Journal of
Synchrotron
Radiation

ISSN 0909-0495

Received 3 June 2008
Accepted 28 July 2008

# Inner-shell chemical shift of DNA/RNA bases and inheritance from their parent purine and pyrimidine

**Feng Wang,[a]* Quan Zhu[a,b] and Elena P. Ivanova[c]**

[a]Center for Molecular Simulation, Faculty of Information and Communication Technologies, Swinburne University of Technology, Hawthorn, Melbourne, Victoria 3122, Australia, [b]College of Chemical Engineering and State Key Laboratory of Biotherapy, Sichuan University, Chengdu 610065, People's Republic of China, and [c]Environment and Biotechnology Centre, Faculty of Life and Social Sciences, Swinburne University of Technology, Hawthorn, Melbourne, Victoria 3122, Australia. E-mail: fwang@swin.edu.au

Inner-shell electronic structures, properties and ionization spectra of DNA/RNA bases are studied with respect to their parent pyrimidine and purine species. Density functional theory B3LYP/aug-cc-pVTZ has been employed to produce the geometries of the bases, whereas LB94/et-pVQZ//B3LYP/aug-cc-pVTZ is used to calculate site-related Hirshfeld charges and core (vertical) ionization energies, as well as inner-shell spectra of C1*s*, N1*s* and O1*s* for DNA/RNA bases and their parent pyrimidine and purine species. The site-dependent variations of properties indicate the changes and inheritance of chemical environment when pyrimidine and purine become substituted. In general, although the changes are site-dependent, they are also ring-dependent. Pyrimidine bases change less significantly with respect to their parent pyrimidine than the purine bases with respect to their parent purine. Pyrimidine bases such as uracil, thymine and cytosine inherit certain properties from their parent pyrimidine, such as the Hirshfeld charge distributions and the order of core ionization energy level *etc*. No particular sites in the pyrimidine derivatives are engaged with a dramatic chemical shift nor with energy crossings to other sites. For the core shell spectra, the purine bases inherit very little from their parent purine, and guanine exhibits the least similarities to the parent among all the DNA/RNA bases.



**Keywords: pyrimidine bases; purine bases; inner-shell ionization spectra; chemical shift; property and spectral inheritance.**

## 1. Introduction

When a photoabsorption process occurs in biological systems such as living cells, it may cause certain (photobiological) effects on the system. Induced molecular changes are recognized and modified by various enzymes in the intracellular environment. The repairable and/or unrepairable changes, or damages, may cause macroscopic changes such as cell death or mutation in the biological systems. Initiating processes are clearly the same as those studied in atomic physics or photochemistry (Usami *et al.*, 1991). The intrinsic properties of the model molecules, which are usually hidden in the complex medium of a real biological system, can be understood in an isolated environment in the gas phase. The demands for gas phase information arise from the anticipation that many biological phenomena can be traced to the fundamental properties of molecular constituents. Experiments in the gas phase lead to large amounts of data providing insight into the physicochemical origin of properties of biological molecules (Sponer *et al.*, 2001; Wang, 2005; Segala *et al.*, 2006; Saha *et al.*, 2008; Plekan *et al.*, 2008). Both laser spectroscopy and computational modelling have significantly contributed to elucidating the structures and dynamics of these biomolecules and their solvated complexes in the gas phase (Wang, 2005; Fujii *et al.*, 2003, 2004; Li *et al.*, 2007; Wang *et al.*, 2005; Saha *et al.*, 2008; Plekan *et al.*, 2008).

As the hydrogen-bonded purine (Pu) and pyrimidine (Py) bases store genetic code in DNA and RNA, investigation of the nucleic acid bases is significant in life science when considering radiation damage to DNA/RNA (Shukla & Leszczynski, 2007; Callis, 1983). Novel experimental and theoretical studies of the DNA/RNA bases still envisage providing detailed information towards our understanding of the molecular structures and therefore properties and functions. Nucleic acid bases adenine (A), guanine (G), cytosine (C), thymine (T) and uracil (U) represent some of the most important building blocks of life. A large fraction of biologically relevant damage in living cells can be traced to structural

and chemical modifications of cellular DNA and RNA. Studies have revealed that radiation can induce damage in living cells and create exceeding amounts of low-energy electrons along the ionization track (Sanche, 2002). Electrons with energies of the order of tens of eV can immediately induce reactions, both directly in DNA/RNA and indirectly through the environment (*e.g.* medium).

It has been demonstrated for a large range of energetic ions that a striking correlation exists between the inactivation of various cells and inner-shell ionization cross section in carbon, nitrogen and oxygen atoms of the DNA/RNA molecules (Chetioui *et al.*, 1994). The relative biological efficiency of ultra-soft X-rays to induce cellular inactivation (Herve du Penhoat *et al.*, 1999; Fayard *et al.*, 2002), exchange type chromosomal aberrations (Gobert *et al.*, 2004; Griffin *et al.*, 1998) and double strand break (DSB) induction (Yokoya *et al.*, 1999) was found to be strongly correlated to the number of core events and, moreover, the DSB repair data indicate a significant decrease in the repairability of DSB when it is triggered by core ionization. In addition, such core events usually exhibit approximate atom-localized electronic states, so that the electronic structure about specific elements can be studied even in the presence of many other elements. It is well known that binding energies and orbital wavefunctions of electrons are element specific in the core region. The core-electron binding-energy shift with chemical environment contained in the orbital wavefunctions makes it possible to differentiate atoms of the same element located at the inequivalent sites in a molecule, giving a simpler and more novel picture of the molecular structure on a specific element than the largely delocalized valence orbitals of small energy splitting (Wang, 2005; Alagia *et al.*, 2005).

Recently, in a study to explore the impact to the inner shell of the amino group ($-NH_2$) attachment in adenine with respect to the parent Pu (Saha *et al.*, 2008), it was anticipated that the N1*s* ionization spectra between the molecule pair would exhibit significant differences owing to the extra N atom in adenine, whereas the C1*s* spectra were expected to be very similar owing to almost the same carbon frames (Saha *et al.*, 2008). The results, however, are surprising: similarities are observed for the N1*s* spectra of A and Pu, whereas the C1*s* spectra of the molecule pair are significantly different. Such a finding from core-shell structures of A and Pu leads to questions, *i.e.* is this trend also true for all Pu bases including G? What changes would the inner-shell structures and spectra of the Py bases T, C and U have?

In the present study, inner-shell chemical shifts of all Pu and Py bases are investigated against the unsubstituted Pu and Py to reveal the impact of modifications to the parent purine and pyrimidine in the C1*s*, N1*s* and O1*s* spectra in the gas phase. The simulated X-ray photoelectron spectra (XPS) of T and A are compared favourably with the most recently available experiment using synchrotron-sourced XPS (Plekan *et al.*, 2008). As a result, the model has been further applied to study inner-shell spectra and their electronic structures of other Pu and Py bases to reveal the trends of chemical shifts in the inner shell with respect to their corresponding parent Pu and Py.

## 2. Computational details

All geometry optimizations of the DNA/RNA bases are carried out using the B3LYP/aug-cc-pVTZ model incorporated in the *Gaussian03* computational chemistry package (Frisch *et al.*, 2004), followed by harmonic vibrational analysis. The obtained geometries of the bases exhibit minimum energy structures without any virtual frequencies. Single-point calculations using density function theory (DFT) based on the LB94/et-pVQZ model (van Leeuwen & Baerends, 1994; Chong *et al.*, 2004) are then employed to produce the core orbital energies. The core-shell ionization potential (IP) energies are calculated using the 'meta-Koopman theorem' without further modification and scaling, which is similar to the LB94/TZ2P model used previously (Saha *et al.*, 2008). The DFT single-point calculations are performed using the *Amsterdam Density Functional* (ADF) computational chemistry package (Baerends *et al.*, 2006). A number of molecular properties such as Hirshfeld charges (Hirshfeld, 1977) and dipole moments are calculated using the same model as the IP energies.

## 3. Results and discussion

The ground electronic states of the Pu and Py bases, A, G, T, C, U, and their parents Pu and Py, are all in closed shells with doubly occupied orbitals. While Pu, Py, U and T exhibit $C_s$ point group symmetry, the amino group ($-NH_2$) of A, G and C is reported to be pyramidal owing to the partial $sp^3$ pyramidalization of the amino nitrogen (Leszczynski, 1992; Hobza & Sponer, 1999; Fonseca-Guerra *et al.*, 2006; Jones *et al.*, 2006; Saha *et al.*, 2008). The non-planarity of A was observed by the vibrational transition moment direction measurement (Dong & Miller, 2002). However, a previous study indicated that the amino group non-planarity of A does not cause the apparent binding ionization energy shift but leads to distortions of a group of related orbital wavefunctions (Downton & Wang, 2006).

The chemical structures and numbering of the five DNA/RNA bases and their parents Pu and Py are given in Fig. 1. The DNA/RNA bases can be considered as the derivatives of unsubstituted Pu and Py. For example, U can be considered as the product when the pair of C—H bonds at positions C(2) and C(4) of Py are reproduced by a pair of C═O bonds. T is considered as the product when the C(5)—H bond of U is replaced by the C(5)—C(5′)$H_3$ fragment, whereas cytosine is produced when the C(4)═O bond of U is replaced by a C(4)—N(4′)$H_2$ fragment. Minimal geometric relaxations are obtained among the base derivatives with respect to their unsubstituted Pu and Py, accordingly. For example, the hexagon and pentagon ring perimeters, $R_6$ and $R_5$ (summations of all bonds constituting the rings) (Wang *et al.*, 2005), which are given in Fig. 1 with the structures of the bases, exhibit only small variations from Pu and Py, depending on the numbers and positions of the double bonds. For example, the $R_6$ (8.29 Å) of G extends 0.16 Å from the $R_6$ (8.13 Å) of Pu owing to the saturation of the N(1)═C(6) double bond in G.

**Table 1**
Comparison of calculated and experimental core IPs for thymine (in eV).

| Site | Mean† | LB94/et-pVQZ | CEBE/TZP‡ | ADC(4)/6-31G§ | Expt¶ | ΔIP†† |
|---|---|---|---|---|---|---|
| C(5′) | 291.50 | 289.94 | 291.18 | 290.95 (+1.60) | 291.00 | −1.06 |
| C(5) | 292.10 | 290.60 | 290.95 | 290.67 (+1.60) | 291.00 | −0.40 |
| C(6) | 293.68 | 291.70 | 292.27 | 292.29 (+1.60) | 292.30 | −0.60 |
| C(4) | 294.94 | 292.87 | 293.79 | 294.19 (+1.60) | 294.20 | −1.33 |
| C(2) | 295.88 | 293.74 | 294.75 | 295.36 (+1.60) | 295.20 | −1.46 |
| N(3) | 408.26 | 404.87 | 406.56 | 406.52 (+1.20) | 406.70 | −1.83 |
| N(1) | 408.62 | 405.25 | 406.91 | 406.80 (+1.20) | 406.70 | −1.45 |
| O(4′) | 539.58 | 534.35 | 537.25 | 537.24 (+1.00) | 537.30 | −2.95 |
| O(2′) | 539.61 | 534.47 | 537.45 | 537.46 (+1.00) | 537.30 | −2.83 |

† Simple mean of IPs based on the models of HF/aug-cc-pVTZ and B3LYP/aug-cc-pVTZ. ‡ CEBE from Chong *et al.* (Takahata *et al.*, 2006). § The IPs based on ADC(4)/6-31G have been shifted by −1.60, −1.20 and −1.00 eV for C1*s*, N1*s* and O1*s*, respectively (Plekan *et al.*, 2008). ¶ Experimental data in the gas phase (Plekan *et al.*, 2008). †† Derivation of the IPs using the LB94/et-pVQZ model from the experimental data in the gas phase.

**Table 2**
Comparison of calculated and experimental core IPs for adenine (in eV).

| Site | Mean† | LB94/et-pVQZ | CEBE‡ | ADC(4)/6-31G§ | LB94/TZ2P¶ | Expt†† | ΔIP‡‡ |
|---|---|---|---|---|---|---|---|
| C(5) | 291.93 | 290.42 | 290.96 | 291.19 (+1.27) | 290.57 | 291.00 | −0.58 |
| C(2) | 293.16 | 291.08 | 291.66 | 291.94 (+1.27) | 291.23 | 292.50 | −1.42 |
| C(4) | 293.43 | 291.50 | 292.15 | 292.47 (+1.27) | 291.66 | 292.50 | −1.00 |
| C(8) | 293.60 | 291.63 | 292.13 | 292.66 (+1.27) | 291.77 | 292.50 | −0.87 |
| C(6) | 293.95 | 291.89 | 292.64 | 292.96 (+1.27) | 292.05 | 292.50 | −0.61 |
| N(1) | 406.26 | 402.83 | 404.23 | 404.27 (+1.32) | 402.87 | 404.40 | −1.57 |
| N(3) | 406.41 | 402.97 | 404.37 | 404.40 (+1.32) | 403.03 | 404.40 | −1.43 |
| N(7) | 406.71 | 403.20 | 404.78 | 404.99 (+1.32) | 403.25 | 404.40 | −1.20 |
| N(6′) | 407.23 | 403.85 | 405.86 | 405.62 (+1.32) | 403.91 | 405.70 | −1.85 |
| N(9) | 408.44 | 404.94 | 406.74 | 406.63 (+1.32) | 405.00 | 406.70 | −1.76 |

† Simple mean of IPs based on the models of HF/aug-cc-pVTZ and B3LYP/aug-cc-pVTZ. ‡ CEBE from Chong *et al.* (Takahata *et al.*, 2006). § The IPs based on ADC(4)/6-31G have been shifted by −1.27 and −1.32 eV for C1*s* and N1*s*, respectively (Plekan *et al.*, 2008). ¶ Results with the LB94/TZ2P model (Saha *et al.*, 2008). †† Experimental data in the gas phase (Plekan *et al.*, 2008). ‡‡ Derivation of the IPs using the LB94/et-pVQZ model from the experimental data in the gas phase.

No apparent changes in the pentagon ring of the Pu bases are observed and therefore the perimeters of this ring change very little ($\Delta R_5$ = 0.03 Å). Half of the six bonds in Py are double bonds, which gives an $R_6$ of 8.10 Å (Py). When U is produced, the number of double bonds contained in the hexagon ring reduces to one, leading to the extension of $R_6$ by 0.25 Å, from 8.10 Å in Py to 8.35 Å in U. Replacing the H atom on C(5) of U by a methyl group to produce T does not change the numbers and positions of double and single bonds in the Py ring directly; as a result, the perimeter $R_6$ of T retains the same value as that of U ($R_6$ = 8.35 Å). Finally, formation of C produces two double bonds in the Py ring and therefore the $R_6$ of C is given by 8.26 Å, between Py and U (and T).

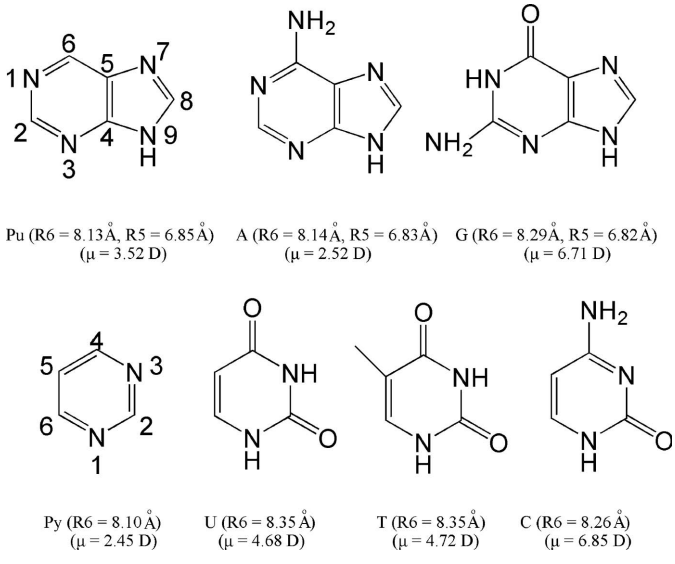


**Figure 1**
Structure, numbering and dipole moment of DNA/RNA bases with their parent purine and pyrimidine.

To validate the computational model, the simulated inner-shell binding-energy spectra of A and T are compared with available results including a recent synchrotron-sourced experimental spectra in the gas phase (Plekan *et al.*, 2008). Tables 1 and 2 compare the inner-shell IP results produced in the present study (LB94/et-pVQZ//B3LYP/aug-cc-pVTZ) of T and A, respectively, with available results such as the core-electron binding energy (CEBE) (Takahata *et al.*, 2006), the fourth-order algebraic diagrammatic construction, ADC(4)/6-31G (Plekan *et al.*, 2008), and the experimental results in the gas phase (Plekan *et al.*, 2008). The results in the second column of the tables are estimated using the averaged orbital energies (negative) generated from the HF/aug-cc-pVTZ and B3LYP/aug-cc-pVTZ calculations. Although the mean IPs so generated do not exhibit a strict theoretical basis, they are able to provide a quick estimation for experimental measurement as found by Powis *et al.* (2003). The IPs produced by the simple model of LB94/et-pVQZ in the present study using the meta-Koopman theorem fit the experimental results, with good agreement without any scaling. For example, the largest IP deviation in A is −1.85 eV on the N(6′) site whereas the maximum IP deviation of T is −2.95 eV on the O(4′) site of T. Given the fact that the IPs are generated in such simple calculations without taking into consideration a number of effects such as orbital relaxation and relativistic effects *etc.*, the present method is very attractive.

Figs. 2 (thymine) and 3 (adenine) report the agreement between the simulated inner-shell spectra of the base pair and recent synchrotron-sourced XPS in the gas phase (Plekan *et al.*, 2008). The simulated spectral lines (solid lines) are produced using the IPs calculated in the LB94/et-pVQZ model (column 3 in Tables 1 and 2). The energy shifts of 1.00 eV, 1.60 eV and 2.90 eV in the C1*s*, N1*s* and O1*s* spectra of thymine, respectively, in Fig. 2 have been applied in the spectra in order to compensate the systematic errors, such as orbital relaxation, relativistic effects and basis set superposition errors, introduced by the model. Lorentzian shape functions with full widths at half-maximums (FWHMs) of

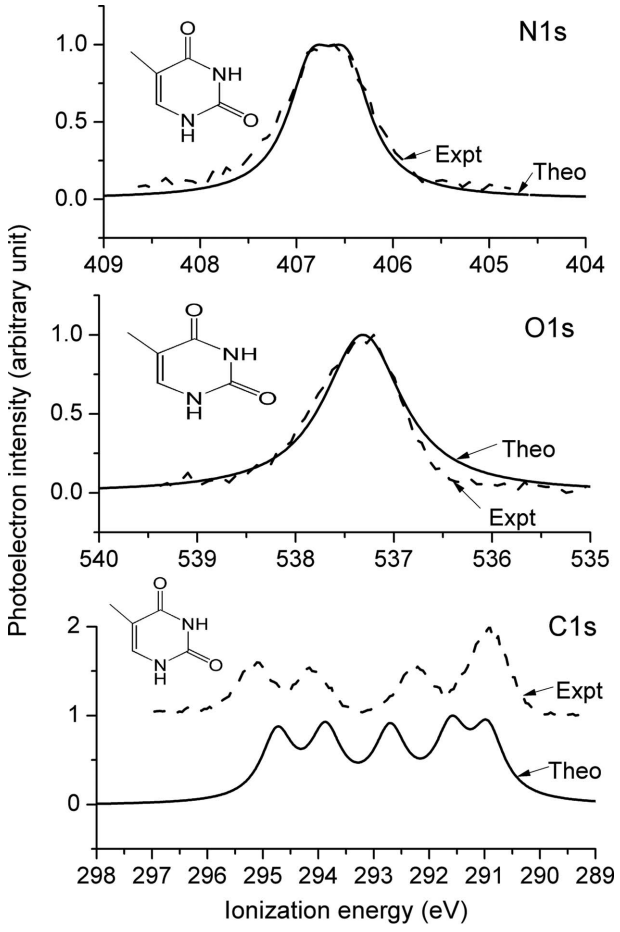


**Figure 2**
Comparison of N1*s*, O1*s* and C1*s* XPS of thymine between experiment and calculation. The theoretical results have been constructed by Lorenzians of 0.59 eV, 0.95 eV and 0.70 eV FWHM and the ionization energy shifted as a whole by 1.60 eV, 2.90 eV and 1.00 eV for N1*s*, O1*s* and C1*s*, respectively.

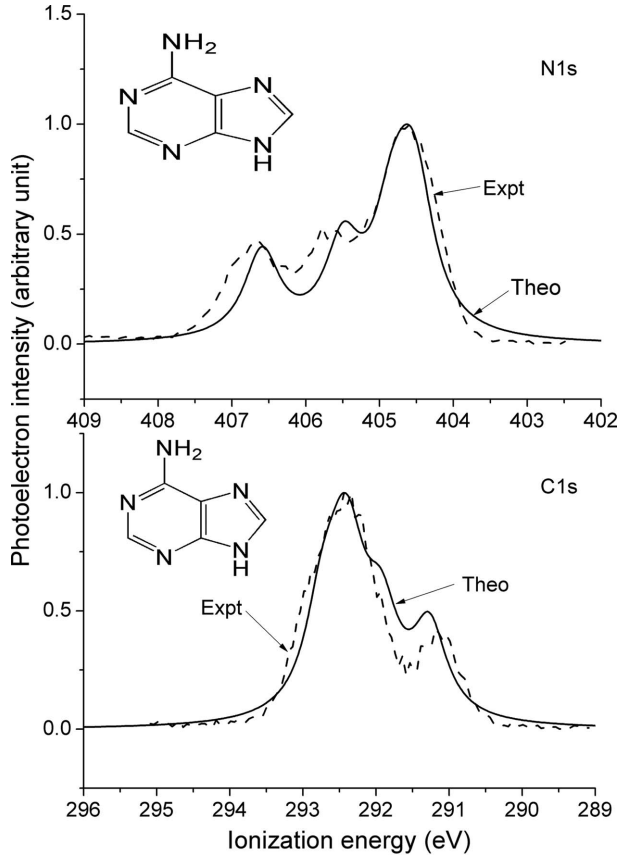


**Figure 3**
Comparison of N1*s* and C1*s* XPS of adenine between experiment and calculation. The theoretical results have been constructed by Lorenzians of 0.59 eV and 0.57 eV FWHM and the ionization energy shifted as a whole by 1.65 eV and 0.85 eV for N1*s* and C1*s*, respectively.

0.59 eV, 0.70 eV and 0.95 eV are employed to simulate the thymine N1*s*, C1*s* and O1*s* spectra, respectively, according to the experimental conditions (Plekan *et al.*, 2008). It is noted that the theoretically produced spectra (Plekan *et al.*, 2008) of thymine were simulated using the superior ADC(4)/6-31G model with energy shifts of 1.60 eV, 1.20 eV and 1.0 eV for the C1*s*, N1*s* and O1*s* spectra, respectively (Plekan *et al.*, 2008).

For the simulated XPS of adenine given in Fig. 3, the energies of N1*s* and C1*s* have been shifted by 1.65 eV and 0.85 eV, respectively, and the spectra were simulated using FWHMs of 0.59 eV and 0.57 eV, respectively, from the experiment (Plekan *et al.*, 2008). The agreement between the observed XPS and simulated spectra is excellent at the given level of theory. The LB94/et-pVQZ model is therefore applied to calculate the inner-shell binding-energy spectra of other DNA/RNA bases in the present work.

Table 3 reports the site-based Hirshfeld charges of all seven Pu and Py bases calculated in the present study. As observed previously (Saha *et al.*, 2008), all the N sites (and O site in G) of the Pu bases possess negative charges whereas the positive charges are deposited onto the C sites. This is also true for the Py bases except for the C(5) sites and the branch connecting at C(5)–C(5′)$H_3$ in T which exhibits opposite charge to the rest of the C sites in the Py bases. In Py, the C(5) site is located at one end of the triangle formed by N(3), N(1) and C(5) in the hexagon ring to balance the negatively charged N sites of the triangle, forming a conjugated Py ring. This negatively charged C(5) site has been passed into other Py bases of U, T and C.

Table 4 displays the inner-shell IPs for Py bases, including Py, U, C and T, together with some other available results. The IPs of the Py bases exhibit an appropriate variation tendency depending on the chemical surroundings. To further demonstrate the trends of the IPs, Fig. 4(*a*) provides an energy diagram for N1*s* and C1*s* sites of the Py bases. Apparent site-dependent changes with respect to the unsubstituted Py are observed. The N1*s* pair in the unsubstituted Py is situated in a similar chemical environment, hence the N(1) and N(3) sites are almost energy degenerated as they are both engaged with the C—N═C chains in Py. The C(4) and C(6) sites in Py are also in a similar situation with almost degenerated energies. In U and T, the equivalency of the nitrogen pair no longer exists and the N═C double bonds are saturated. As a result, the core orbital ionization energies of the N pair split with small energy difference, reflecting the alternation of the chemical environment. In C, the energy gap of the N(1) and N(3) sites is

**Table 3**
Site-dependent Hirshfeld charges $Q$(H) of purine and pyrimidine bases (in atomic units).

| Site | Py | U | T | C | Site | Pu | A | G |
|---|---|---|---|---|---|---|---|---|
| O(2′) | | −0.33 | −0.34 | −0.35 | O(6′) | | | −0.33 |
| O(4′) | | −0.32 | −0.31 | | N(2′) | | | −0.18 |
| N(4′) | | | | −0.16 | N(6′) | | −0.16 | |
| N(1) | −0.18 | −0.08 | −0.09 | −0.09 | N(9) | −0.08 | −0.08 | −0.08 |
| N(3) | −0.18 | −0.09 | −0.09 | −0.21 | N(1) | −0.19 | −0.21 | −0.10 |
| C(2) | 0.08 | 0.21 | 0.21 | 0.19 | N(3) | −0.19 | −0.22 | −0.23 |
| C(4) | 0.04 | 0.17 | 0.17 | 0.12 | N(7) | −0.21 | −0.22 | −0.20 |
| C(6) | 0.04 | 0.04 | 0.02 | 0.04 | C(2) | 0.07 | 0.07 | 0.16 |
| C(5) | −0.04 | −0.06 | −0.01 | −0.08 | C(6) | 0.04 | 0.12 | 0.16 |
| C(5′) | | | −0.10 | | C(4) | 0.10 | 0.09 | 0.08 |
| | | | | | C(8) | 0.08 | 0.05 | 0.06 |
| | | | | | C(5) | 0.03 | 0.01 | 0.00 |

**Table 4**
Calculated core IPs of the pyrimidine bases (in eV).

| Atom | Mean IPs† | LB94/et-pVQZ | CEBE/TZP (gas phase)‡ |
|---|---|---|---|
| Py | | | |
| N(1) | 406.94 | 403.46 | |
| N(3) | 406.94 | 403.46 | |
| C(2) | 293.42 | 291.45 | |
| C(4) | 293.05 | 291.14 | |
| C(6) | 293.05 | 291.14 | |
| C(5) | 291.92 | 290.35 | |
| U | | | |
| O(2′) | 539.75 | 534.66 | 537.66 |
| O(4′) | 539.61 | 534.43 | 537.33 |
| N(1) | 408.79 | 405.45 | 407.14 |
| N(3) | 408.33 | 404.98 | 406.67 |
| C(2) | 296.02 | 293.93 | 295.07 |
| C(4) | 295.05 | 293.02 | 294.12 |
| C(6) | 294.00 | 292.02 | 292.85 |
| C(5) | 291.97 | 290.44 | 291.14 |
| C | | | |
| O(2′) | 538.56 | 533.42 | 536.45 |
| N(1) | 408.15 | 404.67 | 406.34 |
| N(4′) | 407.51 | 404.09 | 406.13 |
| N(3) | 406.02 | 402.71 | 404.11 |
| C(2) | 294.72 | 292.61 | 293.72 |
| C(4) | 294.32 | 292.21 | 293.11 |
| C(6) | 293.66 | 291.54 | 292.32 |
| C(5) | 291.69 | 290.11 | 290.71 |
| T | | | |
| O(2′) | 539.61 | 534.47 | 537.45 |
| O(4′) | 539.58 | 534.35 | 537.25 |
| N(1) | 408.62 | 405.25 | 406.91 |
| N(3) | 408.26 | 404.87 | 406.56 |
| C(2) | 295.88 | 293.74 | 294.75 |
| C(4) | 294.94 | 292.87 | 293.79 |
| C(6) | 293.68 | 291.70 | 292.27 |
| C(5) | 292.10 | 290.60 | 290.95 |
| C(5′) | 291.50 | 289.94 | 291.18 |

† Mean IPs from the negative orbital energies of the HF/aug-cc-pVTZ and B3LYP/aug-cc-pVTZ models. ‡ CEBE from Chong *et al.* (Takahata *et al.*, 2006).

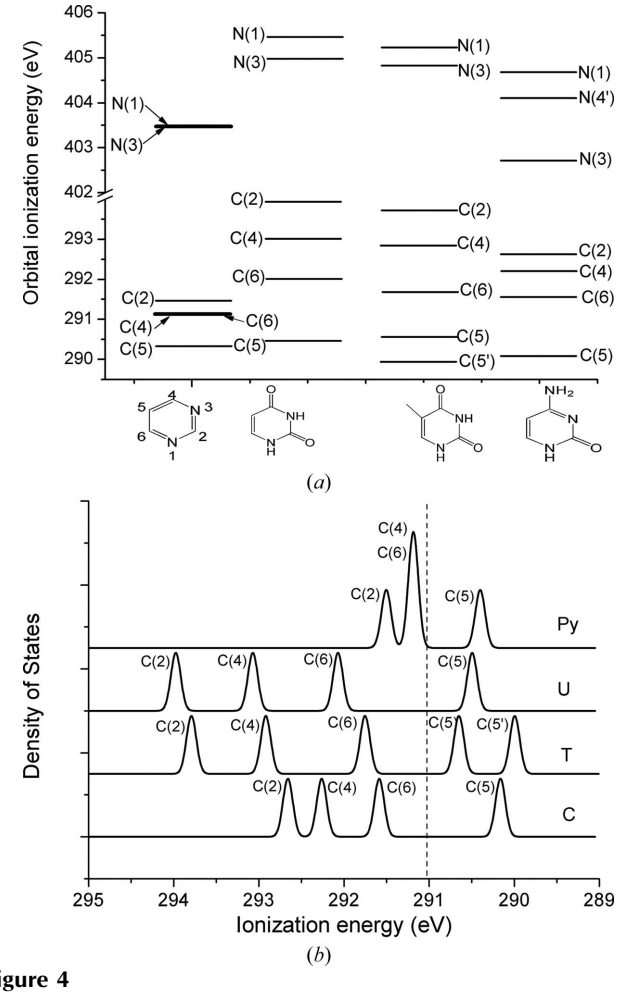


**Figure 4**
Energy correlation diagram of pyrimidine and its bases in the core-shell (*a*) and C1*s* ionization energy spectra of pyrimidine, uracil, thymine and cytosine (*b*).

further widened as they associate with imino [N(3)] and amino [N(1)] nitrogen atoms. The site energy of N(4′) in the additional amino group, $-NH_2$, shows closer energy to the amino N(1) site in the diagram. The results confirm the previously observed N1*s* energy splitting in other bases, and that the N sites separation is associated with a single/double (imino/amino) bond of the sites, rather than on/out of the Py ring (Harada *et al.*, 2006; Wang, 2006; Saha *et al.*, 2008). No crossing in the core energies is observed in the N1*s* site among the Py bases. Similar trends are presented in the C1*s* spectra of the Py bases.

The C1*s* spectra of U and T, owing to the two C═O groups at the C(2) and C(4) positions, exhibit a significant expansion to the larger energy end of the spectra, whereas the C1*s* spectrum of cytosine shows a limited energy expansion. Even though their spectra (Fig. 4*b*) indicate apparent differences, the degree of chemical shift is similar to all C sites, and the chemical shift does not engage with any level crossings in the C1*s* spectra of the Py bases. As noted in the atom-based Hirshfeld charge analysis, the site energy for C(5) is apparently separated from other C atoms at the lower energy end for Py and its bases.

The N1*s* and C1*s* site-dependent IPs of the Pu bases have changed more significantly than the Py counterparts, as the spectra of the Pu bases are not just associated with the spectral peak shifts but also the site-dependent energy-level crossing of the spectral peaks. As a result, structurally, the inner-shell structures and spectra of the Pu bases are more complex and therefore contain richer information than their Py counterparts. Table 5 compares the inner-shell IPs of the Pu bases with

**Table 5**
Calculated core IPs for the purine bases (in eV).

| Atom | Mean IPs† | LB94/et-pVQZ | CEBE/TZP |
|---|---|---|---|
| Pu‡ | | | |
| N(9) | 408.72 | 405.29 | 407.16 |
| N(3) | 407.05 | 403.61 | 405.19 |
| N(7) | 406.90 | 403.47 | 405.09 |
| N(1) | 406.74 | 403.29 | 404.84 |
| C(8) | 293.96 | 292.02 | 292.67 |
| C(4) | 293.82 | 291.96 | 292.73 |
| C(2) | 293.38 | 291.43 | 292.09 |
| C(6) | 292.97 | 291.05 | 291.71 |
| C(5) | 292.39 | 290.85 | 291.58 |
| A§ | | | |
| N(9) | 408.44 | 404.94 | 406.74 |
| N(6′) | 407.23 | 403.85 | 405.86 |
| N(7) | 406.71 | 403.20 | 404.78 |
| N(3) | 406.41 | 402.97 | 404.37 |
| N(1) | 406.26 | 402.83 | 404.23 |
| C(6) | 293.95 | 291.89 | 292.64 |
| C(8) | 293.60 | 291.63 | 292.13 |
| C(4) | 293.43 | 291.50 | 292.15 |
| C(2) | 293.16 | 291.08 | 291.66 |
| C(5) | 291.93 | 290.42 | 290.96 |
| G§ | | | |
| O(6′) | 539.08 | 533.91 | 536.71 |
| N(9) | 408.41 | 404.94 | 406.65 |
| N(1) | 408.32 | 404.80 | 406.42 |
| N(2′) | 407.87 | 404.41 | 406.46 |
| N(3) | 406.69 | 403.36 | 404.72 |
| N(7) | 406.48 | 403.02 | 404.53 |
| C(2) | 295.41 | 293.20 | 293.74 |
| C(6) | 294.59 | 292.49 | 293.30 |
| C(4) | 293.63 | 291.72 | 292.20 |
| C(8) | 293.25 | 291.40 | 291.79 |
| C(5) | 291.79 | 290.35 | 290.68 |

† Mean IPs from the negative orbital energies of the HF/aug-cc-pVTZ and B3LYP/aug-cc-pVTZ models. ‡ CEBE calculated by Chong *et al.* (Saha *et al.*, 2008). § CEBE from Chong *et al.* (Takahata *et al.*, 2006).

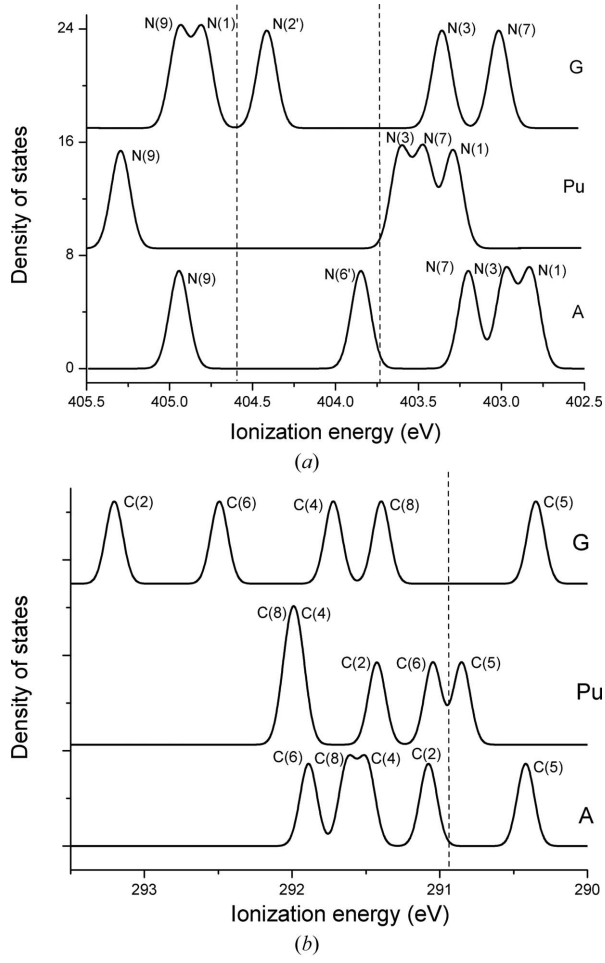


**Figure 5**
N1*s* ionization energy spectra (*a*) and C1*s* ionization energy spectra (*b*) of purine, guanine and adenine.

other available results in the gas phase. As indicated earlier, the IPs of adenine calculated using the simple DFT-LB94/et-pVQZ model agree well with the recently available XPS experiment with the largest deviation of −1.85 eV [at N(6′)], which is almost at the same degree of accuracy as for the more sophistic ACD(4)/6-31G model. The latter leads to an error of 1–2 eV (see Table 2). The present calculations agree reasonably well with the more computationally demanding DFT CEBE/TZP calculations of Chong *et al.* (Takahata *et al.*, 2006). The CEBE method, which is a ΔDFT method, engages with $n + 1$ times state calculations (where $n$ is the number of core hole states), compared with the present method. The major difference between the present and the CEBE method is due to the orbital relaxation ignored in the meta-Koopman theorem, apart from the $V_{xc}$ functional and basis set differences. The available XPS experimental IPs of adenine are not fully resolved, so that detailed site-dependent IPs of the base are expected to be obtained from theory. For example, all C1*s* sites in the experiment XPS of adenine contribute to the peak centred at 292.50 eV, except for the C(5) site, which is approximately 1.50 eV split from the main C1*s* peak, whereas the N1*s* sites split into three peaks, among which two peaks represent sites N(6′) and N(9), and the third peak consists of N(1), N(3) and N(7) (Plekan *et al.*, 2008).

The inner-shell of guanine behaves very differently from that of adenine. Figs. 5(*a*) and 5(*b*) present the simulated N1*s* and C1*s* spectra of the Pu bases: A, G and unsubstituted Pu. In general, any substitution to the unsubstituted Pu results in a red shift (lower energy) to the sites without direct substitutions or bond character changes, such as N(9), N(3) and N(7) for N1*s* spectra and C(4), C(5) and C(8) for C1*s* spectra. Our previous study on the amino group $–NH_2$ attachment to Pu indicated certain similarities in the N1*s* spectra between A and Pu, whereas the C1*s* spectra of Pu and A are surprisingly different (Saha *et al.*, 2008). This finding is not applicable to G. The present study further reveals that the N1*s* and C1*s* spectra of G have little similarities with either A or Pu, indicating substantial structural differences between G and A (and Pu). Fig. 6 provides the orbital-based information underlying the changes associated with the Pu bases. In this figure, the single-bonded amino group, $–N(2')H_2$, in G is pushed further to the blue side (larger IP) in G than in A [N(6′)], to join the amino N sites of N(9) and N(1). The imino N sites of N(3) and N(7) are well separated from the amino N1*s* sites, leading to the N(7) site of G possessing the lowest IP energy in the N1*s* spectrum. This implies that this site can be the most active site from the energy point of view, as found by Sponer *et al.* (2004)

in a study of metal–guanine binding positions. In Fig. 5(*a*) the N1*s* ionization spectra of Pu and its bases can be divided into three bands: the higher-energy band assigned to amino N sites in the Pu ring, the middle band representing the N atom in the $-NH_2$ group, and the lower-energy band representing imino N sites.

The most significant change in the C1*s* spectrum of G with respect to Pu is the extension of the C1*s* spectral band, ranging from 1.17 eV in Pu to 2.85 eV in G, owing to the significant blue shift of the C(2) and C(6) sites in G (see Fig. 5*b*). The C(2) site connected with the amino $-NH_2$ attachment shifts more significantly than the C(6) site associated with the C═O attachment in G. The peak positions in G are very different from in its parent Pu too. For example, the IP energy order of the C1*s* sites in Pu is C(5) < C(6) < C(2) < C(4) < C(8), whereas this order in G is given by C(5) < C(8) < C(4) < C(6) < C(2). There is little similarity in the two C1*s* spectra of A and G. Perhaps one of the most apparent differences between the purine bases A and G and parent Pu is the separation of the C(5) peak from the rest of the C1*s* sites in A and G, which indicates that the C(5) site has a very similar chemical environment in A and G. As a result, in Fig. 5(*b*) the lower IP energy peak owing to C(5) can be easily separated from other C atoms in the Pu rings of A and G.

**Table 6**
N1*s* and C1*s* site IP shifts (LB94/et-pVQZ) with respect to Py and Pu (in eV).

| Site | U(Py)† | T(Py)† | C(Py)† | Py | Site | A(Pu)‡ | G(Pu)‡ | Pu |
|---|---|---|---|---|---|---|---|---|
| N(1) | 1.99 | 1.79 | 1.21 | 403.46 | N(9) | −0.35 | −0.35 | 405.29 |
| N(3) | 1.52 | 1.41 | −0.75 | 403.46 | N(3) | −0.64 | −0.25 | 403.61 |
| C(2) | 2.48 | 2.29 | 1.16 | 291.45 | N(7) | −0.27 | −0.45 | 403.47 |
| C(4) | 1.88 | 1.73 | 1.07 | 291.14 | N(1) | −0.46 | 1.51 | 403.29 |
| C(6) | 0.88 | 0.56 | 0.40 | 291.14 | C(8) | −0.39 | −0.62 | 292.02 |
| C(5) | 0.09 | 0.25 | −0.24 | 290.35 | C(4) | −0.46 | −0.24 | 291.96 |
| | | | | | C(2) | −0.35 | 1.77 | 291.43 |
| | | | | | C(6) | 0.84 | 1.44 | 291.05 |
| | | | | | C(5) | −0.43 | −0.50 | 290.85 |

† Site-dependent IP shifts for Py bases with respect to the N1*s* and C1*s* IPs of unsubstituted Py. ‡ Site-dependent IP shifts for Pu bases with respect to the N1*s* and C1*s* IPs of unsubstituted Pu.

Table 6 reports the core ionization energy shift of the DNA/RNA bases with respect to their parent Py and Pu. The directions of such shifts are very different in the Py and Pu bases. The nitrogen sites and carbon sites of the Py bases are dominated by blue shift towards the larger energy end of the spectra with respect to the parent Py. For example, in T and U, all core sites shift up against the corresponding sites in their Py parent. The core sites of cytosine, however, differentiate themselves from those of T and U. In cytosine, the N(3) and C(5) sites shift against the direction of other N and C sites in red shift (lower energy) with respect to the same sites of their parent Py. The Pu bases are opposite to the Py bases. The core shell energy shifts are dominated by red shift towards the lower energy end of the spectra, when the parent Pu becomes substituted. Usually the sites which connect to a fragment are associated with an opposite energy shift: a blue shift (larger energy). For example, in adenine, only C(6) is associated with an amino group, $-NH_2$, and therefore the C(6) site shifts to the blue end of the spectrum. In G, three sites, C(6), N(1) and C(2), which form a chain C(6)–N(1)–C(2), do not change their energies in the same direction as the other sites in G.

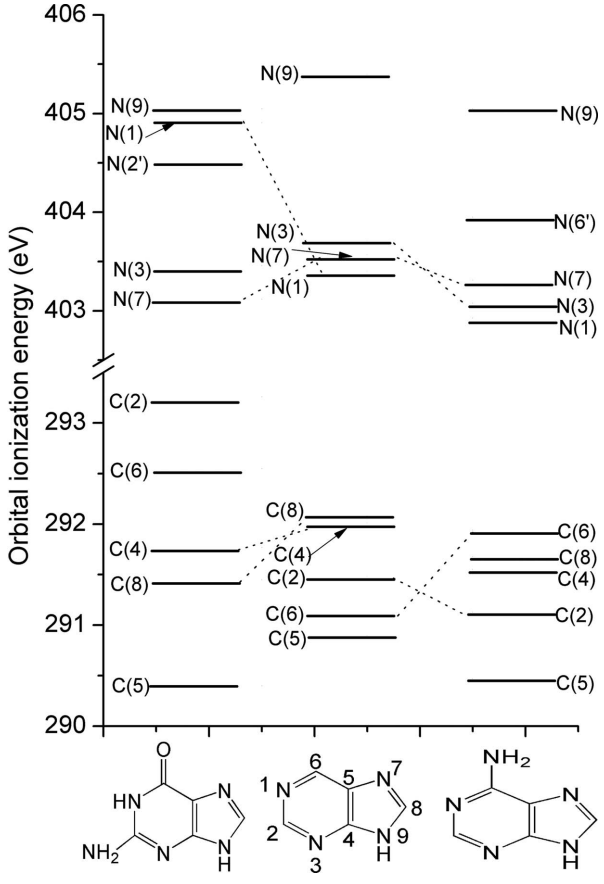


**Figure 6**
Energy correlation diagram of purine, adenine and guanine in the core shell.

## 4. Conclusions

Site-dependent changes in the inner-shell spectra and structures of DNA/RNA bases with respect to their parent Py and Pu are investigated using the DFT LB94/et-pVQZ//B3LYP/aug-cc-pVTZ model. Good agreement with the recently available synchrotron-sourced XPS in thymine and adenine indicates the suitability of the model for studying inner-shell structures. It is found in the study that the changes in inner-shell properties of these DNA/RNA bases with respect to their parent Py and Pu can be divided into Py bases and Pu bases. More significant changes are found to relate to the Pu bases.

Although the geometric parameters such as the perimeters of the hexagon and pentagon rings experience small variations, depending on the number of bonds and the bonding characters, some properties exhibit certain consistency within the single ring bases (Py bases) and the double ring bases (Pu bases). Hirshfeld charges, for example, of Py have been inherited by its derivatives. Similarly, the same positive and

negative charge distributions of Pu are also passed into A and G. Other properties, such as dipole moment and core shell energies, are quite different. The dipole moment of the Py derivatives, U (4.68 D), T (4.72 D) and C (6.85 D), are significantly larger than their parent Py (2.45 D) but, in the Pu case (3.52 D), adenine has an apparently smaller dipole moment (2.52 D) but G exhibits a very large dipole moment of 6.71 D. The site-dependent core ionization energy shifts in Py bases are dominated by blue shifts (larger) with respect to their corresponding Py sites, with the N(3) and C(5) sites of cytosine being the exceptions; whereas the core energy shifts in the Pu bases are opposite to the shifts in the Py bases: they are dominated by the red shifts (smaller) with respect to their parent Pu sites, except for the sites, such as C(6) in adenine, and N(1), C(2) and C(6) in guanine, which are associated with blue energy shifts. These sites in the Pu bases are dominated by the substituted sites.

The spectral shifts of the bases with respect to their parent Py and Pu are single-ring and double-ring related. In the Py bases, when the derivatives such as U, T and C are produced, the chemical environment of the sites alters noticeably, leading to the apparent spectrum range expansion towards the blue energy end, depending on the bonding situation. The degrees of the site-dependent shift in Py bases do not vary significantly so that there are no level crossings in the N1*s* spectra, nor in the C1*s* spectra. The Pu bases exhibit a very different nature however. With respect to the N1*s* and C1*s* spectra of the parent Pu, adenine inherits certain features in the N1*s* spectra as previously found (Saha *et al.*, 2008), but little has been inherited by G. In conclusion, if the DNA/RNA bases have more or less spectral and structural properties inherited from their parent Py and Pu species, G is the species with the least inheritance from its parent Pu in the inner shell.

This work is supported by a Dean's Collaborative Grant of Faculty of Information and Communication Technologies, Swinburne University of Technology. It is also supported by an award under the Merit Allocation Scheme on the APAC National Facility at the ANU. FW thanks Dr K. Prince for providing the digital traces of their recent XPS of thymine and adenine.

## References

Alagia, M., Baldacchini, C., Betti, M. G., Bussolotti, F., Carravetta, V., Ekstrom, U., Mariani, C. & Stranges, S. (2005). *J. Chem. Phys.* **122**, 124305.

Baerends, E. J. *et al.* (2006). *ADF* (version 2006.01), http://www.scm.com/.

Callis, P. R. (1983). *Annu. Rev. Phys. Chem.* **34**, 329–357.

Chetioui, A., Despiney, I., Guiraud, L., Adoui, L., Sabatier, L. & Dutrillaux, B. (1994). *Int. J. Radiat. Biol.* **65**, 511–522.

Chong, D. P., Lenthe, E., van Gisbergen, S. J. A. & Baerends, E. J. (2004). *J. Comput. Chem.* **25**, 1030–1036.

Dong, F. & Miller, R. E. (2002). *Science*, **298**, 1227–1230.

Downton, M. T. & Wang, F. (2006). *Mol. Simulat.* **32**, 667–673.

Fayard, B. *et al.* (2002). *Radiat. Res.* **157**, 128–140.

Fonseca-Guerra, C., Bickelhaupt, F. M., Saha, S. & Wang, F. (2006). *J. Phys. Chem. A*, **110**, 4012–4020.

Frisch, M. J. *et al.* (2004). *Gaussian03*, revision D.01. Gaussian, Wallingford, CT, USA.

Fujii, K., Akamatsu, K., Muramatsu, Y. & Yokoya, A. (2003). *Nucl. Instrum. Methods Phys. Res. B*, **199**, 249–254.

Fujii, K., Akamatsu, K. & Yokoya, A. (2004). *J. Phys. Chem. B*, **108**, 8031–8035.

Gobert, F. N., Lamoureux, M., Herve du Penhoat, M. A., Ricoul, M., Boissiere, A., Touati, A., Abel, F., Politis, M. F., Fayard, B., Guigner, J. M., Martins, L., Testard, I., Sabatier, L. & Chetioui, A. (2004). *Int. J. Radiat. Biol.* **80**, 135–145.

Griffin, C. S., Hill, M. A., Papworth, D. G., Townsend, S., Savage, J. R. K. & Goodhead, D. G. (1998). *Int. J. Radiat. Biol.* **73**, 591–598.

Harada, Y., Takeuchi, T., Kino, H., Fukushima, A., Takakura, K., Hieda, K., Nakao, A., Shin, S. & Fukuyama, H. (2006). *J. Phys. Chem. A*, **110**, 13227–13231.

Herve du Penhoat, M. A., Fayard, B., Abel, F., Touati, A., Gobert, F., Despiney-Bailly, I., Ricoul, M., Sabatier, L., Stevens, D. L., Hill, M. A., Goodhead, D. T. & Chetioui, A. (1999). *Radiat. Res.* **151**, 649–658.

Hirshfeld, F. L. (1977). *Theor. Chim. Acta*, **44**, 129–138.

Hobza, P. & Sponer, J. (1999). *Chem. Rev.* **99**, 3247–3276.

Jones, D. B., Wang, F., Winkler, D. A. & Brunger, M. J. (2006). *Biophys. Chem.* **121**, 105–120.

Leeuwen, R. van & Baerends, E. J. (1994). *Phys. Rev. A*, **49**, 2421–2431.

Leszczynski, J. (1992). *Int. J. Quantum Chem. Quantum Biol. Symp.* **19**, 43–55.

Li, X., Bowen, K. H., Haranczyk, M., Bachorz, R. A., Mazurkiewicz, K., Rak, J. & Gutowski, M. (2007). *J. Chem. Phys.* **127**, 174309.

Plekan, O., Feyer, V., Richter, R., Coreno, M., de Simone, M., Prince, K. C., Trofimov, A. B., Gromov, E. V., Zaytseva, I. L. & Schirmer, J. (2008). *Chem. Phys.* **347**, 360–375.

Powis, I., Rennie, E. E., Hergenhahn, U., Kugeler, O. & Bussy-Socrate, R. (2003). *J. Phys. Chem. A*, **107**, 25–34.

Saha, S., Wang, F., MacNaughton, J. B., Moewes, A. & Chong, D. P. (2008). *J. Synchrotron Rad.* **15**, 151–157.

Sanche, L. (2002). *Mass. Spectrom. Rev.* **21**, 349–369.

Segala, M., Takahata, Y. & Chong, D. P. (2006). *J. Mol. Struct.* **758**, 61–69.

Shukla, M. K. & Leszczynski, J. (2007). *J. Biomol. Struct. Dynam.* **25**, 93–118.

Sponer, J. E., Sychrovsky, V., Hobza, P. & Sponer, J. (2004). *Phys. Chem. Chem. Phys.* **6**, 2772–2780.

Sponer, J., Leszczynski, J. & Hobza, P. (2001). *J. Mol. Struct.* **573**, 43–53.

Takahata, Y., Okamoto, A. K. & Chong, D. P. (2006). *Int. J. Quantum Chem.* **106**, 2581–2586.

Usami, N., Kobayashi, K., Maezawa, H., Hieda, K. & Ishizaka, S. (1991). *Int. J. Radiat. Biol.* **60**, 757–768.

Wang, F. (2005). *J. Mol. Struct.* **728**, 31–42.

Wang, F. (2006). *Micro Nano Lett.* **1**, 23–24.

Wang, F., Downton, T. & Kidwani, N. (2005). *J. Theor. Comput. Chem.* **4**, 247–264.

Yokoya, A., Watanabe, R. & Hara, T. (1999). *J. Radiat. Res.* **40**, 145–158.