its stability in forming a complex with guanine; (2) the model linking these properties with substituent energy effects is linear. A model based on NPA charges does not achieve this high linear correlation for C^{6X}:G. In summary, the models built in this work can be used to estimate the interaction energies of Watson−Crick type C^{6X}:G and C^{5X}:G base pairs. Since only monomeric data are required this represents a typical saving in CPU time from a given number of days to that same number in hours.

Acknowledgment. This work is supported by European Union "Individual Fellowship" (Marie Curie, FP6-2004-Mobility, Proposal No. 021966-DUPLEX).

References and Notes

(1) Watson, J. D.; Crick, F. H. C. *Nature* **1953**, 737.
(2) Brenner, S.; Jacob, F.; Meselson, M. *Nature* **1961**, *190*, 576.
(3) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984.
(4) Morris, S. M. *Mutat. Res.* **1993**, *297*, 39.
(5) Hobza, P.; Sponer, J. *Chem. Rev.* **1999**, *99*, 3247.
(6) Hobza, P.; Zahradnik, R.; Mueller-Dethlefs, K. *Collect. Czech. Chem. Commun.* **2006**, *71*, 443.
(7) Guerra, C. F.; Bickelhaupt, F. M.; Snijders, J. G.; Baerends, E. J. *Chem.−Eur. J.* **1999**, *5*, 3581.
(8) Grunenberg, J. *J. Am. Chem. Soc.* **2004**, *126*, 16310.
(9) Guerra, C. F.; Wijst, T. v. d.; Bickelhaupt, F. M. *Struct. Chem.* **2005**, *16*, 211.
(10) Guerra, C. F.; Wijst, T. v. d.; Bickelhaupt, F. M. *Chem.−Eur. J.* **2006**, *12*, 3032.
(11) Meng, F.; Wang, H.; Xu, W.; Liu, C.; Wang, H.; Xu, W.; Liu, C. *Chem. Phys.* **2005**, *308*, 117.
(12) Meng, F.; Liu, C.; Xu, W. *Chem. Phys.* Letters **2003**, *373*, 72.
(13) Meng, F.; Wang, H.; Xu, W.; Liu, C. *Int. J. Quantum Chem.* **2005**, *104*, 79.
(14) Kawahara, S.; Uchimaru, T. *Eur. J. Org. Chem.* **2003**, 2577.
(15) Kawahara, S.-i.; Kobori, A.; Sekine, M.; Taira, K.; Uchimaru, T. *J. Phys. Chem. A* **2001**, *105*, 10596.
(16) Kawahara, S.-i.; Sekine, M.; Taira, K.; Kobayashi, H.; Uchimaru, T. *Nucleic Acids Res. Supp.* **2002**, 191.

(17) Kawahara, S.-i.; Taira, K.; Sekine, M.; Uchimaru, T. *Nucleic Acids Symp. Ser.* **2000**, 237.
(18) Kawahara, S.-i.; Uchimaru, T. *THEOCHEM* **2002**, *588*, 29.
(19) Kawahara, S.-i.; Uchimaru, T. *Eur. J. Org. Chem* **2003**, 2577.
(20) Kawahara, S.-i.; Uchimaru, T.; Taira, K.; Sekine, M. *J. Phys. Chem. A* **2001**, *105*, 3894.
(21) Kawahara, S.-i.; Uchimaru, T.; Taira, K.; Sekine, M. *J. Phys. Chem. A* **2002**, *106*, 3207.
(22) Kawahara, S.-i.; Uchimaru, T.; Sekine, M. *THEOCHEM* **2000**, *530*, 109.
(23) Kawahara, S.-i.; Wada, T.; Kawauchi, S.; Uchimaru, T.; Sekine, M. *J. Phys. Chem. A* **1999**, *103*, 8516.
(24) O'Brien, S. E.; Popelier, P. L. A. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 764.
(25) O'Brien, S. E.; Popelier, P. L. A. *J. Chem. Soc., Perkin Trans.* **2002**, *2*, 478.
(26) O'Brien, S. E.; Popelier, P. L. A. *Can. J. Chem.* **1999**, *77*, 28.
(27) Popelier, P. L. A. *J. Phys. Chem. A* **1999**, *103*, 2883.
(28) Popelier, P. L. A.; Chaudry, U. A.; Smith, P. J. *J. Chem. Soc., Perkin Trans.* **2002**, *2*, 1231.
(29) Popelier, P. L. A. *Atoms in Molecules: an Introduction*; Pearson Educ., London, GB, 2000.
(30) Bader, R. F. W. *Atoms in Molecules: A Quantum Theory*; Clarendon Press, Oxford, GB, 1990.
(31) Popelier, P. L. A. *J. Phys. Chem. A* **1999**, *103*, 2883.
(32) Popelier, P. L. A.; Smith, P. J. *Eur. J. Med. Chem.* **2006**, *41*, 862.
(33) Chaudry, U. A.; Popelier, P. L. A. *J. Phys. Chem. A* **2003**, *107*, 4578.
(34) Chaudry, U. A.; Popelier, P. L. A. *J. Org. Chem.* **2004**, *69*, 233.
(35) Smith, P. J.; Popelier, P. L. A. *J. Comp.-Aided Mol. Des.* **2004**, *18*, 135.
(36) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.
(37) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.
(38) Foresman, J. B.; Frisch, A. *Exploring Chemistry with Electronic Structure Methods*, 2nd ed.; Gaussian Inc.: Pittsburgh, PA, 1996.
(39) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Vreven, J. T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision C.02; Wallingford, CT, 2003.
(40) Popelier, P. L. A.; Joubert, L. *J. Am. Chem. Soc.* **2002**, *124*, 8725.
(41) Joubert, L.; Popelier, P. L. A. *Phys. Chem. Chem. Phys.* **2002**, *4*, 4353.
(42) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553.
(43) Sponer, J.; Florian, J.; Leszczynski, J.; Hobza, P. *J. Biomol. Struct. Dyn.* **1996**, *13*, 827.
(44) Bader, R. F. W.; Slee, T. S.; Cremer, D.; Kraka, E. *J. Am. Chem. Soc.* **1983**, *105*, 5061.
(45) Howard, S. T.; Lamarche, O. *J. Phys. Org. Chem.* **2003**, *16*, 133.
(46) *MORPHY98: a program written by P.L.A. Popelier with a contribution from R.G.A. Bone*; UMIST: Manchester, England, EU (1998).
(47) Darlington, R. B. *Regression and Linear Models*; McGraw-Hill: New York, 1990.
(48) Massart, D. L.; Vandeginste, B. G. M.; Buydens, L. M. C.; Jong, S. D.; Lewi, P. J.; Smeyers-Verbeke, J. *Handbook of Chemometrics and Qualimetrics. Part A.*; Elsevier: Amsterdam, the Netherlands, 1997.
(49) SPSS Inc., http://spss.com, Chicago, 2000.
(50) Guerra, C. F.; Bickelhaupt, F. M. *Angew. Chem., Int. Ed.* **1999**, *38*, 2942.
(51) Meng, F.; Wang, H.; Xu, W.; Liu, C. *Int. J. Quantum Chem.* **2005**, *104*, 79.
(52) Popelier, P. L. A.; Joubert, L. *J. Am. Chem. Soc.* **2002**, *124*, 8725.
(53) Rafat, M.; Popelier, P. L. A. *J. Comput. Chem.* **2007**, *28*, 292.
(54) Pendás, A. M.; Francisco, E.; Blanco, M. A. *J. Comput. Chem.* **2005**, *26*, 344.