Valence Orbital Response to Methylation of Uracil

Zejin Yang[1,2*], Wenning Pang[3], Patrick Duffy[4] 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]Polarized Physics Laboratory, Department of Physics, Tsinghua University, Beijing 100084, China;

[4]Chemistry Department, Kwantlen Polytechnic University, Richmond Campus, 8771 Lansdowne Road, Richmond, British Columbia, Canada V6X 3V8

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

Abstract

Molecular orbital signatures of the methyl group in thymine are identified using information from both coordinate and momentum spaces, in comparison with the RNA base uracil. The influence of the methyl group in thymine on the electronic structure of the uracil ring is thoroughly investigated using the dual space analysis method. The B3LYP/cc-PVTZ//B3LYP/TZVP calculations employed in the present work show that the attachment of the methyl group may be identified as orbital signatures locally, that is, $9a'$ (the methyl core orbital), $15a'$ in the inner valence shell of thymine and two orbitals ($2a''$ and $25a'$) in the outer valence shell of thymine. The fact that all the molecular orbitals (MOs) demonstrate small differences from the corresponding MOs of uracil is characteristic of the outer valence electron delocalization in thymine. Large changes in orbital energies do not directly lead to large changes in orbital momentum distributions, however, indicating a more anisotropic nature of the orbital wavefunctions. In addition, the highest occupied molecular orbitals (HOMOs) and next HOMOs (NHOMOs) of the two species do not appear to be significantly affected by the attachment of the methyl group, although the presence of the methyl group does result in certain degree of wavefunctions distortion. Furthermore, the current orbital ionization energies show good agreement with experimental data, and all of the spectroscopic pole strengths (PS) calculated from the OVGF method are larger than 0.85 which means that the current results are credible. Examination of the Raman spectrum reveals that the methyl substitution has enhanced the intensity of two antisymmetric stretch vibrations and one symmetric stretch vibration which is far lower than the corresponding hydrogen atom vibration in frequency.

1 Introduction

Study of the variation in chemical reactivity among nucleotides, nucleosides, H-bonded and stacked dimers of purines and pyrimidine bases, and other fragments of nucleic acids is of great importance to modern biology and biochemistry[1, 2], because the two nucleic acids deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) are the informational molecules of all living organisms. Besides storing and transmitting information, RNA forms structural and functional parts of units such as the ribosome and in some systems has a similar catalytic function to ribosomes. Therefore, detailed studies of the electronic structure[3-10] of nucleobases is essential for gaining an understanding of the influence of the mutation and repair mechanisms on the biological function of the nucleic acids. Moreover, the current interest in charge transfer in DNA is not restricted to its role in biology, since the advent of molecular electronics has stimulated interest in the possibility to exploit this species in fundamental mesoscopic electronic devices[11].

Nucleic acid analogs have also been used with proteins that bind to specific sites on DNA but do not catalyze reactions with it. The effects of analog substitutions can arise directly from lost interactions because of missing or added groups on the modified bases themselves or from indirect effects as a result of conformational changes in the DNA introduced by the presence of the analogs[12]. Base analogs can be used in the sequencing of DNA. Thymine and uracil, which are considered to be pyrimidine analogs, differ only by a methyl group linked to the ring in thymine but absent in uracil. However, thymine is a DNA base, whereas uracil serves as an RNA base. From a structural point of view thymine is a methylated derivative of uracil. One possible reason for its existence is that the methyl group in thymine may be able to prevent some mutagens from chemically modifying the DNA bases. An important

aspect in the repair mechanism of DNA, especially in base mismatches, is the ability to distinguish between strands. The methyl group in thymine makes it possible to distinguish it from uracil, which can be formed by the deamination of the cytosine base. If uracil were a DNA base, the deaminated cytosine could not be distinguished from uracil. Recent studies[13, 14] found roles for the methyl group as both an electron donor and an acceptor, and that which role it will take depends on whether the molecular environment is saturated or unsaturated.

Due to their biological importance as the chromosphere of DNA, a significant amount of work has been dedicated to the study of DNA and RNA bases. Ab initio quantum-mechanical calculations with inclusion of electron correlation have contributed significantly to our understanding of the molecular interaction of DNA and RNA bases. The molecular electronic structure and chemical bonding mechanism of the isolated bases contributes to a variety of properties and phenomena[15] such as the structure and hydrogen bonding of the base pairs, the nature of base stacking, the interaction between metal ions and nucleobases, solvent effects and non-planarity of isolated nucleobases and other monomer properties. However, many of the quantum mechanical (QM) methods and spectroscopic measurements concentrate on the energetic properties of species. However, energetic properties, such as the total energy and orbital energies of a molecule, are often more sensitive to isotropic properties rather than wavefunctions themselves. Because small energy changes in molecular orbitals do not necessarily result in similarly minor changes in anisotropic properties and wavefunctions, there is a need to use other properties which are able to differentiate the electronic structural signatures of conformational isomers and/or tautomeric species in addition to just the energetic properties. However, measurement of those anisotropic properties provides a challenging task when designing new

experimental techniques. For example, it took ten years[15] for an experiment[16] to show that the isolated adenine molecule is in fact non-planar rather than non-planar. The latter is still taught in biology textbooks[17]. Recently, studies of the photodynamics along the main decay paths of thymine after excitation to the lowest π,π^* state have been studied with MS-CASPT2 and semiclassical CASSCF dynamics calculations[18]. Photodynamic simulations of thymine focusing on relaxation into the first excited singlet have also been made[19].

Recently, studies of the effect of the attachment of a functional group has been a topic of biochemical interest, because the methylation of nucleotides and nucleosides has been thought to contribute to mutagenesis and carcinogenesis[1, 21, 22]. These studies have revealed the bonding mechanism at different chemical environments. Some examples include the orbital signature of the methyl group in L-alanine (by EMS[4]), the methylation of zebularine (by a quantum mechanical study incorporating an interactive three dimentional format[9]), the inner valence shell structures and spectra of the attachment of amino fragments to purine (by DFT[20]), and a comprehensive investigation of the inner valence shell of DNA/RNA bases from their parents purine and pyrimidine[5].

Electron momentum spectroscopy (EMS) and photoelectron spectra (PES) have been a powerful tool for the exploration of biological and molecular structures. Examples include the valence shell electronic structure of pyrimidine (by EMS [7]), and the valence shell photoelectron spectra of molecules including uracil and methyluracils[23], cytosine, thymine and adenine[24], and purine and pyrimidine[25]. Tautomerism in cytosine and uracil has been studied experimentally and theoretically[26]. In that study the authors have also summarized research progress on the study of tautomers up to that point. The core level electron excitation and

ionization spectra of thymine and adenine have also been investigated by photoabsorption and photoemission spectroscopy, together with *ab initio* calculations[27], and the ionization energies of aqueous nucleic acids of the pyrimidine nucleosides have been studied by PES and *ab initio* calculations[28]. In addition, Electron propagator methods have been used for the calculation of the vertical ionization energies of the five most stable tautomers of cytosine and to two oxo forms of 1-methylcytosine[29]. Finally, a thorough investigation of the Near-Edge X-ray absorption fine structure of DNA nucleobase thin films in the nitrogen and oxygen K-edge region has been carried out[30, 31].

DFT has been used extensively in the study of various properties and processes for nucleobases, for example the photoionization dynamics of uracil[32], the electron affinities and ionization potentials of nucleotides[33, 34], the core-electron binding energies of pyrimidine and purine[35], and the tautomerism of uracil[36, 37]. In addition, the infrared spectrum of cytosine has been studied employing Hartree-Fock methods[38], as well as the binding energies and Dyson orbitals for some DNA/RNA bases[29, 39-41], and the photoabsorption and photoemission (excitation and ionization) spectra of thymine and adenine have been investigated using the algebraic-diagrammatic construction (ADC) method[27].

In this paper, the calculated molecular orbital wavefunctions were studied in both configuration space into momentum space. The extra information obtained by examining the orbitals in the second space should allow us to obtain greater insight into the nature of the species' chemical bonding, since information in coordinate and momentum space allows for more definite identification of the signatures of orbital electron charge redistribution and therefore chemical bonding.

2 Method and Computational Details

Ground state geometries for thymine and uracil have been optimized using the DFT-based hybrid B3LYP functional with Dunning's augmented correlation corrected triple zeta plus polarization Gaussian basis set aug-cc-pVTZ. Single point calculations after the geometry optimizations were performed using the B3LYP functional and a triple-zeta valence polarised Gaussian basis set (TZVP)[42]. The B3LYP/TZVP model has been proved reliable for the calculation of a variety of properties and of orbital momentum distributions for molecular systems[3, 4, 43-48]. which were examined using electron momentum spectroscopy (EMS)[49]. The valence vertical ionization potentials are produced using both the OVGF/TZVP method/basis set, and the SAOP/et-pVQZ[50] functional/basis set. All B3LYP and OVGF calculations were performed using the Gaussian 03[51] computational chemistry package. The SAOP/et-pVQZ calculations were carried out using the ADF[52] package. Molecular orbitals (MOs) obtained in coordinate space (based on the B3LYP/TZVP model) were transformed into momentum space to give theoretical momentum distributions (TMD). It should be noted that the TMDs so calculated had embedded within them a number of approximations, such as the Born-Oppenheimer approximation, the independent particle approximation and the plane wave impulse approximation (PWIA). Within these approximations, the EMS cross-section is proportional to the target-ion overlap, which is the one electron Dyson orbital[49],

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

Here, $\vec{p}$ is the momentum of the target electron at the instant of ionisation. The Dyson orbital $\phi_j(p)$ in momentum space is approximated by the Kohn-Sham (KS) orbitals of the ground electronic states. The total valence momentum distributions of

the species are calculated by summing the valence orbital momentum distributions. The Fourier transforms were carried out using modifed HEMS code[53].

3 Results and Discussion

3.1 Properties of thymine and uracil

Rigid planar structures (C_S symmetries) of thymine and uracil in the gas phase have been found experimentally[25] and confirmed by the present calculations. Using a (lower) C_1 symmetry for the structures resulted in calculated total energies for thymine and uracil that were identical within the precision of the calculations. This results supports the necessity of studying anisotropic properties such as wavefunctions, dipole moment *etc*, to gain a better understanding of factors affecting wavefunction quality. The ground state of thymine and uracil is X^1A'.

Detailed studies of the two bases as regards the structural relaxation in thymine associated with the methyl group have been reported elsewhere[5]. The most significant geometrical changes in thymine with respect to the methyl group attachment are the out-of-plane changes to the backbone atoms. In uracil, all of the atoms are confined in the same plane. When a hydrogen atom at C(5) is replaced by a methyl group, the overall symmetry remains the same and the two hydrogen atoms H(7) and H(8) are symmetric with respect to the backbone plane.

3.2 Molecular orbital information in coordinate space

The ground electronic states of thymine and uracil have 66 and 58 doubly occupied MOs, respectively; the methyl group on thymine adds eight electrons, which form four doubly occupied orbitals: one in the core and three in the valence regions. a detailed study of the core orbital energies has been reported previously[7]; here we

only present valence orbital energies (see Tables 1 (a) and (b)). The core molecular orbital order and atom assignment for thymine and uracil is:

$1a'(O'_{(2)})2a'(O'_{(4)})3a'(N_{(1)})4a'(N_{(3)})5a'(C_{(2)})6a'(C_{(4)})7a'(C_{(6)})8a'(C_{(5)})9a'(C'_{(5)})$

$1a'(O'_{(2)})2a'(O'_{(4)})3a'(N_{(1)})4a'(N_{(3)})5a'(C_{(2)})6a'(C_{(4)})7a'(C_{(6)})8a'(C_{(5)})$

This shows clearly that thymine and uracil show same energy orders. Structurally, the primary difference between thymine and uracil is due to the methyl group. Therefore, thymine will have four more MOs than uracil. One of those methyl MOs in thymine can be easily identified as the $9a'(C'_{(5)})$ in the core shell and the other three MOs due to the extra methyl group can be found in the valence shell.

We are aware of no study that reports the IPs in the entire valence space for the pyrimidine bases. Table 1 gives the valence MO energies of thymine and uracil. The core orbitals are highly localized on the non-hydrogen atoms, and as a result the extra core MO of thymine does not seem to affect the other core MOs significantly, except for that of C(5), which directly connects to the methyl group. The detailed core orbital energies are from previous studies[7] and used the LB94/et-PVQZ functional/basis set. Because the species have Cs symmetry, when the hydrogen atom bonded with the C(5) atom is replaced by a methyl group, it causes subtle changes in all bonds of the species and all the orbital energies show greater or lesser shifts in energy. Though close, none of the orbital energies are degenerate as all the atoms in the six member pyrimidine rings are located at nonequivalent sites.

In Figure 2 (a) and (b), the inner (E>17 eV) and outer (E<17 eV) valence orbital energy diagrams of thymine and uracil are presented. In general, the orbital energy shifts in the inner valence shells of the species are small except for the methyl signature MOs of thymine. The largest orbital energy shift in the inner valence shell is found to be the pair of $18a'$ in thymine and $16a'$ in uracil. A full discussion of the

methyl signature orbitals will be given in the next section. The outer valence orbitals shows some larger and some smaller shifts in energy between the two molecules, most likely due to the delocalization of electrons. In the outer valence shells, the anti-symmetric (a″) orbitals of the species are formed by the p_z AOs of the non-hydrogen atoms and so are likely used in π-bonds, leaving the molecular plane (σ_h) as the nodal plane. Symmetry also dictates that, when the methyl group attaches to uracil on atom C(5), the a″ symmetry of the MOs will not correlate to a different symmetry such as a′ since the molecular plane remains in the species.

The outer valence shells of the species might be divided roughly into three zones; the middle zone, which contains the methyl signature MOs in thymine, exhibits significant changes. Therefore, Figure 2(b) demonstrates that the attachment of the methyl group affects the electronic states of the species in a more complex manner than just as a simple addition. When compared to uracil, the orbitals in the valence space (inner and outer) of thymine may be grouped into four types: not very affected by the methyl group; methyl-disturbed orbitals; methyl-affected orbitals; and methyl-dominated signature orbitals. The individual orbitals have been listed in Table 1. The methyl-dominated signature orbitals are a direct result of the addition of the methyl group, and so are referred to as signature orbitals of the methyl group. The methyl-affected orbitals are those with an apparent energy shift in the immediate vicinity of the signature orbitals; they are usually located in the immediate (geographic) vicinity of the signature orbitals. The methyl-disturbed orbitals are those that show only a slight energy shift. These orbitals are usually influenced directly by the methyl-affected orbitals in their geographic vicinity, and only in a secondary way by the signature orbitals. The orbitals that are not very affected are those with only minor energy shifts and very similar MDs. They are located in a region sufficiently

geographically far away from the methyl signature orbitals and so show only very small energy perturbations. For these orbitals, therefore, the orbital relaxation to accommodate the insertion of the methyl group seems negligible.

3.3 Molecular orbital momentum distributions

Figures 3-6 give the orbital MDs for the inner and outer valence shells for thymine (T) and uracil (U). As indicated in Figure 2, the four inner most valence MOs of T and U, i.e., 10 a′-13 a′ for T, and 9 a′-12 a′ for U, exhibit only very small shifts in their orbital energies. That they are relatively unchanged by the addition of the methyl group is further supported by the shapes of their orbital MDs in momentum space (Figure 4). The orbital MDs of the innermost valence MOs of both T and U are almost identical, with negligible perturbation in the small momentum region of each orbital pair. Orbitals 20 a′ (T) and 27 a′ (U) also show very similar orbital energies and spectra shapes. Selected MDs for the relatively unaffected orbitals are shown in Figure 4: are 10 a′ (T) and 9 a′ (U) with distributions dominated by contributions from s electrons, 20 a′ (T) and 18 a′ (U) with distributions dominated by p electrons mixed with some *s*-electron components, and 27 a′ (T) and 24 a′ (U) with distributions dominated by *p*-dominated electrons. From the relatively unaffected MOs it is clearly seen that all of the different kinds of orbitals (*s*-, *p*-, and *sp*-electron dominated MDs) show very weak dependence on the attachment of the methyl group. The influence of the methyl group on orbitals 10 a′-13 a′, 20 a′ and 27 a′ for T, and 9 a′-12 a′, 18 a′ and 24 a′ for U may therefore be taken to be insignificant.

Orbitals 27 a′ (T) and 24 a′ (U) belong to NHOMO and contribute to chemical reactions. The energy- and charge-based analysis clearly shows that the methylation has caused little influence to the frontier orbital, as also demonstrated by the very

minor change in orbital energy, the indistinguishable MDs, and very similar charge distributions. The electron charge distribution further confirms that the NHOMO is not active in chemical reactions, as only half of the typical charge density in the pyrimidine ring is present in this orbital, which is formed mainly by $2p_z$ electrons and more likely is used to form π bonds.

Figure 5 shows selected methyl perturbed orbitals, as chosen by their small orbital energy shifts and shape variations. The MDs for these orbitals show significant changes in shape at low momentum (P<0.5 a.u.). Examples of this phenomenon include the 17 a′ (T) and 15 a′ (U) orbitals, which show different shapes at low momentum; similar types of shape differences are found for the 5 a″ in T and 4 a″ in U (NHOMO) orbital pair. The 21 a′ (T) and 19 a′ (U) show similar shapes but with different intensities at about 0.5 a.u., which roughly corresponds to the long range region in coordinate space. It is interesting to note that the 21 a′ (T) and 19 a′ (U) orbital pair is dominated by contributions from p_z electrons but shows a′ symmetry.

Figure 6 shows some of the methyl-affected orbitals, i.e. those with significant orbital energy shifts caused by the presence of the methyl group. For these orbitals, it is evident that the methyl group has contributed much to the charge distributions and therefore the MDs of the orbitals, which exhibit significant variations in shape and intensity. For these orbitals, the change in shape of the MDs extends to intermediate momentum regions -- about 1.0-1.5 a.u.. Orbitals 14 a′ and 16 a′ in T show significant shifts relative to orbitals 13 a′ and 14 a′ in U. Because of the insertion of methyl signature orbital 15 a′, the resulting orbitals 14 a′ and 16 a′ shift towards lower and higher energy respectively. The direct influence of the inserted orbital 15 a′ is to cause marked changes in the Dyson orbital distributions. For example, orbital 14 a′ (T) shows much greater intensity than orbital 13 a′ (U), although the shapes of the MDs

still seem relatively similar. This orbital is primarily for the bonding hydrogen atom at C(5). The only difference in shape may be due to the change in mixture of the s and p orbitals in U compared to T. Another example of significant shape change is found when comparing the 2a″ in U to the 3a″ in T. As a result of the insertion of the methyl group, the 3a″ orbital experiences a double influence from the signature 2a″ and 25a′ orbitals, resulting in an influence on the electron distribution that extends to about 2.0 a.u.. In contrast to the case for the 14a′ in T and 13a′ in U pair, where methyl insertion caused the intensity to increase significantly at low momentum region (p<0.50 a.u.) and high momentum (p>1.15 a.u.), a decrease in intensity at intermediate momentum regions (0.50-1.15 a.u.) is evident here. The *p*-dominated profile also varies with the insertion of the methyl group, changing to an *sp*-hybrid distribution. This change may be confirmed from the charge distributions, which change to p_z to p_x or p_y together with partial *s* contributions.

The HOMO, 6a″ in T and 5a″ in U, should be the most important frontier orbitals in chemical reactions. The MDs clearly show that the peak has shifted towards the higher momentum region as a whole and the intensity shows a slight increase with the methylation of U. The same *p*-dominated profiles are still evident for both orbitals, but the intensity of the T peak has decreased greatly at low momentum (p<1.0 a.u.). Analysis of the current HOMO and the NHOMO above reveals that large orbital energy shifts in the HOMO has caused large changes in their MDs or wavefunctions, as is consistent with the charge distributions of methyl in T, whereas the small orbital energy shifts in the NHOMO is a predictor of small changes in MDs or wavefuncions (as has been confirmed by the charge distributions). The current HOMO and NHOMO variations show some similarities when compared to the corresponding orbitals of L-alanine[4] but the current NHOMO shows very little

influence from the methyl group, whereas the L-alanine NHOMO shows obvious variations on methylation.

3.4 Signatures of the methyl MOs in the valence shell

As indicated in the previous sections, the methyl group has four signature MOs in the valence shell, of which one of them is in the inner valence shell (i.e.15 a′) of thymine. The two methyl MO signatures in the outer valence shell are identified from the information from both coordinate space (orbital energies) and momentum space (orbital MDs) as 2 a″ and 25 a′.

Figure 3 displays the signature MOs of the methyl group of T, which do not have any correlated MOs in uracil. Therefore, these three MOs (15 a′, 2 a″ and 25 a′) are purely MO signatures of the methyl group. As clearly indicated by their orbital symmetries and orbital MDs in Figure 3, the 2 a″ MO is anti-symmetric with respect to the molecular plane and arises from *p*-electron contributions related to the methyl carbon atom, whereas the MD of 2 a″ is *s*-dominated (as evidenced by the low momentum intensity), with some *p*-contributions at higher momentum. The 15 a′ MO has both *s* and *p* electron contributions, and orbital 25 a′ has *p*-electron contributions only. Clearly, these three orbitals show a significant charge distribution located on the methyl group. Orbital 25 a′ is distributed on both sides of the molecular plane, indicating *p* electron contributions. In contrast, the 25 a′ MO, which is again concentrated in the area of the methyl group and its vicinity is contained largely within the pyrimidine plane, indicating a strong *s* contribution and more of a σ bonding character.

Comparison of the current three methyl-dominated orbitals with those from previous L-alanine and glycine results[4] shows quite obvious similar tendencies,

meaning that contributions to the charge distributions from the methyl group show similar orientation in space. For example, the 15a′ orbital in T gives a similar *sp*-hybrid MD to the 19a orbital in L-alanine. In addition, the current 25a′ orbital shows *p*-dominated MDs similar to those of the 20a orbital of L-alanine, and the current 2a″ orbital shows an *sp*-hybrid MD which is similar to that of the 12a orbital of L-alanine. From the similar MDs formed by methylation one could conclude that distributions of methyl charge exhibit great similarities between the two molecules, but the intensity of the T orbitals is greater than those of L-alanine, which may be due to strong interactions of the methyl group with the pyrimidine ring.

3.5 Raman spectroscopy

The current work shows the Raman shows the spectrum in the region of the most variation between the spectra of the two molecules, which is the range from 3000 to 3750cm^{-1}, in Figure 7. (The variation from 0-3000 cm^{-1} is small and thus will not be discussed further here.) Figure 7 shows four sharp peaks for U; the number of peaks increases to six when one hydrogen atom located at C(5) is replaced by a methyl group. Clearly, only one peak shows little change because of the substitution: the symmetric stretch vibration of the hydrogen atom located at N(3). The other two small changes visible may be ascribed to the perturbation caused by the insertion of the methyl group; they are one symmetric stretch that has increased in intensity but that has not changed in frequency (the hydrogen located at N(1)), and a symmetric stretch of the hydrogen on C(6) that has shifted towards a slightly lower frequency and decreased slightly in intensity. The substitution of a hydrogen atom located at C(5) has resulted in the corresponding vibration frequency shifting to a lower frequency and being split into three peaks. The three hydrogen atoms on the methyl group

interact with the pyrimidine ring, and as a result the frequency shift is greater than 100 cm^{-1}, as can be clearly seen from the figure. Analysis of the three vibrations due to the methyl hydrogens also reveals that the maximum peak intensity arises out of the symmetric stretch vibration of the three hydrogen atoms in the methyl group, although that peak occurs at the lowest frequency of the three. The other two peaks correspond to the antisymmetric stretches of H(7) and H(8), and of H(7,8) and H(9). The H(7,8) and H(9) antisymmetric stretch vibration occurs at a lager frequency and smaller peak intensity when compared to the antisymmetric stretch of atoms H(7) and H(8). Generally speaking, the methylation has added two anti-symmetric stretch vibrations and the corresponding symmetric stretch vibration has been greatly increased in intensity and shifted towards lower frequency.

4 Conclusions

The electronic structures of thymine and uracil in their ground electronic states (X^1A') have been studied using the B3LYP/cc-pVTZ//B3LYP/TZVP models. In this study, the DNA base thymine has been treated as if the RNA base uracil had a methyl group attached, and the information was analyzed in both configuration space and momentum space.

This study has shown that the signature MOs of the methyl group spread into all shells including the core, inner valence, and outer valence. The deeply rooted effect on the electronic structure of thymine compared to uracil, and the fact that the frontier MOs (HOMO and NHOMO) do not show significant changes in electronic structure when comparing the two species indicates that more comprehensive information is needed to study the characteristics of thymine and uracil. It has also been shown that large energetic changes such as might arise from orbital energies can cause significant

changes to the orbital wavefunctions and, as a result, the full picture of molecules lies in the comprehension of both its energy and wavefunction. Finally, changes in orbital wavefunctions may sometimes be more easily assessed in momentum space via momentum distributions because of the increased sensitivity of the momentum distribution to large-r discrepancies between wavefunctions.

The present work has also identified the methyl signature MOs of thymine and identified how they interact with those of uracil in both configuration and momentum spaces. In configuration space, the attachment of the methyl group shows that the methyl group has deep roots in the electronic structure of thymine and causes the orbital energy spectrum to shift, in particular the outer valence orbital energies. The distortion of the wavefunction as a result of the methyl addition is somewhat reduced, however, as the pyrimidine ring acts as a kind of buffer. In momentum space, significant attachment-related changes can be identified on an individual molecular orbital basis: the four signature MOs of $9a'$, $15a'$, $2a''$ and $25a'$, though many other MOs also experience some degree of change. The present work provides a comprehensive and innovative method based on dual space analysis to study electronic structures of DNA and RNA bases.

Raman spectroscopy provides more insight into the effect of the methyl insertion on uracil. Analysis showed that the methyl group has resulted in a significant change to the vibration of the substituted hydrogen atom, whereas the other vibrations have not been affected as much in the current spectroscopic ranges. In addition, little influence is visible in the range from 0-3000 cm^{-1}.

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. Swinburne University Supercomputing Facilities is also gratefully acknowledged. One of the authors, ZY, thanks the Ministry of Education (China) for a Postgraduate Research Scholarship to Australia, Swinburne University of Technology (SUT) for hospitality, and financial support from the National Natural Science Foundation of China (Grant No: 11104247).

References

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

[2] J. Sponer, P. Hobza, and J. Leszczynski, *Computational molecular biology* (Elsevier, Amsterdam, 1999).

[3] C. T. Falzon and F. Wang, J. Chem. Phys. **123**, 214307 (2005).

[4] C. T. Falzon, F. Wang, and W. N. Pang, J. Phys. Chem. B **110**, 9713 (2006).

[5] F. Wang, Q. Zhu, and E. Ivanova, J. Synchrotron Rad. **15**, 624 (2008).

[6] F. F. Chen and F. Wang, molecules **14**, 2656 (2009).

[7] C. G. Ning, K. Liu, Z. H. Luo, et al., Chem. Phys. Lett. **476**, 157 (2009).

[8] S. Saha and F. Wang, J. Phys. Conf. Ser. **185**, 012040 (2009).

[9] L. Selvam, V. Vasilyev, and F. Wang, J. Phys. Chem. B **113**, 11496 (2009).

[10] Q. Zhu, F. Wang, and E. Ivanova, J. Synchrotron Rad. **16**, 545 (2009).

[11] Y. Ito and E. Fukusaki, J. Mol. Cata. B-Enzymatic **28**, 155 (2004).

[12] C. R. Ailen and R. I. Gumport, *Base analogs in study of restriction enzyme-DNA interactions in methods in enzymology, protein-DNA interactions* (Harcourt Brace Jovanovich, Academic Press 1991).

[13] G. Tasi, F. Mizukami, and I. Palinko, J. Mol. Struct. **401**, 21 (1997).

[14] L. J. Sathre, N. Berrah, J. D. Bozek, et al., J. Am. Chem. Soc. **123**, 10729 (2001).

[15] J. Sponer and P. Hobza, Collect. Czech. Chem. Commun. **68**, 2231 (2003).

[16] F. Dong and R. E. Miller, Science **298**, 1227 (2002).

[17] W. Saenger, *Principles of nucleic acid structure* (Pringer, New York, 1988).

[18] D. Asturiol, B. Lasorne, M. A. Robb, et al., J. Phys. Chem. A **113**, 10211 (2009).

[19] J. J. Szymczak, M. Barbatti, J. T. S. Hoo, et al., J. Phys. Chem. A **113**, 12686 (2009).

[20] S. Saha, F. Wang, J. B. MacNaughton, et al., J.Synchrotron Rad. **15**, 151 (2008).

[21] Q. Zhu and P. R. LeBreton, J. Am. Chem. Soc. **122**, 12824 (2000).

[22] N. S. Kim and P. R. LeBreton, J. Am. Chem. Soc. **118**, 3694 (1996).

[23] D. M. P. Holland, A. W. Potts, L. Karlsson, et al., Chem. Phys. **353**, 47 (2008).

[24] A. B. Trofimov, J. Schirmer, V. B. Kobychev, et al., J. Phys. B: At. Mol. Opt. Phys. **39**, 305 (2006).

[25] A. W. Potts, D. M. P. Holland, A. B. Trofimov, et al., J. Phys. B: At. Mol. Opt. Phys. **36**, 3129 (2003).

[26] V. Feyer, O. Plekan, R. Richter, et al., J. Phys. Chem. A **113**, 5736 (2009).

[27] O. Pleken, V. Feyer, R. Richter, et al., Chem. Phys. **347**, 360 (2008).

[28] P. Slavicek, B. Winter, M. Faubel, et al., J. Am. Chem. Soc. **131**, 6460 (2009).

[29] O. Dolgounitcheva, V. G. Zakrzewski, and J. V. Ortiz, J. Phys. Chem. A **107**, 822 (2003).

[30] K. Fujii, K. Akamatsu, and A. Yokoya, J. Phys. Chem. B **108**, 8031 (2004).

[31] G. Vall-llosera, B. Gao, A. Kivimaki, et al., J. Chem. Phys. **128**, 044316 (2008).

[32] D. Toffoli, P. Decleva, F. A. Gianturco, et al., J. Chem. Phys. **127**, 234317 (2007).

[33] S. D. Wetmore, R. J. Boyd, and L. A. Eriksson, Chem. Phys. Lett. **322**, 129 (2000).

[34] N. Russo, M. Toscano, and A. Grand, J. Comput. Chem. **21**, 1243 (2000).

[35] Y. J. Takahata, A. K. Okamoto, and D. P. Chong, Int. J. Quantum Chem. **106**, 2581 (2006).

[36] S. X. Tian, C. F. Zhang, Z. J. Zhang, et al., Chem. Phys. **242**, 217 (1999).

[37] E. S. Kryachko, M. T. Nguyen, and T. Zeegers-Huyskens, J. Phys. Chem. A **105**, 1288 (2001).

[38] J. S. Kwiatkowski and J. Leszczynski, J. Phys. Chem. **100**, 941 (1996).

[39] O. Dolgounitcheva, V. G. Zakrzewski, and J. V. Ortiz, Int. J. Quantum Chem. **90**, 1547 (2002).

[40] O. Dolgounitcheva, V. G. Zakrzewski, and J. V. Ortiz, J. Phys. Chem. A **106**, 8411 (2002).

[41] M. Preuss, W. G. Schmidt, K. Seino, et al., J. Comput. Chem. **25**, 112 (2004).

[42] N. Godbout, D. R. Salahub, J. Andzelm, et al., Can. J. Chem. **70**, 560 (1992).

[43] F. Wang, M. J. Brunger, I. E. McCarthy, et al., Chem. Phys. Lett. **382**, 217 (2003).

[44] M. Downton and F.Wang, Chem. Phys. Lett. **384**, 144 (2004).

[45] F. Wang and M. Downton, J. Phys. B: At. Mol. Opt. Phys. **37**, 557 (2004).

[46] S. Saha, F. Wang, C. T. Falzon, et al., J. Chem. Phys. **123**, 124315 (2005).

[47] F. Wang, J. Mol. Struct.: THEOCHEM **728**, 31 (2005).

[48] F. Wang, M. T. Downton, and N. Kidwani, J. Theor. Comput. Chem. **4**, 247 (2005).

[49] I. McCarthy and E. Weigold, Rep. Prog. Phys. **54**, 789 (1991).

[50] P. R. T. Schipper, O. V. Gritsenko, S. J. A. Gisbergen, et al., J. Chem. Phys. **112**, 1344 (2000).

[51] 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).

[52] F. M. G. t. B. Velde, E. J. Baerends, C. F. Guerra, et al., SCM Theoretical Chemistry Vrije University Amsterdam the Netherlands ADF (2008.01).

[53] P. Duffy, M. E. Casida, C. E. Brion, et al., Chem. Phys. **165**, 183 (1992).

[54] P. R. Callis, Annu. Rev. Phys. Chem. **34**, 329 (1983).

[55] M. K. Shukla and J. Leszczynski, J. Biomol. Struct. Dynam. **25**, 93 (2007).

Table 1 Classification of orbital types in thymine and uracil

type	thymine		uracil	
	Inner	outer	inner	outer
relatively unaffected	10a′,11a′,12a′,13a′	27a′	9a′,10a′,11a′,12a′	24a′
	20a′		18a′	
methyl-disturbed	17a′,19a′	21a′,23a′,26a′	15a′,17a′	19a′,21a′23a′
		1a″,5a″		1a″,4a″
methyl-affected	14a′,16a′,18a′	22a′,24a′	13a′,14a′,16a′	20a′,22a′
		3a″,4a″,6a″		2a″,3a″,5a″
methyl-dominated	15a′	2a″,25a′		

Table 1 (a) Ionization potentials of thymine (eV)

This work			EXP.		Other works	
Et-pVQZ		TZVP				6-311++G**
Sy.	SAOP	OVGF (PS)	PES[1]	PES[54]	P3[54]	OVGF[1]
6 a''	10.56	9.24 (0.89)	9.19	9.10	9.14	8.85
27 a'	10.96	10.52 (0.88)[b]	10.14	10.0	9.95	10.46
26 a'	11.71	10.19 (0.88)[b]	10.89[a]	10.8-11.0	10.43	11.36
5 a''	11.73	11.06 (0.88)[b]	10.45[a]	10.4	10.99	10.46
4 a''	13.39	12.62 (0.87)	12.27	12.3	12.52	12.52
25 a'	13.86	13.85 (0.90)	13.31			13.92
24 a'	14.30	13.79 (0.87)[d]	14.90			13.83
3 a''	14.32	14.02 (0.89)[d]	15.75			13.81
23 a'	14.91	14.78 (0.88)[e]				15.02
2 a''	14.97	15.06 (0.89)[e]				14.85
22 a'	15.30	15.38 (0.88)				15.49
21 a'	16.05	16.22 (0.85)[f]				16.35
1 a''	16.46	16.25 (0.88)[f]				16.08
20 a'	17.26	17.58 (0.88)	16.86			17.37
19 a'	18.35	18.76 (0.86)	18.03			
18 a'	18.57		18.03			
17 a'	20.91					
16 a'	21.30					
15 a'	22.22					
14 a'	24.88					
13 a'	27.76					
12 a'	29.11					
11 a'	30.87					
10 a'	31.68					

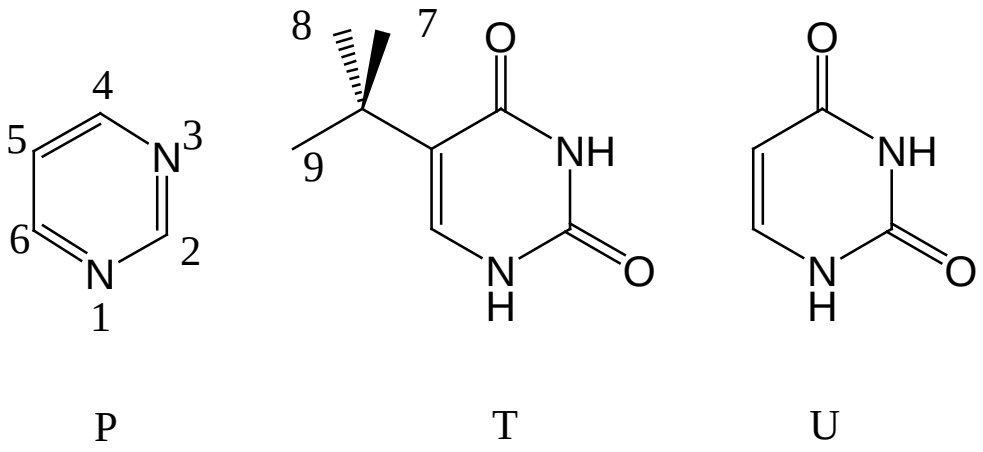
[a]The experimental order of 26 a' and 5 a'' is opposite to that of the current SAOP results.

[b]OVGF gives the orbital order 5 a'', 27 a' and 26 a', in contrast to SAOP, which gives 27 a', 26 a' and 5 a''. [d,e,f]Orbitals 3 a'' and 24 a', 2 a'' and 23 a', 1 a'' and 21 a' are swapped compared to the corresponding SAOP orbital order.

Table 1 (b) Ionization potentials of uracil (eV)

This work			Exp.		Other work	
Et-pVQZ		TZVP				6-311++G**
Sy.	SAOP	OVGF (PS)	PES[23]	PES[55]	P3[54]	OVGF[23]
$5\,a''$	10.95	9.65 (0.89)	9.46	9.45-9.6	9.54	9.26
$24\,a'$	11.07	10.62 (0.88)		10.02-10.23 [a]	10.15	10.54
$4\,a''$	11.86	10.26 (0.88)	10.08, 10.44	10.5-10.6[a]	10.52	10.55
$23\,a'$	11.88	11.21 (0.88)	10.90	10.9-11.2	11.12	11.50
$3\,a''$	13.82	13.04 (0.87)	12.53	12.5-12.7	12.91	12.82
$2\,a''$	14.70	14.08 (0.85)	13.5			14.14
$22\,a'$	14.77	14.81 (0.89)	14.1			14.71
$21\,a'$	14.94	15.01 (0.89)	14.5			15.05
$20\,a'$	15.69	15.82 (0.88)				15.87
$19\,a'$	16.18	16.36 (0.88)[c]	15.4[b]			16.15
$1\,a''$	16.57	16.31 (0.85)[c]				
$18\,a'$	17.43	17.79 (0.88)				16.45
$17\,a'$	18.38	18.76 (0.86)				
$16\,a'$	19.42					
$15\,a'$	21.07					
$14\,a'$	21.77					
$13\,a'$	24.55					
$12\,a'$	27.91					
$11\,a'$	29.31					
$10\,a'$	30.98					
$9\,a'$	31.85					

[a]Exp.[26] obtained the opposite order of $24\,a'$ and $4\,a''$ compared to the current SAOP method, and [b]experimental measurement assigned $19\,a'$ as $1\,a''$.
[c]The only orbital order difference between OVGF and SAOP is that $19\,a'$ is swapped with $1\,a''$.



P T U

Fig. 1 Schematic numbering structures for thymine and uracil derived from the nucleic acid pyrimidine.

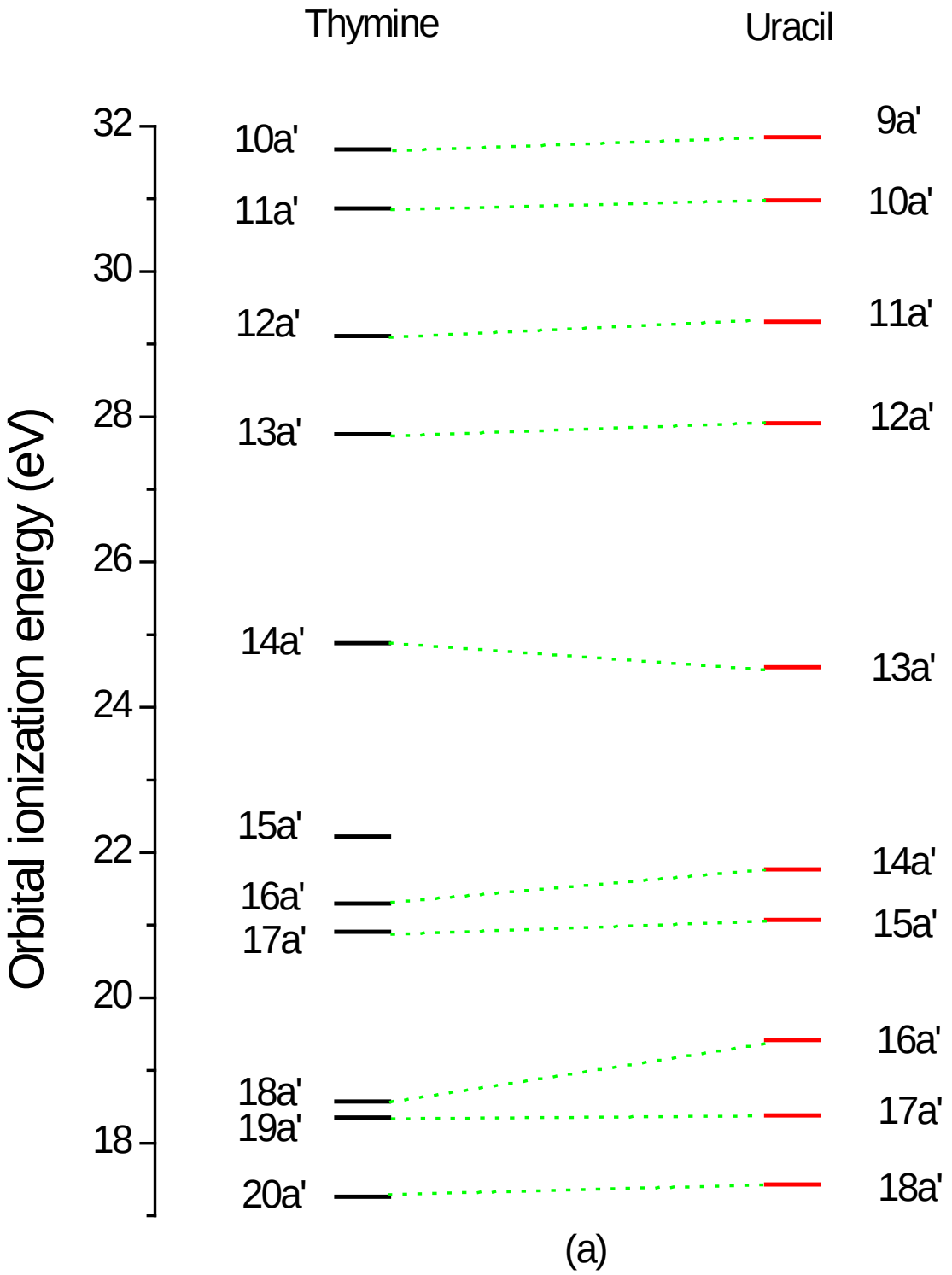


Fig. 2 (a) The orbital correlation diagram of the inner shell valence orbitals of thymine and uracil in their ground state calculated using the SAOP/et-pVQZ model.

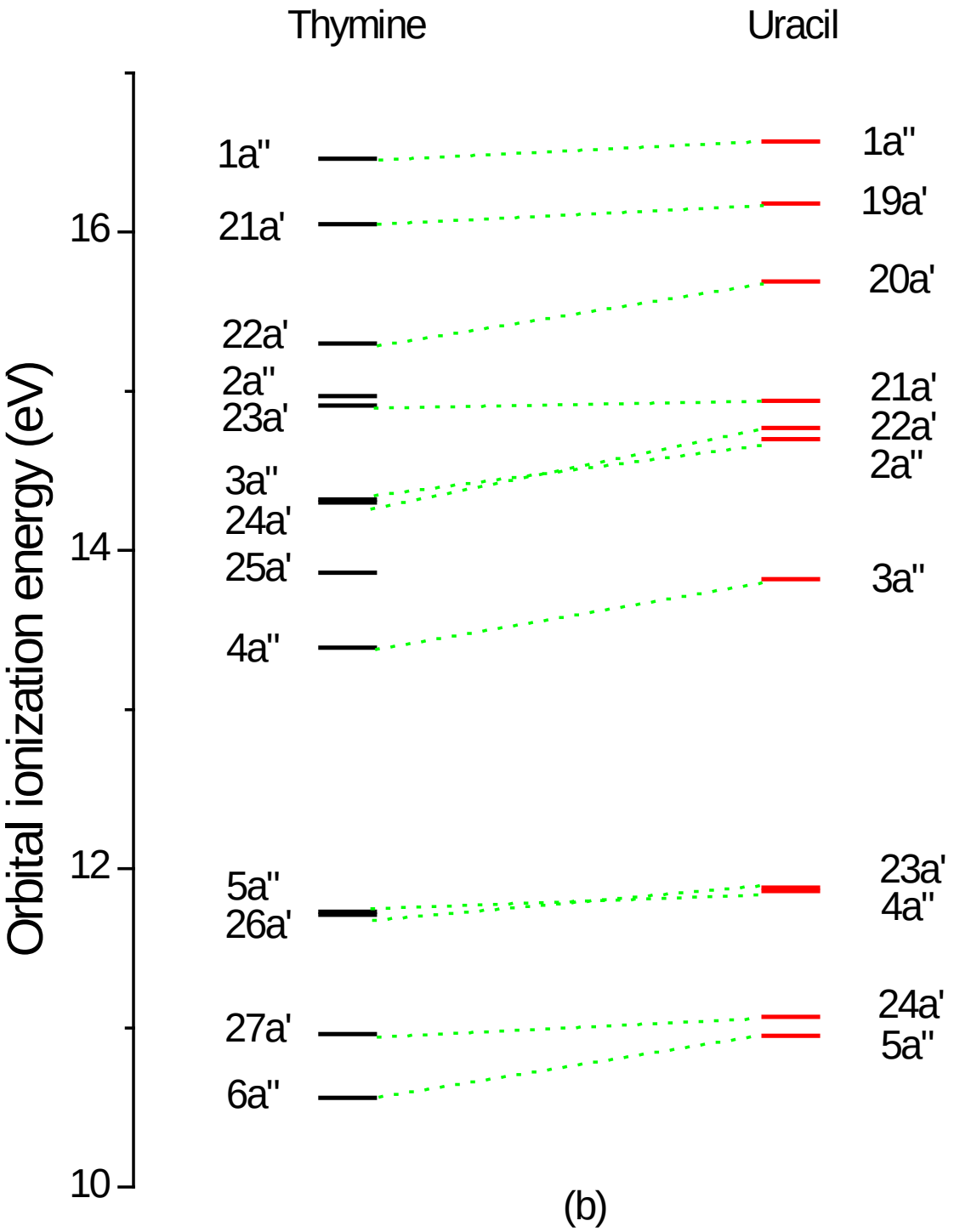
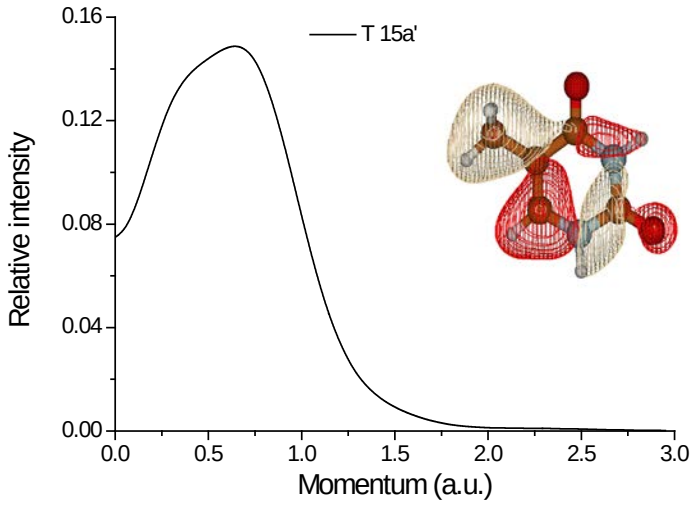
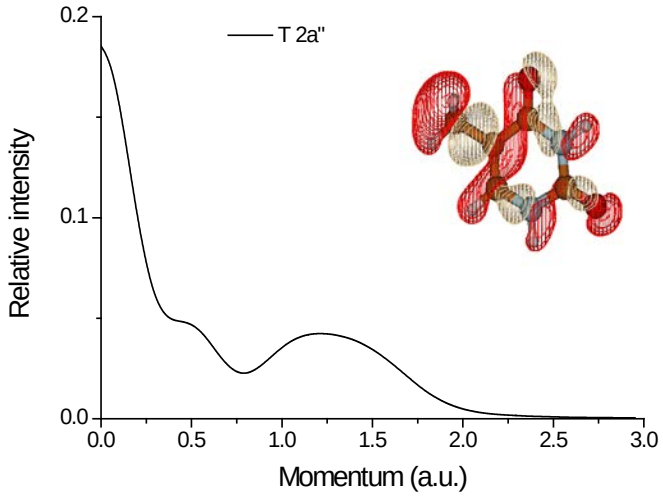


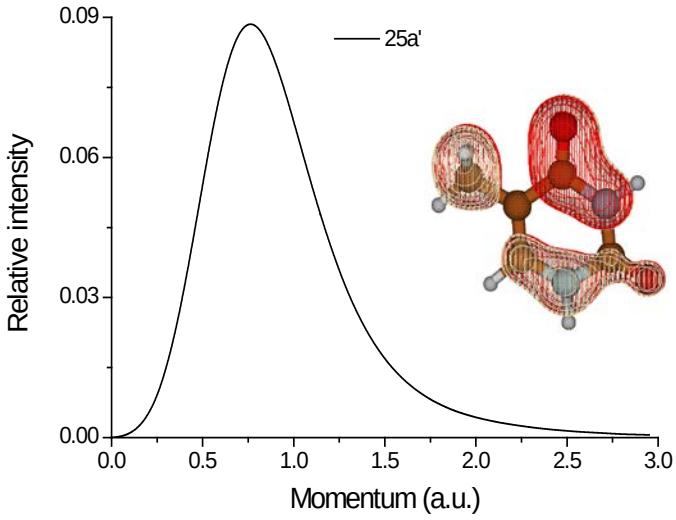
Fig. 2 (b) The orbital correlation diagram of the outer shell valence orbitals of thymine and uracil in their ground state calculated using the SAOP/et-pVQZ model.



(a)



(b)



(c)

Fig. 3 Molecular orbital signatures of the methyl group of thymine

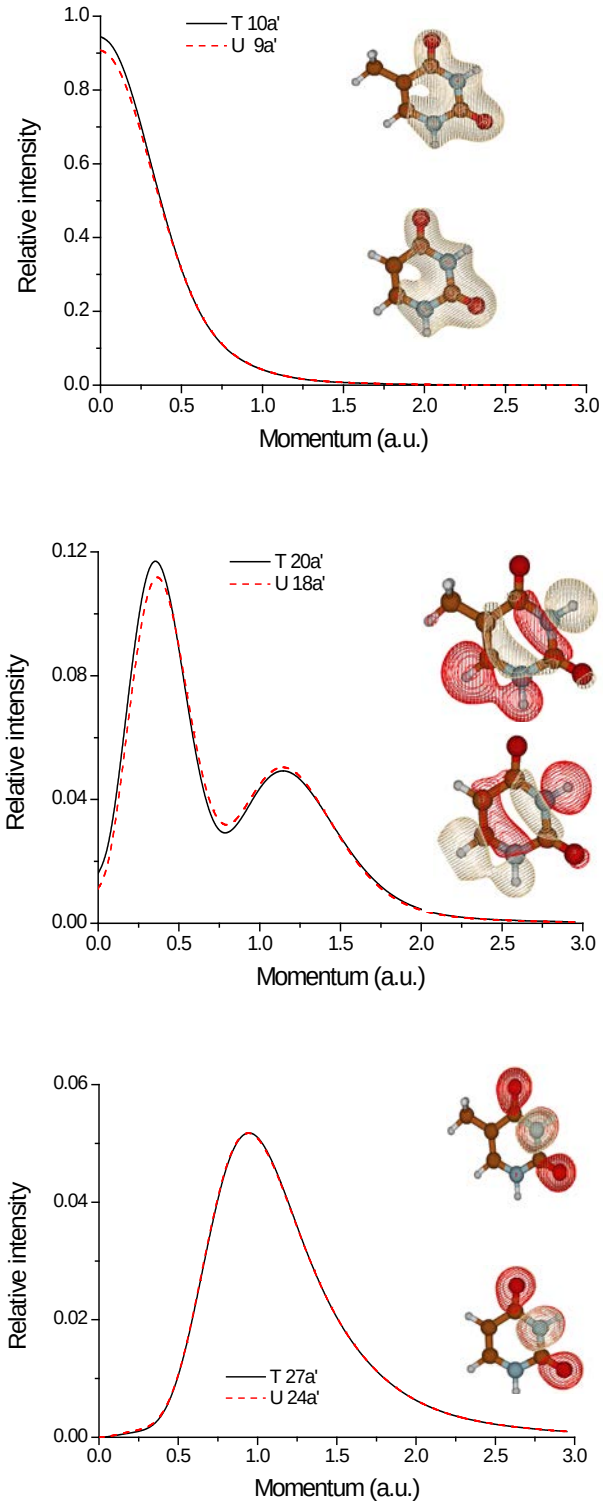


Fig. 4 Relatively unaffected molecular orbitals of the methyl group of thymine and uracil.

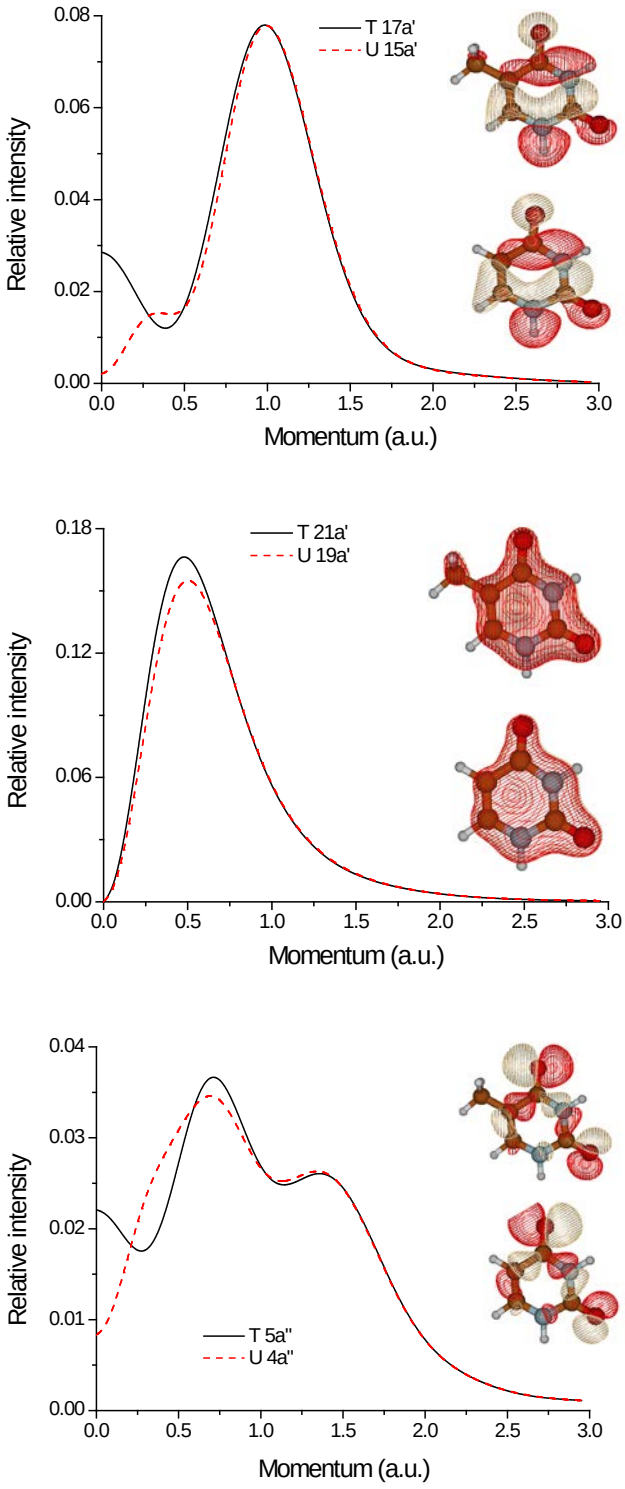


Fig. 5 Disturbed molecular orbitals of the methyl group of thymine and uracil.

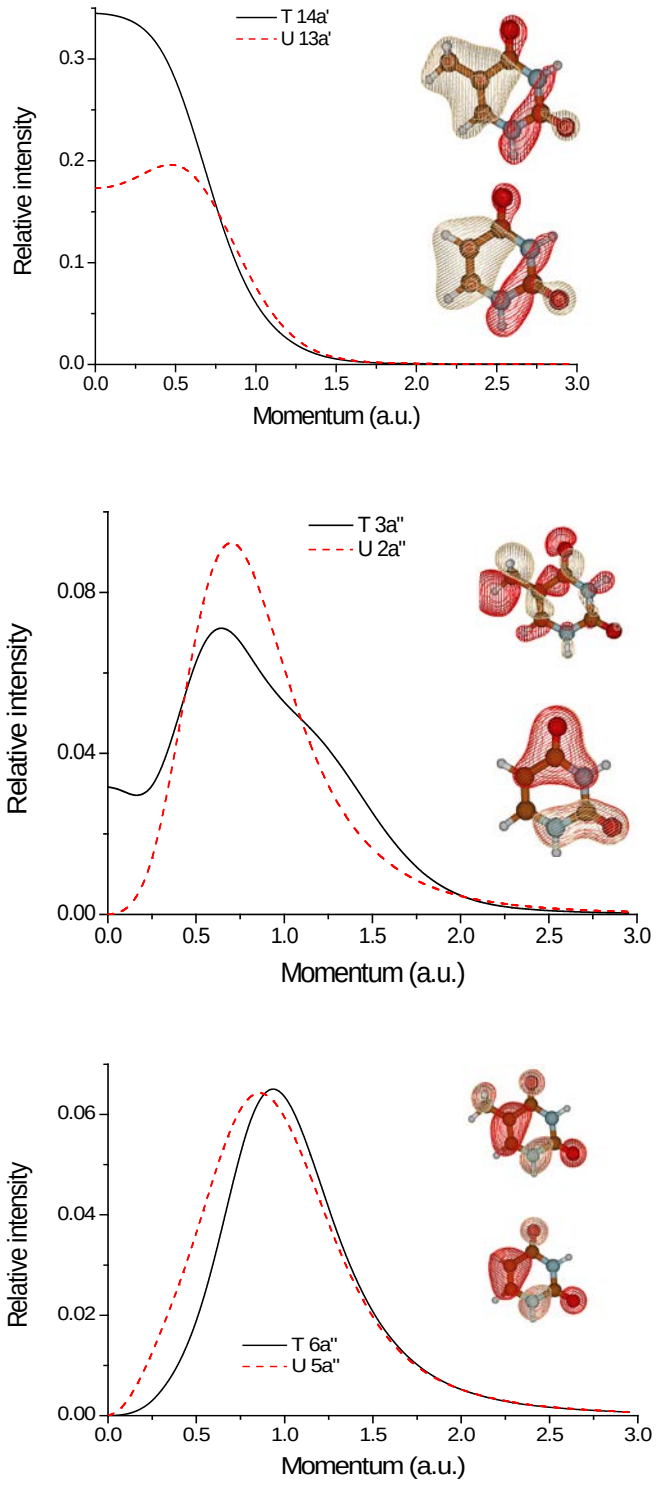


Fig. 6 Affected molecular orbitals of methyl group of thymine and uracil.

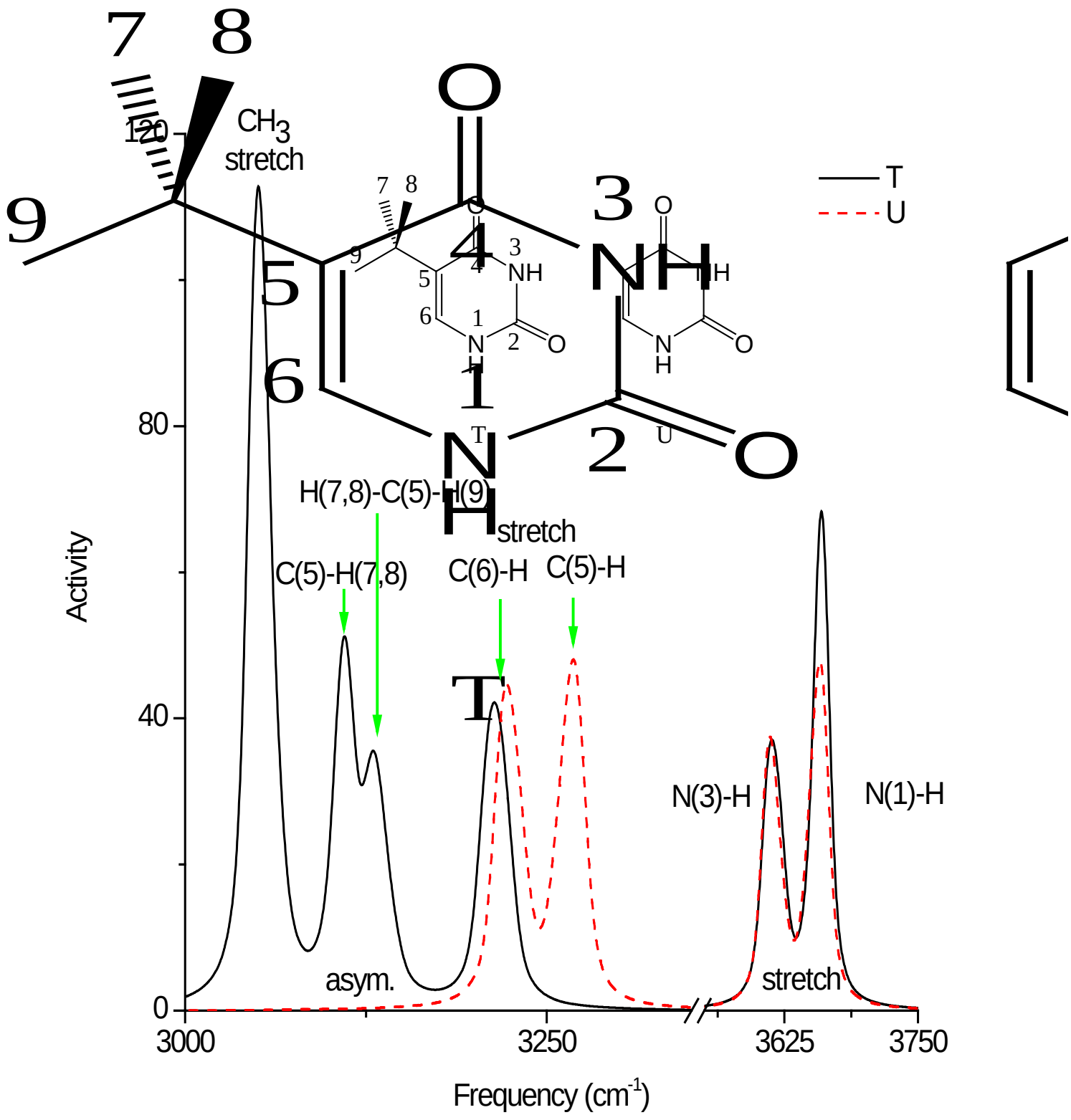


Fig. 7 Raman spectroscopy of thymine and uracil