# Inheritance and correlation of nucleic acid pyrimidine bases

Zejin Yang[1,2*], Patrick Duffy[3] and Feng Wang[2†]

[1]School of Science, Zhejiang University of Technology, Hangzhou, 310023, China;

[2]eChemistry Laboratory, Faculty of Life and Social Sciences, Swinburne University of Technology, Hawthorn, Melbourne, Victoria, 3122, Australia.

[3]Department of Chemistry, Kwantlen Polytechnic University, Richmond Campus, 8771 Lansdowne Road, Richmond, B.C., V6X 3V8, Canada

[*] yzjscu@163.com
[†] fwang@swin.edu.au

## Abstract

Valence electronic structures of pyrimidine (P, $C_4N_2H_4$) and nucleic acid (NA) pyrimidine bases, such as cytosine (C, $C_4N_3OH_5$), thymine (T, $C_5N_2O_2H_6$) and uracil (U, $C_4N_2O_2H_4$) are simulated quantum mechanically using density functional theory (DFT) methods, including B3LYP/aug-cc-pVTZ, B3LYP/TZVP and SAOP/et-pVQZ, together with OVGF/TZVP. The agreement with most recent experimental results is excellent. The highest occupied molecular orbital (HOMO) and the next HOMO (NHOMO) of pyrimidine are conclusively assigned as $7b_2$ and $2b_1$, respectively, using combined quantum mechanical calculations and symmetries of the orbitals through the resolved orbital momentum distributions. From the ionization energy spectra and valence orbital momentum distributions, it is found that the NA bases, i.e., cytosine, thymine and uracil exhibit a larger degree of similarity to each other than to pyrimidine, although they do inherit certain properties from pyrimidine. An interactive three-dimensional (3D) PDF technique has been used to display the properties where 3D may make a difference (Figure 3).

# 1. Introduction

A good understanding of biological systems requires knowledge of the molecular structures and interactions of the building blocks of life[1,2]. As functional building blocks of life, nucleic acid (NA) pyrimidine bases are of great importance in life science[3,4]. Electron collision experiments on biologically important molecules have suggested that resonant mechanisms induced by nonthermal low-energy electrons could be related to NA lesions such as single/double strand breaks[5]. This is significant because one of the principal effects on cells of high-energy radiation such as that encountered in radiation therapy or from cosmic rays is the ionization of DNA[6]. The loss of an electron (ionization) from DNA generates an electron "hole" (a radical cation), located most often on its nucleobases, that migrates reversibly through duplex DNA by hopping until it is trapped in an irreversible chemical process[7].

The study of electron collisions with nucleic acid bases is important to the understanding of the mechanism of DNA/RNA damage. Recently, valence electronic structures of pyrimidine have been reported by electron momentum spectroscopic (EMS) experiments[8]. In addition to the EMS study, the valence shell photoelectron spectra of pyrimidine, uracil, the methyluracils[9], cytosine, thymine and adenine[10], and purine and pyrimidine[11-17] have also been reported. The experimental results provide a solid foundation for a high level theoretical study of the properties and functionalities of these molecules, and build upon our previous studies that focused on the inner-shell structures of these species[11].

The NA bases and compounds related to them have been the subject of many theoretical and experimental investigations. For example the photoionization dynamics of uracil[18], the near edge x-ray absorption fine structure spectra of

pyrimidine[17], the electron affinities (EAs) and ionization potentials (IPs) of some nucleotides[19, 20], the core-electron binding energies of pyrimidine and purine[21], the tautomerism of uracil[22,23], have all been investigated by density functional theory (DFT). Moreover, molecular structure and vibrational infrared spectra of cytosine and its thio and seleno analogues have been studied by both the DFT and Hartree-Fock (HF) methods[24]. Electron propagator theory (including OVGF) has been applied to evaluate the binding energies of molecular fragments found in nucleic acids[25], as well as the ionization energies and Dyson orbitals of cytosine and 1-methylcytosine[15], and thymine and methylated uracil[26]. Ground– and excited–state properties including ionization energies have also been calculated for the NA base molecules by employing a plane-wave calculation using ultrasoft pseudopotentials[27]. One example of this is the study of the core level photoemission spectra of cytosine and uracil, which were measured in vapour and interpreted via theoretical calculations[28]. The algebraic-diagrammatic construction (ADC) method has also been a powerful tool for the study of the excitation and ionization spectra of molecules[29] but suffers from computational costs.

Subtle energy differences among biomolecular conformers such as those caused by angle changes and space orientation, cannot be sufficiently resolved by energy dominant spectroscopic or computational means, as the energy differences may be even smaller than the uncertainties inherent in the methods. However, momentum distributions of individual orbitals provide additional information and unique insight in the differentiation of the structural differences. Combining position space and momentum space information affords an even more comprehensive understanding of the electronic structures and inheritances of the nucleic acid pyrimidine bases. In this

study, we concentrate on the differentiation and correlation of the NA pyrimidine bases that arises out of their ionization spectra and orbital momentum distributions.

## 2. Methods and computational details

Ground state geometries for pyrimidine (P), cytosine (C), thymine (T) and uracil (U) have been optimized using a DFT-based hybrid B3LYP functional with Dunning's augmented correlation corrected triple zeta plus polarization Gaussian basis set, i.e., B3LYP/aug-cc-pVTZ. The geometry output from that calculation was then input into (and re-optimized by) a single point calculation that used the B3LYP/TZVP model. The TZVP Gaussian basis set is triple zeta in quality and has valence polarisation functions[30]. Orbital densities in coordinate space were also obtained using the B3LYP/TZVP model. However, the valence vertical ionization potentials were produced using the OVGF/TZVP model and the SAOP/et-pVQZ[31] model.

All B3LYP and OVGF calculations were produced using the Gaussian 03[32] computational chemistry package. The SAOP/et-pVQZ model is a part of the ADF[33] computational chemistry package. Molecular orbitals (MO) obtained in coordinate space (based on the B3LYP/TZVP model) were directly mapped into momentum space as theoretical momentum distributions (TMD). The resulting TMDs have embedded within them a number of approximations, such as the Born-Oppenheimer approximation, the independent particle approximation and the plane wave impulse approximation (PWIA)[34,35]. Given these approximations, the target-ion overlap is a one electron property,

$$\sigma \propto \int d\Omega \left|\phi_j(\vec{p})\right|^2.$$

Here, $\vec{p}$ is the momentum of the target electron at the instant of ionisation. Orbital $\phi_j(p)$ is the Kohn-Sham (KS) orbital in momentum space[36]. The Fourier transform is calculated via modified HEMS code[35].

Molecular structures (Figure 1) and orbital charge distributions were produced using Molden[38] with a contour of 0.03 a.u. A new 3D feature[37] in the portable document format (PDF) has been used to display them. Double-clicking the structures on a computer will activate the 3D models (Adobe Acrobat 8.1 or above is required).

# 3. Results and discussion

### 3.1 3D molecular structures and dipole moment

Chemical structures and nomenclature for the NA pyrimidine bases are given in Figure 1. Geometric details in their ground electronic states have been reported previously[11]. As shown in Table 1, the dipole moments of the species calculated using the B3LYP/TZVP model agree well with those produced by the B3LYP/aug-cc-pVTZ model in an earlier study[11] and with experimental results[39-41]. Pyrimidine exhibits a small total dipole moment of 2.43 D, in good agreement with 2.45 D as calculated using the B3LYP/aug-cc-pVTZ model, and with the experiment result, 2.33 D[39]. The dipole moment of cytosine is calculated to be 6.73 D, which makes it the largest in the group of pyrimidine bases in this study (the dipole moment was calculated to be 6.85 D using B3LYP/aug-cc-pVTZ[11]). The dipole moments of thymine and uracil are calculated to be 4.51 D and 4.56 D, respectively.

The calculated values of the dipole moments are consistent with their molecular symmetries. Here, the symmetry of the pyrimidine bases is $C_{2v}$ (pyrimidine) > $C_s$ (thymine, uracil) > $C_1$ (cytosine), and the order of dipole moment is, therefore, the opposite: $\mu$(pyrimidine) < $\mu$(thymine, uracil) < $\mu$(cytosine). In Table 1, the x, y and z components of the dipole moments are also presented (the molecular plane is the xy-

plane and the z-axis is perpendicular to the molecular plane). For example, pyrimidine, which possess $C_{2v}$ symmetry ($C_2$ axis viewed as the *y* axis), exhibits zero dipole moment along the *x* and *z* axes in the *xz* plane (backbone atoms), whereas thymine and uracil ($C_s$ symmetry) possess a symmetry plane (the molecular plane) with zero dipole moment along the *z* axis perpendicular to the molecular plane.

### 3.2 Valence ionization potentials and frontier orbitals of pyrimidine

Table 2 compares valence electron configurations and vertical ionization potentials (VIPs) obtained from various theoretical and experimental studies for (a) pyrimidine, and (b) cytosine. Good agreement is observed when comparing our OVGF results in the outer valence space with the OVGF and SAOP results from the other studies, with the exception that our first and second VIPs are reversed in order. Our first and second calculated VIPs are also reversed in order when compared to experiment, Agreement between the SAOP model and experiment improves as the ion hole moves inwards. In the inner valence shell (above 19 eV), OVGF results are not reported, since the OVGF model may no longer be used for calculation of IPs in that region of the spectrum. For comparison purposes, other available calculations including OVGF and HF with 6-311++G** basis sets[8] are also listed in the table.

The OVGF and SAOP models, when combined, ensure that the calculated IPs are valid and reliable, as the models are in complement. The OVGF model is able to produce outer valence ionization energies of a molecule, whereas the SAOP model, through “meta-Koopman theorem,” produces ionization potentials for the entire valence space; those in the inner-valence space appear to be in particularly good agreement with experiment. It is also important to note that experimental ionization energy measurements are averaged values, because of conversion between conformers) or other uncertainties inherent in the experiment. The theoretical methods employed

to obtain the computational results in the present study are not strictly accurate either, of course, as a number of effects such as orbital relaxation, relativistic effects and basis set superposition errors *etc*. have not been accounted for. Nevertheless, the results obtained in the present study confirm that the cumulative effect of the errors inherent in the theoretical approximations alluded to previously is relatively small. In addition, apart from the SAOP model and the OVGF results from the other work, no model could reproduce a complete valence space ionization spectrum in relatively good agreement with experiment (i.e. at least with the correct IP ordering).

The orbital orders in the ground state configurations predicted by the theoretical models may vary [8], as the degrees of inclusion of the electron exchange-correlation energies varies with respect to specific models as do the basis sets employed. The energy difference between the IPs of the frontier orbitals, i.e., the highest occupied molecular orbitals (HOMO, $7b_2$) and the next HOMO, NHOMO ($2b_1$) of pyrimidine is very small and so the orbital ordering is often switched in experiments or calculations. The latter depends on the model. For example, the SAOP/et-pVQZ, B3LYP/6-311++G** and OVGF/6-311++G** models[8] produce an order of ionization potentials (IP)($7b_2$) > IP($2b_1$), whereas the opposite order has been obtained using the OVGF/TZVP and the HF/aug-cc-pVTZ models[8]. However, examining both our calculations and the experimental orbital MDs, we are able to determine which orbital, $7b_2$ or $2b_1$, is the HOMO of pyrimidine. This will be discussed in the next section.

**3.3 Valence ionization spectra**

Although experimental photoelectron spectra for cytosine, thymine and uracil have been available for some time[9,10], EMS studies of pyrimidine have only been published recently[8]. This is due to a variety of factors. First, accurate theoretical calculation and subsequent interpretation of the valence shell of this class of

biomolecules still presents a challenge. Second, apart from the outermost valence space, valence ionization potentials for these pyrimidine bases are not available. Third, the delocalized nature of the valence space, coupled with small energy differences between orbitals, makes experimental measurements difficult to interpret due to limitations imposed by the requirement for finite experimental resolution. All of the above combined to place significant limits on the conclusions that could be made when interpreting theoretical caculations in combination with experimental results.

To validate the SAOP/et-pVQZ model in the valence space, the ionization spectrum of pyrimidine has been simulated against the recently available EMS binding energy spectra of pyrimidine. In order to best reproduce the experimental binding energy spectrum, an energy resolution with the full width at half maximum (FWHM) of 1.0 eV is employed (the experimental resolution is 0.68 eV). Figure 2 compares the simulated IP spectrum of pyrimidine with the EMS experimental binding energy spectrum[8]. The agreement between theoretical simulation and experimental measurement is good, especially in the outer valence shell region. The simulated spectrum reproduces the major features of the IP spectrum of pyrimidine well, although an energy shift of as large as 0.13 eV is found in the spectral peaks at approximately 13.4 eV and 17.5 eV. Small errors in the calculated spectrum in the region above 25 eV may be reduced with a global shift towards higher energy; the observed discrepancy may be the result of satellite shake-ups and shake-offs[42]. For outer valence shell IPs, the present SAOP model reproduces the ionization spectrum with satisfactory accuracy.

The simulated ionization energy spectra of the four pyrimidine bases are also shown in Figure 2 (note that the LUMO “peaks” in the simulated spectra are for reference purposes only). The binding energy spectra of the various molecules clearly

show that the compounds are related. Obviously, the bases have different valence electronic structures with varied numbers of valence orbitals, for example, 15 for pyrimidine and 21 for cytosine. The HOMO-LUMO energy gaps of the four bases are 3.81 eV, 3.93 eV, 3.96 eV and 4.15 eV for pyrimidine, cytosine, thymine and uracil, respectively. For all nucleic acid bases, i.e., cytosine, thymine and uracil, the HOMOs and LUMOs are very different from those of pyrimidine. Certain similarities in the ionization energy spectra of the pyrimidine bases in the region greater than 20 eV indeed indicate that the nucleic bases possess some common structures, whereas the mid-valence region between 15 and 20 eV reveals the most significant molecule-dependent chemical bonding and interactions, as we observed in an earlier study[37].

### 3.4 What we learn from the orbital MDs

Figure 3 compares the simulated orbital MDs for the experimentally resolved orbitals $7b_2$ (HOMO), $2b_1$ (NHOMO) and $9a_1$. Note that in the theoretically simulated orbital MDs, the instrumental resolution of $\delta\theta = 1.20°$ and $\delta\phi = 0.60°$ with an impact energy of 1500 eV plus binding energies[8] are also incorporated into the simulation using a Monte Carlo resolution folding program developed earlier[43]. The binding energies used in the resolution folding routines were taken from the EMS study (9.8, 10.5 and 15.7 eV, respectively[8]). All theoretical cross-sections presented in the figure 3 below have been scaled to have identical maximum vertices, thus permitting comparison for shape rather than for absolute cross-section.

Shape agreement between theory and experiment for the three experimentally resolved orbitals is excellent, as shown in Figure 3. More importantly, the orbital symmetry of the HOMO and NHOMO of pyrimidine results in a unique shape for their orbital momentum distributions (MDs) in momentum space. The shape

agreement between theoretical and experimental MDs provides conclusive evidence that the HOMO and NHOMO orbital assignment for pyrimidine in Table 2(a) is $7b_2$ and $2b_1$, respectively, in agreement with the present SAOP model. Given this success, the present study proceeded to simulate other valence orbitals of pyrimidine (and other bases) which are neither experimentally resolved, nor experimentally measured. All simulated valence orbital MDs for pyrimidine, cytosine, thymine and uracil have been screened in the present study. Unlike the inner shell structures[11], valence structures of the pyrimidine bases are not related by simple orbital energy correlation diagrams. However, orbitals can be categorized as (a) very similar, (b) very different, or (c) somewhere between the two categories. The majority of orbitals belong to categories (b) and (c).

Figure 4 shows the (perfect resolution) outermost valence orbital MDs, that is, the HOMO (a) and NHOMO (b), for pyrimidine, cytosine, thymine and uracil. It can be seen from this figures, that the DNA/RNA bases exhibit more similarities to each other than to pyrimidine, as is also seen in their valence ionization spectra in Figure 2. Differences between the HOMOs and NHOMOs ofthe DNA/RNA (C, T and U) bases and pyrimidine (P) may indicate some link to the reasons why pyrimidine does not appear in normal DNA/RNA. That is, the HOMOs and NHOMOs of the C, T and U bases in Figure 4 (a) and (b), respectively, show similarities among the bases, except for the HOMO and NHOMO of pyrimidine which are very different from its nucleic acid bases. In addition, the present study unambiguously concludes that the HOMO of pyrimidine is $7b_2$ and the NHOMO of pyrimidine is $2b_1$, which is supported by the EMS experiment[8] of pyrimidine. However, the HOMO ($7b_2$) orbital MDs of pyrimidine in Figure 4(a) indeed appear similar to the NHOMO of the bases (28a(C), 27 a′ (T) and 24 a′ (U)). For example, the HOMO MDs of all bases other than

pyrimidine exhibit a *p*-electron dominant bell shape with the maximum cross section peaked at approximately 1.0 a.u. in momentum. The HOMO MDs of pyrimidine, although also being dominated by *p*-electrons, exhibit more anti-bonding character and a maximum cross section for the major peak at approximately 0.5 a.u., and the secondary peak at 1.5 a.u. In Figure 4(b), orbital $2b_1$ of pyrimidine, 27 a′ of thymine and orbital 24 a′ of uracil are dominated by *p*-electrons (bonding), whereas orbital 28a of cytosine is dominated by *sp*-hybrid character. The higher point group symmetry of pyrimidine confines the molecules in a plane and restricts interchanges of their x- and y- components. Such a cross HOMO-NHOMO similarities between pyrimidine and the nucleic acid bases probably led other calculations which employed position space only information to confuse the HOMO-NHOMO assignment for pyrimidine. As a result, the momentum space information supported by the EMS experiment[8] of pyrimidine confirms the HOMO ($7b_2$) and the NHOMO ($2b_1$) conclusively in the present study.

## 4. Conclusions

Valence space ionization energy spectra and electron momentum spectra of the pyrimidine bases, that is, pyrimidine, cytosine, thymine and uracil have been investigated theoretically, using models that are validated by experiment. Vertical ionization spectra of the bases are produced using the SAOP/et-pVQZ and OVGF/TZVP models. It is found that for the highest point group symmetry species, pyrimidine, quantum mechanical models do not provide a conclusive assignment for the ionization states of HOMO and NHOMO, although their ionization energies can be accurately reproduced. In order to unambiguously assign the symmetries of the orbitals, theoretical calculations need to combined with an appropriate experiment

(such as EMS), since the shape of the orbital momentum distributions shows the symmetries of the orbitals conclusively. This connection between the orbital energy and corresponding orbital momentum distribution has also been used in the study of the most stable conformer for tetrahydrofuran[44].

In a similar outcome to that reported for the inner-shell study earlier[11], the valence ionization spectra of the pyrimidine bases exhibit certain similarities due to their similar aromatic hexagon ring structures. In addition, the valence ionization spectra of the NA bases, that is, cytosine, thymine and uracil, exhibit a larger degree of similarities in their ionization spectra, including the HOMO-LUMO regions, which may be a result of the substituted functional groups such as $-NH_2$, $-CH_3$ and C=O, which lower the molecular symmetry. Orbital momentum distributions reveal that pyrimidine exhibits less similarity to the NA bases than the NA bases do to each other. The high symmetry of pyrimidine restricts interchanges between the x- and y-components of its orbitals, whereas the NA bases (cytosine, thymine and uracil) are more flexible because they do not possess such symmetry restriction. This may be a part of the reason that pyrimidine does not appear in most of RNA and DNA polymers.

## ACKNOWLEDGMENTS

This work is supported by the Australian Research Council (ARC) under the Discovery Project (DP) Scheme and an award under the Merit Allocation Scheme on the National Computational Infrastructure Facility at the ANU. The Swinburne University Supercomputing Facilities are also gratefully acknowledged. One of the authors, ZY, thanks a Postgraduate Research Scholarship to Australia from the Ministry of Education (China), Swinburne University of Technology (SUT) for

hospitality, and the financial support from the National Natural Science Foundation of China (Grant No: 11104247).

## REFERENCES

[1] F. Wang, M. T. Downton, and N. Kidwani, J. Theo. Comp. Chem. **4**, 247 (2005).
[2] C. T. Falzon, F. Wang, and W. N. Pang, J. Phys.Chem. B **110**, 9713 (2006).
[3] M. K. Shukla and J. Leszczynski, J. Biomol. Struct. Dynam. **25**, 93 (2007).
[4] P. R. Callis, Annu. Rev. Phys. Chem. **34**, 329 (1983).
[5] B. Boddaiffa, P. Cloutier, D. Hunting, M. A. Huels, and L. Sanche, Science **287**, 1658 (2000).
[6] E. C. Friedberg, G. C. Walker, S. W, R. D. Wood, R. A. Schultz, and T. Ellenberger, *DNA Repair and Mutagenesis* (ASM Press, Washington, D. C., 2005).
[7] S. Kanvah, J. Joseph, G. B. Schuster, R. N. Barnett, C. L. Cleveland, and U. Landman, Accounts of Chemical Research **43**, 280 (2009).
[8] C. G. Ning, K. Liu, Z. H. Luo, S. F. Zhang, and J. K. Deng, Chem. Phys. Lett. **476**, 157 (2009).
[9] D. M. P. Holland, A. W. Potts, L. Karlsson, I. L. Zaytseva, A. B. Trofimon, and J. Schirmer, Chem. Phys. **353**, 47 (2008).
[10] A. B. Trofimov, J. Schirmer, V. B. Kobychev, A. W. Potts, D. M. P. Holland, and L. Karlsson, J. Phys. B: At. Mol. Phys. **39**, 305 (2006).
[11] F. Wang, Q. Zhu, and E. Ivanova, J. Synchrotron Rad. **15**, 624 (2008).
[12] P. Slavicek, B. Winter, M. Faubel, S. E. Bradforth, and P. Jungwirth, J. Am. Chem. Soc. **131**, 6460 (2009).
[13] J. Peeling and F. E. Hruska, CAN. J. CHEM. **56**, 1555 (1978).
[14] K. Fujii, K. Akamatsu, and A. Yokoya, J. Phys. Chem. B **108**, 8031 (2004).
[15] O. Dolgounitcheva, V. G. Zakrzewski, and J. V. Ortiz, J. Phys.Chem. A **107**, 822 (2003).
[16] A. W. Potts, D. M. P. Holland, A. B. Trofimov, J. Schirmer, L. Karlsson, and K. Siegbahn, J. Phys. B: At. Mol. Phys. **36**, 3129 (2003).
[17] G. Vall-llosera, B. Gao, A. kivimaki, M. Coreno, J. A. Ruiz, M. d. Simone, H. Agren, and E. Rachlew, J. Chem. Phys. **128**, 044316 (2008).
[18] D. Toffoli, P. Decleva, F. A. Gianturco, and R. R. Lucchese, J. Chem. Phys. **127**, 234317 (2007).
[19] S. D. Wetmore, R. J. Boyd, and L. A. Eriksson, Chem. Phys. Lett. **322**, 129 (2000).
[20] N. Russo, M. Toscano, and A. Grand, J. Comput. Chem. **21**, 1243 (2000).
[21] Y. J. Takahata, A. K. Okamoto, and D. P. Chong, Int. J. Quantum Chem. **106**, 2581 (2006).
[22] S. X. Tian, C. F. Zhang, Z. J. Zhang, X. J. Chen, and K. Z. Xu, Chem. Phys. **242**, 217 (1999).
[23] E. S. Kryachko, M. T. Nguyen, and T. Zeegers-Huyskens, J. Phys. Chem. A **105**, 1288 (2001).
[24] J. S. Kwiatkowski and J. Leszczynski, J. Phys. Chem. **100**, 941 (1996).
[25] O. Dolgounitcheva, V. G. Zakrzewski, and J. V. Ortiz, Int. J. Quantum Chem. **90**, 1547 (2002).

[26] O. Dolgounitcheva, V. G. Zakrzewski, and J. V. Ortiz, J. Phys. Chem. A **106**, 8411 (2002).
[27] M. Preuss, W. G. Schmidt, K. Seino, J. Furthmuller, and F. Bechstedt, J. Comput. Chem. **25**, 112 (2004).
[28] V. Feyer, et al., J. Phys. Chem. A **113**, 5736 (2009).
[29] J. V. Ortiz, et al., Chem. Phys. **347**, 360 (2008).
[30] N. Godbout, D. R. Salahub, J. Andzelm, and E. Wimmer, CAN. J. CHEM. **70**, 560 (1992).
[31] P. R. T. Schipper, O. V. Gritsenko, S. J. A. Gisbergen, and E. J. Baerends, J. Chem. Phys. **112**, 1344 (2000).
[32] M. J. T. Frisch, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. J. A.; Vreven, 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.; Kelna, 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, A.; 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.; Ak-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 Inc. Wallingford CT (2003).
[33] G. t. B. Velde, F. M.; Baerends, E. J.; Guerra, C. F.; Gisbergen, S. J. A. v.; Snijders, J. G.; Ziegler, T., SCM Theoretical Chemistry Vrije University Amsterdam the Netherlands ADF 2008.01.
[34] E.Weigold and I.E.McCathy, *Electron Momentum Spectroscopy* (Kluwer Academic/Plenum Publishers, New York, 1999).
[35] P. Duffy, M. E. Casida, C. E. Brion, and D. P. Chong, Chem. Phys. **165**, 183 (1992).
[36] P. Duffy, D. P. Chong, M. E. Casida, and D. R. Salahub, Phys. Rev. A **50**, 4707 (1994).
[37] L. Selvam, V. Vasilyev, and F. Wang, J. Phys. Chem. B **113**, 11496 (2009).
[38] G. Schaftenaar and J. H. Noordik, J. Comput.-Aided Mol. Design **14**, 123 (2000).
[39] M. E. Martin, M. L. Sanchez, M. A. Aguilar, and F. J. O. d. Valle, J. Mol. Struct. (Theochem) **537**, 213 (2001).
[40] W. C. Schneider and I. F. Halvestadt, J. Am. Chem. Soc. **70**, 2626 (1948).
[41] G. Bakalarski, P. Grochowski, J. S. Kwiatkowski, B. Lesyng, and J. Leszczynski, Chem. Phys. **204**, 301 (1972).
[42] C. G. Ning, Y. R. Huang, S. F.Zhang, J. K. Deng, K. Liu, Z. H. Luo, and F. Wang, J. Phys. Chem. A **112**, 11078 (2008).
[43] F. Wang, P. Duffy, and D. P. Chong, *"Nanoscale interaction and their applications: Essays in Honour of Ian McCarthy". Eds. F. Wang and M. J. Brunger*, Kerala, India, 2007).
[44] T. C. Yang, G. L. Su, C. G. Ning, J. K. Deng, F. Wang, S. F. Zhang, X. G. Ren, and Y. R. Huang, J. Phys. Chem. A **111**, 4927 (2007).

[45] N. S. Kim, Q. Zhu, and P. R. LeBreton, J. Am. Chem. Soc. **121**, 11516 (1999).

Table 1 Comparison of dipole moment of the nucleic acid pyrimidine bases (D)[a].

| Contents | Pyrimidine | Cytosine | Thymine | Uracil |
|---|---|---|---|---|
| Symmetry | $C_{2V}$ | $C_1$ | $C_S$ | $C_S$ |
| State | $X^1A_1$ | $X^1A$ | $X^1A'$ | $X^1A'$ |
| $\mu_x$ | 0 | 4.47 | 3.46 | -1.18 |
| $\mu_y$ | 2.43 | 5.03 | -2.89 | -4.40 |
| $\mu_z$ | 0 | 0.21 | 0 | 0 |
| $\mu_{Total}$ | 2.43 | 6.73 | 4.51 | 4.56 |
| Exp. | 2.33[39] | 6.0-6.5[41] | 4.13[40] | 4.16[40] |
| Ref[11] | 2.45 | 6.85 | 4.72 | 4.68 |

[a] Calculations are based on the B3LYP/TZVP model.

Table 2 (a) Comparison of ionization potentials of pyrimidine (eV),

| This work | | | EXP.[8] | Other work[8] | | |
|---|---|---|---|---|---|---|
| et-pVQZ | | TZVP | | Aug-cc-pVTZ | 6-311++G** | |
| Sy. | SAOP | OVGF (PS)* | EMS | HF | B3LYP | OVGF |
| $7b_2$ | 10.43 | 10.50(0.89)[a] | 9.8 | 11.34[d] | 7.29 | 9.83 |
| $2b_1$ | 11.64 | 9.75 (0.89)[a] | 10.5 | 10.32[d] | 8.19 | 10.40 |
| $11a_1$ | 11.66 | 11.37 (0.88)[b] | 11.3 | 12.92 | 8.60 | 11.36 |
| $1a_2$ | 12.44 | 11.22 (0.88)[b] | | 11.56 | 9.04 | 11.28 |
| $10a_1$ | 14.39 | 14.44 (0.88)[c] | 14.1 | 16.00 | 11.55 | 14.49 |
| $6b_2$ | 14.60 | 14.31 (0.81)[c] | | 16.25 | 11.76 | 14.49 |
| $1b_1$ | 15.05 | 14.65 (0.88)[c] | | 15.77 | 12.00 | 14.63 |
| $9a_1$ | 15.90 | 16.06 (0.87) | 15.7 | 17.76 | 13.09 | 16.25 |
| $5b_2$ | 17.06 | 17.31 (0.86) | 17.5 | 19.18 | 14.39 | 17.26 |
| $8a_1$ | 17.68 | 18.09 (0.85) | | 20.16 | 15.00 | 18.25 |
| $7a_1$ | 20.73 | | 20.6 | 24.31 | 18.24 | |
| $4b_2$ | 20.77 | | | 24.41 | 18.27 | |
| $6a_1$ | 24.25 | | 24.5 | 29.32 | 21.95 | |
| $3b_2$ | 26.75 | | 26.2 | 32.62 | 24.61 | |
| $5a_1$ | 29.25 | | 29.3 | 35.82 | 27.25 | |

* Spectroscopic pole strength (PS).

[a,b]OVGF obtained interchanged orbital symmetries when compared to the corresponding SAOP results.

[c]When compared to $10a_1$, $6b_2$ and $1b_1$ from SAOP, OVGF obtained the orbital symmetries are $1b_1$, $10a_1$ and $6b_2$.

[d]HF/Aug-cc-pVTZ obtained interchanged HOMO and NHOMO symmetries when compared to the SAOP method.

Table 2 (b) Comparison of ionization potentials of cytosine (eV).

| This work | | | EXP. | | Other work[45] | |
|---|---|---|---|---|---|---|
| Et-pVQZ | | TZVP | | | HF | |
| Sy. | SAOP | OVGF (PS) | PES[45] | PES[15] | 6-31G | 6-311 ++G** |
| 29 *a* | 10.21 | 9.47 (0.88) | 8.89 | 8.70-9.2 | 9.24 | 9.40 |
| 28 *a* | 10.45 | 9.67 (0.89) | 9.55 | 9.40-10.5 | 10.37 | 10.54 |
| 27 *a* | 10.74 | 9.94 (0.88) | 9.89 | 11.5-12.5 | 11.15 | 11.45 |
| 26 *a* | 10.80 | 12.13 (0.86) | 11.20 | 13.0-13.5 | 11.94 | 12.15 |
| 25 *a* | 12.98 | 13.14 (0.87) | 11.64 | | 13.50 | 13.49 |
| 24 *a* | 13.80 | 14.29 (0.88) | 12.93 | | 14.68 | 14.66 |
| 23 *a* | 14.20 | 15.06 (0.89) | 12.93 | | 16.06 | 16.31 |
| 22 *a* | 14.85 | 15.15 (0.83) | 13.86 | | 16.70 | 16.68 |
| 21 *a* | 15.27 | 15.31 (0.88) | 14.94 | | 17.11 | 17.18 |
| 20 *a* | 15.57 | 16.90 (0.88) | | | 17.23 | 17.06 |
| 19 *a* | 16.59 | 17.78 (0.86) | | | 18.56 | 18.61 |
| 18 *a* | 17.36 | 18.79 (0.86) | | | 19.66 | 19.68 |
| 17 *a* | 18.30 | 18.83 (0.85) | 18.02 | | | |
| 16 *a* | 18.36 | | 18.02 | | | |
| 15 *a* | 20.51 | | | | | |
| 14 *a* | 20.89 | | | | | |
| 13 *a* | 23.91 | | | | | |
| 12 *a* | 26.23 | | | | | |
| 11 *a* | 28.12 | | | | | |
| 10 *a* | 28.80 | | | | | |
| 9 *a* | 30.55 | | | | | |

**Figure Captions**

**Fig. 1** Structures and nomenclature of pyrimidine (P) and nucleic bases cytosine (C), thymine (T) and uracil (U).

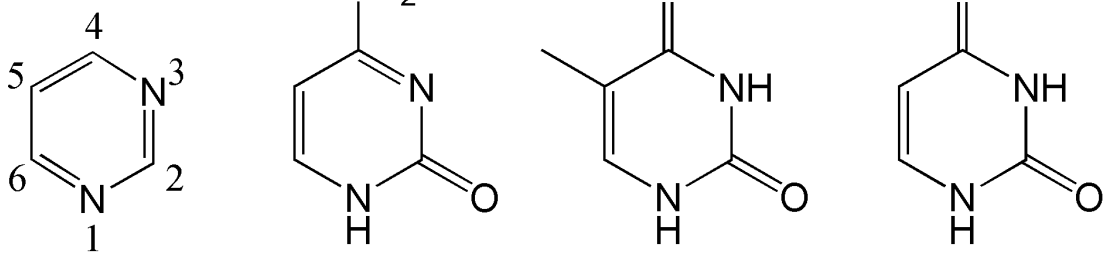
**Fig. 2** Comparison of simulated vertical ionization energy spectra of pyrimidine with experiment[10], and with simulated spectra of cytosine, thymine and uracil in the complete valence space. Here an FWHM of 1.0 eV is employed in the simulation.

**Fig. 3** Comparison of the outer valence orbital momentum distributions of pyrimidine with the experimentally resolved orbitals of $7b_2$, $2b_1$ and $9a_1$, respectively, under the total energy of 1500 eV plus binding energy. Double clicking the orbitals will activate the 3D-pdf feature.

**Fig. 4 (a)** Orbital momentum and charge distributions of the HOMOs of pyrimidine, cytosine, thymine and uracil.

**Fig.4(b)** Orbital momentum and charge distributions of the NHOMOs of pyrimidine, cytosine, thymine and uracil.

Fig.1




P C T U

`

Fig.2

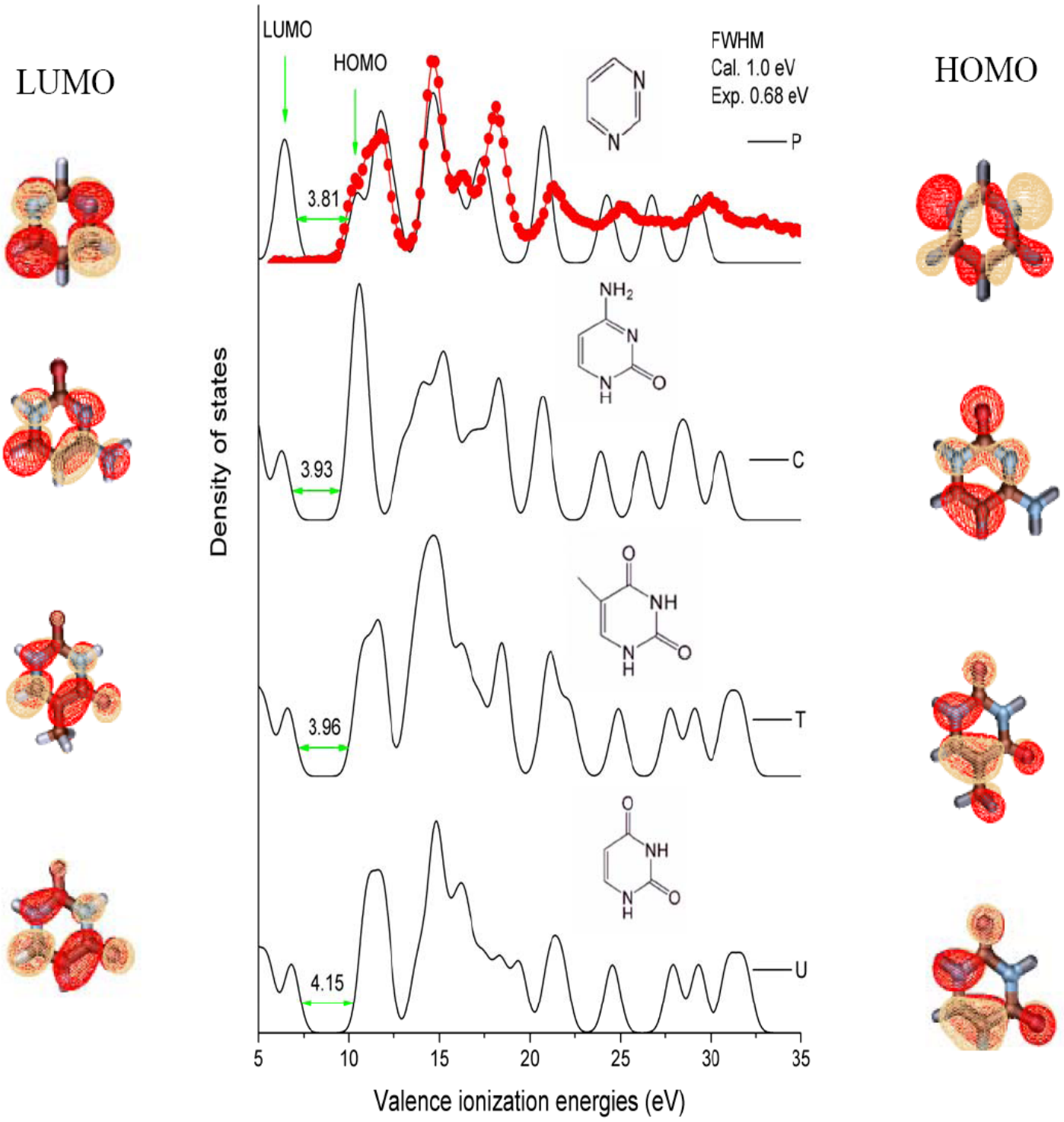
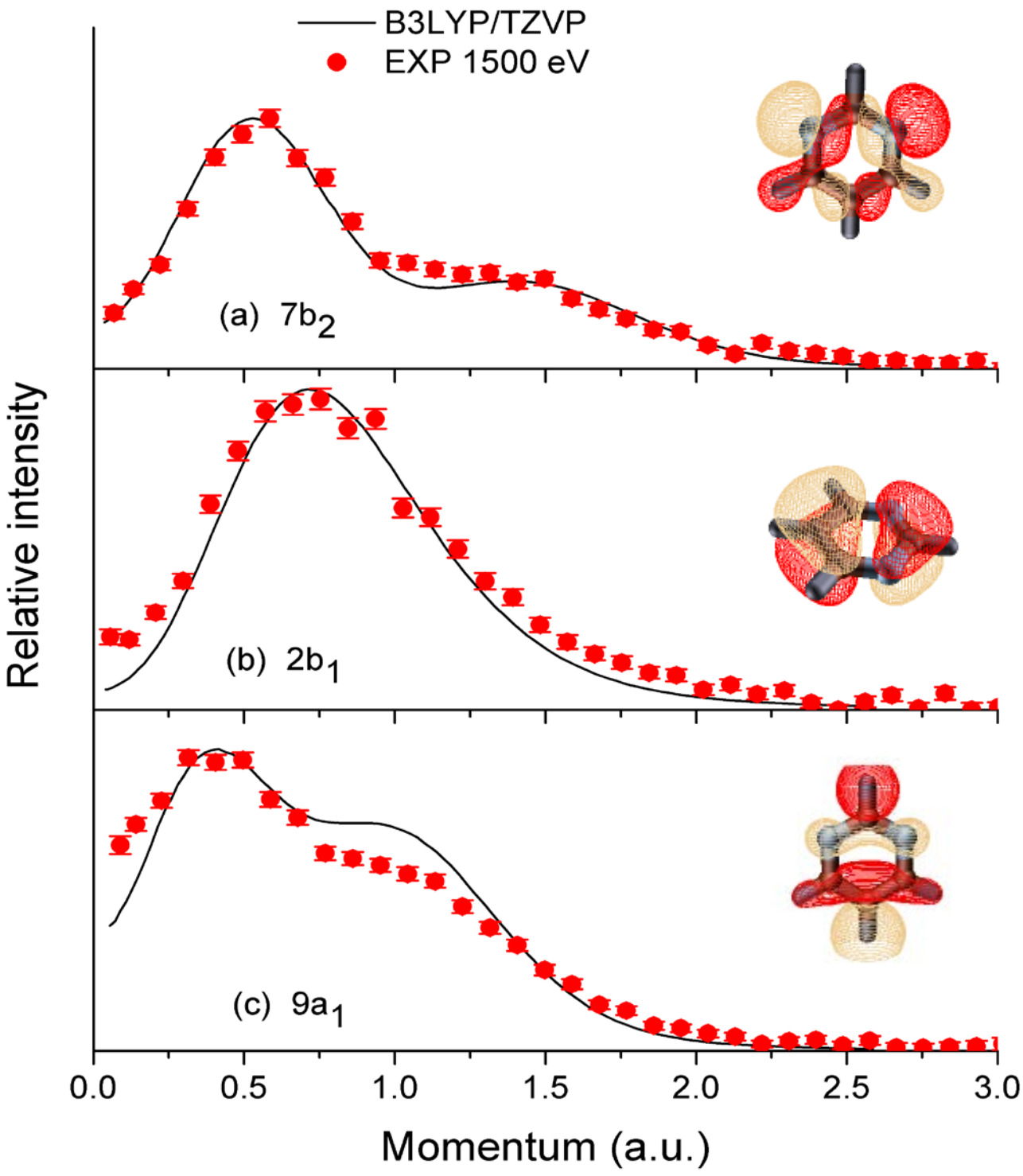

LUMO
HOMO
LUMO
HOMO
FWHM
Cal. 1.0 eV
Exp. 0.68 eV
P
3.81
C
3.93
T
3.96
U
4.15
Density of states
5
10
15
20
25
30
35
Valence ionization energies (eV)

Fig.3


B3LYP/TZVP
EXP 1500 eV
(a) 7b2
(b) 2b1
(c) 9a1
Relative intensity
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Momentum (a.u.)

.

Fig.4(a)

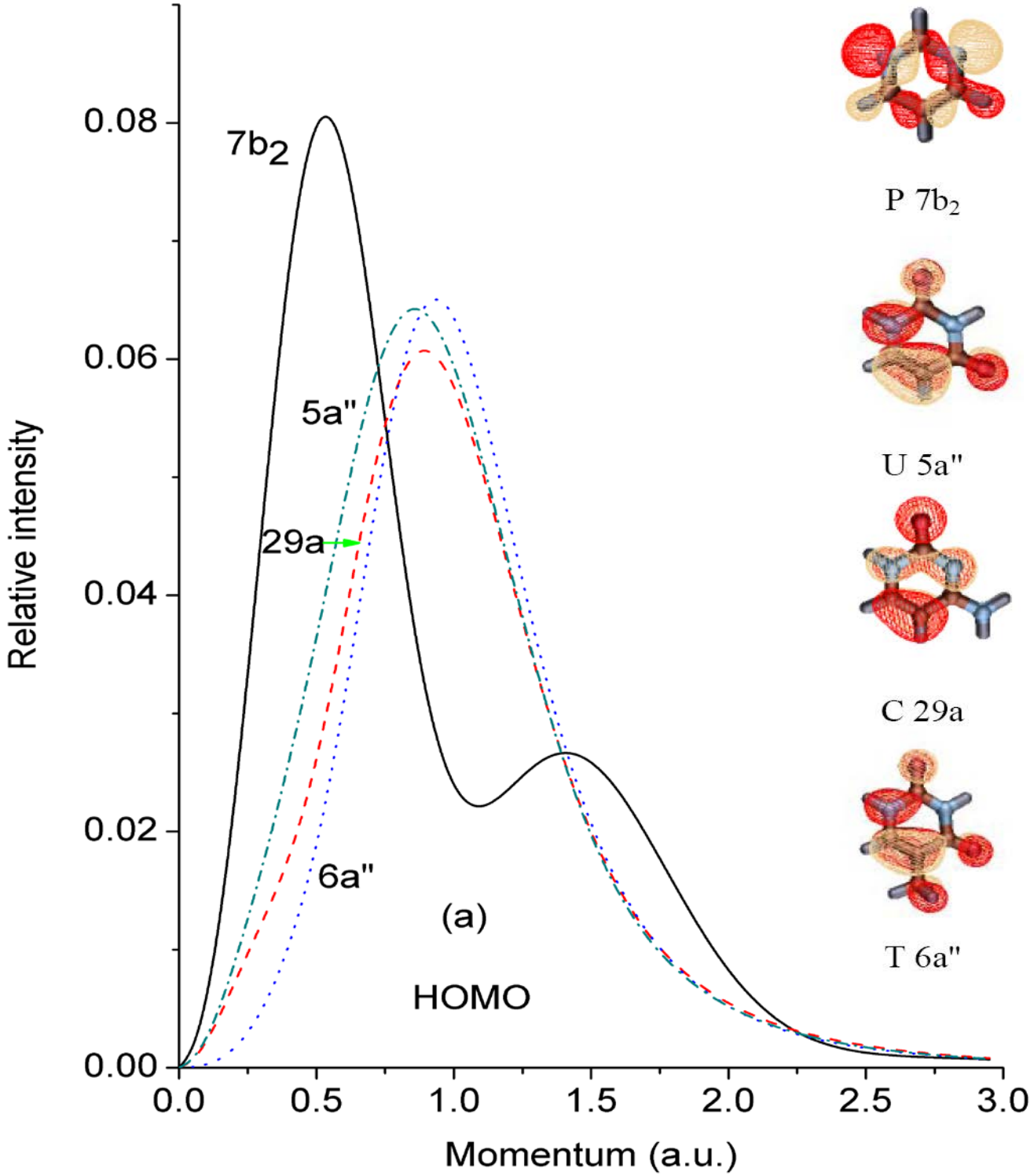
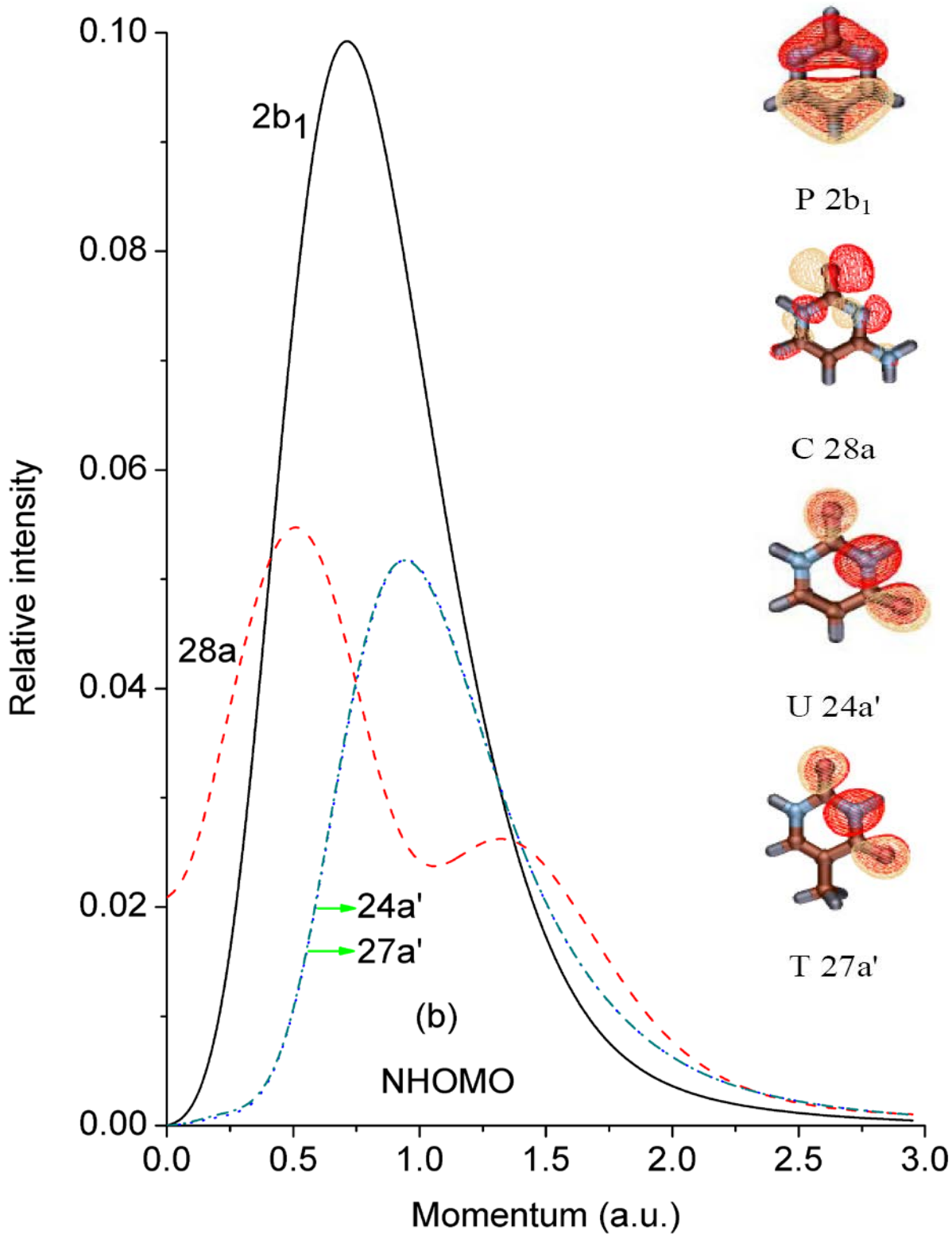

0.08
0.06
0.04
0.02
0.00
Relative intensity
7b2
5a"
29a
6a"
(a)
HOMO
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Momentum (a.u.)
P 7b2
U 5a"
C 29a
T 6a"

Fig. 4(b)


0.10
0.08
0.06
0.04
0.02
0.00
Relative intensity
2b1
28a
24a'
27a'
(b)
NHOMO
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Momentum (a.u.)
P 2b1
C 28a
U 24a'
T 27a'