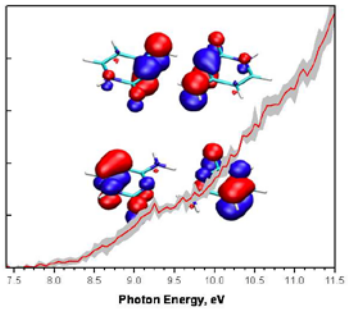
Ionization of cytosine monomer and dimer studied by VUV photoionization and electronic structure calculations.

Oleg Kostko,[1] Ksenia Bravaya,[2] Anna Krylov,[2*] and Musahid Ahmed[1*]

[1] *Chemical Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA*

[2] *Department of Chemistry, University of Southern California, Los Angeles, CA 90089-0482, USA*

*mahmed@lbl.gov; *krylov@usc.edu



Photoionization of the cytosine dimer

Summary

We report a combined theoretical and experimental study of ionization of cytosine monomers and dimers. Gas-phase molecules are generated by thermal vaporization of cytosine followed by expansion of the vapor in a continuous supersonic jet seeded in Ar. The resulting species are investigated by single photon ionization with tunable vacuum-ultraviolet (VUV) synchrotron radiation and mass analyzed using reflectron mass spectrometry. Energy onsets for the measured photoionization efficiency (PIE) spectra are 8.60±0.05 eV and 7.6±0.1 eV for the monomer and the dimer, respectively, and provide an estimate for the adiabatic ionization energies (AIE). The first AIE and the ten lowest vertical ionization energies (VIEs) for selected isomers of cytosine dimer computed using equation-of-motion coupled-cluster (EOM-IP-CCSD) method are reported. The comparison of the computed VIEs with the derivative of the PIE spectra, suggests that multiple isomers of the cytosine dimer are present in the molecular beam. The calculations reveal that the large red shift (0.7 eV) of the first IE of the lowest-energy cytosine dimer is due to strong inter-fragment electrostatic interactions, i.e., the hole localized on one of the fragments is stabilized by the dipole moment of the other. A sharp rise in the CH^+ signal at 9.20±0.05 eV is ascribed to the formation of protonated cytosine by dissociation of the ionized dimers. The dominant role of this channel is supported by the computed energy thresholds for the CH^+ appearance and the barrierless or nearly barrierless ionization-induced proton transfer observed for five isomers of the dimer.

I. Introduction

The structures and properties of many biological molecules are very well characterized. However, despite 50 years of research that have elapsed since Watson and Crick's original postulation for the structure of DNA, the fundamental aspects of the "nuts and bolts" of the molecules, which are the building blocks of life, are not yet fully understood.[1, 2] The ionization of nucleobases is a key step leading to damage and mutation of DNA.[3] The electron hole introduced by ionization or oxidation migrates along the helix through various hopping mechanisms coupled with tautomerization through proton transfer and ultimately leads to distant chemical modifications of the bases, strand cleavage and dissociation of the helix itself. Apart from the evolutionary and carcinogenic effects that this damage induces in living systems, there is also much interest in the electronic properties of the DNA molecules themselves owing to their potential use in molecular electronics.[4, 5] Molecular shape, conformational dynamics and electronic properties (i.e., charge distributions, excited states) of DNA play crucial roles in its selectivity and function. The properties of DNA are determined by the properties of its individual blocks and their complex interactions. Hence, the intrinsic properties of the DNA bases are of fundamental interest.[2]

While the ionization of isolated DNA bases has been studied extensively, there is very little experimental information for the dimers. Moreover, even on the monomer level, the picture is not yet complete owing to a daunting task of disentangling contributions of numerous tautomers and conformers that can be produced under the experimental conditions (especially in the case of cytosine and guanine). Previous ionization studies of cytosine consist of photoionization mass spectrometry (PIMS) measurements of AIEs,[6, 7] photoelectron spectroscopy (PES) at the valence[8-11] and core level,[12] resonance enhanced multiphoton ionization (REMPI) experiments[13, 14] and a number of electronic structure calculations.[9, 15-18] Trofimov et al.[9] measured a valence-shell PES of cytosine. By analyzing photoelectron energy and angular distributions, they concluded that only one tautomer, **C2b** (Fig. 1), is populated under their thermal vaporization conditions. They reported the lowest VIEs of 8.89 (±0.02) eV arising from ionization from a π orbital, as well as 8 other bands, the ones relevant to the present experiment are at

9.55 eV (σ), 9.89 eV (π), and 11.20 eV (σ). A slightly higher value of the lowest VIE (8.94 eV) was reported in an earlier PES study.[10] Although the tautomer ratio of **C1**, **C2a**, **C2b** and **C3a** in the PES study (463K) is expected to be 0.22:0.17:0.38:0.24 (as estimated[9] by the authors using ab initio thermodynamical data from Ref.[19]), Trofimov et. al concluded that their experimental results can be explained by population of **C2b** only. In contrast, a very recent core-level PES study[12] of cytosine suggested that 3 tautomers of cytosine are populated upon thermal vaporization at 450 K, with tautomer **C2(a+b)** (≈ 60 %) being the dominant species based on free energy calculations of Trygubenko et al.[20] and Fogarasi.[19]

A previous PIMS measurement from our group determined the AIE of cytosine monomer to be 8.65 (±0.05) eV,[6] in agreement with 8.68 eV reported in early PIMS work,[7] and within the range (8-9 eV) obtained using one and two color resonant 2 photon ionization spectroscopy by Nir et al.[14] The latter work also reported that two tautomers, one keto (**C1**) and one enol (**C2b**), are prevalent in their laser desorption jet-cooled molecular beam.

Ab initio calculations of the IE's of the biologically relevant tautomer of cytosine have been reviewed by Roca-Sanjuan et al.[16, 17] and Cauet et al.[17] Other tautomers were considered by Wolken et al. who estimated the lowest IEs by DFT, MP2 and CCSD(T) calculations.[15] More reliable values of valence-shell VIEs of the five lowest-energy tautomers obtained using electron propagator methods were reported by Ortiz and coworkers.[21] However, accurate estimates of the AIEs and valence-shell VIEs obtained at the same level of theory (so that ionization-induced relaxation effects can be quantified) for the most stable cytosine tautomers that are likely to be populated in the experiment are still missing. As discussed in Refs. [22] and [23], using computational methods that are capable of describing multiple interacting states of open-shell character is crucial for obtaining reliable results for the ionized systems, and EOM-IP-CC is one such approach.[24-29]

There are no experimental reports for the IEs of the cytosine dimer. A number of groups characterized the dimers by using multiphoton ionization spectroscopy and laser ablation. Nir et al. reported that while two tautomers were populated in the monomer channel, only one isomer **$(C1)_2$HB1**

(Fig. 2) was observed for the dimer, the H-bonding motif being C=O···HNH/NH···N.[13] Dey et al. also reported a mass spectrum for the dimer using laser desorption coupled to molecular beams.[30, 31] 1-methylated cytosine dimers have been formed in a Knudsen cell type field emitting mass spectrometer,[32] however, to the best of our knowledge, the only report of cytosine dimer formation using thermal desorption and detection using PIMS is a recent work from our group.[6] Choi et al.[33] suggested that the cytosine dimers were formed in their He droplet experiment; however, since this was a non mass-selective IR experiment, higher-energy monomeric tautomers could also have given rise to the signal. The results for other ionized dimers of nucleobases are also scarce.[1, 34-36]

To alleviate this paucity in ionization of nucleobases, we have started a systematic effort using tunable VUV photoionization mass spectrometry and high-level electronic structure calculations. This work presents the first report of the experimental detection of the cytosine dimer using thermal desorption coupled with molecular beams, measurement of the AIE of the cytosine dimer and the appearance energy of the protonated cytosine. The comparison of the VIEs computed by EOM-IP-CCSD method with the experimental spectra presents evidence for presence of several isomers in the molecular beam under our conditions. The computed IEs and comparison of the experimental spectra of the cytosine monomer and the dimer allow us to elucidate the effects of dimerization on cytosine photoionization. We found that IEs of cytosine are strongly affected by the inter-fragment interactions in the dimer, i.e., the lowest IE of the most stable dimer is red-shifted by almost 1 eV. This effect is much larger than previously reported values for similar systems.[22, 23, 37] We also discuss the ionization-induced dissociation of cytosine dimers leading to the formation of the protonated cytosine species. Our results indicate that ionization of the five H-bonded cytosine dimers considered in this study initiates a barrierless proton transfer from one base to another. In these proton-transferred structures, the positive charge is localized on the closed-shell protonated fragment, whereas the unpaired electron resides on the deprotonated moiety. By comparison of the measured dependence of the $\mathbf{CH^+}$ signal on the photon energy to the computed energy thresholds, we demonstrate that the $\mathbf{CH^+}$ formation can be ascribed to the dissociation of proton-transferred cytosine dimers. Ionization-induced barrierless proton transfer in hydrogen-bonded dimers might have important

implications for the mechanism of hole migration through the DNA molecule. The proton transfer in the H-bonded pairs results in the separation of the unpaired electron and the positive charge between the strands and may result in hole trapping. [38] In a subsequent paper, we will address the ionization-induced proton transfer between the complementary pairs of nucleobases.

The structure of the paper is as follows. The next section briefly outlines the experimental and theoretical methods (a complete description is given in Ref.[22]). The experimental spectra and computed IEs and energetic parameters are presented in **Results** section. Analysis of inter fragment interactions on cytosine ionization and interpretation of experimental spectra results are given in **Discussion**. Our main results and concluding remarks are summarized in **Conclusions** section.

II. Methods

The theoretical and experimental methods have been described in detail in our paper on photoionization of thymine and adenine.[22] Here we give only the essential parameters for the cytosine study. The experiments are performed on a molecular beam apparatus coupled to a 3 meter VUV monochromator on the Chemical Dynamics Beamline at the ALS. The thermal vaporization source[6] was heated to around 600 K to generate cytosine monomers and dimers in a supersonic jet expansion. In the present experiments, the backing pressure was 35 kPa of Ar through a 100 μm diameter nozzle. The time-of-flight spectra were recorded for the photoionization energy range between 7.4 and 11.5 eV. The typical step size for the PIE scans is 50 meV and a dwell time of 10 s at a repetition rate of 10 kHz.

The equilibrium structures of the neutral monomeric tautomers were optimized with the RI-MP2 method[39-41] using Dunning's cc-pVTZ basis set. The geometries of the neutral and cationic dimers were computed using the long-range corrected ωB97X[42] functional with an empirical dispersion correction (ωB97X -D)[43-45]and the 6-31+G(d,p) basis set. Using these geometries, binding energies and relative energies of the neutral dimers were computed using RI-MP2/6-311+G(d,p) and ωB97X-D/6-311++G(2df,2pd) methods. To verify the structures, the Hessians were computed with the ωB97X-D/6-

31+G(d,p) method. The fine EML(75,302) grid consisting of 75 points in the Euler-Maclaurin radial grid[46] and 302 points in the Lebedev angular grid[47] was used in all DFT calculations. Thermodynamic analysis was performed within the rigid rotor — harmonic oscillator — ideal gas approximation (RR-HO-IG) for the standard conditions (T=298.18 K, p=1 atm) and for T=582 K. Thermodynamic analysis for the cytosine monomers was performed using RI-MP2/6- 311+G(d,p)//RI-MP2/6-311+G(d,p) frequencies and CCSD/cc-pVTZ//RI-MP2/cc-pVTZ relative energies. RI-MP2/6-311+G(d,p)//ωB97X-D/6-31+G(d,p) relative energies were used for the dimer thermodynamic analysis along with frequencies computed with ωB97X-D/6-31+G(d,p) method. Since zero-point energy is included in the enthalpy term in the Q-Chem code for the RR-HO-IG calculations, the non-ZPE corrected electronic energies were used in the Gibbs energy calculations.[48]

VIEs of the monomers and dimers were computed with the EOM-IP-CCSD method[24-29] and the cc-pVTZ and 6-311+G(d,p) basis sets, respectively. The frozen natural orbitals (FNO)[49] approximation was used for the IE calculations of the dimers with the virtual space truncated using 99.50% natural population cut-off criterion. Performance of FNO approximation for the cases of both vertical and adiabatic ionization energies was extensively discussed in Ref. 49. The errors in IEs introduced by truncation of active virtual orbital space according to the 99.5 % population criterion relative to the full virtual space results were found to be less than 0.1 eV for similar molecular systems, including thymine, guanine and uracil dimer.[22, 49] The core electrons were frozen in all EOM-IP-CCSD calculations. The lowest AIEs were obtained as the difference between the EOM-IP-CCSD energy of the first ionized state at the cation geometry and the CCSD energy of the reference state at the geometry of the neutral species. Optimized geometries, relevant total energies, and harmonic frequencies are given in Supplementary Materials. All calculations were performed using the Q-CHEM electronic structure program.[50]

III. Results

As mentioned above, the analysis of the experimental spectra is complicated by the presence of multiple cytosine tautomers and even larger amount of the dimer isomers. Moreover, ionization-induced defragmentation of larger clusters contributes to the signal of the smaller ones giving rise to "fill-in" and "drop-out" effects. Therefore, interpretation of the spectra requires taking into account contributions from multiple isomers and ionization-induced dimer dissociation channels. With this in mind, we organized this section as follows. The first subsection (**Cytosine monomers: structures, relative energies and populations)** addresses the selection of cytosine tautomers that can be populated under our experimental conditions and discusses their properties relevant for the analysis of inter-fragment interactions in the dimers. We then proceed to describe the structures of the selected cytosine dimers, their binding and relative energies, and provide estimates of their relative populations assuming thermal equilibrium conditions (**Cytosine dimers: structures, binding energies and populations)**. The third subsection presents the mass spectra of the ionized species and discusses the dissociation and defragmentation channels (**Experimental mass spectrum and thermal fragmentation)**. We then discuss the measured PIE spectra for the monomer and the dimer, as well as appearance energy curve for protonated cytosine (**Experimental PIE curves**). Subsection E (**IEs of cytosine monomers and dimers)** summarizes computational results on cytosine ionization and includes EOM-IP-CCSD estimates of 1st AIE and lowest VIEs computed for the selected cytosine tautomers and dimers. In the last subsection (**Ionization-induced proton transfer and dissociation**), we describe several ionization-induced dimer dissociation channels and present the respective energy thresholds for dissociation products formation computed with ωB97X-D.

A Cytosine monomers: structures, relative energies and populations

We considered the five lowest energy isomers of the cytosine monomer, four of them lying within 1.35 kcal mol^{-1} and the fifth one being 2.75 kcal mol^{-1} above the most stable tautomer (Fig. 1). Tautomerization strongly affects the dipole moments of cytosine, i.e., the computed dipole moments vary from 2.32 D for the **C3b** tautomer to 6.21 D for the biologically relevant **C1** tautomer. As discussed

below, these changes in dipole moment reverse the relative stability of the dimers formed by different tautomeric forms of cytosine (as compared to the monomers) and also explain the magnitude of the IEs shifts of the dimers relative to the monomer values.

Populations of different tautomeric forms based on the free energy calculations were previously reported for several temperatures (Table 1). The four lowest-energy tautomers were found to be significantly populated with the fifth tautomer (**C3b**) having only minor contribution to the overall gas-phase cytosine species. The estimated relative tautomers populations in the gas phase for the thermal vaporization conditions computed in this work are given in the last column of Table 1. In agreement with previous calculations, our results show considerable populations of the four lowest tautomers. Note that both vibrational enthalpy and entropy favor the **C1** tautomer and Gibbs free energy difference between **C1** and **C2b** is reduced to ~ 0.4 kcal/mol.

Thermal populations, however, are not directly related to the populations in the molecular beam because of the several non-equilibrium steps involved in the experiment. Moreover, the distribution of the isomers in the beam can be affected by tautomerization kinetics and initial non-thermal populations of tautomers.[51, 52] Yang and Rodgers,[51] who studied the unimolecular and bimolecular tautomerization of cytosine using MP2 methods, made an intriguing suggestion that the relative populations of the tautomers produced by thermal vaporization depend on intermolecular hydrogen bonding interactions present in the condensed phase. Kosenkov et al.[52] have taken this a step further and, using a kinetic approach based on ab-initio calculated rate constants, suggested that upon thermal vaporization at 490 K, the dimers constitute 28% of the total population where 29% and 39% are due to the **C3a/b** and **C1** monomer forms, respectively.

In this work we are concerned about the qualitative composition of the gas-phase mixture, i.e. whether a particular tautomer or dimer isomer can be thermodynamically populated. Thus, we do not attempt to predict precise populations in the molecular beam and rely on simple Maxwell-Boltzmann estimations as a guideline.

B Cytosine dimers: structures, binding energies and populations

The representative structures of the dimers and the corresponding binding energies are shown in Fig. 2. We focus on the lowest-energy H-bonded structures only, as the T-shaped and stacked manifolds were reported to be 7-10 kcal mol^{-1} higher.[53, 54] Moreover, the geometry optimization of several stacked CC isomers without symmetry constrains collapsed to H-bonded structures. Reoptimized with the ωB97X-D functional, the stacked structures from Ref.[53] (isomers 9 and 11) have one imaginary frequency corresponding to the tilt motion of one of the fragments, and thus are not true minima at this level of theory. The only stable stacked structure we found corresponds to isomer 4 from Ref. [53] and lies 5.9 kcal mol^{-1} above the lowest-energy H-bonded dimer at the RI-MP2/6-311+G(d,p)//ωB97X-D/6-31+G(d,p) level of theory. Thus, its population is likely to be negligible giving rise to only minor contribution to the overall signal of the dimer.

The absence of other stable minima for the stacked cytosine dimers is in striking contrast to the thymine dimer,[22] possibly due to more polar character of the cytosine monomer resulting in stronger electrostatic interactions in the dimers, which prevail over weak dispersion interactions contributing to the stability of the stacked structures. However, we cannot rule out artifacts caused by the limitations of the ωB97X-D functional in describing the interplay between the two different types of interactions in relatively weakly bound systems. An early MD study of free energy surface of the cytosine dimer also reported only T-shaped and H-bonded isomers as stable minima.[54]

Sampling the full configurational space of the cytosine dimer is beyond the scope of this work and we have chosen to focus on a few representative structures with different inter-fragment arrangements and H-bonding patterns composed of tautomers **C1** and **C2b** (see Fig. 2). These are: the **$(C1)_2$HB1** structure with the O···H/N···H bonding; **(C1C2b)HB2** and **(C1C2b)HB1** that have the N···H/O···H bonding; **$(C1)_2$HB2** with the N···H/H···N bonding; and **$(C2b)_2$HB1** and **$(C2b)_2$HB2** that have the N···H/H···N bonding. Binding energies calculated by ωB97X-D and RI-MP2 are given in Table 2. The RI-MP2/6-311+G(d,p)//ωB97X-D/6-31+G(d,p) relative energies are shown in Fig. 2. The **$(C1)_2$HB1** dimer is

the most stable followed by five other isomers which lie within 4.5 kcal mol^{-1}. Although DFT and MP2 yield different energy ordering of the cytosine tautomers,[55, 56] the DFT errors cancel out in the calculations of binding energies (which are dominated by electrostatic interactions), and the DFT values agree well with those computed with RI-MP2 (Table 2). For the **$(C1)_2$HB1**, and **$(C1)_2$HB2** H-bonded dimers, our binding energies of 23.6 and 21.7 kcal mol^{-1} are close to interaction energies calculated by Kabelac and Hobza[54] who reported the values of 20.0 and 19.3 kcal mol^{-1}, respectively. A recently reported BSSE-corrected binding energy of 19.51 kcal mol^{-1} for the **$(C1)_2$HB2** dimer[57] is also in agreement with our value. Interestingly, despite a higher stability of the **C2b** monomer relative to **C1**, the dimers formed by the **C1** tautomers are lower in energy due to favorable dipole-dipole interactions of the two units (the respective dipole moments are 6.21 D and 3.19 D for **C1** and **C2b** (Fig. 1)).

The relative populations of the cytosine dimers estimated using the RR-HO-IG approximation based on the RI-MP2 relative energies are shown in Fig. 2. All of the H-bonded dimers are expected to have notable population under thermal equilibrium conditions (T=582K) with the dominant contributions from the **$(C1)_2$HB1, (C1C2b)HB1** and **(C1C2b)HB2** isomers.

To date only the **$(C1)_2$HB1** isomer has been identified in a molecular beam by de Vries and co-workers, who pioneered the study of DNA bases and their clusters using multiphoton ionization spectroscopies in conjunction with supersonic jets and mass spectrometry.[1] They produced cytosine dimers using laser desorption and presented strong evidence for only one isomer – **$(C1)_2$HB1** – being present in their molecular beam.[13] According to Kabelac and Hobza,[54] this is the most populated isomer based on their molecular dynamics/quenching calculations (albeit of a 298K ensemble). The absence of other isomers could be explained[1] by a number of reasons stemming from the detection scheme in a REMPI experiment: (1) poor absorption in the first excited state; (2) ionization energy higher than accessible by two photons; (3) fragmentation of the cation; (4) lifetimes of the excited state being too short for second photon absorption.

Not only does ionization induce significant changes in the structural parameters of the dimers, but it also changes the chemical structure of the fragments. Geometry optimization of dimer cations

converged to the structures with the proton being transferred from one base to another for all the dimers, except for **$(C1)_2$HB2** (Fig. 3). This structure has one very short hydrogen bond (1.550 Å), and a barrier for proton transfer along this bond is expected to be small (the structure was verified by frequency calculations and no imaginary frequencies were found). Proton transfer results in structures in which the positive charge is localized on a closed-shell protonated fragment, whereas the unpaired electron resides on the deprotonated neutral moiety. This process is accompanied by large relaxation energies, which amounts to ~15-20 kcal mol^{-1} for the dimers considered in this study (See Supporting Materials).

Dissociation of the proton-tranferred dimers can give rise to the signal of protonated monomer in the resulting mass spectra.

C Experimental mass spectrum and thermal fragmentation

In this subsection we present our experimental results which consist of VUV photoionization mass spectrometry of cytosine and cytosine dimers. Fig. 4 shows a log scale plot of a mass spectrum of a molecular beam of thermally vaporized cytosine recorded at a heater temperature of 603 K and photon energy of 10 eV. The most prominent peak is the cytosine monomer (m/z 111) followed by protonated cytosine (m/z 112). Peaks at m/z 222 and 223 arise from the cytosine dimer and protonated cytosine dimer, respectively. There are a number of peaks below the parent ion, which arise from fragmentation of cytosine at higher temperatures. The intensity plots of the most important ones are shown in Fig. 5. The peak corresponding to m/z 112 includes some contribution from isotopes in the cytosine monomer, however the majority of the signal is due to protonated cytosine, which most likely originates from the dissociative ionization of the cytosine dimer. Fragments at m/z 40, 68, 95 and 110 have been observed previously in electron impact[58] and VUV photoionization.[59] Since in this experiment, Ar (m/z=40) is being used as the carrier gas, there will be some metastable or Rydberg ionization of Ar and this is confirmed by the slight dependence of its signal versus temperature shown in Fig. 5. m/z 68 probably arises from elimination of NCOH from the enol tautomers **C2a/b**, and/or **C3a/b**, since this involves breakage of two single bonds. Plekan et al.[59] discussed the formation of m/z 68 and 69 in the context of

dissociative ionization in the VUV and electron impact (EI) experiments, however we see very little evidence of m/z 69. It is also important to note that the mass spectra arising from fragmentation at 10 eV are mostly due to thermal energy bond breaking in contrast to dissociative ionization, hence the results are reflective of neutral cytosine. m/z 95 originates from NH_2 elimination, and Rice et al.[58] suggested that m/z 68 could be due to the further HCN elimination from this fragment. This channel can operate at higher temperatures employed in our work. m/z 109, which arises from H_2 elimination from cytosine, is a very prominent peak at higher temperatures and has been observed previously upon EI ionization of a hydrated cytosine beam by Kim et al.[60] As mentioned earlier, Kosenkov et al.[52] estimated that upon thermal vaporization at 490K, dimers constitute 28% of the total population. As clearly seen from the dimer contribution in Figure 5, our data disagree with this prediction. In our experiment we see only 3% dimer contribution (dimer+protonated monomer) around ~ 550 K, and only when we reach a temperature of 640 K, the dimer population is around 41%. This increase in dimer population is due to an increase in the concentration of monomers at higher temperatures. In principle, Vant-Hoff type plots could be extracted from the temperature dependence shown in Fig. 5, however we have refrained from performing such an analysis here since it is believed that the molecular beams give rise to a highly non-equilibrium environment. A previous attempt[31] at generating association constants from such experiments have been shown to be subsequently wrong.[61] However, qualitatively our results suggest that it appears that cytosine, protonated cytosine and the cytosine dimer are generated with sufficient concentration and stability for us to extract meaningful photionization efficiency curves and these are presented next.

D Experimental PIE curves

Fig. 6 (a) and (b) show the photoionization efficiency curves for cytosine monomer and dimer, and Figure 6 (c) shows the signal from m/z 112 recorded at a heater temperature of 582 K from 7.4 to 11.5 eV. The insets show an expanded region at the onset for photon energy range of 7.5 to 9.0 eV. To extract more information from the PIE curves, we undertook an analysis pioneered by Berkowitz.[62] A derivative of the smoothed PIE signal yields a pseudo photoelectron spectrum. In the absence of the

partial photoionization cross section for that particular mass channel, the resulting spectrum only provides the information of band origins and maxima of band heads. Fig. 6 (d) shows such a curve for the cytosine monomer. Three previously reported photoelectron spectra[8, 9] are also shown in that figure and are normalized relative to the first band head maximum at 8.9 eV. The near perfect fit of our first peak to the PES ones gives us confidence in using this technique for interpreting the experimental photoionization results. Fig. 6 (e) shows a similar spectrum for the cytosine dimer. The S/N is worse compared to the monomer spectrum, however a number of bands can be clearly identified in the spectrum. Finally, Fig. 6 (f) shows the isotopically corrected plot for m/z 112, protonated cytosine. Energy onsets for the measured photoionization efficiency (PIE) spectra are 8.60±0.05 eV and 7.6±0.1 eV for the monomer and the dimer, respectively, and provide an estimate for the adiabatic ionization energies.

E IEs of cytosine monomer and dimers

AIE and VIEs computed with EOM-IP-CCSD for several cytosine tautomers are given in Table 3. As one can see, tautomerization affects both AIE and VIEs. For example, AIE for **C2a** and **C3a** differ by 0.2 eV, and the fourth ionized state of **C1** is red-shifted by more than 1.3 eV relative to **C2a**. The relaxation energy, i.e., the VIE-AIE energy difference, varies from 0.11 to 0.34 eV for different tautomers. Large variations in VIEs are due to changes in the nature and the order of ionized states upon tautomerization. These effects are discussed in details in a forthcoming paper on ionization of individual DNA bases.[63]

The AIE and the ten lowest VIEs for the dimers are presented in Table 4 and the corresponding molecular orbitals are shown in Figs. 7-9. Both the AIE and VIEs of the dimers are red-shifted relative to the monomer. Moreover, the ionization-induced relaxation is much larger in the dimers than in the monomer due to large geometry changes caused by proton transfer between the bases in the ionized dimers. Ionization energies of the cytosine dimers depend strongly on their structures. Even for the dimers formed by the same monomers, the shifts in VIEs can be as large as 0.45 eV (first VIE for **$(C1)_2$HB1** and **$(C1)_2$HB2** dimers).

Further analysis of dimerization effects on ionization of cytosine requires detailed characterization of the electronic structure of the ionized dimers. The **$(C1)_2$HB1** dimer is composed of two non-equivalent (due to non-symmetric structure) fragments, which results in the localized ionized states, i.e., the eight lowest ionized states of the dimer correspond to the four lowest ionized states of each of the **C1** fragments (see Fig. 7). The **$(C1)_2$HB2** dimer has C_{2h} symmetry and is formed by two equivalent (by symmetry) **C1** fragments. Consequently, the hole is equally delocalized between the two fragments (Fig. 7). The electronic structure of this type of dimers can be described in terms of DMO-LCFMO (dimer molecular orbitals – linear combination of fragment molecular orbitals) framework.[37, 64] DMO-LCFMO assumes that the dimer MOs are in-phase and out-of-phase combination of the monomer's MOs, and the states shown in Fig. 7 are of this type. However, the MOs describing the 5th to 8th ionized state of this dimer slightly deviate from this model (Fig. 7), i.e., even though the MOs are the in-phase and out-of-phase combinations of the fragment molecular orbitals (FMO), the shapes of the FMOs are slightly different in the pairs of states. Thus, the corresponding shifts in VIEs relative to the monomer should be considered with caution. The first and third ionized states of the **$(C1)_2$HB2** dimer are of non-Koopmans character and are derived from ionization from the orbitals corresponding to the first and second ionized state of the monomer (See Supporting Information). Due to this mixed character, the shift of the dimer VIE relative to the monomer is not well defined for these multi configurational states.

The electronic structure of the ionized states of the **$(C2b)_2$HB1** dimer is similar to that of the **$(C1)_2$HB1** isomer: the eight lowest ionized states correspond to the four ionized states of each fragment, with the exception of the two lowest states of the dimer cation for which the MOs are significantly delocalized (Fig 8). Similarly to **$(C1)_2$HB2, $(C2b)_2$HB2** has C_{2h} symmetry and its ionized states can be interpreted in terms of DMO-LCFMO (Fig. 10), i.e., the MOs corresponding to the eight lowest ionized states of this dimer are in-phase and out-of-phase MOs describing the four lowest ionized states of the **C2b** monomers.

The **C1C2b** heterodimers present an interesting and more complex case. Despite the different IEs of the fragments, the electron hole is significantly delocalized (Fig. 9). The MOs describing the eight

lowest ionized states for the **(C1C2b)HB1** dimer are combination of the MOs corresponding to the 1^{st} - 4^{th} ionized state of **C2b** and the 1^{st} - 5^{th} ionized states of **C1**. For the **(C1C2b)HB2** dimer, the MOs for 1^{st} - 4^{th} ionized states of the monomer are involved. The 4^{th} and 5^{th} ionized states of this dimer are of non-Koopmans character and mainly involve ionization from the MOs corresponding to the 2^{nd} ionized state of **C2b** and the 3^{rd} and 4^{th} ionized states of **C1**. Thus, the VIE shifts of the dimer relative to the monomer for the states with delocalized hole and multiconifigurational character cannot be defined for the **(C1C2b)HB1** and **(C1C2b)HB2** dimers.

F Proton transfer and dissociation

One of the possible channels for photo-induced dynamics in the **CC** dimers is a proton between the bases. For five of the dimers considered in this work no minimum on the cation potential energy surface, which correspond to the structure without proton transfer, was found. Therefore, one can expect, that the ionization-induced proton transfer in these **CC**dimers is a barrierless or nearly barrierless process (Fig. 3).The existence of efficient channels of ionization-induced proton transfer in hydrogen-bonded base pairs and related systems was previously reported by both theory[65-68] and experiment.[34, 69] The dissociation of the proton-transferred dimer can give rise to the $\mathbf{CH}^+$ signal. The dimers considered here are formed by two cytosine tautomers and have different H-bonding patterns (Fig. 3). Thus, their dissociation may yield different tautomeric forms of the $\mathbf{CH}^+$ and $\mathbf{C}^+$ species. Appearance energies for different isomers of $\mathbf{CH}^+$ and $\mathbf{C}^+$ computed as the energy differences between the ground states of the neutral dimer and the corresponding products (and corrected for ZPE) are given in Table 5. Although the threshold energies were computed with DFT, we expect that the errors in relative energy order of the **C1** and **C2b** tautomers cancel out in a similar way as for binding energies (see Section III B).

IV. Discussion

A Effects of dimerization

As shown above by the experimental data and ab initio calculations, dimerization strongly affects both the IEs and the character of the ionized states. Here we analyze how the inter-fragment interactions in the different types of the H-bonded cytosine dimers affect their ionized states in order to explain the origin of the strong shifts in VIEs due to dimerization. The dimers considered here can be classified into two distinct groups: ones with equivalent fragments (**$(C1)_2$HB2** and **$(C2b)_2$HB2**) and those with structurally (**$(C1)_2$HB1** and **$(C2b)_2$HB1**) or chemically (**(C1C2b)HB1** and **(C1C2b)HB2**) non-equivalent fragments. Below we show that the VIE shifts can be qualitatively explained by inter-fragment electrostatic interactions and the character of corresponding MOs.

The first type are the isomers composed of the same tautomers that are geometrically non-equivalent (**$(C1)_2$HB1** and **$(C2b)_2$HB1**). The VIEs shifts in these dimers relative to monomers can be explained by inter-fragment electrostatic interactions: the dipole moment of one fragment destabilizes the MOs of another leading to a large drop in IEs. As a model system, consider the **$(C1)_2$HB1** dimer (Fig. 7). Different relative orientation of the monomers results in different VIEs shifts for the ionized states localized on one of the two monomers: the ionized states localized on one of the monomers are affected more (-0.68 – -0.55 eV) than those localized on the other (-0.44 – 0.16 eV). The same trend is observed for the **$(C2b)_2$HB1** dimer (Fig. 8). However, the MOs describing ionized states are partially delocalized in that case due to a lower value of the dipole moment of **C2b** relative to **C1**. This also explains smaller VIE shifts in the **$(C2b)_2$HB2** dimer as compared to the **$(C1)_2$HB1** isomer**.**

The shifts in VIEs of the heterodimers can also be explained by the inter-fragment electrostatic interactions (Fig. 9). For example, the first ionized state in both **(C1C2b)HB1** and **(C2C2b)HB2** is localized on the **C2b** fragment, however, the respective VIE is 0.08 eV higher than that of isolated **C1.** This points to a strong destabilization of the **C2b** HOMO by the large dipole moment of the **C1** fragment. The analysis of the VIEs shifts for higher ionized states is complicated by the delocalized character of the corresponding MOs.

The second group is the dimers with equivalent fragments. Their electronic structure as well as the magnitude of the VIE shifts in different states can be described by DMO-LCFMO to a large extent.

We illustrate this by considering the **$(C2b)_2HB2$** dimer as an example (see Fig. 8). In agreement with DMO-LCFMO, the ionized states of this dimer described by the MOs that are out-of-phase combination of the monomer MOs are stabilized in the dimer cation giving rise to the VIEs shifts of -0.55 – -0.25 eV, and the magnitude of the shifts correlate well with the overlap of the respective FMOs. However, the VIE shifts for the states described by the MOs that are in-phase combinations are smaller or even positive (-0.47 – +0.06 eV). These positive shifts cannot be explained by DMO-LCFMO, which predicts symmetric splitting between the pairs of the ionized states. Same-magnitude shifts of different sign for ionization from bonding/antibonding pairs of orbitals were observed in a variety of stacked dimers.[22, 23, 37] Thus, the deviation is likely to be due to the stronger perturbation of the orbitals introduced by hydrogen bonding. The observed red shifts in VIEs for the states described by the in-phase combination of the FMOs suggest that the interaction of two or more FMOs for each fragment needs to be considered.

It is worth noting that the magnitudes of the shifts of VIEs for second type of systems are comparable to those for the cytosine dimers with non-equivalent fragments, in contrast to what was observed for thymine dimer in our earlier work.[22] This can be explained by a more delocalized character of the cytosine MOs and, consequently, larger overlap between the fragments MOs.

The above analysis demonstrates that simple considerations accounting for electrostatic inter-fragment interactions and character of corresponding MOs can be used for a qualitative explanation and prediction of the dimerization effects on cytosine VIEs.

B Monomers IEs: Theory vs. experiment

In addition to providing the basis for the analysis of the dimerization effects in the ionized states of cytosine, the computed IEs of the monomers and the dimers can also be used as a reference for the interpretation of the experimental spectra. The onset of the PIE spectrum shown in Fig. 6 (a) represents the AIE and is at 8.60 eV. Previously our group has reported an IE of 8.65(±0.05) eV,[6] and a very early literature value is 8.68 eV.[7] As discussed above, thermal vaporization of cytosine can populate at least

four tautomers under our conditions. The AIEs and VIEs for the five tautomers of cytosine forms computed with EOM-IP-CCSD are shown in Table 3.

For the lowest energy **C2b** tautomer (Fig. 1), the computed AIE (8.54 eV) is in excellent agreement with the PIE onset energy of 8.60 eV shown in Figure 6 (a) and provides validation of the higher accuracy of PIE's to extract adiabatic onset information. Contrary to that, vertical energy determinations are much more reliable in photoelectron measurements, as can be seen from comparison of the computed VIEs with the band maxima of the differentiated PIE and PES spectra. Let us analyze the first band in the derived PIE of the monomer (shown in Fig. 10) to obtain an estimate of the 1st VIE. This peak's maximum is at 8.9 eV and the subsequent bands are centered at 9.6 and 10.3 eV, with another band extending out to 11.5 eV. The maximum of the first band agrees perfectly with that of the PES from Trofimov et al.[9] for both their 0 and 90 degree angle resolved PES obtained with thermal vaporization at 190^{O} C and photon energy of 40 eV. The results of Yu et al.[8] recorded at heater temperature of 224^{O} C and photon energy of 21.1 eV agree well after an energy correction of 0.27 eV to lower energies. This offset of 0.27 eV is probably due to an incorrect calibration of the absolute energy scale in their experiments. However, the onset in both PES's are much lower and the photoelectron spectra tail off to around 8.0 eV, which is 0.5 eV below our derived onset. This is probably due to the energy width of the light sources used in the photoelectron experiments.

The maximum of the first band at 8.9 eV (Fig. 10) agrees well with the calculated values for the first vertical IE (see Table 3) for all 5 tautomers. The second band centered at 9.6 eV agrees well with the calculated 2nd IE of **C1**, **C2a** and **C2b** as well as the 3rd IE of **C1**. Following these bands, there is a broad peak stretching from 9.8 to 10.8 eV. Ionization to the 2nd state of the **C3a** and **C3b** cations, the 3rd state of the **C2a**, **C2b**, **C3a** and **C3b** cations and the 4th state of the **C1** cation can contribute to this peak. Finally, the enhancement of the signal around 11.3 eV could arise from the 4th state of **C2a** and **C2b**, while signal around 11 eV could be due to the 4th state of **C3b**. Without Franck-Condon and photoionization cross sections calculations, it is difficult to determine relative contributions of the different states to the overall spectrum. However, the comparison of the spectrum with the superimposed stick spectrum representing

the different states of the tautomers in Fig. 10 does suggest that under our conditions we are populating all the low-lying tautomers. Our desorption temperature of around 582 K is much higher than those employed in the photoelectron spectroscopy measurements due to the need to produce sufficient vapor pressure of cytosine to induce clustering in our beam. A number of theoretical calculations have been performed to estimate population distributions of cytosine upon thermal vaporization and are displayed in Table 1[15, 19, 20, 51] along with our calculations. The 582 K experimental temperature suggest that we have higher contributions of the **C1, C2a** and **C3a/b** tautomers (Table 1) compared to the photoelectron measurements of Trofimov et al.[9] performed at ~ 463 K and could explain the difference in the peak heights in our spectrum compared to the photoelectron measurement. There could also be fill in effects from the dissociation of the dimer to give rise to ionized cytosine monomer in our derived PIE spectrum. A more quantitative comparison is not possible since we would need to normalize our spectrum taking into account the photoionization cross sections, which are unknown.

C Dimers IEs: Theory vs. experiment.

Dimerization results in dramatic changes in the shape of the photoionization curves and the underlying energetics. The onset shifts down to ~7.6 eV relative to 8.60 eV in the monomer (Fig. 6 (a) and (b)). Thus, AIE is red-shifted by 1.0 eV. Interestingly, an early work on base pairing in cytosine, which employed Koopmans calculations, suggested that upon formation of the CC pair, the IE of cytosine increases by 0.58 eV relative to the monomer.[70] The second band in the monomer spectrum, which extends from around 9.2 eV to around 9.8 eV in our spectra (10.5 eV in the PES spectra[9]) is very much depleted in the dimer spectrum (the strong depletion of the dimer signal around 9.5 eV will be discussed below). Fig. 11 displays the derivative of the PIE spectrum for cytosine dimer along with the calculated vertical IEs for the 5 isomers as a stick spectrum. The relevant energies are presented in Table 4. Adiabatic IEs of the five H-bonded **$(C1)_2$HB2**, **$(C2b)_2$** and **(C1C2b)** dimers agree well with the experimental onset. We anticipate poor FC factors for the proton-transferred structures, which can explain the onset of 0.3 eV above the AIE of the dominant isomer, as well as the gentle and slow rise in the

spectrum. Moreover, the onset of 7.31 eV predicted for the **$(C1)_2$HB1** isomer is not visible under our experimental conditions since our scans only extend to 7.4 eV due to the limitations of the beamline configuration. The first VIE for **$(C1)_2$HB1** is 8.10 eV suggesting poor FCFs due to large geometry changes upon ionization. The 1st and the 2nd ionized states of the mixed **(C1C2b)** and **(C1C2b)** dimers as well as the **$(C2b)_2$** dimers explain the band at 8.5 eV, however, the latter isomers are expected to be less populated (Fig. 2). The calculated vertical IE's for the 2nd to the 5th ionized states of **$(C1)_2$HB1** and the 4th ionized state of **$(C1)_2$HB2** fall in nicely at the peak of the bands centered around 9 eV.

The depletion of the signal at 9.2-9.5 eV in the derived PIE spectrum can be explained by dissociation of the dimer cations producing cation radicals (**C1**$^+$ and **C2b**$^+$) and protonated monomers (**C1H**$^+$**(N3)**, **C2bH**$^+$**(N1)** and **C2bH**$^+$**(N3)**) (Table 5). The computed thresholds for all considered channels of the **CH**$^+$ formation lie within 0.1 eV of the observed at 9.2 eV rise in the **CH**$^+$ signal (shown in Fig. 6 (f)). We observed a similar behavior in the photoionization spectrum and appearance energies of the protonated species for thymine.

The peaks around 9.5 eV in the derived PIE could arise from ionization of multiple isomers. A broad feature at 10.25 eV can be ascribed to the 8th ionized state of the **$(C1)_2$HB1** and **$(C1)_2$HB2** isomers. The band at 11 eV could be attributed to the 7th ionized state **$(C2b)_2$HB1**. The rise of the signal at 11.25 eV can be explained by ionization to the 8th ionized state of the **$(C2b)_2$** dimers and the **(C1C2b)HB1** isomer, as well as by the 9th ionized state of **$(C1)_2$HB1**. In summary, there is evidence for the presence of multiple cytosine dimer isomers, and the spectra could not be explained by the presence of only the most stable **$(C1)_2$HB1** isomer. The comparison of the computed dissociation energy thresholds with both the dimer PIE spectrum and the **CH**$^+$ signal curve points to the efficient channel for intradimer proton transfer and dissociation at energy above ~9.1 eV. This is also supported by the presence of multiple ionized states in the 9.2-9.5 eV photon energy region, e.g. the 5th ionized state of **$(C1)_2$HB2** and the 4th ionized state of the **(C1C2b)HB1** dimer, which are not observed in the experiment.

V. Conclusions

This work demonstrates strong effects upon dimerization on cytosine ionization. The interaction between the fragments in the dimers affects both the character of ionized states and IEs. By using VUV single photon ionization mass spectrometry we determined the first experimental AIE for the cytosine dimer to be 7.6±0.1 eV. The onset in the dimer PIE spectra is red-shifted by ~1 eV relative to the monomer. The computed EOM-IP-CCSD AIE and VIEs for the selected cytosine dimers range between 7.31-7.64 eV. The calculations provide an insight into the origin of the shifts and the character of ionized states, aiding the interpretation of the experimentally derived PIE spectra. The electronic structure analysis reveals that the origin of this large red shift is in the electrostatic interactions between the fragments. The largest shift (0.7 eV) was predicted for the lowest-energy dimer, **$(C2b)_2$HB1**, in which the hole localized on one of the fragments is stabilized by a large dipole moment (6.21 D) of the "neutral" fragment.

Both experimental and theoretical results suggest that a number of tautomers and H-bonded dimers are present in the molecular beam, however, more quantitative analysis would require calculations of FCFs and ionization cross-sections.

The computed energy thresholds for the ionization-induced dimer dissociation forming the $\mathbf{CH^+}$ species show that this channel can be efficient at photon energies above ~9.1 eV, which explains strong rise in the measured $\mathbf{CH^+}$ signal at 9.2 eV. The large yield of the protonated species is consistent with the barrierless (or almost barrierless) proton transfer observed for the H-bonded cytosine dimers considered in this study.

Future experiments using sophisticated two color IR-VUV spectroscopy, ion-electron coincidence spectroscopy and mass analysed threshold ionization will allow unambiguous identification of the various species present in our molecular beam. Nevertheless, the results presented here are a necessary first step towards an unequivocal molecular-level understanding of dynamics of photoionization of DNA bases.

VI. Acknowledgements

This work is conducted under auspices of the iOpenShell Center for Computational Studies of Electronic Structure and Spectroscopy of Open-Shell and Electronically Excited Species supported by the National Science Foundation through the CRIF:CRF CHE-0625419+0624602+0625237 grant and support by the Director, Office of Energy Research, Office of Basic Energy Sciences, Chemical Sciences Division of the U.S. Department of Energy under contract No. DE-AC02-05CH11231.

VII. Electronic supplementary information

Figure of EOM-IP-CCSD/cc-pVTZ//RI-MP2/cc-pVTZ VIEs for the **C1** and **C2b** tautomers with corresponding molecular orbitals. Table of relative energies for neutral and cation radical cytosine dimers and cation relaxation energies computed at CCSD/6-311+G(d,p) and EOM-IP-CCSD/6-311+G(d,p) levels of theory. Atomic coordinates for cytosine dimers considered in this work in neutral and cationic states (XYZ file format). Harmonic frequencies (ωB97X-D/6-31+G(d,p)) and thermodynamic parameters for cytosine dimers.

VIII. References

1 M. de Vries, in *Radiation Induced Molecular Phenomena in Nucleic Acids,* ed. M. K. Shukla and J. Leszczynski, Springer, 2008, p. 323.

2 M. S. de Vries and P. Hobza, *Annu. Rev. Phys. Chem.*, 2007, **58**, 585-612.

3 M. K. Shukla and J. Leszczynski, in *Radiation Induced Molecular Phenomena in Nucleic Acids*, ed. M. K. Shukla and J. Leszczynski, Springer, 2008, p. 1.

4 F. C. Grozema and L. D. A. Siebbeles, *Int. Rev. Phys, Chem.*, 2008, **27**, 87-138.

5 M. Taniguchi and T. Kawai, *Physica E*, 2006, **33**, 1-12.

6 L. Belau, K. R. Wilson, S. R. Leone and M. Ahmed, *J. Phys. Chem. A*, 2007, **111**, 7562-7568.

7 V. M. Orlov, A. N. Smirnov and Y. M. Varshavsky, *Tetrahedron Lett.*, 1976, **48**, 4377-4378.

8 C. Yu, S. Peng, I. Akiyama, J. Lin and P. R. Lebreton, *J. Am. Chem. Soc.*, 1978, **100**, 2303-2307.

9 A. B. Trofimov, J. Schirmer, V. B. Kobychev, A. W. Potts, D. M. P. Holland and L. Karlsson, *J. Phys. B: At. Mol. Opt. Phys.*, 2006, **39**, 305-329.

10 N. S. Hush and A. S. Cheung, *Chem. Phys. Lett.*, 1975, **34**, 11-13.

11 D. Dougherty, E. S. Younathan, R. Voll, S. Abdulnur and S. P. McGlynn, *J. Electron Spectrosc. Relat. Phenom.*, 1978, **13**, 379-393.

12 V. Feyer, O. Plekan, R. Richter, M. Coreno, G. Vall-Ilosera, K. C. Prince, A. B. Trofimov, I. L. Zaytseva, T. E. Moskovskaya, E. V. Gromov and J. Schirmer, *J. Phys. Chem. A*, 2009, **113**, 5736-5742.

13 E. Nir, I. Hunig, K. Kleinermanns and M. S. de Vries, *Phys. Chem. Chem. Phys.*, 2003, **5**, 4780-4785.

14 E. Nir, M. Muller, L. I. Grace and M. S. de Vries, *Chem. Phys. Lett.*, 2002, **355**, 59-64.

15 J. K. Wolken, C. Yao, F. Turecek, M. J. Polce and C. Wesdemiotis, *Int. J. Mass Spectrom.*, 2007, **267**, 30-42.

16 D. Roca-Sanjuan, M. Rubio, M. Merchan and L. Serrano-Andres, *J. Chem. Phys.*, 2006, **125**, 084302.

17 E. Cauet, D. Dehareng and J. Lievin, *J. Phys. Chem. A*, 2006, **110**, 9200-9211.

18 E. Cauet and J. Lievin, in *Advances in Quantum Chemistry, Vol 52*, Elsevier Academic Press Inc, San Diego, 2007, vol. 52, pp. 121-147.

19 G. Fogarasi, *J. Phys. Chem. A*, 2002, **106**, 1381-1390.

20 S. A. Trygubenko, T. V. Bogdan, M. Rueda, M. Orozco, F. J. Luque, J. Sponer, P. Slavicek and P. Hobza, *Phys. Chem. Chem. Phys.*, 2002, **4**, 4192-4203.

21 O. Dolgounitcheva, V. G. Zakrzewski and J. V. Ortiz, *J. Phys. Chem. A*, 2003, **107**, 822-828.

22 K. Bravaya, O. Kostko, M. Ahmed and A. I. Krylov, *(in press)*, 2009.

23 A. A. Golubeva and A. I. Krylov, *Phys. Chem. Chem. Phys.*, 2009, **11**, 1303-1311.

24 P. A. Pieniazek, S. E. Bradforth and A. I. Krylov, *J. Chem. Phys.*, 2008, **129**, 074104.

25 A. I. Krylov, *Annu. Rev. Phys. Chem.*, 2008, **59**, 433-462.

26 J. F. Stanton and J. Gauss, *J. Chem. Phys.*, 1994, **101**, 8938-8944.

27 P. A. Pieniazek, S. A. Arnstein, S. E. Bradforth, A. I. Krylov and C. D. Sherrill, *J. Chem. Phys.*, 2007, **127**.

28 M. Kamiya and S. Hirata, *J. Chem. Phys.*, 2006, **125**.

29 S. Pal, M. Rittby, R. J. Bartlett, D. Sinha and D. Mukherjee, *Chem. Phys. Lett.*, 1987, **137**, 273-278.

30 M. Dey, J. Grotemeyer and E. W. Schlag, *Z. Naturforsch., A: Phys. Sci.*, 1994, **49**, 776-784.

31 M. Dey, F. Moritz, J. Grotemeyer and E. W. Schag, *J. Am. Chem. Soc.*, 1994, **116**, 9211-9215.

32 I. K. Yanson, A. B. Teplitsky and L. F. Sukhodub, *Biopolymers*, 1979, **18**, 1149-1170.

33 M. Y. Choi, F. Dong and R. E. Miller, *Phil. Trans. R. Soc. A*, 2005, **363**, 393-412.

34 N. Gador, E. Samoylova, V. R. Smith, A. Stolow, D. M. Rayner, W. G. Radloff, I. V. Hertel and T. Schultz, *J. Phys. Chem. A*, 2007, **111**, 11743-11749.

35 E. Nir, K. Kleinermanns and M. S. de Vries, *Nature*, 2000, **408**, 949-951.

36 C. Plutzer, I. Hunig, K. Kleinermanns, E. Nir and M. S. de Vries, *Chemphyschem*, 2003, **4**, 838-842.

37 P. A. Pieniazek, A. I. Krylov and S. E. Bradforth, *J. Chem. Phys.*, 2007, **127**, 044317.

38 S. Steenken, *Free Radic. Res. Commun.*, 1992, **16**, 349-379.

39 M. Feyereisen, G. Fitzgerald and A. Komornicki, *Chem. Phys. Lett.*, 1993, **208**, 359-363.

40 F. Weigend and M. Haser, *Theor. Chem. Acc.*, 1997, **97**, 331-340.

41 R. A. Distasio, R. P. Steele, Y. M. Rhee, Y. H. Shao and M. Head-Gordon, *J. Comput. Chem.*, 2007, **28**, 839-856.

42 J. D. Chai and M. Head-Gordon, *J. Chem. Phys.*, 2008, **128**, 084106.

43 S. Grimme, *J. Comput. Chem.*, 2004, **25**, 1463-1473.

44 S. Grimme, *J. Comput. Chem.*, 2006, **27**, 1787-1799.

45 J. D. Chai and M. Head-Gordon, *Phys. Chem. Chem. Phys.*, 2008, **10**, 6615-6620.

46 C. W. Murray, N. C. Handy and G. J. Laming, *Mol. Phys.*, 1993, **78**, 997-1014.

47 V. I. Lebedev, *Zh. Vychisl. Mat. Fiz.*, 1975, **15**, 48.

48 Several popular ab initio packages include ZPE in the vibrational enthalpy term, and using ZPE-corrected energy differences between isomers will result in double-counting, which appears to be a common pitfall in such calculations.

49 A. Landau, K. Khistyaev, S. Dolgikh and A. I. Krylov, *J. Chem. Phys.*, **(In press)**.

50 Y. Shao, L. F. Molnar, Y. Jung, J. Kussmann, C. Ochsenfeld, S. T. Brown, A. T. B. Gilbert, L. V. Slipchenko, S. V. Levchenko, D. P. O'Neill, R. A. DiStasio, R. C. Lochan, T. Wang, G. J. O. Beran, N. A. Besley, J. M. Herbert, C. Y. Lin, T. Van Voorhis, S. H. Chien, A. Sodt, R. P. Steele, V. A. Rassolov, P. E. Maslen, P. P. Korambath, R. D. Adamson, B. Austin, J. Baker, E. F. C. Byrd, H. Dachsel, R. J. Doerksen, A. Dreuw, B. D. Dunietz, A. D. Dutoi, T. R. Furlani, S. R. Gwaltney, A. Heyden, S. Hirata, C. P. Hsu, G. Kedziora, R. Z. Khalliulin, P. Klunzinger, A. M. Lee, M. S. Lee, W. Liang, I. Lotan, N. Nair, B. Peters, E. I. Proynov, P. A. Pieniazek, Y. M. Rhee, J. Ritchie, E. Rosta, C. D. Sherrill, A. C. Simmonett, J. E. Subotnik, H. L. Woodcock, W. Zhang, A. T. Bell, A. K. Chakraborty, D. M. Chipman, F. J. Keil, A. Warshel, W. J. Hehre, H. F. Schaefer, J. Kong, A. I. Krylov, P. M. W. Gill and M. Head-Gordon, *Phys. Chem. Chem. Phys.*, 2006, **8**, 3172-3191.

51 Z. B. Yang and M. T. Rodgers, *Phys. Chem. Chem. Phys.*, 2004, **6**, 2749-2757.

52 D. Kosenkov, Y. Kholod, L. Gorb, O. Shishkin, D. M. Hovorun, M. Mons and J. Leszczynski, *J. Phys. Chem. B*, 2009, **113**, 6140-6150.

53 P. Jurecka, J. Sponer and P. Hobza, *J. Phys. Chem. B*, 2004, **108**, 5466-5471.

54 M. Kabelac and P. Hobza, *J. Phys. Chem. B*, 2001, **105**, 5804-5817.

55 M. Piacenza and S. Grimme, *J. Comput. Chem.*, 2004, **25**, 83-98.

56 J. R. Sambrano, A. R. de Souza, J. J. Queralt and J. Andres, *Chem. Phys. Lett.*, 2000, **317**, 437-443.

57 P. K. Sahu, R. K. Mishra and S. L. Lee, *J. Phys. Chem. A*, 2005, **109**, 2887-2893.

58 J. M. Rice and G. O. Dudek, *J. Am. Chem. Soc.*, 1967, **89**, 2719-2725.

59 O. Plekan, V. Feyer, R. Richter, M. Coreno, M. de Simone and K. C. Prince, *Chem. Phys.*, 2007, **334**, 53-63.

60 S. K. Kim, W. Lee and D. R. Herschbach, *J. Phys. Chem.*, 1996, **100**, 7933-7937.

61 P. Hobza and J. Sponer, *Chem. Phys. Lett.*, 1996, **261**, 379-384.

62 J. Berkowitz, *J. Chem. Phys.*, 1978, **69**, 3044-3054.

63 K. Bravaya, O. Kostko, M. Ahmed, S. Dolgikh, A. Landau and A. I. Krylov, *(to be submitted)*, 2009.

64 P. A. Pieniazek, J. VandeVondele, P. Jungwirth, A. I. Krylov and S. E. Bradforth, *J. Phys. Chem. A*, 2008, **112**, 6159-6170.

65 H. Y. Chen and I. Chao, *Chemphyschem*, 2004, **5**, 1855-1863.

66 J. Bertran, L. Blancafort and M. Noguera, in *Computational studies of DNA and RNA*, ed. J. Sponer and F. Lankas, Springer, 2006, p. 411.

67 J. Rak, J. Makowska and A. A. Voityuk, *Chem. Phys.*, 2006, **325**, 567-574.

68 H. S. Park, S. H. Nam, J. K. Song, S. M. Park and S. Ryu, *J. Phys. Chem. A*, 2008, **112**, 9023-9030.

69 N. J. Kim, H. M. Kim and S. K. Kim, *Int. J. Mass Spectrom.*, 2007, **261**, 32-37.

70 A. O. Colson, B. Besler and M. D. Sevilla, *J. Phys. Chem.*, 1992, **96**, 9787-9794.

Table 1. Relative populations of cytosine tautomers at different temperatures.

Isomer	Fogarasi[a]		Yang & Rodgers[b]		Wolken et al.[c]	Trygubenko et al.[d]		This work[e]
	470 K	570 K	490 K	570 K	473 K	298 K	582 K	582K
C1	0.59	0.78	0.39	0.48	0.72	0.35	0.58	0.69
C2a	0.45	0.52	0.47	0.53	0.47	0.30	0.54	0.57
C2b	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
C3a	0.64	0.86	0.09	0.15	0.12	0.01	0.10	0.47
C3b	-	-	0.02	0.04	-	0.001	0.03	0.13

(a) Ref.[19] geometries: CCSD/TZP; energies: CCSD(T)/cc-pVTZ; frequencies: MP2/TZP

(b) Ref.[51] geometries: MP2/6-31G*; energies: MP2/6-311+G(2d,2p); frequencies: MP2/6-31G*

(c) Ref.[15] geometries: MP2/6-31+G(d,p); energies: CCSD(T)/aug-cc-pVTZ; frequencies: B3LYP/6-31+G(d,p)

(d) Estimated from free energies in Ref.[20] geometries: RIMP2/TZVPP; energies: CCSD(T)/extrapolation to CBS; frequencies: HF/6-31G(d,p)

(e) Energies: CCSD/cc-pVTZ//RI-MP2/cc-pVTZ; frequencies: RI-MP2/6-311+G(d,p)//RI-MP2/6-311+G(d,p).

Table 2. Binding energies (D_e, kcal mol^{-1}) of the H-bonded cytosine dimers.

Isomer	ωB97X-D /6-31+G(d,p)	RI-MP2/6-311+G(d,p) //ωB97X-D/6-31+G(d,p)	ωB97X-D/6-311++G(2df,2pd) //ωB97X-D/6-31+G(d,p)
$(C1)_2$ HB1	24.50	23.60	23.58
$(C1)_2$ HB2	22.38	21.34	21.69
$(C2b)_2$ HB1	15.68	15.72	14.5
$(C2b)_2$ HB2	16.86	14.43	16.17
(C1C2b) HB1	20.74	19.71	19.87
(C1C2b) HB2	18.84	17.64	18.40

Table 3. The lowest AIE and VIEs for the selected cytosine tautomers calculated at the EOM-IP-CCSD/cc-pVTZ//IP-CISD/6-31+G(d) and EOM-IP-CCSD/cc-pVTZ//RI-MP2/cc-pVTZ levels, respectively. All energies in eV.

State	C1	C2a	C2b	C3a	C3b
AIE	8.67	8.52	8.54	8.71	8.68
VIE1	8.78	8.86	8.86	8.90	8.88
VIE2	9.55	9.58	9.62	9.89	10.01
VIE3	9.65	10.12	10.02	10.28	10.19
VIE4	10.06	11.38	11.34	10.74	11.06
VIE5	12.28	11.91	11.94	12.78	12.75
VIE6	13.27	13.52	13.48	13.28	13.31

Table 4. The lowest AIE and VIEs of the cytosine dimers calculated at the EOM-IP-CCSD/6-311+G(d,p)//ωB97X-D/6-31+G(d,p) level of theory. All energies in eV.

state	$(C1)_2$ HB1	$(C1)_2$ HB2	$(C2b)_2$ HB1	$(C2b)_2$ HB2	(C1C2b) HB1	(C1C2b) HB2
AIE	7.31	7.63*	7.48	7.64	7.49	7.63
VIE1	8.10	8.55	8.37	8.31	8.28	8.44
VIE2	8.84	8.56	8.51	8.39	8.48	8.62
VIE3	8.93	8.72	9.22	9.28	9.09	9.23
VIE4	9.07	8.90	9.43	9.68	9.2	9.26
VIE5	9.11	9.31	9.58	9.72	9.67	9.46
VIE6	9.51	9.50	9.8	9.81	9.79	9.62
VIE7	9.79	9.63	10.95	11.09	9.92	9.91
VIE8	10.22	10.18	11.28	11.32	11.35	11.12
VIE9	11.47	11.98	11.53	11.58	11.71	11.59
VIE10	12.27	12.02	11.66	11.66	11.73	11.98

* Corresponds to the H-transferred structure. The AIE for the cation structure without proton transfer is 8.27 eV.

Table 5. ωB97X-D/6-311++G(2df,2pd)//ωB97X-D/6-31+G(d,p) threshold energies for cytosine dimer ionization-induced dissociation including ZPE correction (ωB97X-D/6-31+G(d,p)) . All energies in eV.

Species	Parent dimer			
	$(C1)_2$HB1	**$(C1)_2$HB2**		
C1H$^+$(N3)	9.08	9.33		
	$(C2b)_2$HB1	**(C1C2b)HB2**		
C2bH$^+$(N1)	9.07	9.19		
	$(C2b)_2$HB2	**(C1C2b)HB1**		
C2bH$^+$(N3)	9.23	9.27		
	$(C1)_2$HB1	**$(C1)_2$HB2**		
C1$^+$	9.72	9.66		
	$(C2b)_2$HB1	**$(C2b)_2$HB2**	**(C1C2b)HB1**	**(C1C2b)HB2**
C2b$^+$	9.2	9.22	9.4	9.39

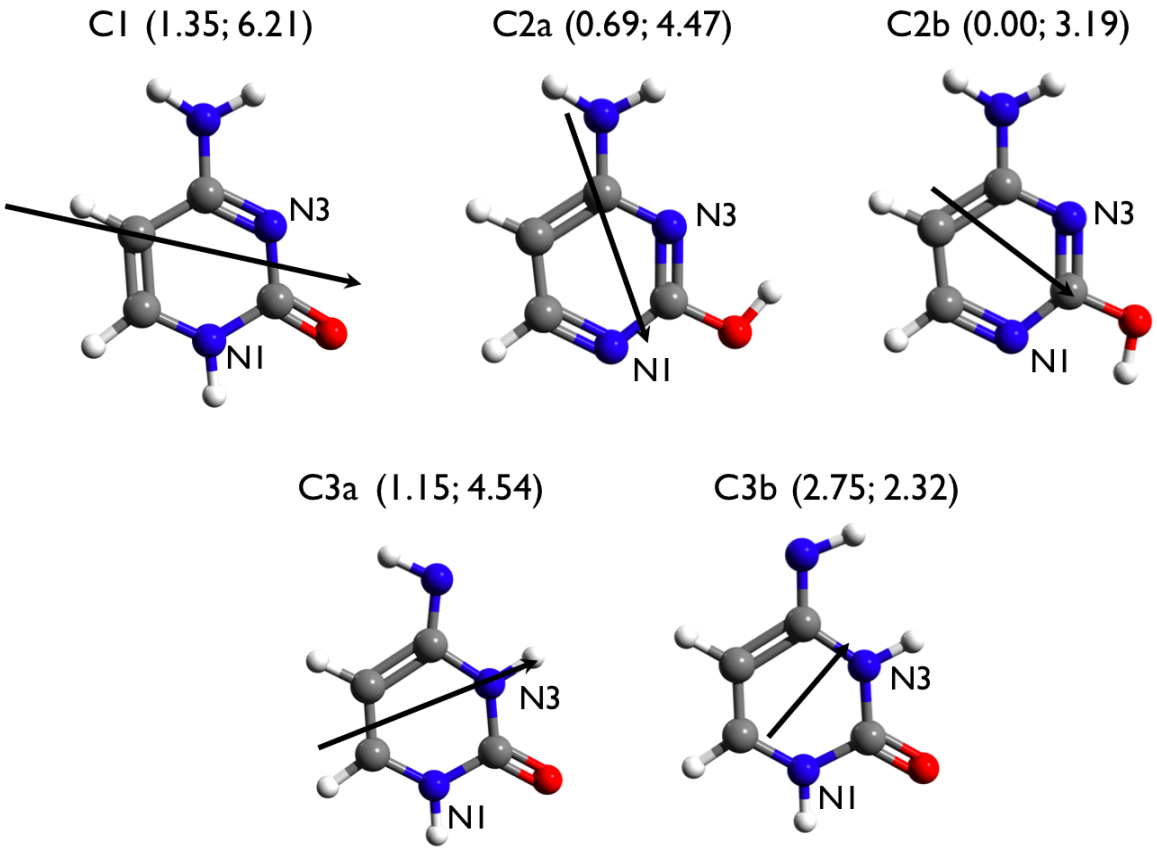


Figure 1. Calculated structures of cytosine tautomers. The CCSD/cc-pVTZ//RI-MP2/cc-pVTZ relative energies (kcal mol^{-1}) and RI-MP2/cc-pVTZ dipole moments (Debye) are shown in parentheses, respectively.

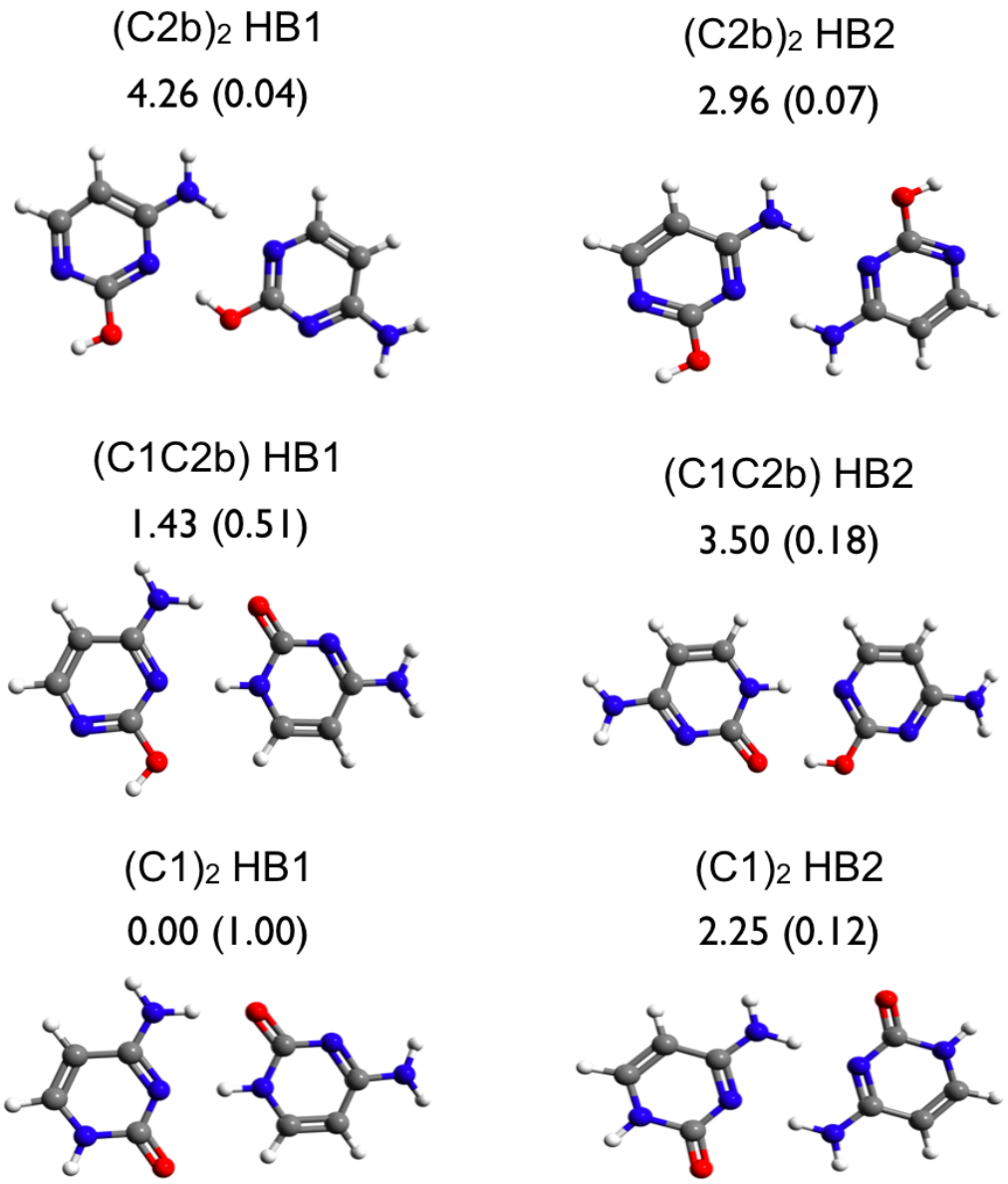


Figure 2. The structures of the H-bonded cytosine dimers and relative energies (kcal mol^{-1}) computed by the RI-MP2/6-311+G(d,p)//ωB97X-D/6-31+G(d,p) methods. Relative populations with respect to the lowest energy dimer estimated from thermodynamic parameters obtained within the RR-HO-IG approximation are given in parentheses.

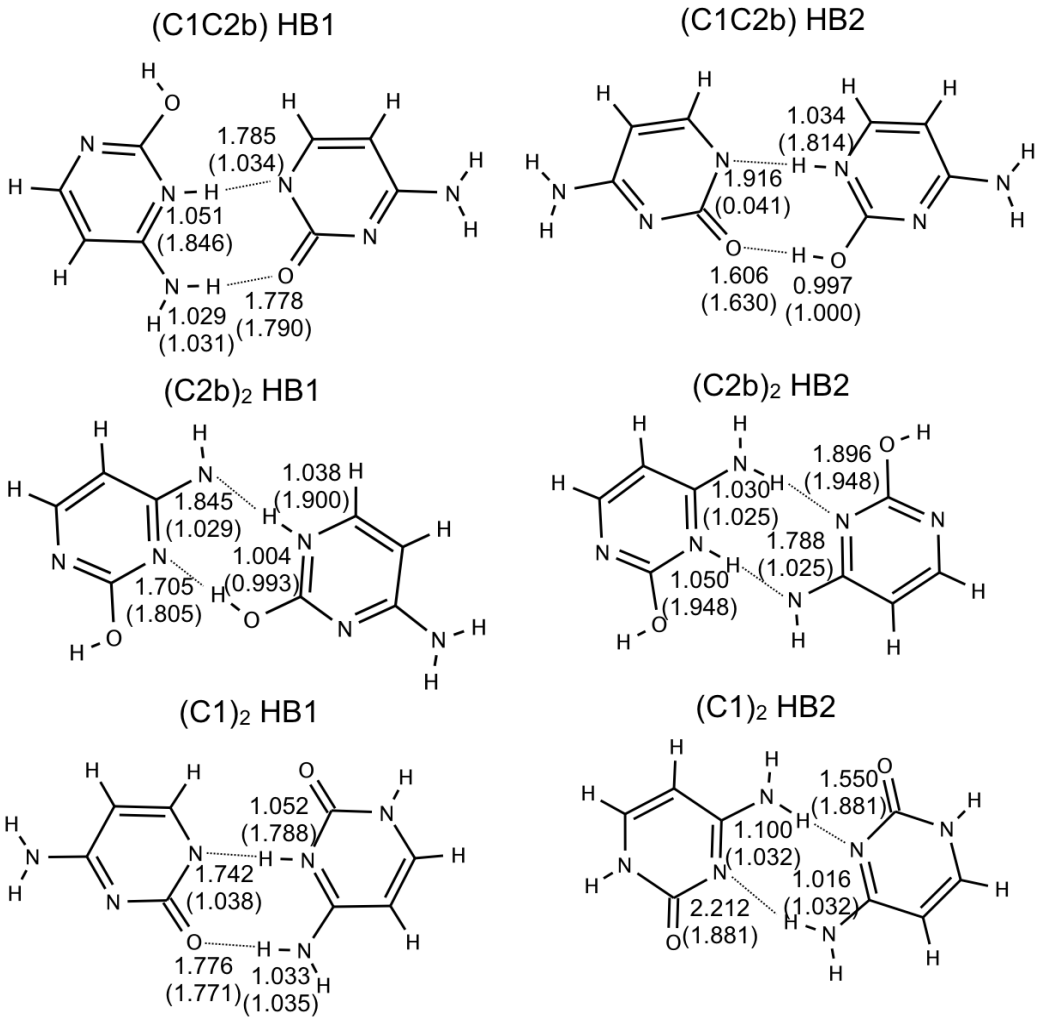


Figure 3. Ionization-induced geometry changes for cytosine dimers (ωB97X-D/6-31+G(d,p)). Important structural parameters for the cations and the neutrals are given in parentheses.

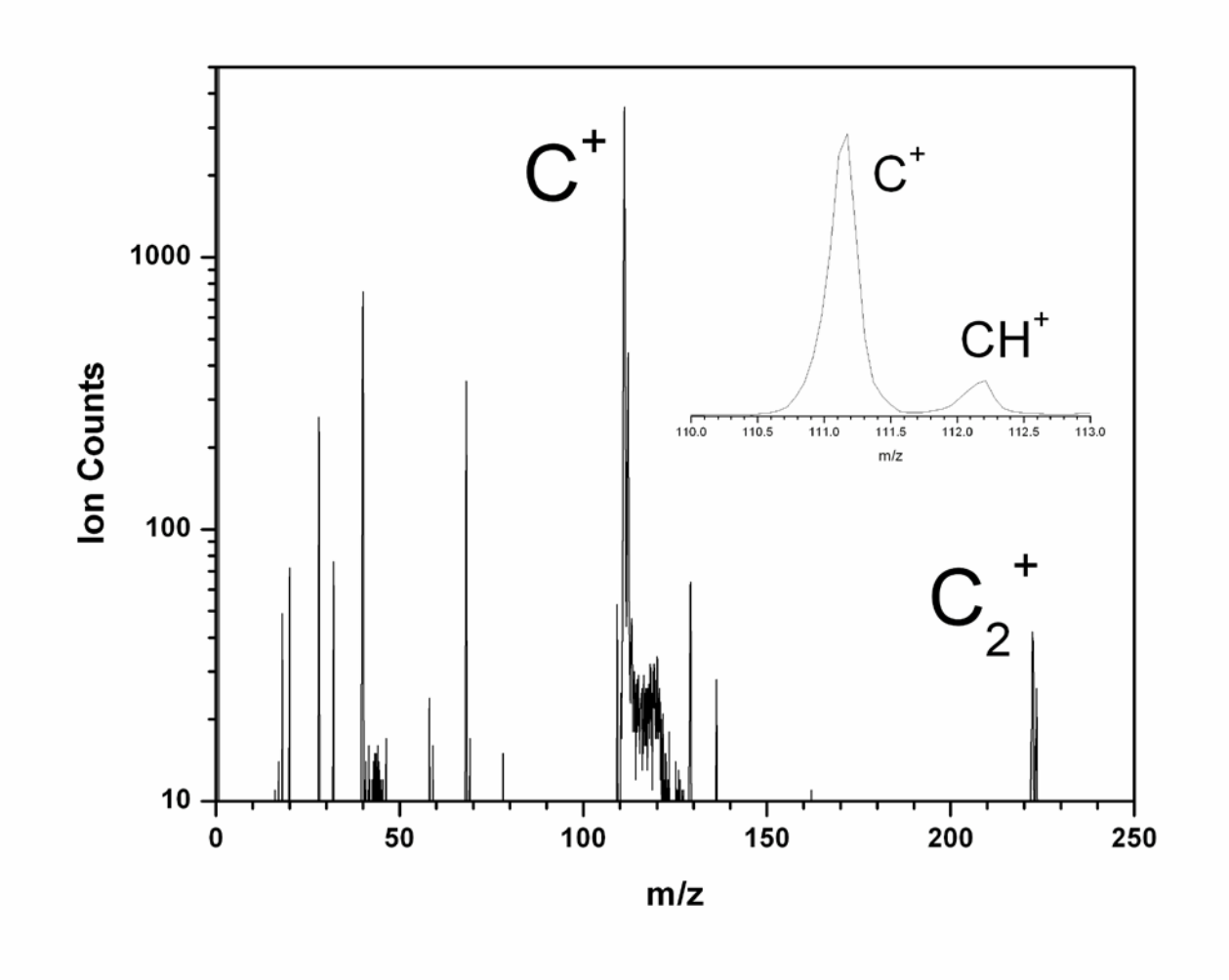


Figure 4. The mass spectrum of cytosine recorded at 10 eV photon energy and a heater temperature of 603 K. The inset shows a region of 110-113 m/z to reveal protonated cytosine.

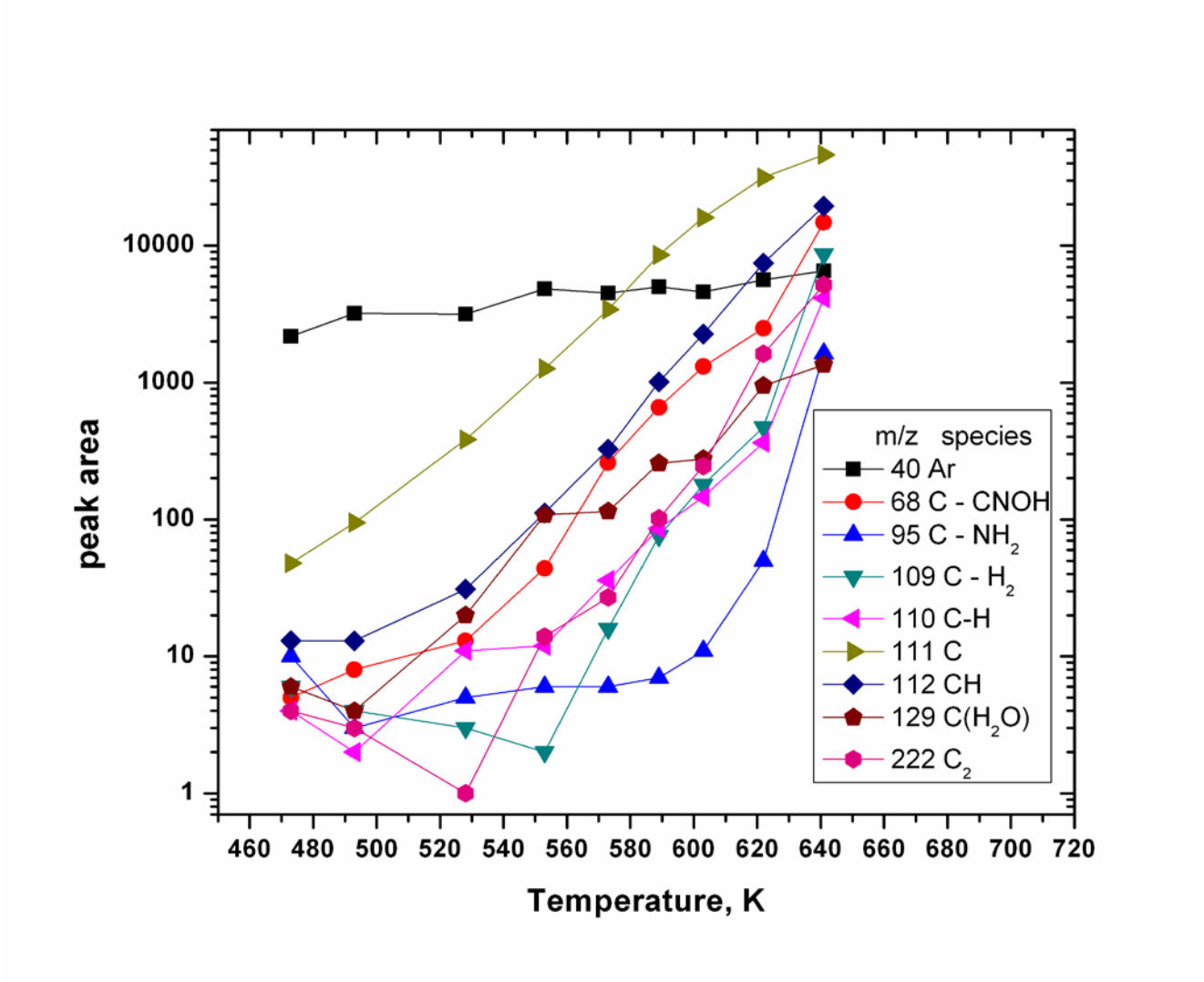


Figure 5. The intensity plots of the major mass spectral peaks recorded at 10 eV photon energy as a function of heater temperature. The inset shows m/z and suggested chemical species and fragments.

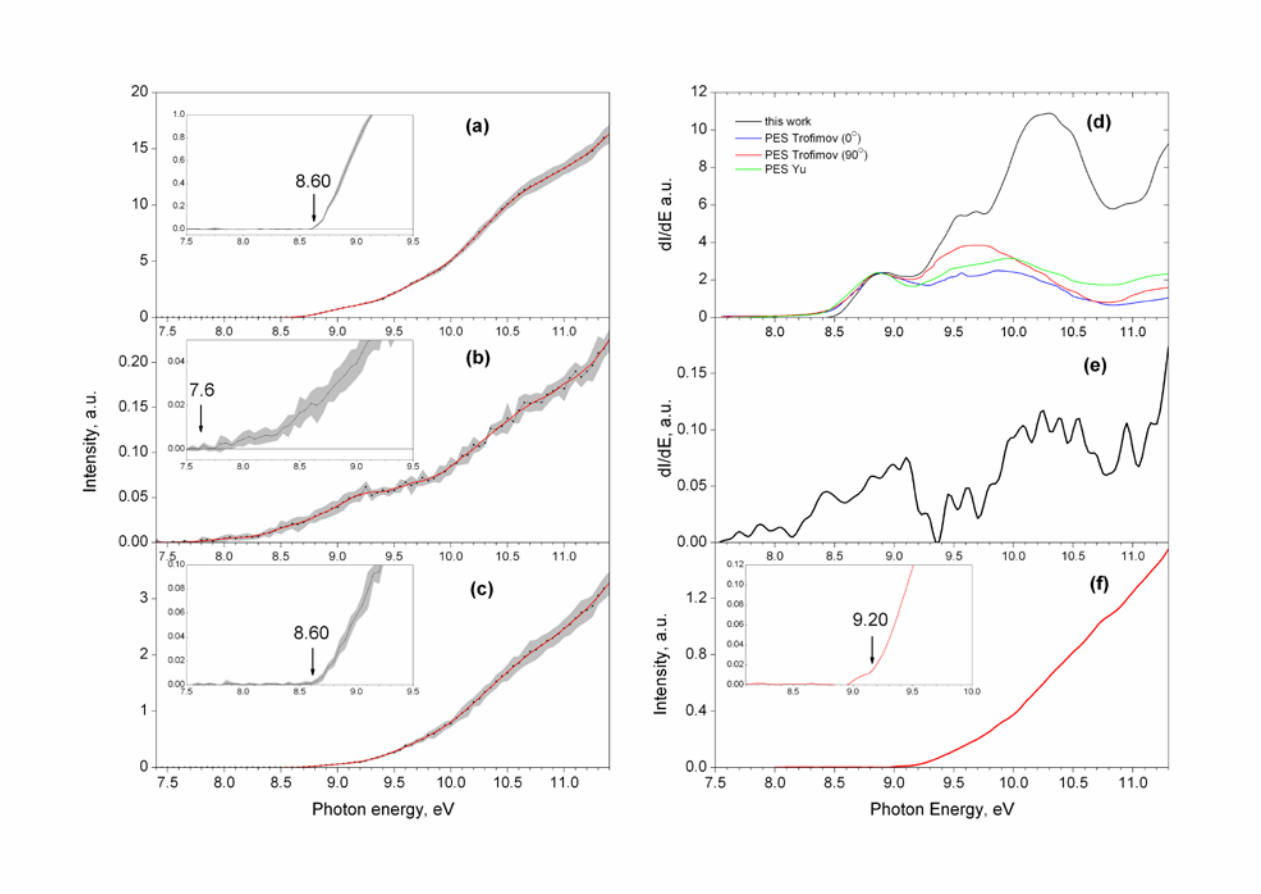


Figure 6. PIE curves for: (a) cytosine monomer (m/z 111); (b) cytosine dimer (m/z 222); (c) m/z 112. The shaded area is the mean standard deviation of the results generated from 12 scans. The red line is obtained by applying a 5 point adjacent averaged smoothing routine. The insets in (a), (b), and (c) show expanded region of the onset between 7.5 and 9.5 eV. (d) Derivative of 5 point smoothed signal from cytosine monomer shown in (a), red and blue lines are the PES's from Trofimov et al.,[9] green line is the PES from Yu et al.[8] (offset by 0.27 eV). (e) Derivative of 5 point smoothed signal from cytosine dimer shown in (b). (f) Isotope corrected PIE from m/z 111 showing true appearance energy of the protonated cytosine at m/z 112.

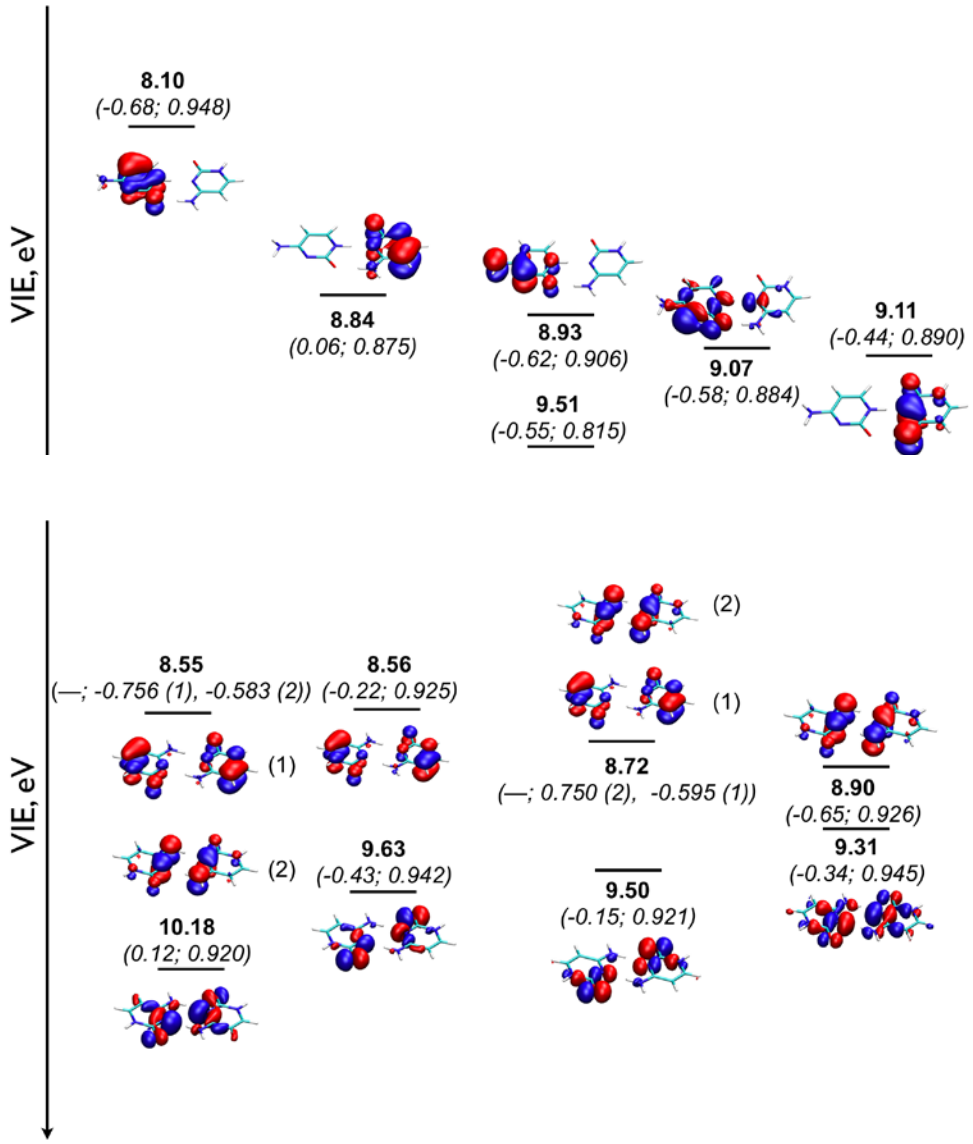


Figure 7. EOM-IP-CCSD/6-311+G(d,p) VIEs (eV) and the respective MOs for the H-bonded **C1** homodimers: **$(C1)_2$HB1** (upper panel) and **$(C1)_2$HB2** (lower panel). The shifts in VIEs relative to the monomer values (eV) and the leading EOM amplitudes are given in parentheses.

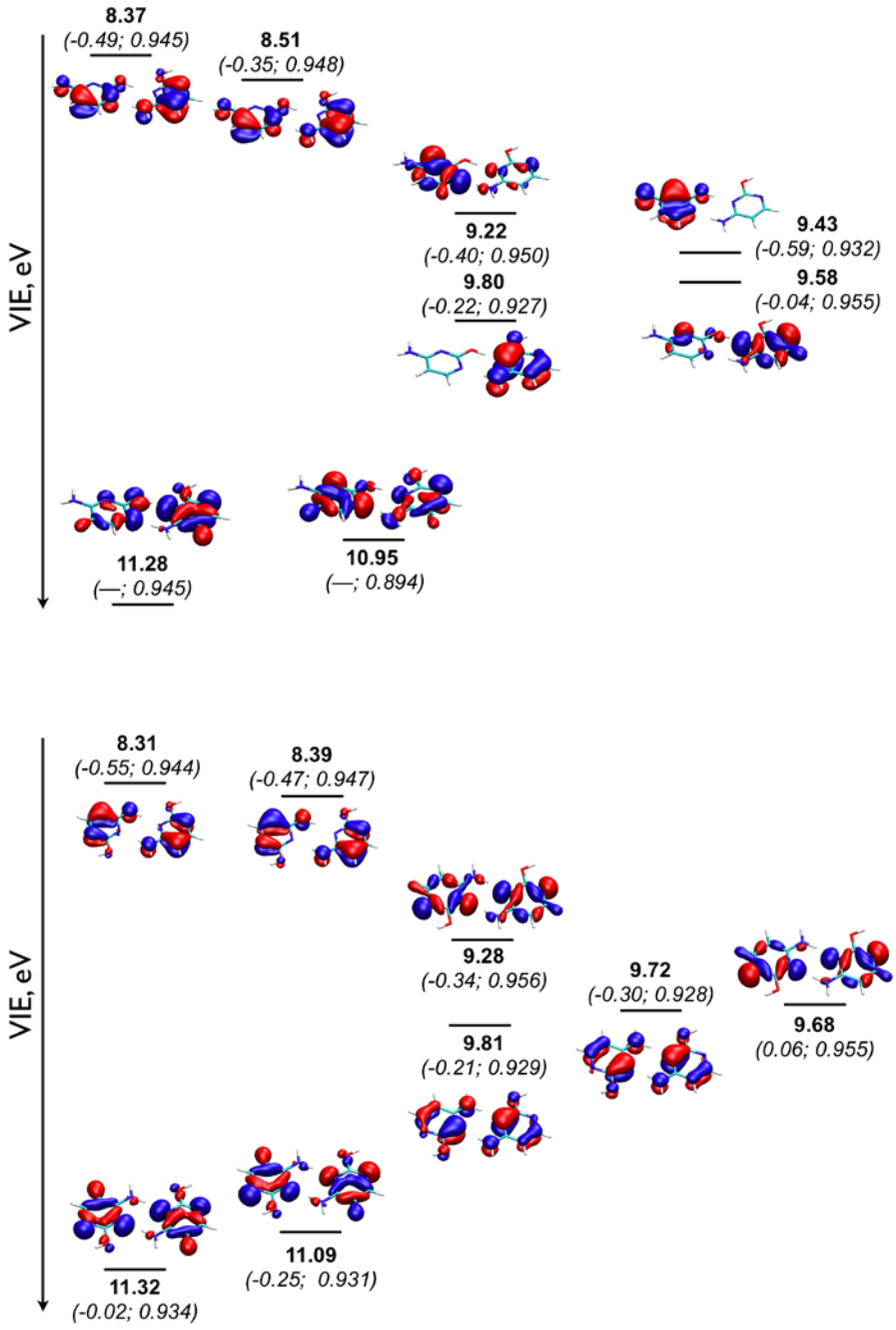


Figure 8. EOM-IP-CCSD/6-311+G(d,p) VIEs (eV) and the respective MOs for the H-bonded **C2b** homodimers: **$(C2b)_2$HB1** (upper panel) and **$(C2b)_2$HB2** (lower panel). The shifts in VIEs relative to the monomer values (eV) and the leading EOM amplitudes are given in parentheses.

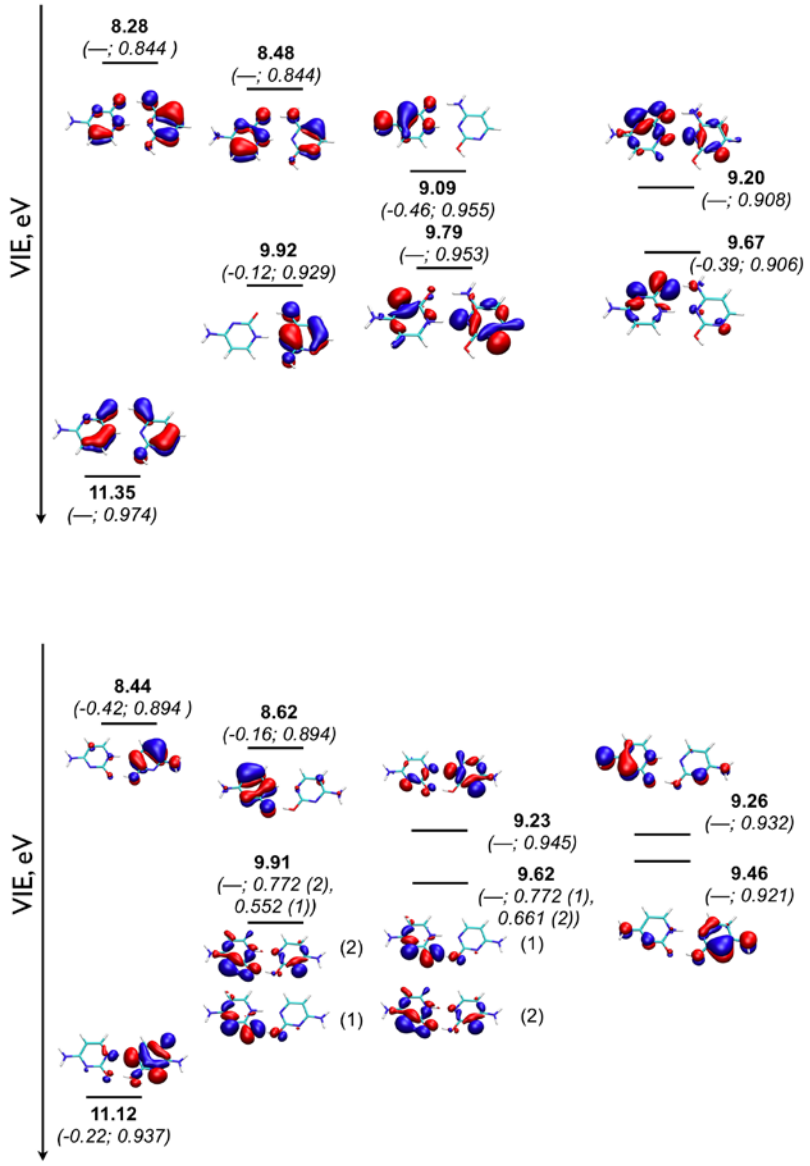


Figure 9. EOM-IP-CCSD/6-311+G(d,p) VIEs (eV) and the respective MOs for the H-bonded **C1** and **C2b** heterodimers: **(C1C2b)HB1** (upper panel) and **(C1C2b)HB2** (lower panel). The shifts in the VIEs shifts relative to the monomer (eV) as well as the leading EOM amplitudes are given in parentheses.

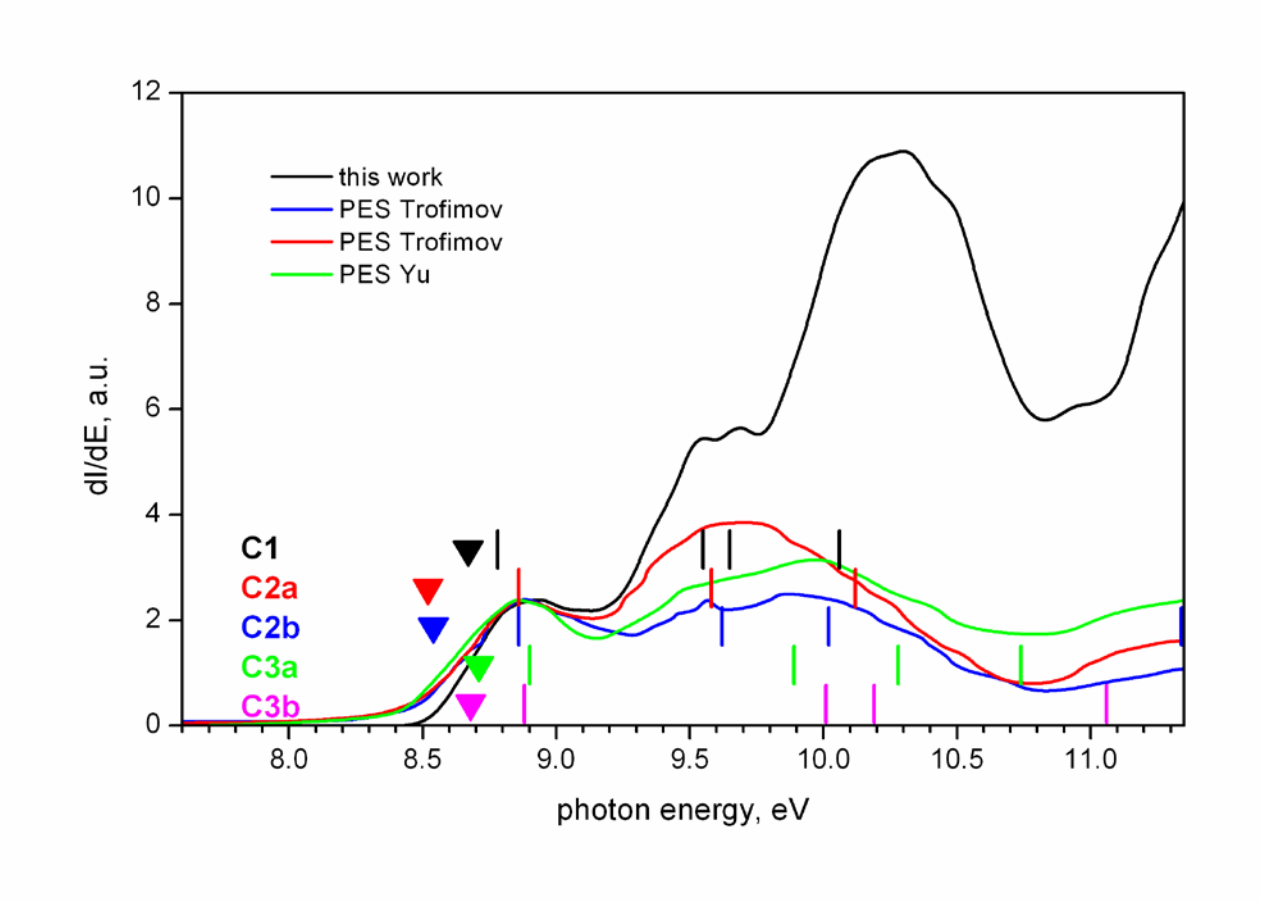


Figure 10. Derivative of the PIE curve of cytosine monomer. Also shown are PES's from Trofimov et al.[9] recorded at 40 eV photon energy, heater temperature of 463 K and two detection geometries (blue 0^{O} and red 90^{O}); green is from Yu et al.[8] recorded at heater temperature of 497 K and 21 eV photon energy. The data of Yu et al. was offset by 0.27 eV to fit the first peak from this work and from Trofimov et al. AIEs are shown by inverted triangles. The vertical lines show the calculated VIEs.

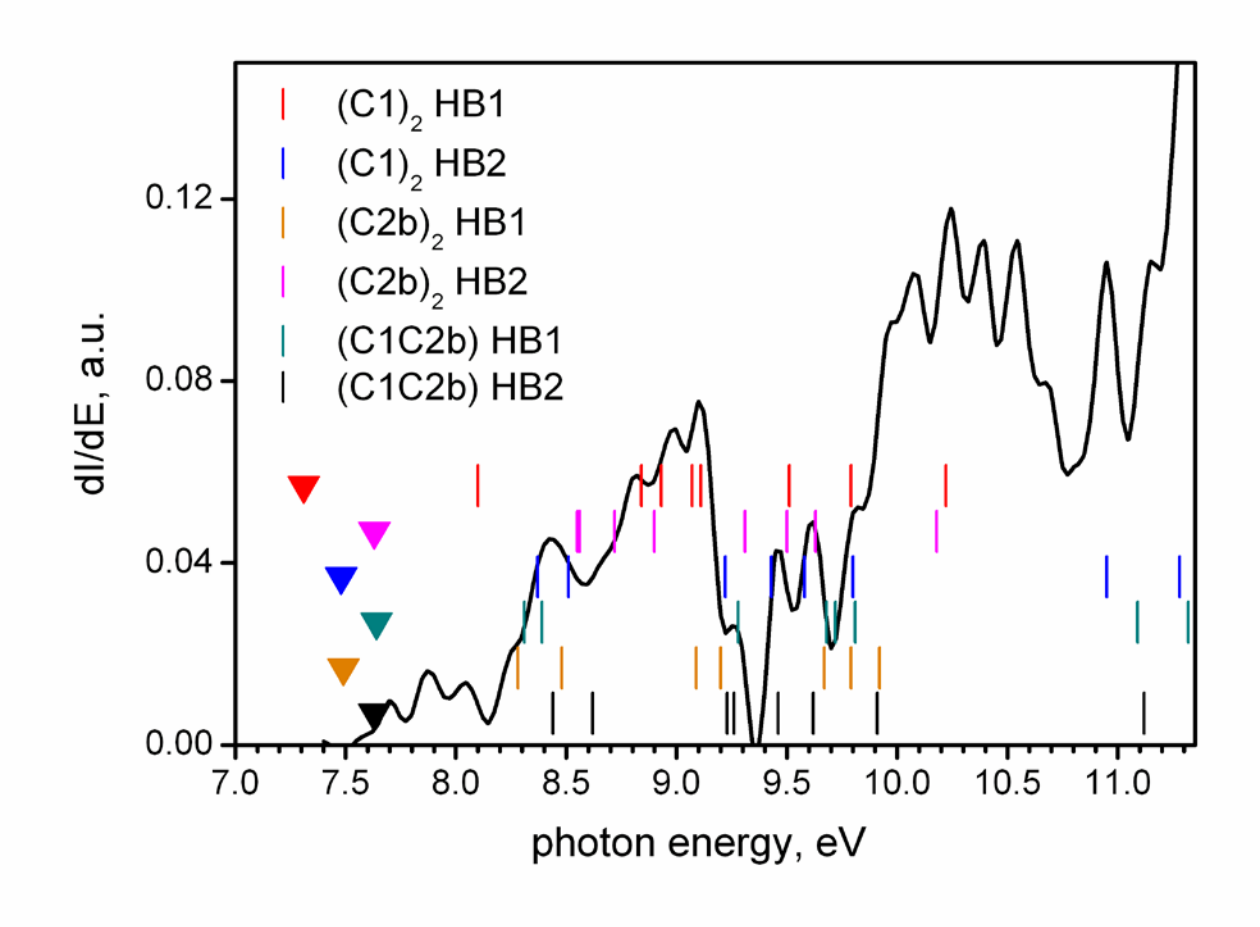


Figure 11. Derivative of the PIE curve for the cytosine dimer. The vertical lines show the calculated VIEs and AIEs are shown by inverted triangles.