

Density Functional Study on Core Ionization Spectra of Cytidine and Its Fragments

Alexander Thompson,[1] Saumitra Saha,[1] Feng Wang,*[1] Takashi Tsuchimochi,[2,†] Ayako Nakata,[2,††] Yutaka Imamura,[2] and Hiromi Nakai*[2]

[1]Centre for Molecular Simulation, Swinburne University of Technology, P. O. Box 218, Hawthorn, Melbourne, Victoria, Australia 3122

[2]Department of Chemistry and Biochemistry, School of Advanced Science and Engineering, Waseda University, 3-4-1 Okubo, Shinjuku-ku, Tokyo 169-8555

Received March 6, 2008; E-mail: fwang@swin.edu.au, nakai@waseda.jp

The sugar–base correlation of cytosine (base) and deoxyribose (sugar) moieties of cytidine is investigated based on their inner-shell electronic structural information. A recently developed density functional theory (DFT) model, CV-B3LYP, with a Gaussian-type basis set of 6-311G**, and the DFT-LB94 model with a Slater-type basis set of TZ2P are employed to calculate inner shell ionization energies. The results reveal that the corresponding geometry of cytidine is not significantly different from its fragments, i.e., cytosine and deoxyribose. Changes in charge distribution of cytidine with respect to cytosine and deoxyribose concentrate on the local C sites in the base pyrimidine ring and sugar ring, as indicated by the atomic Hirshfeld charges. The O-K, N-K, and C-K spectra of cytidine inherit the aromatic signature in cytosine, suggesting that the role of the aromatic ring is a buffer to diffuse the changes brought in by the addition of the deoxyribose moiety. Formation of cytidine, however, substantially changes the C-K spectra of the deoxyribose moiety. In general, the correlated O-K, N-K, and C-K sites of cytidine exhibit small red shifts with respect to the cytosine base, whereas the O-K and C-K sites of cytidine show blue shifts in comparison with those of deoxyribose.

DNA and RNA are polymers of (deoxy)ribonucleotides, which as building units of life consist of nucleosides and phosphate fragments. A nucleoside is a combination of a sugar or (deoxy)ribose and a base. Identification of the sugar–base correlation in nucleotides (RNA/DNA) is one of the most important and challenging steps toward NMR assignment of oligonucleotide responses.[1] Interactions between the DNA/RNA bases and the (deoxy)ribose moiety through a β-N_1-glycosidic C–N linkage are fundamental to the understanding of the double helix and conductivity of DNA fibers.[2] Intrinsic dynamics often observed in the functionally relevant RNA/DNA structure elements seriously complicates experimental assignment procedures, especially in larger systems. Nucleoside derivatives are often involved in important functions in cellular metabolism and are used to synthesize enzyme inhibitors, and antiviral agents, as well as anticancer agents.[3]

Understanding of site-selective X-ray initiated molecular photolysis is an important gateway for the development of biotechnology and evolving technology in fabrication of microelectronic devices into the nanometer range.[4,5] Core orbitals are essentially localized and exhibit properties of theoretical and practical significance.[6] Binding energy spectra of inner-shell electrons depend on the charge distribution in a molecule, and the ability of the neighboring atoms to screen the positive charge introduced through ionization.[7] Hence core-ionization energies express the ability of a molecule to accept charge at a specific site,[7,8] which provides valuable information to study charge distribution and chemical bonding of molecules.[9] Splitting core levels, which are called chemical shift in the molecular or crystal environment,[10] have been exploited using electron spectroscopy for chemical analysis[9] (ESCA) and provided a useful source of information to study fragment correlation.[11]

Observation of the energy splitting in the core shell relies on the development of high-resolution inner-shell spectroscopy. Recent state-of-the-art third generation synchrotron sourced facilities have made it possible to study inner-shell electronic structures of biological molecules. For example, near-edge X-ray absorption fine structure (NEXAFS) is able to distinguish elements and sites of excitation/ionization owing to the localized nature of the core orbitals.[10,12] Recent measurements of O-K, N-K, and C-K inner-shell chemical shifts in DNA bases using synchrotron-sourced soft X-ray absorption spectroscopy (XAS) and X-ray emission spectroscopy (XES) demonstrated different chemical environments of these elements in the species.[2,4,13,14] When the photoabsorption process occurs in biological systems such as living cells, it may cause significant photobiological effects. Induced molecular changes are recognized and modified by various enzymes in intracellular environments. The repairable and unrepairable changes

† Present address: Department of Chemistry, Rice University, Houston, Texas 77005, USA

†† Present address: Department of Applied Chemistry, Graduate School of Engineering, The University of Tokyo, Tokyo 113-8656

Published on the web February 13, 2009; doi:10.1246/bcsj.82.187

or damage may cause macroscopic changes, such as cell death or mutation in the biological system.[15] For example, an Auger study of bromine substituted in DNA indicated that a major part of the biological enhancement (e.g., irradiation sensitivity) was caused by inner-shell photoionization.[15] Quantitative treatment of NEXAFS spectra for comparatively large molecules, such as nucleosides and nucleotides, remains challenging.[4,14,16–20] These experiments need the support of highly accurate theoretical simulations in order to relate the spectra to their electronic structures and to the chemical environment of the excited/ionized atoms.[21] However, when accurate theoretical calculations are not available, the analyses in the inner-shell are sometimes estimated using simple models, which are either lacking a solid physical basis or are unable to reveal electron correlations and other effects. The ΔSCF approach has been also applied for the estimation of the inner-shell energies.[22,23] In addition, a simplified picture called building block principle,[10,24] in which a large molecule is seen as an assembly of similar pieces,[21] would lead to incorrect spectral assignment.[4] A cytidine ($C_9H_{13}N_3O_5$, 4-amino-1-β-D-ribofuranosyl-2(1*H*)-pyrimidinone) is produced when a cytosine ($C_4H_5N_3O$) combines with a ribose ($C_5H_{10}O_4$) to form a glycosidic linkage. The available experimental X-ray spectra of polycytidine[2,3] provide an excellent ground for the study of sugar–base correlation using updated theoretical techniques. In this work, we employ the LB94[25] model and a recently developed novel hybrid functional, CV-B3LYP[26] for core ionized states, to study cytidine (rC) and its fragments, the cytosine base and the deoxyribose (dR) moiety, in order to explore electronic structural information of sugar–base correlation from the core shells of the species.

Computational Details

Molecular geometries of rC, cytosine, and dR were optimized at the B3LYP[27]/6-311G**[28] level using the Gaussian03 computational chemistry package.[29] Figure 1 provides the structures of rC, cytosine, and dR, together with the conventional atomic labels of rC and atomic Hirshfeld charges (in parentheses). The orbital energies are calculated using DFT models with exchange-correlation potential, V_{xc}, being as CV-B3LYP and LB94. The single point CV-B3LYP/6-311G** calculations have been performed using the modified version of the GAMESS computational chemistry package,[30] whereas the LB94/TZ2P calculations are performed using the Amsterdam Density Functional (ADF) computational chemistry package.[31] Chong et al.[32] showed that the Kohn–Sham orbital energies can be interpreted as the vertical ionization potentials (VIP) of the species. Using this theoretical interpretation, some recent studies have found that the DFT LB94/TZ2P orbital energies obtained from a single electronic calculation were able to produce a good estimate of the vertical ionization energies in the core shell for molecules.[33,34]

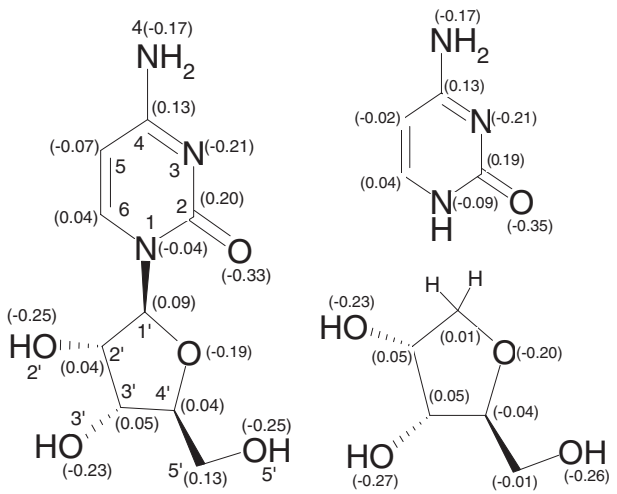

Figure 1. Molecular structures and atomic labels of cytidine and its fragments, cytosine and deoxyribose. The numbers in parentheses are Hirshfeld charges.

Based on the electronic structures of the species, distributions of charge densities according to the Hirshfeld scheme[35] using the LB94/TZ2P wave functions have also been calculated. Hirshfeld charge is a hypothetical "promolecule" with electron density $\sum\rho_B$ and is constructed by the superposition of spherically symmetric charge densities ρ_B of an isolated atom B. The Hirshfeld atomic charge Q_A^H is obtained by subtraction of the resulting partial electron density associated with atom A from the corresponding nuclear charge Z_A as[35]

$$Q_A^H = Z_A - \int \omega_A(\mathbf{r})\rho(\mathbf{r})d\mathbf{r} \quad (1)$$

where the weight function $\omega_A(\mathbf{r})$ is defined as

$$\omega_A(\mathbf{r}) = \frac{\rho_A(\mathbf{r})}{\sum_B \rho_B(\mathbf{r})} \quad (2)$$

The atomic Hirshfeld charge calculations are incorporated in the ADF program.

Analysis of the structure of the sugar pentose ring in dR and rC is assisted by the calculation of the sugar puckering amplitude (ϕ_{max}) and pseudorotational phase angle (ϕ_p) of the ring:

$$p = \frac{(\phi_4 - \phi_0) - (\phi_3 - \phi_1)}{2 * \phi_2 * (\sin(\pi/5) + \sin(2\pi/5))} \quad (3)$$

$$\phi_p = \begin{cases} \tan^{-1} p + \pi & if \quad \varphi_2 < 0 \\ \tan^{-1} p + 2\pi & if \quad p < 0 \\ \tan^{-1} p \end{cases} \quad (4)$$

$$\phi_{max} = \left| \frac{\phi_2}{\cos \phi_p} \right| \quad (5)$$

where ϕ_0–ϕ_4 are defined as follows: $\phi_0 = \angle C(4')OC(1')C(2')$, $\phi_1 = \angle OC(1')C(2')C(3')$, $\phi_2 = \angle C(1')C(2')C(3')C(4')$, $\phi_3 = \angle C(2')C(3')C(4')O$, and $\phi_4 = \angle C(3')C(4')OC(1')$. Calculations were performed using PROSIT,[34] an online tool to perform pseudorotational analysis for nucleosides and nucleotides.

Results and Discussion

Geometries and Atomic Hirshfeld Charges. The molecular geometries of rC and its fragments cytosine and dR optimized at the B3LYP/6-311G** level are listed in Table 1, together with their dipole moments. Experimental crystal structure of β-cytidine at 123 K observed using X-ray diffraction[36] is also listed in this table for reference as there are

Table 1. Geometric Parameters and Dipole Moments of Cytidine and Its Fragments Based on B3LYP/6-311G** Model

	Geometrical parameters	Fragment		Cytidine	Exptl[a]
		Cytosine	Deoxyribose		
Bond lengths/Å	N(1)–C(1′)			1.451	
Perimeter/Å	R_6[b]	8.299	—	8.290	8.239
	R_5[c]	—	7.492	7.499	7.456
Bond angle/°	∠C(1′)N(1)C(6)	—	—	120.48	121.9
Dihedral angle/°	∠HH′N(4)C(4)	−151.26	—	−154.29	
	∠C(4)C(2)N(3)O	−179.76	—	−178.64	−176.1
	∠C(4)C(5)C(6)N(1)	−0.07	—	0.08	1.2
	∠C(5)C(6)N(1)C(2)	0.04	—	−1.70	−0.5
	∠C(6)N(1)C(2)N(3)	0.30	—	2.43	−2.6
	∠C(2)N(1)C(1′)C(2′)	—	—	111.98	79.2
	∠N(1)C(1′)OC(4′)	—	—	−158.88	−113.1
	∠C(6)N(1)C(1′)C(2′)	—	—	−60.00	−100
	∠C(1′)C(2′)C(3′)C(4′)	—	−37.03	−31.94	38.4
	∠C(1′)OC(4′)C(3′)	—	−27.62	14.35	17.2
	ϕ_0 = ∠C(4′)OC(1′)C(2′)	—	−37.03	−35.32	
	ϕ_1 = ∠OC(1′)C(2′)C(3′)	—	39.63	41.44	
	ϕ_2 = ∠C(1′)C(2′)C(3′)C(4′)	—	−27.62	−31.94	
	ϕ_3 = ∠C(2′)C(3′)C(4′)O	—	7.14	12.49	
	ϕ_4 = ∠C(3′)C(4′)OC(1′)	—	18.60	14.35	
	ϕ_p[d]	—	133.97	141.35	
	ϕ_{max}[e]	—	39.78	40.90	
Dipole moment/D		6.28	2.71	5.72	

a) X-ray diffraction data of crystal structure of the nucleoside β-cytidine at 123 K. See Ref. 38. b) R_6 is the perimeter of the hexagon ring defined in Ref. 37. c) R_5 is the perimeter of the pentagon ring defined in Ref. 37. d) Pseudorotational phase angle ϕ_p estimated by eq 4. e) Puckering amplitude ϕ_{max} estimated by eq 5.

no gas-phase experimental results available at the moment. The agreement between the calculation and available experimental data are reasonably good. It is not a surprise that the bond lengths such as the perimeters agree very well with the derived bond lengths from the experiment: The perimeters of the hexagon and pentagon rings, R_6 and R_5,[38] respectively, exhibit negligible relaxation, compared with their cytosine and dR moieties. For example, R_6 and R_5 of rC are given as 8.290 and 7.499 Å, respectively, which exhibits a small difference of 0.009 and 0.007 Å from R_6 (8.299 Å) of cytosine and R_5 (7.492 Å) of dR. The length of the β-*N*-glycosidic bond (N(1)–C(1′)) is given as 1.451 Å, which is apparently longer than the N–C single bonds in rC, such as the N(4)–C(4) (1.363 Å) bond between the amino group (–NH_2) and pyrimidine ring, as well as the N(1)–C(6) bond of 1.360 Å and the N(3)–C(2) bond of 1.369 Å. However, this *N*-glycosidic bond is not significantly longer than its neighbor, the 1.448 Å N(1)–C(2)(=O) bond in the pyrimidine ring of rC, as given by the same B3LYP/6-311G** calculation. This bond in cytosine is given as 1.423 Å by the same model.

Apparent angular changes in rC are not observed in the aromatic ring but in the sugar pentagon ring. For example, the aromatic ring has been approximately confined in a plane in both rC and cytosine with negligible changes in dihedral angles, except for the pyramidization of the amino group (–NH_2) which lifts out of the pyrimidine plane with the dihedral angles of −154.29 and −151.26° for rC and cytosine, respectively. The pentagon sugar ring retains a similar degree of puckering in rC and in dR, with the pseudorotational angle slightly increasing in rC from 133.97 to 141.35°. When the aromatic base connects with the sugar ring in rC, the C(1′)–N(1)–C(6) angle is given by 120.48°. The dihedral angle formed by the base ring and sugar ring in rC is 111.98°. The differences in the dihedral angles between cytosine and dR are observed: The dihedral angle of C(6)N(1)C(1′)C(2′), which approximately defines the orientation between cytosine (base) and the sugar via the glycosidic bond N(1)–C(1′), is different in gas phase (calculations) and crystalline phase (experiment). This angle is flexible in gas phase with free rotation, whereas in crystal, this angle is restricted. As a result, a large difference between the calculated (−60°) and the measured (−100°) is observed. The dihedral angle of C(1′)OC(4′)C(3′) also exhibits a large difference, which may be attributed to the flexibility of the sugar ring in gas phase.

As indicated before,[39] the dipole moment of a DNA base is dominated by the relative location of the amino group (–NH_2). The dR ring connecting to the opposite site (N(1), see Figure 1) of the cytosine aromatic ring to form rC, contributes to balance the electron charge distributions from the amino moiety and therefore, the total dipole moment of rC (5.72 D) is reduced, compared to that of cytosine (6.28 D).

It is suggested that the Hirshfeld partition of electron density,[40,41] hence molecule fragmentation,[42] is one of the most appropriate molecular fragmentation procedures. For this reason, atomic site specific Hirshfeld charges, Q^H, of rC, cytosine, and dR, which are calculated using the LB94/TZ2P//B3LYP/6-311G** model, are also given in Figure 1 (numbers in parentheses). Atomic Hirshfeld charges ("heavy" atoms

Table 2. Core-Orbital Energies of Cytidine Calculated by B3LYP/6-311G**, LB94/TZ2P, and CV-B3LYP/6-311G** (in eV)

Spectrum	Site	B3LYP	LB94	CV-B3LYP	Exptl
O-K spectrum	O(3′)	520.97 (−13.2)	535.20 (1.6)	535.68 (1.5)	
	–O–	520.86 (−13.3)	535.84 (1.4)	535.57 (1.4)	534.2
	O(5′)	520.76 (−13.4)	535.60 (1.5)	535.47 (1.3)	
	O(2′)	520.26 (−13.9)	535.75 (1.0)	534.97 (0.8)	
	O=C(2)[c]	518.74 (−13.8)	533.48 (1.0)	533.46 (1.0)	532.5
N-K spectrum	N(1)	390.97 (−11.0)	404.75 (2.8)	403.69 (1.7)	402.0
	N(4)	390.11 (−11.9)	404.02 (2.0)	402.84 (0.8)	
	N(3)	388.73 (−11.8)	402.63 (2.1)	401.47 (1.0)	400.5
C-K spectrum	C(2)	279.79 (−10.2)	292.64 (2.6)	290.56 (0.6)	290.0
	C(4)	279.30 (−8.8)	292.39 (4.1)	290.08 (2.0)	288.1
	C(1′)	279.20 (−8.9)	292.17 (4.3)	289.96 (1.9)	
	C(3′)	278.71 (−7.3)	291.88 (5.4)	289.48 (3.5)	
	C(4′)	278.62 (−7.4)	291.83 (6.0)	289.38 (3.4)	
	C(6)	278.59 (−7.4)	292.00 (5.9)	289.37 (3.4)	286.0
	C(2′)	278.54 (−7.5)	291.36 (5.7)	289.30 (3.3)	
	C(5′)	278.51 (−7.5)	291.71 (5.8)	289.27 (3.3)	
	C(5)	277.06 (−7.6)	290.07 (5.4)	287.83 (3.1)	284.7
Mean discrepancy		(−10.3)	(3.4)	(2.0)	

a) The core-orbital energies are shown as absolute values of orbital energies. b) The experimental ionization potentials of O-K, N-K, and C-K in H_2O, NH_3, and CH_4 are 539.78, 405.56, and 290.91 eV in Ref. 43. c) The core-orbital energies correspond to oxygen in C=O.

only) indicate the atom-based contribution to the molecular charge distribution, which reveals fragment related electronic structural information. As seen from this figure, all the N and O sites in the three species exhibit negative Q^H, whereas the C sites exhibit either positive or negative Hirshfeld charges in the species, depending on the position and the role of the particular carbon site.

Apparent changes in Hirshfeld charges of rC with respect to its cytosine and dR fragments are localized along the *N*-glycosidic bond, N(1)–C(1′), in rC which connects the cytosine and dR moiety together, and the C(4′)–C(5′) chain of the dR moiety. The Hirshfeld charges on other corresponding "heavy" atoms in the aromatic rings show only negligible relaxation. No changes of signs are observed in the cytosine moiety when rC is formed. For example, Q^H of N(3) and N(4) (the amino N sites) remain the same in rC and cytosine, as −0.21 and −0.17, respectively. The most noticeable change in the aromatic ring happens at the C(5) site, from −0.02 in cytosine to −0.07 in rC. The connection site of N(1) exhibits a less negative Q^H of −0.04 in rC, compared to −0.09 at the same site (N(1)) in cytosine. The most significant change of Hirshfeld charges in rC is the C(4′)–C(5′) bond of the sugar branch, comparing to the same C–C bond of the dR fragment. The Q^H of C(4′) and C(5′) in the dR molecule are given by −0.04 and −0.01, respectively, which become positive in rC, that is, 0.04 and 0.13, respectively. Such significant changes in Q^H indicate the distinctly different roles of this C–C bond in the molecule pair.

Comparison of Theoretical Models for Core–Shell Spectra of Cytidine. This subsection compares three theoretical models, i.e., B3LYP/6-311G**, LB94/TZ2P, and CV-B3LYP/6-311G**, in evaluating the core-orbital energies of rC, which correspond to the vertical ionization potentials (IP). The relativistic effects and basis-set dependence are also examined, which are given in the Appendix. The number of core orbitals in rC is the same as its corresponding number of "heavy" atoms (non-hydrogen atoms): five O1s, three N1s, and nine C1s orbitals. Table 2 compares those core IPs of rC using the B3LYP/6-311G**, LB94/TZ2P, and CV-B3LYP/6-311G** models, as well as the experimental results of polycytidine (poly rC) homopolymer thin film whose details have been given in Ref. 4. The phosphate chain of poly rC film and screening effects may cause small IP differences between gas-phase calculations and experimental results but the differences seem small enough to make the assignments. In comparison between the experimental spectra[4] and theoretical ones, the energy bands are assigned. The theoretical spectra of Figure 2, whose peak positions are estimated by the core-orbital energies with the CV-B3LYP/6-311G** calculations, are constructed by Lorenzian-type functions with the half bandwidth of 0.4 eV. The intensities of all peaks are assumed to be the same in the simulation. Discussion of the assignment are given in the next subsection.

For the O-K shell, the oxygen sites in rC are energetically close so that the O-K spectrum of rC remains unresolved in experiment and could only be assigned to energies of 534.2 and 532.5 eV[4] based on the Gaussian–Lorenzian fit. The vertical IPs using both the CV-B3LYP/6-311G** and LB94/TZ2P models achieved excellent agreement with the measured: the discrepancies between theory and experiment are less than 1.6 eV. It is noted that the B3LYP/6-311G** model resulted in a significantly larger discrepancy of approximately 13 eV, which can be mainly attributed to self-interaction in the B3LYP model.[23,44]

In the N-K spectra of rC, the discrepancies using the LB94/TZ2P model become slightly larger, i.e., 2.0–2.8 eV than using

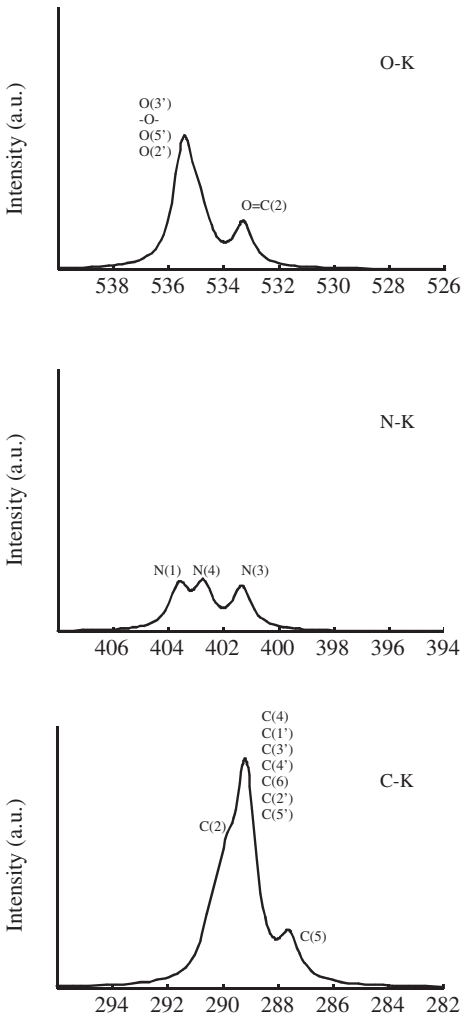

Figure 2. Theoretical binding energy spectra of O-K, N-K, and C-K of cytidine obtained from CV-B3LYP/6-311G** calculation (in eV).

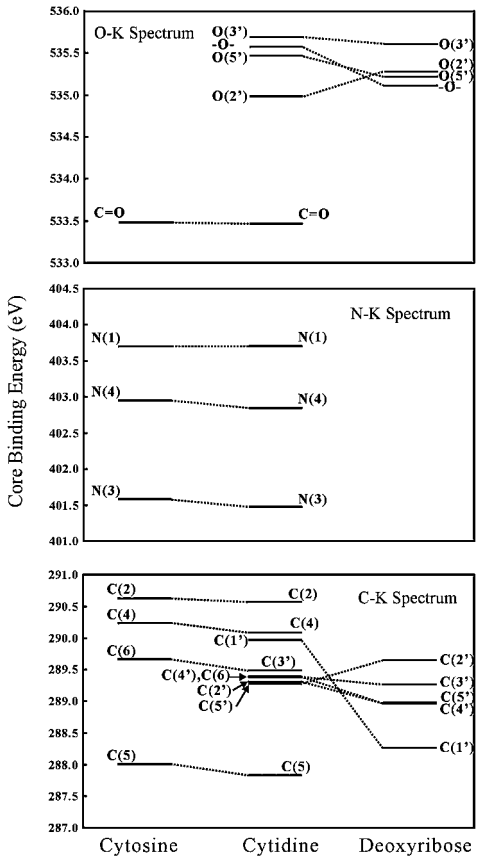

Figure 3. Correlation of core ionization energy diagram of cytidine, cytosine, and deoxyribose based on CV-B3LYP/6-311G** energies.

the CV-B3LYP/6-311G** model which results in an error of 0.8–1.7 eV. However, it is quite acceptable as the target molecule, rC, is quite large, and we are not aware of any better theoretical results than the present ones. For example, in the most recent theoretical and synchrotron-sourced XPS study for adenine, even the ab initio calculations using fourth-order algebraic diagrammatic construction ADC(4) method suffer 1.27 and 1.32 eV for C-K and N-K binding energy discrepancies, respectively.[14] This ADC(4) method is however, not applicable for larger molecules such as cytidine. The assignment for the N-K spectrum is somewhat controversial.[45] This point will be discussed in detail in the next subsection.

The C-K spectrum of rC is more congested due to the closely located energies of the C sites in rC. Amongst the nine C sites, only the energy of the C(5) site in rC, is separable from the other C sites under the experimental energy resolution of 0.7–2.0 eV,[4] and therefore, can be confidently assigned to 284.7 eV.[4] The remaining eight carbon sites of rC form an energy band with a peak position which coincides with the experimentally observed at 288.1 and 286.0 eV.[4] This feature has been revealed well in our simulated C-K spectrum in Figure 2.

The discrepancies of the C-K spectra between our models, LB94/TZ2P and CV-B3LYP/6-311G**, and the experiment are slightly larger than those of the O-K and N-K spectra. However, the opposite is true for the B3LYP/6-311G** model, that is, the B3LYP/6-311G** model performs better in the C-K spectra rather than the O-K and N-K spectra. Nevertheless, the largest discrepancies for LB94 (6.0 eV) and CV-B3LYP (3.5 eV) are still less than that for B3LYP (−10.2 eV). Thus, LB94/TZ2P and CV-B3LYP/6-311G** models still produce the C-K spectra of rC closer to the experimental ones.

The overall mean discrepancies for LB94/TZ2P and CV-B3LYP/6-311G** are 3.4 and 2.0 eV, respectively, which are significantly lower than −10.3 eV for B3LYP. However, what is important in the present study is the inner-shell binding energy shifts, $\Delta\varepsilon$, rather than the energies themselves. It is noted that the models we used ignore a number of effects such as orbital relaxation and relativistic effects. However, it is no doubt that the LB94/TZ2P and CV-B3LYP/6-311G** models can offer more reliable description and enable us to discuss chemical shifts of molecules as large as rC with relatively simple calculations which achieve acceptable accuracy.

Sugar–Base Correlation between Cytidine and Its Fragments. The correlation between rC and the fragments in core–shell binding energies is shown in energy diagrams in Figure 3, which are based on the CV-B3LYP/6-311G** calculations. The O-K shell of rC is dominated by the contributions from the

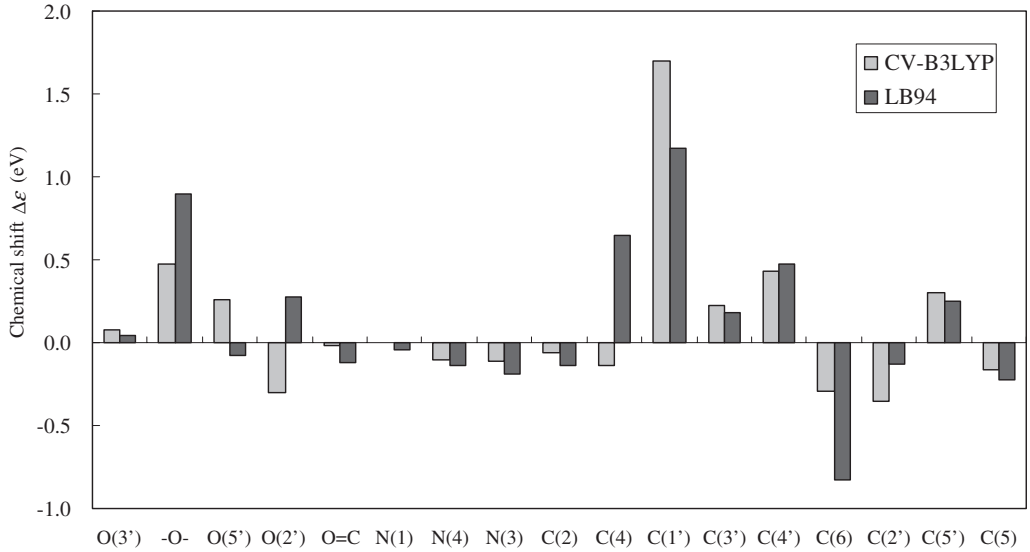

Figure 4. Site-based chemical shifts ($\Delta\varepsilon$ in eV) of rC with respect to the fragments cytosine and dR. The calculations are based on the CV-B3LYP/6-311G** and LB94/TZ2P models, respectively.

O sites of the dR fragment, whereas the N-K shell of rC is solely contributed by the N sites of cytosine species alone. The C-K shell of rC, however, is more complex as it receives contributions from both fragments of cytosine and dR. In general, core–shell binding energies of cytosine are red-shifted (down) in rC whereas those of dR are blue-shifted (up) when rC is formed.

In the O-K shell, there is a clear energy gap of approximate 1.5 eV between the lowest peak of keto C=O oxygen (cytosine) and the rest of single-bonded O sites of the dR fragment. The two peaks in our theoretically simulated O-K spectrum of rC given in Figure 2 reflect this observation. The single-bonded O sites in rC, which are contributed from the pentagon sugar ring, belong to the larger peak in the simulated O-K spectrum.

The N-K spectrum of rC provides an excellent case to study chemical shift in the inner shell. Considerably small blue shift (<0.11 eV) exhibits in the IP of the N sites of rC from cytosine. As shown in Figure 3, the N-K shell of rC is solely contributed from the N sites of the cytosine fragment, and the small shift in energy is due to the interactions caused by the attachment of dR.[2,35] The three N sites of rC are well separated by 1.37 and 0.85 eV for N(3)–N(4) and N(4)–N(1) splitting, respectively, which is similar to the cytosine fragment.[13] The N-sites in rC can be grouped by their chemical bonding information, i.e., saturated (amino) and unsaturated (imino) N-bonds, rather than by their positions of being in and out of the pyrimidine ring.[2,45] Theoretically, there are three peaks in the simulated binding energy spectrum of N1s spectrum as seen in Figure 2, being assigned as N(1), N(4) (amino group), and N(3) sites from left to right. Due to the large energy resolution, the experimentally resolved spectra (Table 2) with two peaks can be assigned to C–N–C and $–NH_2$ (saturated N sites) for the left peak and to C–N=C (unsaturated N site) to the right peak in the spectrum of rC. Figures 2 and 3 have clearly indicated such information in the N-K shell.

The C-K shell contains the most interesting and important information of the species. It is more complex as it receives contributions from both pyrimidine and the ribose rings with a closer energy splitting, which causes difficulties in experimental analysis.[4] The C sites, C(5), C(6), C(4), and C(2) from cytosine are well separated in energy (left column in Figure 3) and the order remains unchanged in rC, whereas the C sites from dR, i.e., C(5′), C(2′), C(4′), C(3′), and C(1′) are much closer in energy than their cytosine counterpart. This observation indicates that the chemical environment of the C sites are quite different in the aromatic pyrimidine ring due to unsaturated bonds associated with N and O atoms, whereas the C sites in the sugar ring are quite similar because of the single bonds in the sugar ring as indicated in Figure 3.

All the C sites of the sugar moiety shift up in rC except for C(2′). Note that due to the degrees of electron correlation included in the models, there are a couple of energy crossings in the C sites. A significant change in energy and position of the C(1′) site is observed. The binding energy of C(1′) shifts approximately 1.7 eV up in the rC spectrum, from the lowest energy in dR to the third highest C site (just under the C(4) site) in rC. The formation of the *N*-glycosidic bond of C(1′)–N(1) in rC causes such an energy change in C(1′) and stablizes this site.

As a result of the formation of the *N*-glycosidic bond, the C-K spectrum of rC consists of closely located energy levels with no greater than 0.9 eV, except C(2) and C(5) (refer to Table 2). The higher shoulder peak of the C-K spectrum (Figure 2) at 290.0 eV[4] is assigned to the carbonyl carbon site, i.e., C(2) (C(2)=O), whereas the lower energy peak of this spectrum at 284.7 eV[4] is assigned to C(5) (C–C(5)=C), which is the only C site in the aromatic ring possessing a negative Hirshfeld charge. The remaining seven C sites in rC, which are associated with either C–N or C–O bonds, contribute to the extended higher energy peak at 288.1 and 286.0 eV.[4]

Site Specific Chemical Shift in Cytidine. The site-based chemical shifts of rC with respect to its component cytosine and dR, which are generated using the CV-B3LYP/6-311G** and the LB94/TZ2P models, are compared in Figure 4. Several interesting issues are observed in this figure. First, as discussed

in the previous subsection, in general the sites correlated to the sugar ring exhibit blue shifts (positive $\Delta\varepsilon$), whereas the sites that stem from the cytosine base display red shifts (negative $\Delta\varepsilon$). However, although energy shifted, all such sites do not exhibit significant energy changes except for the C(1′) site, since the base–sugar interactions are dominated by direct interactions through the glycosidic bond, N(1)–C(1′) so that the C(1′) site energy changes accordingly.

Second, the nitrogen N-K sites are reported to be the sites with the least energy shift among the O-K, N-K, and C-K sites of rC. It is possible as the N-K sites are either in the aromatic ring, N(1) and N(3), or at the farthest position with respect to the sugar moiety. Third, the site-based chemical shifts of rC are consistently reproduced by two different DFT models, except for the C(6), O(2′), and O(5′), at which the CV-B3LYP/6-311G** and the LB94/TZ2P models predict opposite shift directions. Further confirmation for the C(6), O(2′), and O(5′) shifts in rC requires either higher level of theoretical calculations or high-resolution core ionization spectroscopy.

Finally, both maximum and minimum energy shifts for the formation of rC happen at the β-*N*(1)-glycosidic C–N linkage sites of C(1′) and N(1), respectively, as shown by both of the DFT calculations. The significant differences in energy shifts at the C(1′) and N(1) sites indicate substantially different roles of the sites in this C–N bond. The fact that N(1) possesses the least energy changes suggests that the aromatic buffer of the cytosine ring diffuses the energy changes of the site, whereas the singly bonded sugar ring does not have such an aromatic buffer to diffuse the energy changes at C(1′), causing a large energy deposit at this site.

Conclusion

The sugar–base correlation of cytosine and dR moieties of cytidine in the inner-shell was investigated with respect to the ionization energies generated by DFT models. The results reveal that the aromatic fragment (cytosine) acts as a buffer to resist the changes and hereby to stabilize the nucleoside. This is revealed clearly from the small energy shift in the N-K spectra of rC, which entirely from contributions of the cytosine moiety. The dR moiety is more flexible in structure than the base. The singly bonded five member ring can be easily changed through puckering without breaking its bonds. When bonding with a base (cytosine), the carbon sites, in particular the C(1′) site of dR, exhibit blue shifts in order to accommodate such local changes in creating the *N*-glycosidic linkage. As a result, the inner-shell contains important and sensitive site related information with respect to the sugar–base correlation in rC. The present work provides a clear chemical picture of quantum mechanical fragments in molecules, which can be a useful tool for studying base–sugar interaction/correlation and to help the interpretation of experimental results in these challenging areas. The present study will also shed light on ligand-based quantum mechanical drug design, as cytidine and its derivatives are viewed as drug-like ligands.[3]

In order to simulate experimental XAS, XES, and X-ray photoemission spectroscopy (XPS) of rC, based on time-dependent density functional theory calculations of the CVR-B3LYP[26,46] and other models[47–50] are currently under investigation.

Table 3. Relativistic Effects of Core-Orbital Energies of Cytidine Calculated by CV-B3LYP with 6-311G**[a)]

Spectrum	Site	Non-relativistic	Relativistic
O-K spectrum	O(3′)	535.68	535.87
	–O–	535.57	535.76
	O(5′)	535.47	535.67
	O(2′)	534.97	535.16
	O=C(2)[b)]	533.46	533.65
N-K spectrum	N(1)	403.69	403.80
	N(4)	402.84	402.95
	N(3)	401.47	401.57
C-K spectrum	C(2)	290.56	290.61
	C(4)	290.08	290.13
	C(1′)	289.96	290.01
	C(3′)	289.48	289.53
	C(4′)	289.38	289.43
	C(6)	289.37	289.42
	C(2′)	289.30	289.35
	C(5′)	289.27	289.32
	C(5)	287.83	287.88

a) The core-orbital energies are shown as absolute values of orbital energies. b) The core orbital energies correspond to oxygen in C=O.

Appendix

Relativistic Effect and Basis-Set Dependence. Relativistic effects and basis set dependence of the orbital energies have been estimated using the CV-B3LYP model. Table 3 lists non-relativistic and relativistic orbital energies calculated with the 6-311G** basis set. The relativistic effects are estimated by eq 2 in Ref. 51. The corrections are estimated to be approximately 0.19, 0.10, and 0.05 eV for O, N, and C, respectively. The addition of the relativistic effects increases absolute values of orbital energies. Thus, the relativistic corrections slightly enhance the deviations from the experimental values since the non-relativistic orbital energies are already larger than the experimental values. However, the relativistic effects scarcely affect the main tendency of ionization potentials.

Table 4 lists orbital energies computed by 6-311G**, cc-pVTZ, cc-pCVTZ, and primitive basis functions of cc-pCVTZ (p-cc-pCVTZ). The energy differences from the energies produced using the 6-311G** basis set (shown in parentheses) are also small and will not change the trends and conclusions made from the CV-B3LYP/6-311G** model.

This work was supported by an award under the Merit Allocation Scheme of the APAC National Facility at the ANU. F.W. is grateful for the Vice-Chancellor's Strategic Research Initiative Grant of Swinburne University of Technology (SUT). A.T. thanks the Faculty of Information and Communication Technologies of SUT for a summer internship scholarship. S.S. is grateful for the SUT Postgraduate Research Award (SUPRA). H.N. is grateful for a Grant-in-Aid for Scientific Research on Priority Areas "Molecular Theory for Real Systems" "KAKENHI 18066016" from the Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan; Nanoscience Program in the Next Generation Super Computing Project of MEXT; Global Center Of

Table 4. Basis-Set Dependence of Core-Orbital Energies (OE) of Cytidine Calculated by CV-B3LYP[a)]

Spectrum	Site	6-311G**	cc-pVTZ	cc-pCVTZ	p-cc-pCVTZ
O-K spectrum	O(3′)	535.68	535.89 (0.22)	535.84 (0.17)	535.84 (0.17)
	–O–	535.57	535.77 (0.20)	535.71 (0.14)	535.71 (0.14)
	O(5′)	535.47	535.74 (0.27)	535.69 (0.22)	535.69 (0.22)
	O(2′)	534.97	535.23 (0.26)	535.18 (0.22)	535.18 (0.22)
	O=C(2)[b)]	533.46	533.67 (0.21)	533.62 (0.16)	533.62 (0.16)
N-K spectrum	N(1)	403.69	403.78 (0.09)	403.74 (0.05)	403.74 (0.05)
	N(4)	402.84	402.97 (0.13)	402.94 (0.10)	402.95 (0.11)
	N(3)	401.47	401.63 (0.16)	401.59 (0.12)	401.59 (0.12)
C-K spectrum	C(2)	290.56	290.63 (0.07)	290.60 (0.04)	290.61 (0.05)
	C(4)	290.08	290.15 (0.07)	290.12 (0.05)	290.13 (0.05)
	C(1′)	289.96	290.00 (0.04)	289.98 (0.02)	289.99 (0.03)
	C(3′)	289.48	289.52 (0.05)	289.50 (0.03)	289.51 (0.03)
	C(4′)	289.38	289.42 (0.04)	289.40 (0.02)	289.41 (0.03)
	C(6)	289.37	289.42 (0.05)	289.40 (0.03)	289.40 (0.03)
	C(2′)	289.30	289.35 (0.05)	289.33 (0.03)	289.33 (0.03)
	C(5′)	289.27	289.32 (0.05)	289.31 (0.03)	289.31 (0.04)
	C(5)	287.83	287.88 (0.06)	287.86 (0.04)	287.86 (0.04)

a) The core-orbital energies are shown as absolute values of orbital energies. b) The core orbital energies correspond to oxygen in C=O.

Excellence (COE) "Practical Chemical Wisdom" from MEXT, Japan. Y.I. is grateful for a Grant-in-Aid for Young Scientists (Start-up) "KAKENHI 19850027" from Japanese Society for the Promotion of Science (JSPS), MEXT.

References

1 C. Farès, T. Carlomagno, *J. Am. Chem. Soc.* **2006**, *128*, 9856.

2 Y. Harada, T. Takeuchi, H. Kino, A. Fukushima, K. Takakura, K. Hieda, A. Nakao, S. Shin, H. Fukuyama, *J. Phys. Chem. A* **2006**, *110*, 13227.

3 C. J. Torrance, V. Agrawal, B. Vogelstein, K. W. Kinzler, *Nat. Biotechnol.* **2001**, *19*, 940.

4 J. Magulick, M. M. Beerbom, B. Lägel, R. Schlaf, *J. Phys. Chem. B* **2006**, *110*, 2692.

5 R. J. Levis, G. M. Menkir, H. Rabitz, *Science* **2001**, *292*, 709.

6 S. Svensson, *J. Phys. B: At., Mol. Opt. Phys.* **2005**, *38*, S821.

7 T. Karlsen, K. J. Børve, L. J. Sæhre, K. Wiesner, M. Bässler, S. Svensson, *J. Am. Chem. Soc.* **2002**, *124*, 7866.

8 F. Wang, *THEOCHEM* **2005**, *728*, 31.

9 K. Siegbahn, C. Nordling, G. Johansson, J. Hedman, P. F. Hedén, K. Hamrin, U. Gelius, T. Bergmark, L. O. Werme, R. Manne, Y. Baer, *ESCA Applied to Free Molecules*, North-Holland Pub. Co., Amsterdam-London, **1969**.

10 J. Stohr, *NEXAFS Spectroscopy*, Springer-Verlag, Berlin, **1992**.

11 C. T. Falzon, F. Wang, W. Pang, *J. Phys. Chem. B* **2006**, *110*, 9713.

12 M. Yoshiya, I. Tanaka, K. Kaneko, H. Adachi, *J. Phys.: Condens. Matter* **1999**, *11*, 3217.

13 J. MacNaughton, A. Moewes, E. Z. Kurmaev, *J. Phys. Chem. B* **2005**, *109*, 7749.

14 O. Plekan, V. Feyer, R. Richter, M. Coreno, M. de Simone, K. C. Prince, A. B. Trofimov, E. V. Gromov, I. L. Zaytseva, J. Schirmer, *Chem. Phys.* **2008**, *347*, 360.

15 N. Usami, K. Kobayashi, H. Maezawa, K. Hieda, S. Ishizaka, *Int. J. Radiat. Biol.* **1991**, *60*, 757.

16 S. M. Kirtley, O. C. Mullins, J. Chen, J. van Elp, S. J. George, C. T. Chen, T. O'Hhalloran, S. P. Cramer, *Biochim. Biophys. Acta* **1992**, *1132*, 249.

17 J. J. Rehr, A. L. Ankudinov, *J. Synchrotron Radiat.* **2001**, *8*, 61.

18 O. Takahashi, K. Saito, M. Mitani, H. Yoshida, F. Tahara, T. Sunami, K. Waki, Y. Senba, A. Hiraya, L. G. M. Pettersson, *J. Electron Spectrosc. Relat. Phenom.* **2005**, *142*, 113.

19 M. Nyberg, M. Odelius, A. Nilsson, L. G. M. Pettersson, *J. Chem. Phys.* **2003**, *119*, 12577.

20 H. Öström, L. Triguero, K. Weiss, H. Ogasawara, M. G. Garnier, D. Nordlund, M. Nyberg, L. G. M. Pettersson, A. Nilsson, *J. Chem. Phys.* **2003**, *118*, 3782.

21 V. Carravetta, O. Plashkevych, H. Ågren, *J. Chem. Phys.* **1998**, *109*, 1456.

22 O. Takahashi, L. G. M. Pettersson, *J. Chem. Phys.* **2004**, *121*, 10339.

23 Y. Imamura, H. Nakai, *Int. J. Quantum Chem.* **2007**, *107*, 23.

24 L. G. M. Pettersson, H. Ågren, B. L. Schürmann, A. Lippitz, W. E. S. Unger, *Int. J. Qunatum Chem.* **1997**, *63*, 749.

25 R. van Leeuwen, E. J. Baerends, *Phys. Rev. A: At., Mol., Opt. Phys.* **1994**, *49*, 2421.

26 A. Nakata, Y. Imamura, T. Otsuka, H. Nakai, *J. Chem. Phys.* **2006**, *124*, 094105.

27 A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648.

28 R. Krishnan, J. S. Binkley, R. Seeger, J. A. Pople, *J. Chem. Phys.* **1980**, *72*, 650.

29 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao,

H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, J. A. Pople, *Gaussian 03 Revision C.02*, Gaussian, Inc., Wallingford CT, **2004**.

30 M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis, J. A. Montgomery, Jr., *J. Comput. Chem.* **1993**, *14*, 1347.

31 *ADF2005.01*, SCM, Theoretical Chemistry, Vrije Universiteit, Amsterdam, **2005**.

32 D. P. Chong, O. V. Gritsenko, E. J. Baerends, *J. Chem. Phys.* **2002**, *116*, 1760.

33 R. R. da Silva, T. C. Ramalho, J. M. Santos, J. D. Figueroa-Villar, *J. Phys. Chem. A* **2006**, *110*, 1031.

34 S. Saha, F. Wang, J. B. MacNaughton, A. Moewes, D. P. Chong, *J. Synchrotron Radiat.* **2008**, *15*, 151.

35 F. L. Hirshfeld, *Theor. Chim. Acta* **1977**, *44*, 129.

36 L. Chen, B. M. Craven, *Acta Crystallogr., Sect. B* **1995**, *51*, 1081.

37 G. Sun, J. H. Voigt, V. E. Marquez, M. C. Nicklaus, *Nucleosides, Nucleotides Nucleic Acids* **2005**, *24*, 1029.

38 F. Wang, M. T. Downton, N. Kidwani, *J. Theor. Comput. Chem.* **2005**, *4*, 247.

39 S. Saha, F. Wang, C. F. Guerra, F. M. Bickelhaupt, *J. Comput. Meth. Sci. Eng.* **2006**, *6*, 251.

40 F. De Proft, C. Van Alsenoy, A. Peeters, W. Langenaeker, P. Geerlings, *J. Comput. Chem.* **2002**, *23*, 1198.

41 F. De Proft, R. Vivas-Reyes, A. Peeters, C. Van Alsenoy, P. Geerlings, *J. Comput. Chem.* **2003**, *24*, 463.

42 R. G. Parr, P. W. Ayers, R. F. Nalewajski, *J. Phys. Chem. A* **2005**, *109*, 3957.

43 A. A. Bakke, H.-W. Chen, W. L. Jolly, *J. Electron Spectrosc. Relat. Phenom.* **1980**, *20*, 333.

44 G. Tu, V. Carravetta, O. Vahtras, H. Ågren, *J. Chem. Phys.* **2007**, *127*, 174110.

45 F. Wang, *Micro Nano Lett.* **2006**, *1*, 23.

46 A. Nakata, Y. Imamura, H. Nakai, *J. Chem. Phys.* **2006**, *125*, 064109.

47 Y. Imamura, H. Nakai, *Chem. Phys. Lett.* **2006**, *419*, 297.

48 Y. Imamura, T. Ostuka, H. Nakai, *J. Comput. Chem.* **2007**, *28*, 2067.

49 A. Nakata, Y. Imamura, H. Nakai, *J. Chem. Theor. Comput.* **2007**, *3*, 1295.

50 T. Tsuchimochi, M. Kobayashi, A. Nakata, Y. Imamura, H. Nakai, *J. Comput. Chem.* **2008**, *29*, 2311.

51 D. P. Chong, *J. Chem. Phys.* **1995**, *103*, 1842.