Reactivity of Cytosine with Alkylmercury Ions in the Gas Phase: the Critical Role of the Alkyl Chain

Jean-yves Salpin, Violette Haldys, Jean-claude Guillemin, Otilia Mó, Manuel Yáñez, M. Merced Montero-Campillo

Reactivity of cytosine with alkylmercury ions in the gas phase: the critical role of the alkyl chain

Jean-Yves Salpin*[a], Violette Haldys[a], Jean-Claude Guillemin[b], Otilia Mó[c], Manuel Yáñez*[c], M. Merced Montero-Campillo*[c]

[a] Université Paris-Saclay, Univ Evry, CY Cergy Paris Université, CNRS, LAMBE, 91025, Evry-Courcouronnes, France

[b] Univ Rennes, Ecole Nationale Supérieure de Chimie de Rennes, CNRS, ISCR – UMR6226, F-35000 Rennes, France

[c] Departamento de Química, Módulo 13, Facultad de Ciencias, and Institute of Advanced Chemical Sciences (IAdChem), Universidad Autónoma de Madrid, Campus de Excelencia UAM-CSIC, Cantoblanco, 28049 Madrid, Spain

Dedicated to Prof. Helmut Schwarz on the occasion of his 80th birthday.

Number of pages (including Tables, Figures, legends and schemes): 25

Corresponding authors: Jean-Yves Salpin
Tel: 33 1 69 47 76 44 Fax: 33 1 69 47 76 55
e-mail : jeanyves.salpin@univ-evry.fr

Manuel Yáñez
Tel: 34 91 4974953
e-mail : manuel.yanez@uam.es

M. Merced Montero-Campillo
Tel: 34 91 4973462
e-mail : mm.montero@uam.es

ORCID

J-Y. Salpin: https://orcid.org/0000-0003-0979-1251
M. Yáñez: https://orcid.org/0000-0003-0854-585X
O. Mó: https://orcid.org/0000-0003-2596-5987
J.-C. Guillemin: https://orcid.org/0000-0002-2929-057X
M. Merced Montero-Campillo: http://orcid.org/0000-0002-9499-0900

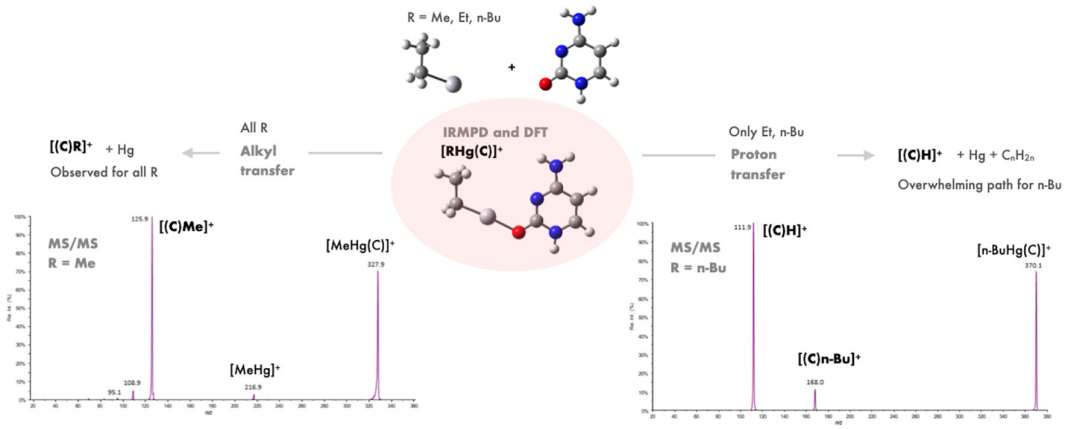
Abstract

The gas-phase reactivity towards cytosine **(C)** of alkylmercury cations $C_nH_{2n+1}Hg^+$, and more particularly CH_3Hg^+, $C_2H_5Hg^+$, *n*-$C_4H_9Hg^+$ and *t*-$C_4H_9Hg^+$, has been studied for the first time by combining tandem mass spectrometry, infrared multiple photon dissociation spectroscopy (IRMPD) and density functional theory (DFT) calculations. Under electrospray conditions, the interaction of **C** with the cations derived from alkylmercury chloride compounds gives rise to a single type of complex of general formula $[RHg(\mathbf{C})]^+$, except for *t*-butylmercury which turned to be unreactive. Subsequent MS/MS experiments showed that $[RHg]^+$ ions (R=Me, Et, n-Bu) exhibit a peculiar reactivity characterized by the transfer of the alkyl group, R, to the nucleobase leading to a $[(\mathbf{C})R]^+$ ion, accompanied by the reduction of the metal and loss of ^0Hg. As the length of the alkyl chain increases (n≥2), a new fragmentation path leading to protonated cytosine is opened, associated with the elimination of a C_n,H_{2n},Hg moiety. This latter process is clearly overwhelming with *n*-$BuHg^+$. The mechanisms associated with both dissociation channels were examined through the use of IRMPD data in the fingerprint region, and by exploring the corresponding potential energy surfaces in the DFT framework.

Keywords:

Alkylmercury compounds; cytosine; mass spectrometry; IRMPD spectroscopy; DFT calculations; pollutants

Table of Contents



1. Introduction

Unlike many transition metals, Hg has no known physiological activity as nutrient or in any other natural function.[1] However, this metal has attracted considerable attention because mercury has become a major environmental contaminant with the advent of the industrial era, and it is, with cadmium and lead, one of the most toxic metals for human beings, causing serious damage to different organs.[2] Mercury may express its toxicity according to different mechanisms, some implying the direct interaction with DNA.[3] The pioneering work of Katz has evidenced a very strong affinity of Hg toward the T-T base pair (T = thymine) and more particularly to the N3 position of the thymine residue.[4,5] The recent discovery that this interaction is not only particularly strong, but also highly selective in clear contrast with other transition metal ions,[6] has recently motivated much research on the interactions between Hg^{II} and DNA, notably to exploit this strong interaction to design mercury-specific sensors.

The high toxicity of mercury is also present in its organometallic forms $[RHg]^+$ (R=alkyl or aryl).[7,8] Among them, the methylmercury cation, CH_3Hg^+, is probably the most ubiquitous (it is naturally found in the environment and in the food chain),[9,10] and owing to its enhanced solubility in water, is a dangerous pollutant. CH_3Hg^+ is strongly neurotoxic, affecting the central nervous system.[11] Its toxicity has been shown to be associated with its interaction with cysteine and selenocysteine, due to the high affinity of Hg to sulfur and selenium.[12,13] Interactions of $[RHg]^+$ ions with DNA double helix have also been hypothesized.[3] However, the detailed mechanisms of the interaction of $[RHg]^+$ ions with DNA building blocks have yet to be clearly characterized. In this context, gas-phase studies may provide useful insights about the mechanisms occurring at the molecular level, especially when these studies combine experimental information and theoretical calculations. Helmut Schwarz has been one of the scientists who clearly demonstrated the important role of the experiment-theory synergy in the study of the gas-phase reactivity of metal ions. Notably, in relation to the present work, he studied the alkylation of amines by methylmetal complexes, the metals being Zn, Cd and Hg.[14,15] Since the mid-nineties, our groups also combine both experimental and theoretical tools, to study a variety of chemical systems involving the interaction of metal ions with organic molecules and biomolecules. We notably examined in great detail the interactions of pyrimidic nucleobases with different metal ions, and compared the unimolecular reactivity observed with copper,[16,17] calcium,[18,19] or heavier metals[17,20,21,22] and recently mercury.[23,24] In the present paper, we keep on exploring the gas-phase reactivity of alkylmercury cations, by considering their interactions with cytosine **(C)**. MS/MS experiments, Infrared Multiple Photon Dissociation (IRMPD) and Density Functional Theory (DFT) calculations were presently used to characterize the structure of both the complexes and resulting product ions, and to explore the key points of the potential energy surfaces of the associated mechanisms.

2. Methodology

2.1 Mass spectrometry.

Complexes were generated in the gas phase by electrospray ionization (turbospray ion source) coupled to a triple-quadrupole instrument (Applied Biosystems/MDS Sciex API 2000). To this end, equimolar mixtures of alkylmercury chloride/cytosine (10^{-4}M /10^{-4}M), prepared in 50/50 methanol/milli-Q water, were prepared and infused in the source with a syringe pump. ESI conditions were as follows: flow rate: 300 µl/h; sprayer probe voltage: 5.0 kV; pressure of GAS1 (nebulizing gas, air): 2.1 bars; pressure of GAS2 (air): 2.1 bars, temperature of GAS2: 100°C; pressure of curtain gas (N_2): 1.4 bars. Cytosine (**C**) and methanol used in this work were purchased from Sigma-Aldrich (Saint-Quentin Fallavier, France) and were used without further purification.

We also recorded low-energy Collision Induced Dissociation (CID) spectra of the complexes of interest by selecting in the first quadrupole (Q1) the precursor ions. Once selected, ions were allowed to collide with nitrogen in the collision cell (Q2), at different collision energies, and the resulting products were analyzed by the second mass filter (Q3). The collision energy was varied from 5 to 20 eV (laboratory frame), by adjusting the difference of potentials between the focusing quadrupole Q0 and Q2. We used nitrogen as collision gas in the second quadrupole at a total pressure of $3x10^{-5}$ mbar, the background pressure being around 10^{-5} mbar as measured by the ion gauge located outside the collision cell. In fact, the actual pressure inside the collision cell for this type of instrument being of several 10^{-2} mbars[25], MS/MS spectra are very likely obtained under a multiple-collision regime, as already discussed in previous works.[21,26]

2.2 Infrared multiple photon dissociation (IRMPD) spectroscopy.

We performed IRMPD experiments in the fingerprint region (900-1900 cm^{-1}) by using the beamline of the free electron laser (FEL) of the Centre Laser Infrarouge d'Orsay (CLIO).[27] The FEL beamline (electron energy set at 44 MeV) was coupled to a Bruker quadrupole ion trap (Esquire 3000+). This coupling has been extensively described previously.[28,29]

Complexes of interest were transferred into the gas phase by electrospraying the water/methanol solutions prepared as described previously (*vide supra*). The ESI source parameters were set as follows: flow rate: 180 µl/h; spray voltage: 4.5 kV; temperature of the transfer capillary: 170 °C.

We used the Bruker Esquire Control (v5.2) software to record the IRMPD spectra. To this end, complexes of interest (or the first generation of fragment ions) were first isolated (we selected the whole isotopic distribution for mercury complexes) and then irradiated for 200-500 ms (with or without attenuation, depending on the ion) during the MS2 (MS3) step. The excitation amplitude was set to 0 to avoid any CID-like process. Mass spectra were acquired by using the following conditions: accumulation time: 20 ms; number of accumulations: 10; *m/z*

range: 50-3000; scan resolution: 13000 Th/s. This acquisition cycle was repeated ten times for each photon wavelength.

IRMPD spectra are obtained by plotting the photofragmentation yield R ($R = -\ln[I_{precursor}/(I_{precursor} + \Sigma I_{products})]$), where $I_{precursor}$ and $I_{products}$ are the integrated intensities of the mass peaks of the precursor and of the product ions, respectively, as a function of the frequency of the IR radiation.

All the *m/z* values discussed in the text correspond to ions incorporating the dominant ^{202}Hg isotope.

2.3 Synthesis.

Methyl, ethyl, *n*-butyl, and *t*-butylmercury chloride have been synthesized as previously reported[30] starting from mercury(II) chloride and methyl magnesium chloride, ethyl magnesium chloride, *n*-butylmagnesium chloride or *t*-butylmagnesium chloride, respectively.[30]

2.4 Computational details.

We carried out a detailed study of the isomers of the cationic forms [MeHg(**C**)]$^+$, [EtHg(**C**)]$^+$, [(**C**)Me]$^+$ and [(**C**)Et]$^+$. For the mechanistic studies, we also obtained some selected structures of [*n*-BuHg(**C**)]$^+$, [(**C**)*n*-Bu]$^+$, [(**C**)H]$^+$ and neutral cytosine, as well as the corresponding hydrocarbon fragments resulting from the reactions. All the equilibrium geometries were obtained with the Gaussian16 software using the B3LYP functional.[31,32,33] The functional was used in combination with the def2-TVZPPD basis set for Hg, which includes a small core pseudopotential to account for relativistic effects, and the Pople basis set 6-31++G(d,p) for the remaining atoms. The harmonic frequencies were calculated at the same level of theory to identify minima and transition states, estimate the energy corrections and obtain the infrared (IR) fingerprints. It is worth mentioning that the method was chosen following the results of a previous theoretical assessment for the computational treatment of mercury compounds[34] for a proper comparison with experimental IR spectra, binding energies and ionization energies. In order to compare the IRMPD and the theoretical vibrational spectra, the computed modes were scaled by a factor of 0.97 and convoluted with a 10 cm^{-1} lorentzian function. Additionally, in order to check the existence of non covalent interactions between Hg and N in possible bidentate structures, the topology of the electron density was analyzed for some particular structures using the QTAIM[35] and NCI methods.[36]

3. Results

3.1 MS and MS/MS study

We first combined electrospray ionization to tandem mass spectrometry to study the interactions taking place between the cations derived from alkylmercury compounds and cytosine.

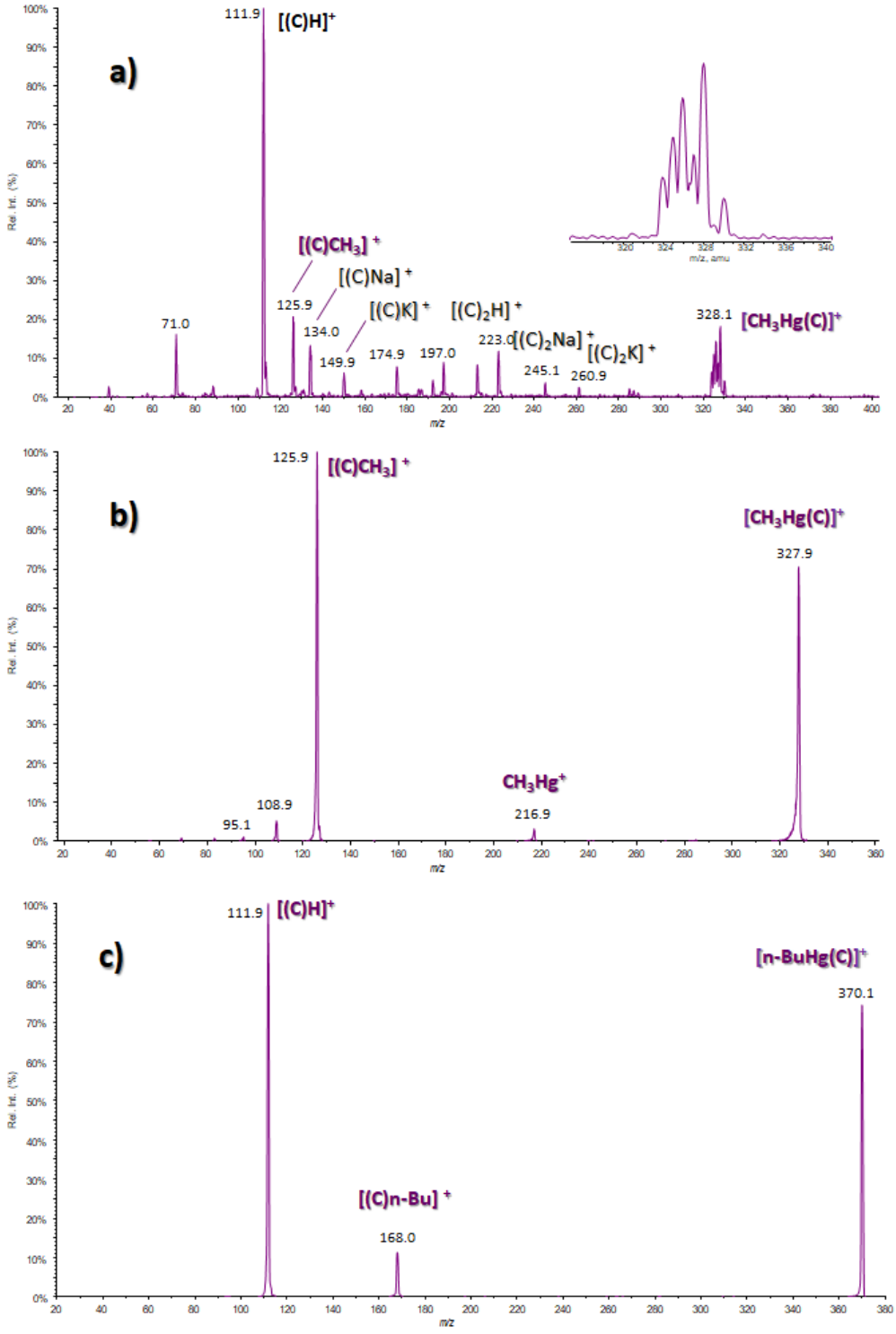
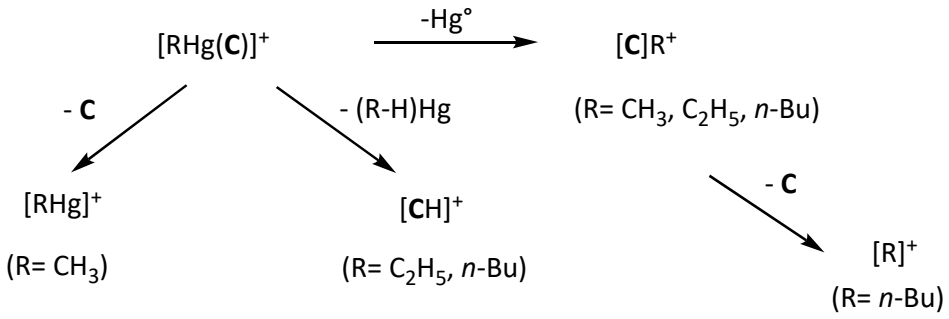


Figure 1 : **a)** Electrospray spectrum of an equimolar mixture of CH_3HgCl and cytosine (10^{-4} M) in a water/methanol mixture (50/50 v/v); low-energy MS/MS spectra of **b)** $[CH_3Hg(\mathbf{C})]^+$ and **c)** [*n*- $C_4H_9Hg(\mathbf{C})]^+$ complexes recorded at a collision energy of 20 and 15 eV, respectively (laboratory frame).

Figure 1a presents a typical electrospray spectrum obtained on our triple-quadrupole instrument, presently for the CH_3HgCl/cytosine system, recorded at a cone voltage (namely the declustering potential; DP) of 20 V. Using a low DP value allows limiting in source fragmentations. As can be seen in Figure 1a, the interaction established leads to a single type of complex of general formula $[CH_3Hg(\mathbf{C})]^+$ (*m/z* 328), resulting from the

simple addition of the $[RHg]^+$ moiety onto the nucleobase. Its abundance intensity is significant albeit low, and quickly drops as the DP parameter is increased. The mercury-containing ions are easily identified by using the characteristic isotopic distribution of this metal (see insert in Figure 1a). The isotopic profiles also indicate the absence of the chlorine atom. In the presence of *n*-BuHgCl, this complex is shifted by 42 mass units (*m/z* 370) (see Figure 1c). We performed additional experiments on a quadrupole ion trap (Bruker Amazon HCT). Using a different instrument and ion source results in similar electrospray spectra, as illustrated by the ESI spectrum displayed in Figure S1a, obtained with ethylmercury chloride and characterized by an abundant $[C_2H_5Hg(\mathbf{C})]^+$ complex (*m/z* 342). Conversely, in spite of the many attempts, by changing the solvent conditions or the metal/nucleobase ratio, we did not manage to observe any complex using the *tert*-butylmercury chloride. Using harsher source conditions results in the fragmentation of the complex, and the formation of the methylated cytosine *m/z* 126 product ions (*vide infra*). The type of complexes observed with organomercury cations is sensibly different from those generated under electrospray conditions when cytosine is mixed with lead nitrate ($[Pb(\mathbf{C})_n\text{-}H]^+$ with n=1-5 and $[Pb(\mathbf{C})_p]^{2+}$ with p=2-4).[26] Deprotonation of cytosine ($[M(\mathbf{C})\text{-}H]^+$) is also the dominant process in presence of alkali earth chloride salts[37], but low abundant $[MCl + \mathbf{C}]^+$ adducts could also be detected. Formation of simple adducts with cytosine has also already been observed with alkali or copper monocations ($[M+\mathbf{C}]^+$ and $[\mathbf{C}\text{-}M\text{-}\mathbf{C}]^+$). [38, 39, 40, 41]

In order to describe the unimolecular reactivity of the $[RHg(\mathbf{C})]^+$ complexes, we recorded a series of MS/MS experiments on different instruments, including in source fragmentations followed by MS/MS spectra of product ions, or monoisotopic selection of precursor ions with different mercury isotopes. With the triple quadrupole, we recorded spectra at different collision energies between 2 to 20 eV in the laboratory frame. Typical MS/MS spectra of the $[CH_3Hg(\mathbf{C})]^+$ (*m/z* 328) and $[n\text{-}BuHg(\mathbf{C})]^+$ (*m/z* 370) complexes are reported in Figure 1b and 1c, respectively. The same behavior upon dissociation has been observed on the ion trap using helium as target gas. An example of ion trap MS/MS spectrum is given in Figure S1b for the C_2H_5HgCl/cytosine system. From these different experiments we could deduce that there are not primary product ions whose *m/z* values are below 100 amu. The fragmentation scheme is summarized both in Scheme 1 and in Table 1.



Scheme 1: fragmentation pattern of the $[RHg(\mathbf{C})]^+$ complexes.

Table 1: Product ions observed during the fragmentation of the different $[RHg(\mathbf{C})]^+$ complexes. *m/z* values are given for the ions including the ^{202}Hg isotope. For R = *t*-C_4H_9, no reaction takes place.

[RHgCl]/C	Precursor ion [RHg(C)]$^+$	Product ions [RHg]$^+$	[C]R$^+$	[CH]$^+$	[R]$^+$
R= CH_3	*m/z* 328	*m/z* 217	*m/z* 126	–	–
R= C_2H_5	*m/z* 342	–	*m/z* 140	*m/z* 112	–
R=*n*-C_4H_9	*m/z* 370	–	*m/z* 168	*m/z* 112	*m/z* 57

The observed fragmentation patterns are remarkably similar to those found for uracil (U) and thymine.[23] The unimolecular reactivity of the [CH_3Hg(**C**)]$^+$ complex is characterized by two distinct processes. The first one corresponds to the elimination of the nucleobase, leading to [CH_3Hg]$^+$ (*m/z* 217). The second and very characteristic process corresponds to the transfer of the methyl group to cytosine, leading to [(**C**)CH_3]$^+$ ions through the loss of Hg°. This is by far the prominent process observed when R=CH_3 (Figure 1b). A new dissociation channel, namely formation of protonated cytosine associated with elimination of C_n,H_{2n},Hg, is opened with the bigger alkyl groups. Alkylation and protonation of cytosine are competitive processes when R= C_2H_5 (Figure S1b), whereas protonation is clearly overwhelming with R= *n*-C_4H_9 (Figure 1c). Consequently, there is an inversion of the alkyl transfer/nucleobase protonation branching ratio as we increase the alkyl chain length. In addition to the spectrum displayed in Figure S1b, we recorded the same day, with exactly the same conditions, the MS/MS spectrum of the [C_2H_5Hg(U)]$^+$ complex (Figure S2), confirming that for uracil the proton transfer and ethyl transfer are also competitive processes.[24] The higher abundance presently observed of protonated cytosine with respect to ethyl transfer is consistent with the higher gas-phase basicity of cytosine as compared to that of uracil.[42]. In fact, as will be shown in the computational studies section, cytosine reactions are in general much more favored than uracil's for the same alkylmercury cations. Finally, in Figure 1b, the ions observed in low abundance below *m/z* 100 come from the subsequent fragmentation of methyl-cationized cytosine [(**C**)CH_3]$^+$ (*m/z* 126). In summary, no matter the alkyl group, the fragmentation channels observed are very specific of alkylmercury cations. In addition, they preserve the integrity of the pyrimidine ring. This situation had been already encountered for the alkali metal complexes,[38,39,40] which dissociate by eliminating the intact nucleobase. The behaviour upon dissociation of [RHg(**C**)]$^+$ complexes is therefore sensibly different from the loss of H,N,C,O observed either during photodissociation of [Cu(**C**)]$^+$ ions,[41] or CID activation of [Pb(**C**)-H]$^+$ complexes.[26]

3.2 Study of the [RHg(C)]$^+$ complexes

Computational study. DFT calculations were used to interpret the IRMPD results of the observed [RHg(**C**)]$^+$ complexes, as well as to understand the reactivity of cytosine towards the different alkylmercury cations. With these aims in mind, we calculated an extensive set of isomers for the different [RHg(**C**)]$^+$ species (R = Me, Et),

taking into account different conformational orientations. Only the most stable [MeHg(**C**)]$^+$ cations are shown in Figure 2, whereas the whole list of energies of the [MeHg(**C**)]$^+$ and [EtHg(**C**)]$^+$ isomers can be found in the Supporting Information (see Tables S1-S2).

Figure 2 contains also the labeling code used for the isomers, in which the cytosine (**C**) ring positions are identified with numbers 1-6 (N1, C2, N3, C4, C5, C6) and characters *a*, *b*, *c*, *d* are related to the different conformers arising from substitution at oxygen O(C2) or nitrogen N(C4), whereas prefix **e**, **i** denotes **e**nol and **i**mine groups, respectively. Some examples to illustrate the nomenclature used are shown in Table S1. It is important to note that the study of the isomers covers not only the rotamers for the most stable forms but also possible tautomers, as oxygen and nitrogen binding sites may lead to different keto/enol and imine/enamine forms.

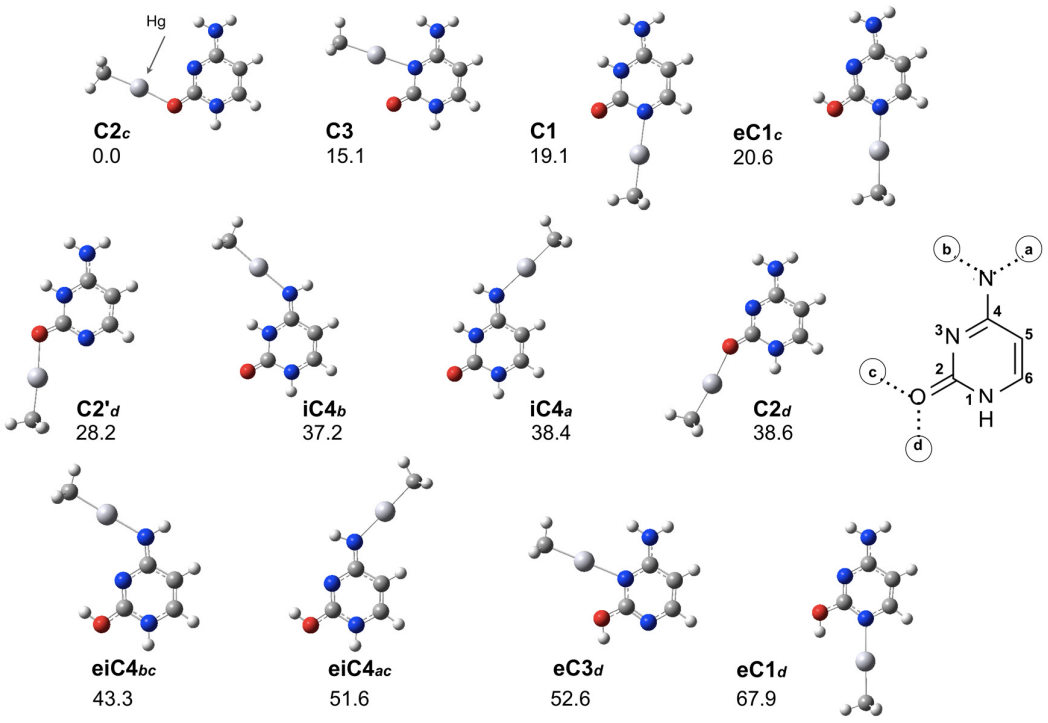


Figure 2 : Most stable methylmercury cytosine cation isomers [CH_3Hg(**C**)]$^+$ along with their relative energies (E+ZPE, kJ/mol) at the B3LYP/6-31++G(d,p)/DEF2-TZVPPD level of theory. See details in the text for the nomenclature adopted and selected examples in Table S1.

The relative energies of the [MeHg(**C**)]$^+$ species in Figure 2 show that, from all basic sites in cytosine, the attachment of methylmercury to the oxygen atom O(C2) of the keto-enamine form of neutral cytosine (see Figure S3) leads to the global minimum of the potential energy surface of the system, **C2*c***. This global minimum is followed in energy by local minima resulting from the attachment of methylmercury at N3 (**C3**, +15.1 kJ/mol) and N1 (**C1**, +19.1 kJ/mol). This preference for oxygen attachment found for [MeHg(**C**)]$^+$ is in line with the results obtained for [MeHg(U)]$^+$, where a keto form involving substitution at the O(C4) is the most stable isomer.[23] The significant gap observed in cytosine between substitution at O(C2) and N3 positions is also close

to that found for uracil at the O(C4) and O(*C*2) positions (+19.9 kJ/mol in terms of free energy). Energies in Figure 2 also reveal that the rotation of the $MeHg^+$ moiety, ongoing from the global minimum **C2*c*** to the **C2*d*** rotamer, has a very significant effect on the stability of the latter that decreases by 38.6 kJ mol^{-1}, as a consequence of the repulsive interactions between the lone pairs of oxygen O(C2) and N3 already analyzed in detail for methylated uracil cations.[43] Consequently, **C2*c*** could be exclusively generated. It should be remarked that instead, the two rotamers of the most stable isomer in $[MeHg(U)]^+$ are practically degenerated (free energy gap of 5.6 kJ/mol), as nitrogen atoms N1 and N3 are protonated in that particular case. For those cases in which N3 is deprotonated, the gap between rotamers for the uracil system is also significant.[23]

IRMPD spectrum of the $[CH_3Hg(C)]^+$ complex. In order to determine the structures that are actually generated in the gas phase, we recorded the IRMPD spectrum of the $[CH_3Hg(\mathbf{C})]^+$ complex. This spectrum, which is associated with the detection of a unique photofragment (methylated cytosine; *m/z* 126), exhibits five distinct features: two significant, albeit low, bands at 1210, 1295 cm^{-1}, one sharp signal at 1480 cm^{-1}, and a broad and intense absorption around 1600 cm^{-1} due to the combination of at least two vibrational modes at 1580 and 1630 cm^{-1} (Figure 3a).

Structural assignment is then achieved by the comparison with the vibrational spectra computed for the various forms. At this point, it is important to remind that the DFT computed spectra presently reported assume a single photon absorption whereas the IRMPD process implies a multiple photon absorption regime.[44,45] Therefore, computed IR spectra may not reproduce the experimental intensities correctly. As can be seen in Figure 3b, almost all the IRMPD bands can be assigned by considering the calculated IR active modes of the global minimum **C2*c*** (see Table S3 of the Supporting Information). Indeed, the signal observed at 1210 cm^{-1} may be interpreted as the CH_3 umbrella bending mode of the CH_3Hg moiety, and a combination of C-H and N-H bending modes of cytosine. The band detected at 1480 cm^{-1} can be attributed to the C4N bond stretch. The most intense signal at 1580 cm^{-1} can be ascribed to the C2=O carbonyl stretch, logically red shifted with respect to an unperturbed carbonyl group, because of the interaction with the CH_3Hg^+ cation. Finally, the strong signal observed at 1630 cm^{-1} may correspond to the NH_2 scissoring or the C5C6 stretch, computed at 1626 and 1647 cm^{-1}, respectively. Examination of Figure 3d shows that the agreement with the spectrum computed for the second most stable structure, **C3**, is not satisfactory as it cannot account for the very broad signal above 1580 cm^{-1}. In addition, the strong absorption computed at 1718 cm^{-1} (the C2=O stretch) is not observed experimentally. Interestingly, the vibrational spectra of the **C2*c*** rotamer, namely **C2*d***, is in very good agreement with the experimental trace (Figure 3c), and this form may also be present. However, given the difference in relative energies, **C2*d*** should be present at a very low relative proportion (~0.01 %) if one assumes a Maxwell-Boltzman distribution at 298 K. Furthermore, it may easily evolve towards the global minimum **C2*c*** as the associated rotational barrier (+2.3 $kJ \cdot mol^{-1}$) is very low (see Table S7 and Figure S7). In summary, IRMPD data and energetics point to the preferential formation of the **C2*c*** structure for the $[CH_3Hg(\mathbf{C})]^+$ complex.

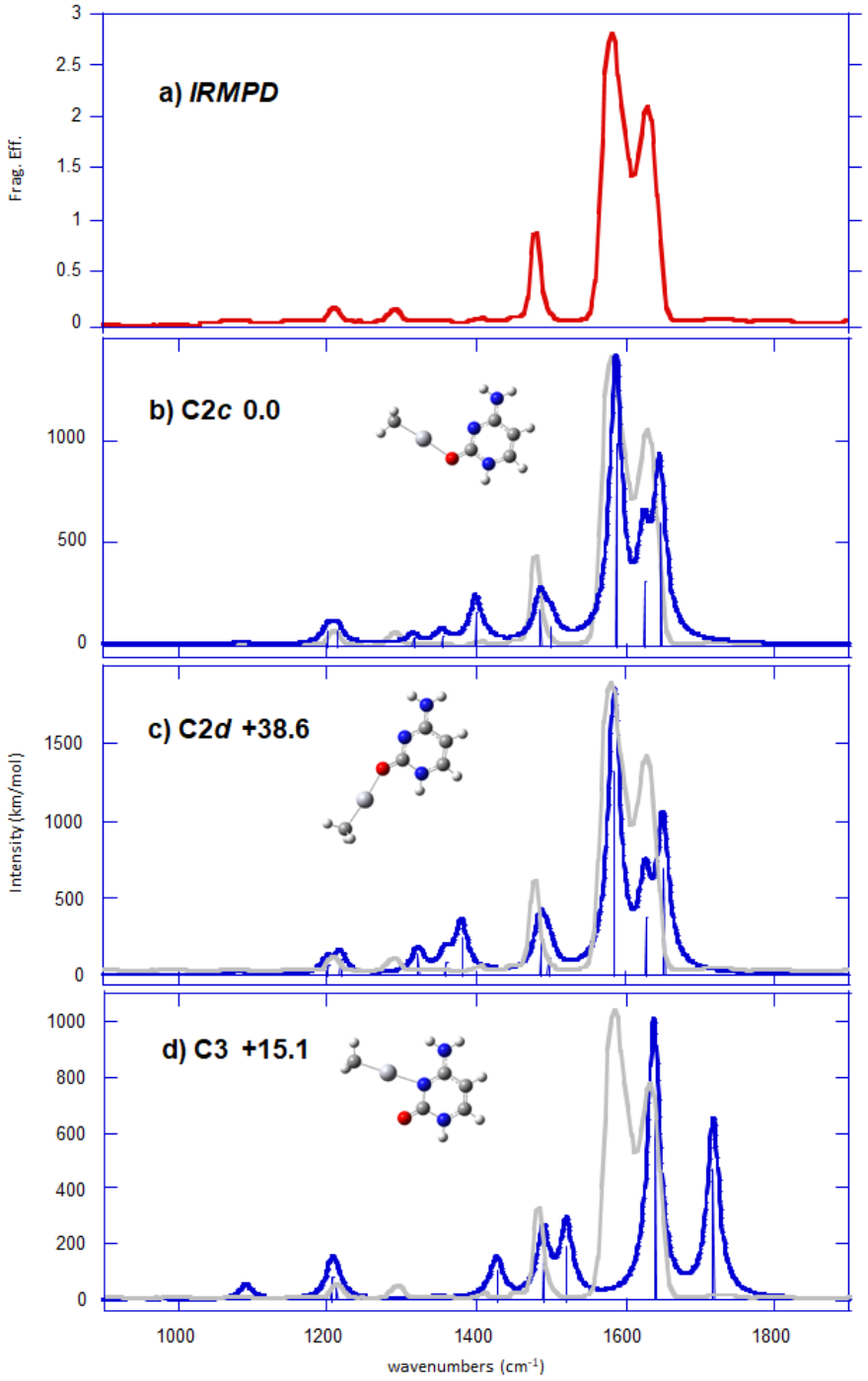


Figure 3: (**a**)IRMPD spectrum obtained for the $[CH_3Hg(C)]^+$ complex compared to DFT-computed IR absorption spectra (**b–d**) of some relevant structures. The experimental IRMPD trace is overlayed in grey. Relative energies in kJ.mol^{-1}.

It is worth mentioning that the IRMPD spectrum presently recorded shares some similarities with the IRMPD spectra recorded for $[M(C)]^+$ complexes, M being the alkali metals.[46] As a matter of fact, the IRMPD spectra of the alkali complexes exhibit notably a very intense and broad signal around 1630-1660 cm^{-1} and a weaker band around 1460-1480 cm^{-1}, their position slightly changing according to the size of the alkali cation. Yang and co-workers concluded that they are associated to a bidentate interaction with the O(C2) and N3 positions of cytosine. The photofragmentation yield of the $[Ag(C)]^+$ ion turned to be lower than those reported for the

alkali complexes,[47] but comparison with DFT calculations also pointed to a N3/O2 interaction. As far as the [Ba(**C**)-H]$^+$ complex is concerned, the N3/O2 binding mode is also evident, but an additional structure involving the interaction with N1 and O2 was also observed.[48] The N1/O2 binding mode was also found to be overwhelming for the [Pb(**C**)-H]$^+$ complex.[26] For these two latter complexes, results suggested that the structures generated by ESI were produced in solution and preserved during the electrospray process leading to the gaseous ions. In our case, the Hg-O(C2) and Hg-N3 distances in [CH_3Hg(**C**)]$^+$ are 2.13 Å and 3.16 Å, respectively, and slightly larger for [EtHg(**C**)]$^+$ (2.15 Å, 3.18 Å). As a reference, for the same period of the periodic table the reported computed values by Yang *et al* for [Cs(C)]$^+$ at the B3LYP/def2-TZVPPD were 2.81 Å and 3.63 Å [46]. The fact that the Hg-N3 distances in [CH_3Hg(**C**)]$^+$ and [EtHg(**C**)]$^+$ complexes are larger than the Hg-O(C2) ones is fully consistent with the fact that both the QTAIM and NCIPLOT topological analyses show no direct *bonding* interactions between Hg and N for the aforementioned complexes. The softer nature of the alkaline atom, whose atomic radius (343 pm) is much larger than that of Hg (150 pm), difference that is reflected in the size of the corresponding cations,[46] and the relative orientation of the N lone pair with respect to the compact Hg cloud might be critical.

3.3 Study of the reaction products of cytosine

In this section, we have gathered the data obtained to characterize the structure of the ions arising from the unimolecular dissociation of the [RHg(**C**)]$^+$ complex. These results include notably IRMPD data obtained for methylated cytosine, and a computational study which aims at proposing mechanisms that could account for the formation of alkylcytosine cations and protonated cytosine.

Structure of methylated cytosine. As the most remarkable process observed upon CID conditions is the alkylation of the nucleobase, we tried to characterize by IRMPD spectroscopy the structure of the ion corresponding to the methylation of cytosine. To this end, photons were introduced in the ion trap during the MS3 step following the CID dissociation of the [CH_3Hg(**C**)]$^+$ complex and the subsequent isolation of the resulting [(**C**)CH_3]$^+$ cation. The IRMPD spectrum obtained with an irradiation time of 1 second is given in Figure 4a.

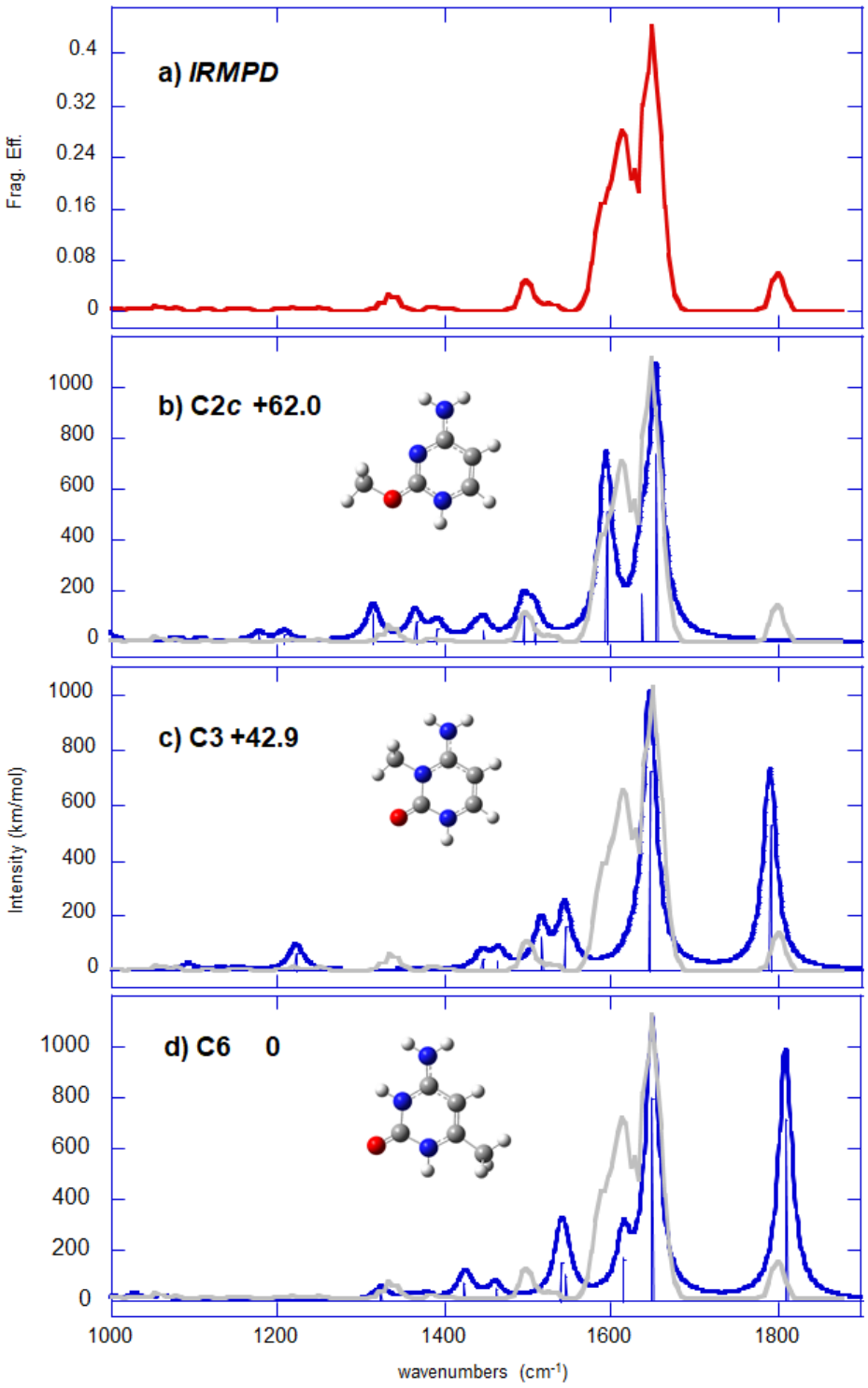


Figure 4: (**a**) IRMPD spectrum obtained for the $[(C)CH_3]^+$ ion compared to DFT-computed IR absorption spectra (**b–d**) of some relevant structures. The experimental IRMPD trace is overlayed in grey. Relative energies in kJ.mol^{-1}.

Four intense photofragments were systematically observed in resonance with the vibrational modes of the cation: *m/z* 109 (-NH_3), 95 (-C_2,H_3,N), 83 (-H,N,C,O) and 69 (-C_2,H_3,N,O). This spectrum exhibits three weak signals at 1335, 1500 and 1800 cm^{-1}, and is dominated by a broad and intense feature between 1550 and 1660 cm^{-1} resulting from the combination of several vibrational modes and notably two distinguishable maxima at

1610 and 1650 cm^{-1}. In order to interpret this spectrum, we carried out an extensive computational study of the $[(C)CH_3]^+$ cation, the structure of which being gathered in the Supporting Information (Figure S4). Figure 4b shows that all the experimental signals but the absorption around 1800 cm^{-1} can be interpreted by the vibrational spectrum computed for the **C2*c*** structure (see Table S6). **C2*c*** is characterized by a methyl group that has been transferred onto the carbonyl of cytosine. The very broad and intense signal is particularly well reproduced and can be attributed to the combination of the carbonyl stretch (1610 cm^{-1}), NH_2 scissoring bending mode (1633 cm^{-1}) and the stretch of the C5=C6 double bond. The signal detected at 1500 cm^{-1} may correspond to both C4N and N3C4 stretches, and the band observed at 1335 cm^{-1} might be ascribed to CH and NH bending modes. The very strong signal is also well reproduced by the rotamer **C2*d*** (Figure S6a), but both forms are unable to reproduce the band detected at 1800 cm^{-1}, which corresponds very likely to an unperturbed C=O stretch. The presence of this signal indicates that there is certainly a mixture of at least two forms, the second form being characterized by a methyl group not located on the carbonyl of cytosine. The computed spectrum of the structure **C3** (Figure 4c), characterized by a methyl group transferred onto the N3 position, shows a C=O stretch in agreement with the experimental signal. The apparent discrepancy between the theoretical and experimental intensities is not surprising given the rapid decrease of the FEL power above 1750 cm^{-1}. On the other hand, the agreement with the strong features around 1600 cm^{-1} is poor when only considering **C3**. These results therefore suggest that a mixture of **C2*c*/C3** structures may be formed experimentally. This is consistent with the fact that a single step is necessary to generate these structures (*vide infra*). The photofragments observed also support this assumption. As a matter of fact, we showed in previous studies that the loss of [H,N,C,O] from various metal/uracil complexes involved specifically both C2=O and N3.[16, 18, 20, 22] The fact that we presently observe a loss of 57 amu as photofragment (presumably CH_3,N,C,O) is coherent with the presence of the methyl group either on N3 or O positions of cytosine.

It is worth mentioning that according to our theoretical study, both **C2*c*** and **C3** structures do not correspond to the global minimum, as they lie 62 and 42.9 kJ/mol, respectively, above the most stable form. The global minimum, **C6**, in fact can be described as a N3 protonated form of 6-methyl-cytosine. Its computed vibrational spectrum is given in the Figure 4d and can account for the signals detected at 1650 and 1800 cm^{-1}. However, it seems reasonable to assume a kinetic control of the fragmentations presently observed, and the preferred formation of **C2c/C3** structures, which require a single exothermic step (*vide infra*). Conversely, the formation of the global minimum would require an extensive reorganization process that should not be favored kinetically. The same comment can be made for the tautomeric form **C1**, for which the agreement with the experimental spectrum is quite satisfactory (Figure S6b). The computational study of such isomerization processes is beyond the scope of the present paper, but this constitutes an open question that could be addressed in future work.

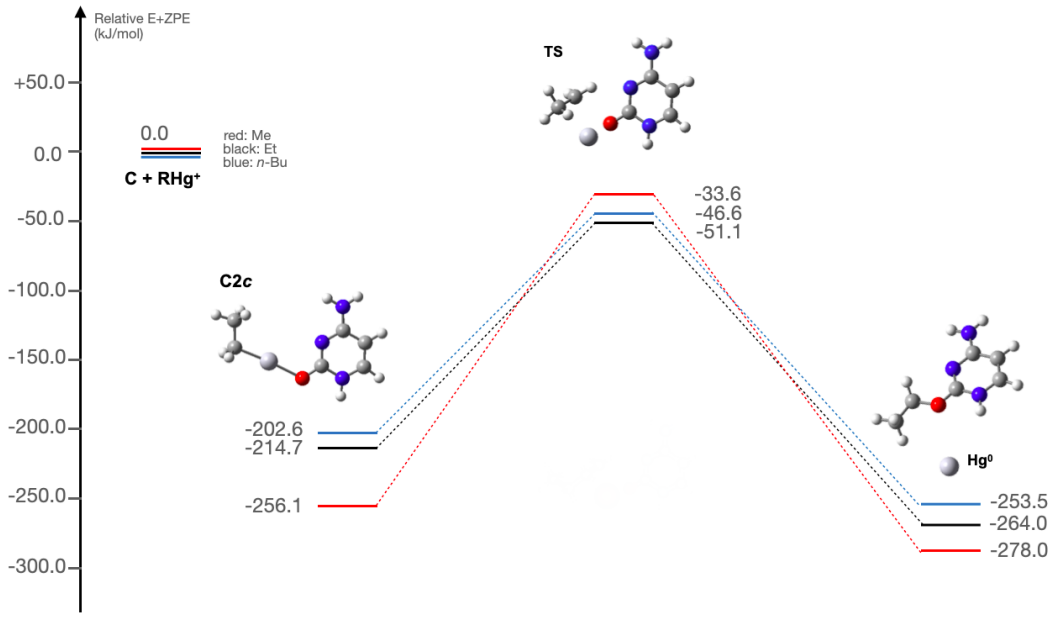


Figure 5: Alkyl transfer path connecting the [RHg(**C**)]$^+$ complexes (R = Me, Et, *n*-Bu) and alkylcytosine plus neutral mercury products respect to free reagents cytosine and alkylmercury cations. Pictures correspond to the particular case in which R = Et. Relative electronic energy plus zero-point corrections are shown in kJ·mol^{-1} at the B3LYP/6-31++G(d,p)/def2-TZVPPD level of theory.

Cytosine alkylation mechanism. Figure 5 illustrates the transition states connecting the very stable [RHg(**C**)]$^+$ complex with the corresponding alkylcytosine products for the different alkyl chains. As shown in the picture, each transition state involves a transfer to O(C2) with a relative energy always clearly below the entrance channel. Interestingly, the lowest activation barrier is obtained for R = Et, whereas the highest corresponds to R = Me, being the one for R = *n*-Bu only slightly higher than for R = Et, indicating a sort of balance between a larger inductive effect but also a larger steric hindrance. The reaction is very favorable in all cases, with products well below -250 kJ/mol, what is also true for the relative free energies (-261.9 kJ/mol (Me), -243.9 kJ/mol (Et), -232.6 kJ/mol (*n*-Bu)). The path shown in Figure 5 is similar to that found for uracil,[23] although the latter presented transition states slightly above the entrance channel in all cases.

For R = Me a different alkyl transfer path would be that of a S$_N$2-like mechanism through O(C2) or N3, similarly to what was found by Schwarz and co-workers for the methylation of ammonia by different methylmetal cations (among them, mercury).[15] We failed in describing such a mechanism for the MeHg$^+$/uracil system, but given the larger reactivity of cytosine, we decided to try again this possibility. Unfortunately, we still could not locate any transition state of this kind through O(C2) or N3; we also repeated the search for uracil, with the same results. However, as commented in previous works, it is worth mentioning that this does not mean that a methyl transfer from methylmercury does not take place, as the [Me(**C**)]$^+$ complex with neutral mercury is very stable. In fact, a simple optimization starting from a structure where the methyl group of the methylmercury cation is oriented towards O2 in cytosine leads to a methyl transfer.

Cytosine protonation mechanism. Figure 6 shows the two energy profiles associated with the formation of protonated cytosine through O(C2) when R = Et and n-Bu, with transition states at -76.6 and -66.2 kJ/mol. We also located for R = Et a second transition state leading to a proton transfer through N3, represented on the right upper corner of the figure and slightly higher in energy (-69.4 kJ/mol) than the O(C2) one. These protonation paths, although largely exothermic with respect to **C** + RHg^+, seem to be slightly unfavored with respect to the corresponding alkylmercury cytosine complexes, as shown in Figure 6. These two mechanisms lead to the structures that were characterized experimentally by IRMPD in a previous study[49] and found as the most stable forms at a high level of calculation.[50] As previously observed for uracil, the transition states associated with proton transfer processes are even lower in energy than the alkyl ones for ethyl and *n*-butyl substituents, though both reactions compete with each other. In this sense it should be observed however that in both processes the barriers are lower in energy than the entrance channel **C** + RHg^+, and the exothermicity of formation of the corresponding **C2c** complexes is in both cases larger enough to overpass the respective barriers.

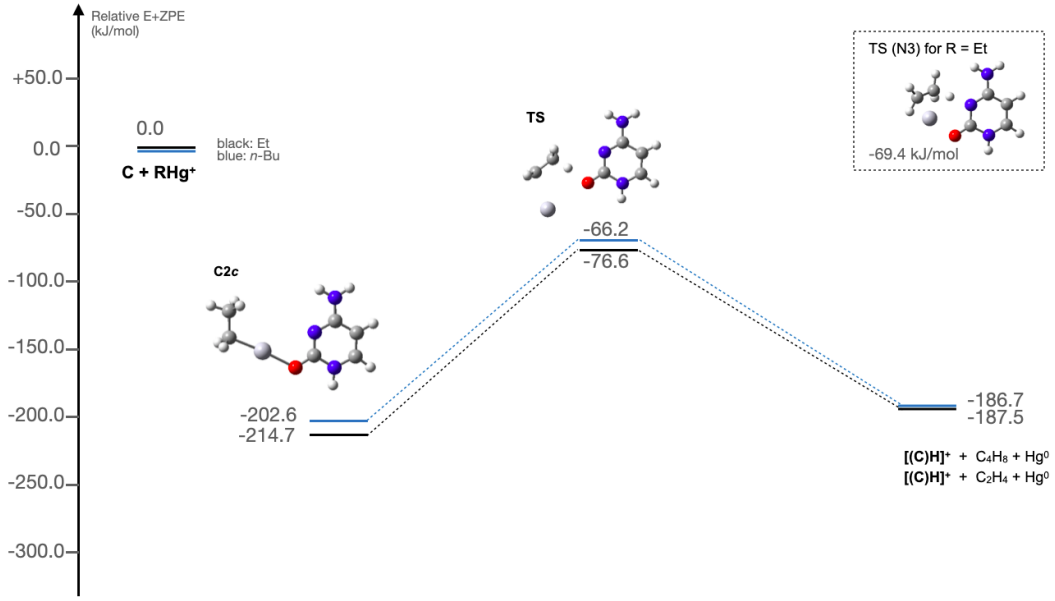


Figure 6: proton transfer path connecting the [RHg(**C**)]$^+$ complexes (R = Et, *n*-Bu) and protonated cytosine plus neutral mercury and hydrocarbon products respect to free reagents cytosine and alkylmercury cations. Relative electronic energy plus zero-point corrections are shown in kJ·mol^{-1} at the B3LYP/6-31++G(d,p)/def2-TZVPPD level of theory.

Finally, a beta-hydride elimination pathway like that studied for uracil in our previous works, would have to overcome a very high energy barrier (see Figure S5).

4. Concluding remarks and future prospect

This work studies in detail the gas-phase reactivity of cytosine towards alkylmercury cations in which the alkyl groups have different length (R = Me, Et, *n*-Bu, *t*-Bu). The first important result is that in all cases, with the

only exception of *t*-Bu which shows no reaction, the nucleobase is able to form a unique type of complex [RHg(**C**)]$^+$. For the particular case of the methyl substituent, the [RHg(**C**)]$^+$ complex has been characterized through its IRMPD spectrum. A comparison of this spectrum with the theoretical IR results clearly indicates that the methylmercury moiety is attached to the oxygen of cytosine, O(C2), resulting in the most stable of all the [MeHg(**C**)]$^+$ isomers according to DFT calculations. The attachment to either N3 or any other substitution pattern gives place to much more unstable isomers. These results are in line with complexes reported in the literature between cytosine and other metals.

The unimolecular reactivity of [RHg(**C**)]$^+$ largely depends on the alkyl chain length. For the methyl group, only the methylation product is observed, whereas ethyl and *n*-butyl groups lead to both alkylated and protonated cytosine, with branching ratios depending on the alkyl group. We have identified the transition states accounting for the formation of both product ions, finding that the proton transfer is kinetically favored with respect to the alkyl transfer.

IRMPD spectra and fragmentations suggest that for methylated cytosine the methyl group would be attached to both the N3 or O atoms of cytosine. The resulting structures are not the most stable, but they could be kinetically favored as they are obtained from the initial complex through a single exothermic step, whereas obtaining the global minimum would require an extensive atomic reorganization through several steps which might not be feasible within the experimental conditions.

A global view of the reactivity of uracil, thymine, cytosine towards alkylmercury reported by our research groups evidences a common role of the alkyl chain in these processes. Further developments are underway with zinc and cadmium to assess the role of the metal onto the observed reactivity.

Acknowledgements

The CLIO team as well as P. Maître, D. Scuderi and V. Steinmetz are warmly acknowledged for their support during the FEL experiments. This work has been also supported by the "Fonds pour le Rayonnement de la Recherche (FRR, Université d'Evry Val d'Essonne) and by Project PID2021-125207NB-C31 and PID2019-110091GB-I00 from the Ministerio de Ciencia e Innovación (MICINN) of Spain. M. M. Montero-Campillo thanks the Ministerio de Universidades for her ARPU (Ayudas para la Recualificación del Profesorado Universitario) fellowship at the Universidade de Vigo (Spain), supported by the Plan de Recuperación, Transformación y Resiliencia. Computational time at Centro de Computación Científica (CCC) of Universidad Autónoma de Madrid is also acknowledged.

Supporting Information

The file contains electrospray mass spectra, MS/MS spectra, theoretical results regarding [EtHg(**C**)]$^+$, [MeHg(**C**)]$^+$, [(**C**)Et]$^+$, [(**C**)Me]$^+$, protonated cytosine and neutral cytosine, experimental and computed IR results for [MeHg(**C**)]$^+$ and [(**C**)Me]$^+$, and the study of the rotational barrier between rotamers **C2*c*** and **C2*d***.

Bibliography

[1] M. Israr, S. Sahi, R. Datta, D. Sarkar, *Chemosphere* **2006**, *65*, 591-598.

[2] T. W. Clarkson, L. Magos, G. J. Myers, *N. Eng. J. Med.* **2003**, *349*, 1731-1737.

[3] I. Onyido, A. R. Norris, E. Buncel, *Chem. Rev.* **2004**, *104*, 5911-5930.

[4] S. Katz, *J. Am. Chem. Soc.* **1952**, *74*, 2238-2245.

[5] S. Katz, *Nature* **1962**, *194*, 569-&.

[6] A. Ono, H. Togashi, *Angew. Chem. Int. Ed.* **2004**, *116*, 4400 –4402.

[7] K. A. Graeme, C. V. Pollack, Jr., *J. Emergency Med.* **1998**, *16*, 45-56.

[8] J. F. Risher, P. Tucker, in *Rev. Environ. Contam. Toxicol., Vol. 240* (Ed.: P. DeVoogt), Springer, New York, **2017**, pp. 105-149.

[9] M. R. Karagas, A. L. Choi, E. Oken, M. Horvat, R. Schoeny, E. Kamai, W. Cowell, P. Grandjean, S. Korrick, *Environ. Health. Perspect.* **2012**, *120*, 799-806.

[10] Y. S. Hong, Y. M. Kim, K. E. Lee, *J. Prev. Med. Public. Health.* **2012**, *45*, 353-363.

[11] M. Aschner, N. Onishchenko, S. Ceccatelli, in *Organometallics in Environment and Toxicology: Metal Ions in Life Sciences, Vol. 7*, The Royal Society of Chemistry, **2010**, pp. 403-434.

[12] J. L. Franco, T. Posser, P. R. Dunkley, P. W. Dickson, J. J. Mattos, R. Martins, A. C. Bainy, M. R. Marques, A. L. Dafre, M. Farina, *Free. Radic. Biol. Med.* **2009**, *47*, 449-457.

[13] P. A. Nogara, C. S. Oliveira, G. L. Schmitz, P. C. Piquini, M. Farina, M. Aschner, J. B. T. Rocha, *Biochim. Biophys. Acta* **2019**, *1863*, 129284.

[14] R. Kretschmer, M. Schlangen, H. Schwarz, *Angew. Chem. Eng. Int. Ed.* **2011**, *50*, 5387-5391.

[15] R. Kretschmer, M. Schlangen, M. Kaupp, H. Schwarz, *Organometallics* **2012**, *31*, 3816-3824.

[16] A. M. Lamsabhi, M. Alcamí, O. Mó, M. Yáñez, J. Tortajada, J.-Y. Salpin, *ChemPhysChem* **2007**, *8*, 181-187.

[17] B. Power, V. Haldys, J.-Y. Salpin, T. D. Fridgen, *Int. J. Mass Spectrom.* **2018**, *429*, 56-65.

[18] C. Trujillo, A. Lamsabhi, O. Mó, M. Yáñez, J.-Y. Salpin, *Int. J. Mass Spectrom.* **2011**, *306*, 27-36.

[19] B. Power, V. Haldys, J.-Y. Salpin, T. D. Fridgen, *J. Mass Spectrom.* **2016**, *51*, 236-244.

[20] S. Guillaumont, J. Tortajada, J.-Y. Salpin, A. M. Lamsabhi, *Int. J. Mass Spectrom.* **2005**, *243*, 279-293.

[21] J.-Y. Salpin, S. Guillaumont, J. Tortajada, A. M. Lamsabhi, *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 359-369.

[22] J.-Y. Salpin, L. Latrous, V. Haldys, A. M. Lamsabhi, *J. Phys. Chem. A* **2018**, *122*, 992-1003.

[23] J.-Y. Salpin, V. Haldys, L. Latrous, J.-C. Guillemin, J. Tortajada, E. Leon, O. Mó, M. Yáñez, M. M. Montero-Campillo, *Int. J. Mass Spectrom.* **2019**, *436*, 153-165.

[24] Á. Pérez-Barcia, M. M. Montero-Campillo, A. M. Lamsabhi, J.-Y. Salpin, M. Yáñez, *Phys. Chem. Chem. Phys.* **2022**, *24*, 20624-20637.

[25] I. V. Chernushevich, A. V. Loboda, B. A. Thomson, *J. Mass Spectrom.* **2001**, *36*, 849-865.

[26] J.-Y. Salpin, V. Haldys, S. Guillaumont, J. Tortajada, M. Hurtado, A. Lamsabhi, *ChemPhysChem* **2014**, *15*, 2959-2971.

[27] R. Prazeres, F. Glotin, C. Insa, D. A. Jaroszynski, J. M. Ortega, *Eur. Phys. J. D* **1998**, *3*, 87-93.

[28] L. MacAleese, A. Simon, T. B. McMahon, J. M. Ortega, D. Scuderi, J. Lemaire, P. Maitre, *Int. J. Mass Spectrom.* **2006**, *249*, 14-20.

[29] B. Chiavarino, M. E. Crestoni, S. Fornarini, J. Lemaire, P. Maitre, L. MacAleese, *J. Am. Chem. Soc.* **2006**, *128*, 12553-12561.

[30] G. A. Russell, P. Ngoviwatchai, H. I. Tashtoush, *Organometallics* **1988**, *7*, 696-702.

[31] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, Williams, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, D. J. Fox, Wallingford, CT, **2016**.

[32] A. D. Becke, *Phys. Rev. A.* **1988**, *38*, 3098-3100.

[33] A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648-5652.

[34] M. M. Montero-Campillo, A. M. Lamsabhi, O. Mó, M. Yáñez, *Theor. Chem. Acc.* **2013**, *132*, 1328.

[35] R. F. W. Bader, *Atoms in Molecules. A Quantum Theory*, Clarendon Press, Oxford, **1990**.

[36] R. A. Boto, F. Peccati, R. Laplaza, C. Quan, A. Carbone, J.-P. Piquemal, Y. Maday, J. Contreras-García, *J. Theor. Comput. Chem.* **2020**, *16*, 4150-4158.

[37] M. Franska, *Eur. J. Mass Spectrom.* **2007**, *13*, 339-346.

[38] P. Wang, M. J. Polce, G. Ohanessian, C. Wesdermiotis, *J. Mass Spectrom.* **2008**, *43*, 485-494.

[39] Z. B. Yang, M. T. Rodgers, *Phys. Chem. Chem. Phys.* **2012**, *14*, 4517-4526.

[40] B. Yang, M. T. Rodgers, *Phys. Chem. Chem. Phys.* **2014**, *16*, 16110-16120.

[41] J. H. Gao, G. Berden, M. T. Rodgers, J. Oomens, *Phys. Chem. Chem. Phys.* **2016**, *18*, 7269-7277.

[42] E. P. L. Hunter, S. G. Lias, *J. Phys. Chem. Ref. Data* **1998**, *27*, 413-656.

[43] J.-Y. Salpin, V. Haldys, V. Steinmetz, E. Léon, M. Yáñez, M. M. Montero-Campillo, *Int. J. Mass Spectrom.* **2018**, *429*, 47-55.

[44] J. Oomens, B. G. Sartakov, G. Meijer, G. von Helden, *Int. J. Mass Spectrom.* **2006**, *254*, 1-19.

[45] C. F. Correia, P. O. Balaj, D. Scuderi, P. Maitre, G. Ohanessian, *J. Am. Chem. Soc.* **2008**, *130*, 3359-3370.

[46] B. Yang, R. R. Wu, N. C. Polfer, G. Berden, J. Oomens, M. T. Rodgers, *J. Am. Soc. Mass Spectrom.* **2013**, *24*, 1523-1533.

[47] M. Berdakin, V. Steinmetz, P. Maitre, G. A. Pino, *Phys. Chem. Chem. Phys.* **2015**, *17*, 25915-25924.

[48] A. F. Cruz-Ortiz, M. I. Taccone, P. Maitre, M. Rossa, G. A. Pino, *ChemPhysChem* **2020**, *21*, 2571-2582.

[49] J.-Y. Salpin, S. Guillaumont, J. Tortajada, L. MacAleese, J. Lemaire, P. Maitre, *ChemPhysChem* **2007**, *8*, 2235-2244.

[50] C. X. Yao, F. Turecek, M. J. Polce, C. Wesdemiotis, *Int. J. Mass Spectrom.* **2007**, *265*, 106-123.

SUPPORTING INFORMATION

* * *

Reactivity of cytosine with alkylmercury ions in the gas phase: the critical role of the alkyl chain

Jean-Yves Salpin*[a], Violette Haldys[a], Jean-Claude Guillemin[b], Otilia Mó[c], Manuel Yáñez*[c], M. Merced Montero-Campillo*[c]

[a] Université Paris-Saclay, Univ Evry, CY Cergy Paris Université, CNRS, LAMBE, 91025, Evry-Courcouronnes, France

[b] Univ Rennes, Ecole Nationale Supérieure de Chimie de Rennes, CNRS, ISCR – UMR6226, F-35000 Rennes, France

[c] Departamento de Química, Módulo 13, Facultad de Ciencias, and Institute of Advanced Chemical Sciences (IAdChem), Universidad Autónoma de Madrid, Campus de Excelencia UAM-CSIC, Cantoblanco, 28049 Madrid, Spain

*jeanyves.salpin@univ-evry.fr, **manuel.yanez@uam.es, ***mm.montero@uam.es

CONTENTS

Figure S5. Transition state of a beta-hydride mechanism connecting $[EtHg(\mathbf{C})]^+$ and product (**C**: HgH^+ : C_2H_4) at the B3LYP/6-31++G(d,p)/def2-TZVPPD level of theory.

Table S1. $[MeHg(\mathbf{C})]^+$ energies of the different isomers at the B3LYP/6-31++G(d,p)/def2-TZVPPD level of theory.

Table S2. $[EtHg(\mathbf{C})]^+$ energies of the different isomers at the B3LYP/6-31++G(d,p)/def2-TZVPPD level of theory.

Table S3. Experimental and computed IR vibrational bands for the $[CH_3Hg(\mathbf{C})]^+$ ion.

Table S4. $[(\mathbf{C})CH_3]^+$ energies of the different isomers at the B3LYP/6-31++G(d,p) level of theory.

Table S5. $[(\mathbf{C})C_2H_5]^+$ energies of the different isomers at the B3LYP/6-31++G(d,p) level of theory.

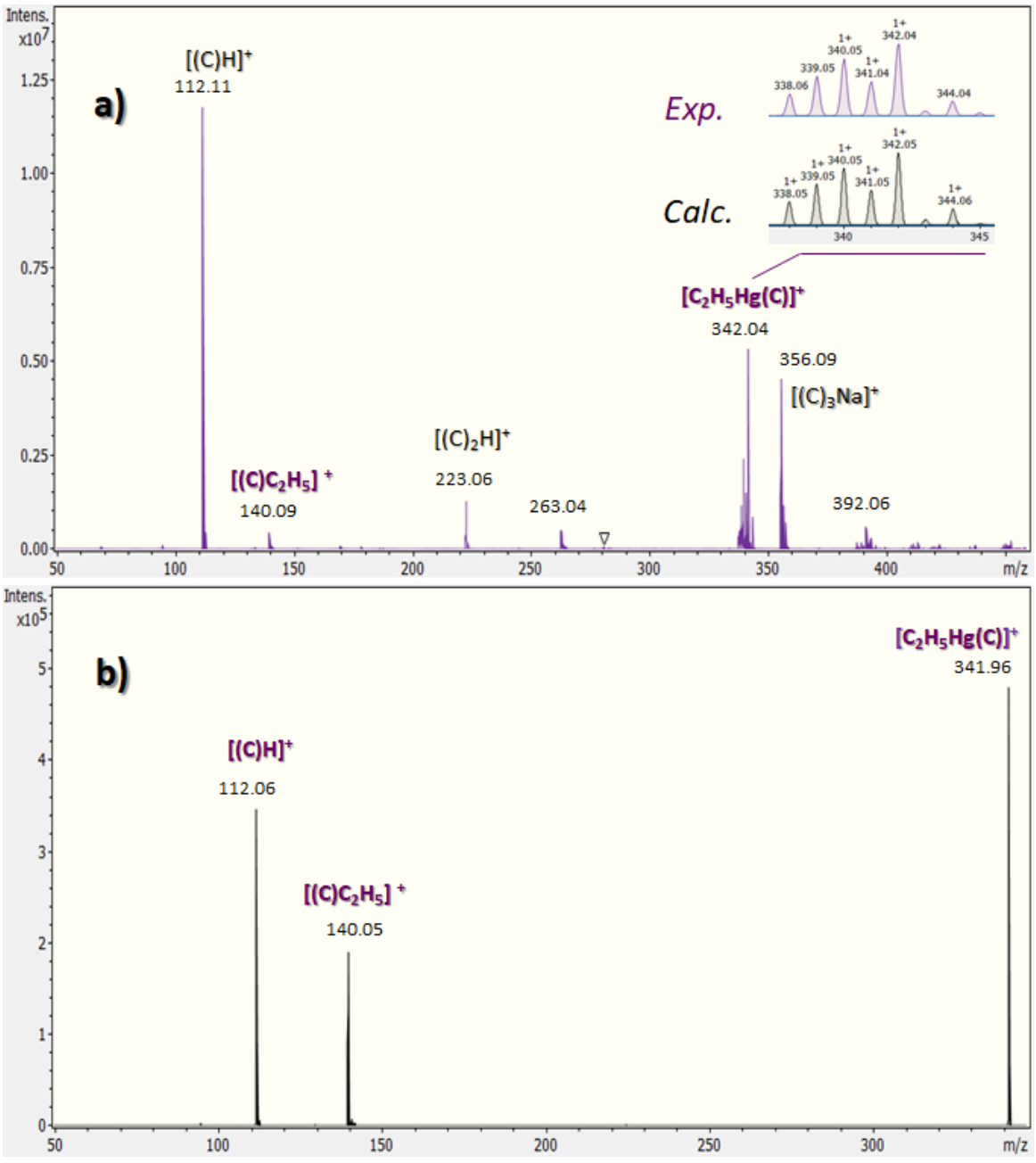
Table S6. Experimental and computed IR vibrational bands for the $[(\mathbf{C})CH_3]^+$ ion.

Figure S6. Computed vibrational spectra of **C2*d*** and **C1** forms of $[(\mathbf{C})CH_3]^+$ cation. The experimental IRMPD spectrum of methylated cytosine is overlayed in grey.

Table S7. Scan of the C2-O-Hg angle through the molecular plane from rotamer **C2*d*** to rotamer **C2*c*** (global minimum) with their corresponding electronic energy values.

Figure S7. Plot of electronic energy values associated to the scan of the C2-O-Hg angle through the molecular plane from rotamer **C2*d*** to rotamer **C2*c*** (global minimum).

Figure S1. a) Electrospray mass spectrum of an equimolar solution of C_2H_5HgCl and cytosine (10^{-4}M) in a water/methanol mixture (50/50 v/v)); **b)** MS/MS spectra of the $[C_2H_5Hg(\mathbf{C})]^+$ (*m/z* 342) complex (fragmentation time of 40 ms and a fragmentation amplitude of 0.7V). Spectra recorded on a Bruker Amazon HCT 3D ion trap.



Complexes were generated in the gas phase by electrospray. To this end, equimolar mixtures of cytosine/C_2H_5HgCl (10^{-4}M in water/methanol 50/50 v/v) were prepared. Typical experimental conditions were as followed: Capillary voltage: 4700 V; End plate offset: -500 V; Dry gas: 6 L/min / Dry gas temperature: 170 °C, Nebuliser gas: 1.04 bar; Cap exit: 125.7 V; Trap Drive 60.3. Flow rate 3 µL/Min, ICC off; Accu time 0.60 ms. For MS/MS spectrum: isolation mass 342 / Isolation width 0.7 Da / Amplitude 0.70 V, Fragmentation time 40 ms, ICC Off ; Accu time 0.60 ms.

Figure S2. MS/MS spectra of the $[C_2H_5Hg(U)]^+$ (U=uracil, *m/z* 343) complex (fragmentation time of 40 ms and a fragmentation amplitude of 0.7V). Spectrum recorded on a Bruker Amazon HCT 3D ion trap.

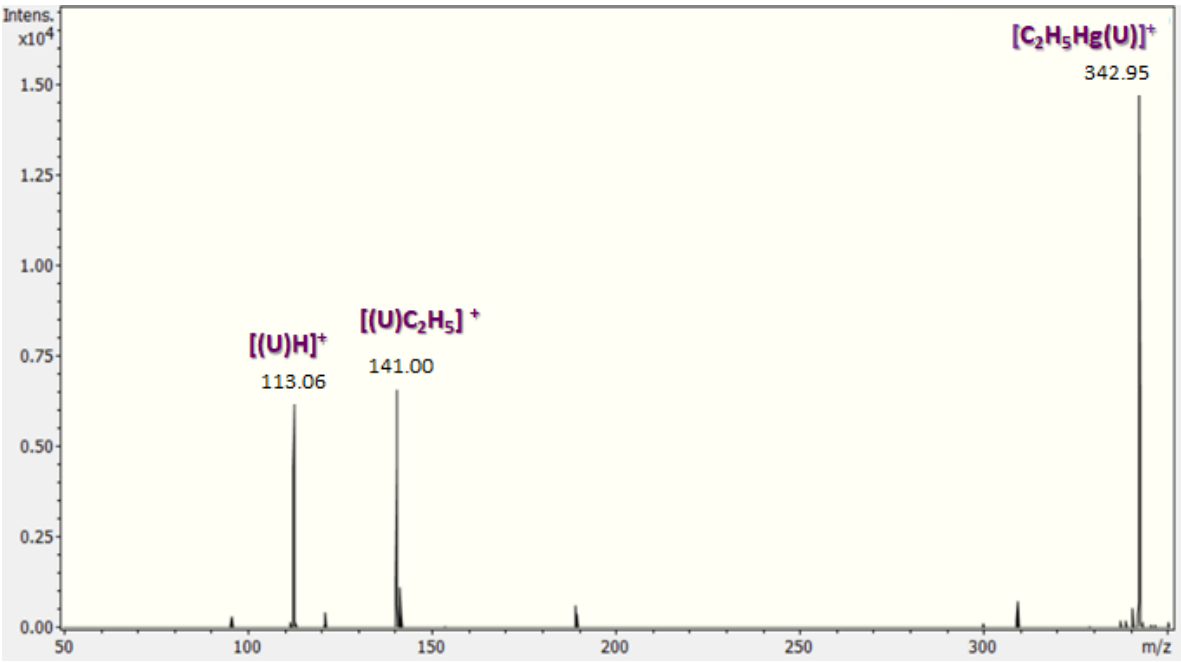
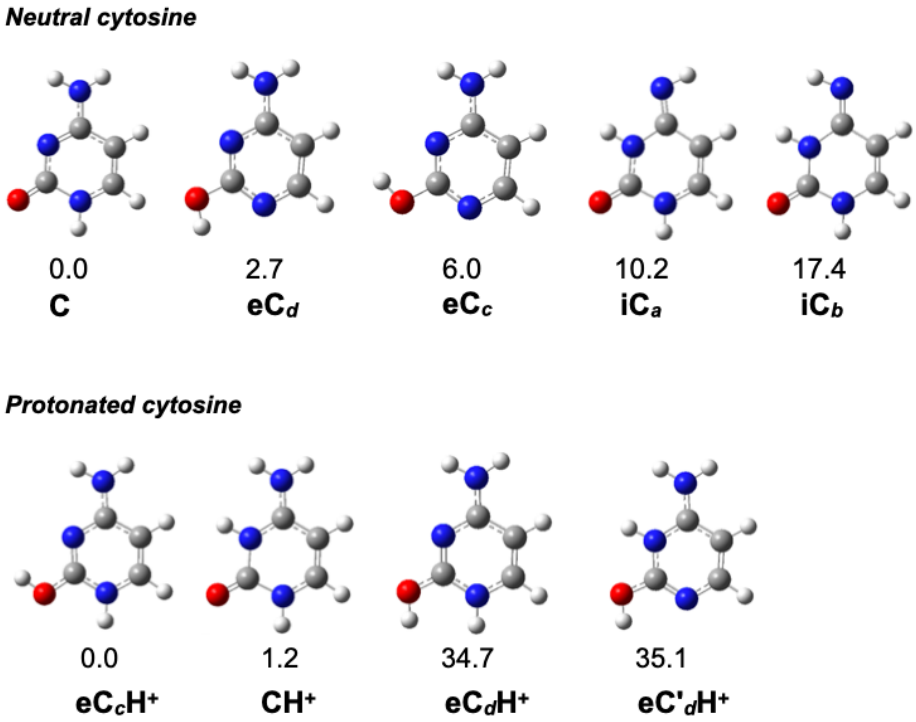


Figure S3. Most stable neutral cytosine and protonated cytosine isomers along with their relative energies (E+ZPE, kJ/mol) computed for this work at the B3LYP/6-31++G(d,p) level.

The keto-enamine isomer (global minimum) of neutral cytosine is labeled as **C**. Prefixes **e**, **i** stand for the **e**nol and **i**mine functional groups for the remaining isomers. Subindexes *a, b, c*, *d* indicate the orientation of the hydrogen atom, as shown in the manuscript or in Figure S4. An apostrophe (') denotes that N3 is protonated *instead of* N1.



These structures have been previously reported at different computational levels: M. Schreiber, L. González, J Comput Chem 2007, 28 (14), 2299-2308; N. Russo, M. Toscano, A. Grand, F. Jolibois, J Comput Chem 1998, 19 (9), 989-1000; L. Shahrokh, R. Omidyan, G. Azimi, Phys Chem Chem Phys 2021, 23, 14, 8916-8925.

Figure S4. Most stable methylated cytosine cation isomers $[(\mathbf{C})CH_3]^+$ along with their relative energies (E+ZPE, kJ/mol) at the B3LYP/6-31++G(d,p) level of theory.

Notation **C**n (n = 1-6) indicates the position of the methyl group in the ring. Labels *a*, *b*, *c*, *d* are relative orientations, as shown in the picture. Prefixes **e**, **i** stand for the **e**nol and **i**mine functional groups (see labeling in neutral cytosine, Figure **S3**). An apostrophe (') denotes that N3 is protonated instead of N1 if both sites are available.

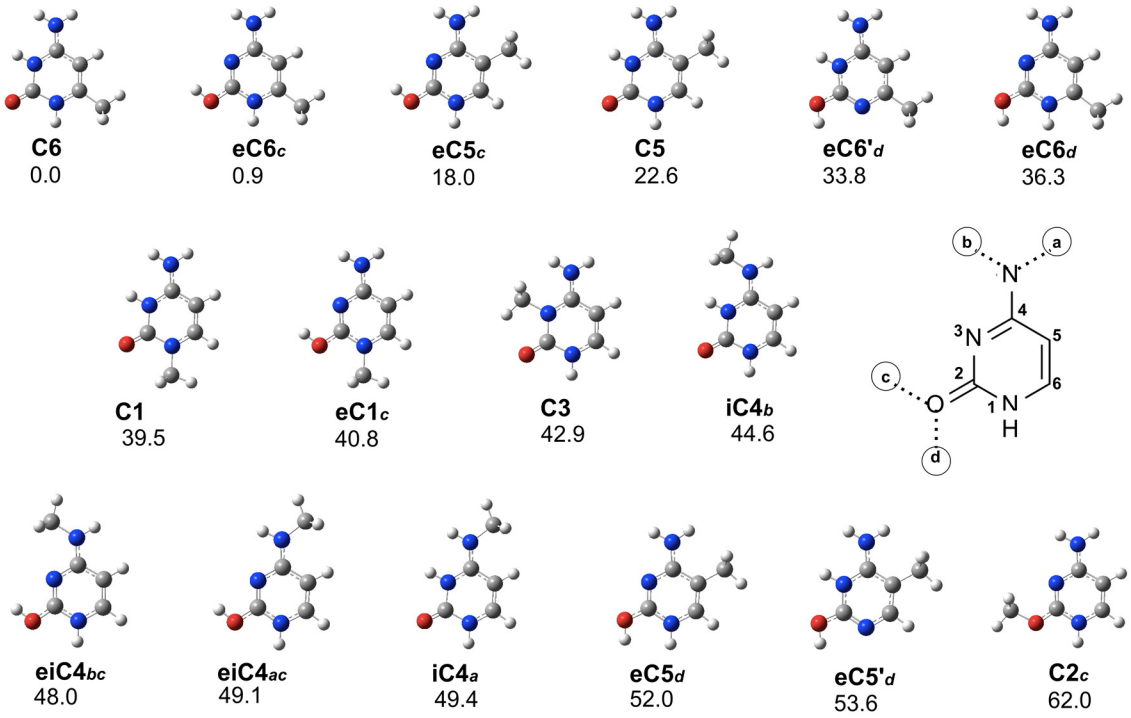


Figure S5. Transition state of a beta-hydride mechanism connecting [EtHg(**C**)]$^+$ and product (**C**: HgH^+ : C_2H_4) at the B3LYP/6-31++G(d,p)/def2-TZVPPD level of theory.

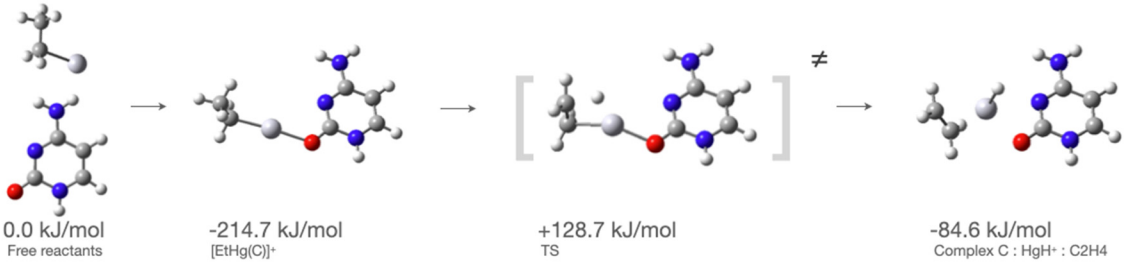
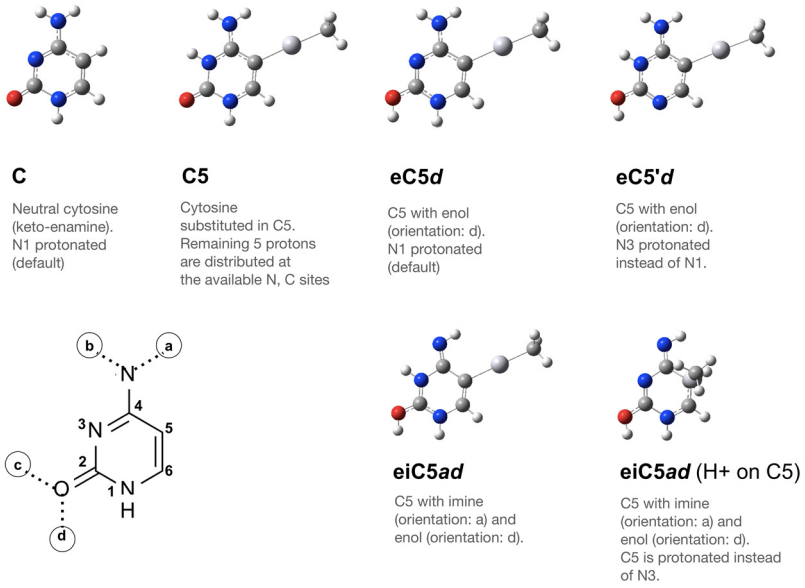


Table S1. [MeHg(**C**)]$^+$ energies of the different isomers at the B3LYP/6-31++G(d,p)/def2-TZVPPD level of theory.

Notation **Cn** (n = 1-6) indicates the position of the methylmercury group in the ring. Labels *a*, *b*, *c*, *d* are relative orientations. Prefix **e**, **i** denotes **e**nol and **i**mine groups. Additional apostrophes are used to note that N3 *instead of* N1 is protonated *if both sites are available*; otherwise, the available N1 and/or N3 are protonated unless indicated explicitly. Some structures are omitted for different reasons: (i) converged to already listed structures; (ii) not converged. Symbol (*) is used for structures presenting imaginary frequencies, which are in most cases very small frequencies associated to the methyl group. Some examples are shown below for the sake of clarity:



CHgMe$^+$	Δ (E+ZPE)	ΔH	ΔG
C1 (H^+ on NH_2)	159.1	160.1	159.0
C1	19.1	19.6	20.9
eC1*c*	20.6	20.7	22.0
eC1*d*	67.9	64.0	79.0
eiC1*ac*	136.4	137.0	136.4
eiC1*ad*	150.4	146.5	161.1
eiC1*bc*	154.1	155.1	153.6
eiC1*bd*	163.5	159.7	174.1
eiC1*ac* (H^+ on N1)	272.2	272.4	266.8
eiC1*bc* (H^+ on N1)	257.7	257.9	255.6
C2*c*	0.0	0.0	0.0
C2*d*	38.6	34.6	48.0
C2*c* (H^+ on NH_2)	170.5	171.5	167.0
C2*d* (H^+ on NH_2)	149.0	147.4	153.6
C2'*d*	28.3	28.9	27.7
iC2*ac*	95.5	96.0	92.1
iC2*ad*	102.4	103.2	97.0

iC2*bc*	114.6	115.3	109.8
iC2*bd*	115.6	116.4	111.3
iC2*ac* (H^+ on N1)	272.4	270.6	276.2
iC2*bc* (H^+ on N1)	259.6	260.1	256.8
iC2'*ad* (H^+ on N3)	241.1	239.7	241.9
iC2'*bd* (H^+ on N3)	264.5	261.0	272.2
C3	15.1	13.2	22.7
eC3*c**	113.2	111.3	121.5
eC3*d*	52.6	53.4	54.1
eiC3*ad*	120.5	121.7	120.6
eiC3*bc*	165.4	164.1	168.6
eiC3*bd**	157.5	156.4	162.7
iC3*a*	165.2	165.9	161.6
iC3*b**	194.5	192.7	199.0
iC3*a** (H^+ on N1)	215.2	213.8	220.4
iC3*b* (H^+ on N1)	237.2	238.0	235.8
eiC3'*bd*	260.3	260.8	258.5
C4*a**	149.1	146.9	152.4
C4'*	201.7	199.7	205.5
iC4*a*	38.4	35.7	45.3
iC4*b*	37.2	36.8	37.7
eiC4*ac*	51.6	51.0	52.0
eiC4*ad*	86.6	86.8	84.4
eiC4*bc*	43.3	42.6	44.3
eiC4*bd*	73.7	73.9	74.0
eiC4'*ac**	117.5	115.7	123.8
eiC4'*ad*	75.6	75.3	71.2
eiC4'*bc*	121.1	121.7	121.1
eiC4'*bd*	76.9	76.4	77.8
iC4*a** (H^+ on N1)	187.5	190.0	181.6
iC4*b** (H^+ on N1)	176.6	178.9	170.2
iC4'*a* (H^+ on N3)	142.5	146.9	132.4
iC4'*b* (H^+ on N3)	113.1	116.8	107.0
iC4*a* (H^+ on N1, deprot. N4)	273.7	274.6	271.9
iC4*b* (H^+ on N1, deprot. N4)	277.6	278.5	274.6
eiC4*ac* (deprot. N4)	183.6	184.5	183.6
eiC4*ad* (deprot. N4)	190.2	191.4	190.2
eiC4*bc* (deprot. N4)	197.8	199.1	195.5
eiC4*bd* (deprot. N4)	200.6	202.0	198.4
eiC4*ac* (H^+ on N1, deprot. N4)	364.9	365.5	364.1
eiC4*ad** (H^+ on N1, deprot. N4)	417.8	418.2	416.6
eiC4*bc* (H^+ on N1, deprot. N4)	346.8	347.2	346.1
eiC4*bd** (H^+ on N1, deprot. N4)	394.7	394.7	393.2

C5 (H^+ on C5)	144.7	145.6	142.5
C5 (H^+ on NH_2)	218.8	219.7	217.8
C5	82.8	83.1	84.5
eC5*c*	83.6	83.7	85.2
eC5*d*	117.7	118.3	118.5
eC5'*c*	159.9	161.6	160.2
eC5'*d*	118.1	118.7	119.4
eiC5*ac*	197.8	196.0	204.9
eiC5*ad**	202.7	201.3	209.5
eiC5*bc*	206.7	207.7	207.1
eiC5*bd*	207.1	208.1	207.6
eiC5*ac* (H^+ on N1)	368.6	369.1	366.3
eiC5*bc* (H^+ on N1)	341.5	342.0	340.8
eiC5'*ac* (H^+ on N3)	382.7	384.5	380.2
eiC5'*ad* (H^+ on N3)	339.5	340.1	338.7
eiC5'*bd* (H^+ on N3)	355.7	357.1	353.9
C5* (H^+ on N1)	244.6	244.0	248.4
C5'* (H^+ on N3)	255.8	258.8	251.6
eC5*c* (H^+ on C5)	136.4	136.7	136.5
eC5*d* (H^+ on C5)	133.8	134.1	133.4
C5 (H^+ on C5, H^+ on NH_2)	312.5	313.9	309.7
eiC5*ac* (H^+ on C5)	237.8	238.3	236.1
eiC5*ad* (H^+ on C5)	282.8	284.4	280.9
eiC5'*ac* (H^+ on C5)	219.0	219.4	219.0
eiC5'*ad* (H^+ on C5)	187.7	187.8	187.0
eiC5'*bc** (H^+ on C5)	227.3	225.6	232.5
eiC5'*bd** (H^+ on C5)	191.1	189.1	196.6
C6 (H^+ on NH_2)	209.2	210.2	205.3
C6	71.7	72.1	71.6
eC6*c*	72.8	72.8	72.9
eC6*d**	108.0	106.3	115.7
eC6'*c*	160.8	162.5	160.7
eC6'*d*	119.1	119.7	120.4
eiC6*ac*	193.4	191.5	200.7
eiC6*ad*	199.0	199.9	197.0
eiC6*bc*	212.3	210.8	219.1
eiC6*bd*	213.6	214.6	212.2
eiC6*ac* (H^+ on N1)	362.7	363.3	358.6
eiC6*bc* (H^+ on N1)	348.0	348.6	342.6
eiC6'*ac* (H^+ on N3)	391.0	392.8	389.0
eiC6'*ad* (H^+ on N3)	347.9	348.6	346.6
eiC6'*bd* (H^+ on N3)	372.5	373.6	369.5
C6'* (H^+ on NH_2)	313.1	312.4	318.2

C6* (H^+ on N1)	234.0	233.4	235.6

Table S2. $[EtHg(C)]^+$ energies of the different isomers at the B3LYP/6-31++G(d,p)/def2-TZVPPD level of theory. but Cn

Notation **Cn** (n = 1-6) indicates the position of the ethylmercury group in the ring. Labels *a*, *b*, *c*, *d* are relative orientations. Prefix **e**, **i** denotes enolic and imino forms. Additional apostrophes are used to note that N3 instead of N1 is protonated if both sites are available; otherwise, the available N1 and/or N3 are protonated unless explicitly indicated. Some structures are omitted for different reasons: (i) converged to already listed structures; (ii) not converged. Symbol (*) is used for structures presenting imaginary frequencies, which are in many cases small frequencies associated to the ethyl group. See some examples in Table **S1**.

Note. For each isomer, different conformations of the ethyl group were considered and calculated. For the sake of brevity, only the most stable minima are included in the list.

CHgEt$^+$	Δ (E+ZPE)	ΔH	ΔG
C1 (H^+ on NH_2)	160.5	161.5	160.2
C1	19.6	20.1	20.8
eC1*c*	21.1	21.2	21.8
eC1*d*	67.1	63.2	79.6
eiC1*ac*	136.2	136.8	134.9
eiC1*ad*	149.0	144.9	161.0
eiC1*bc*	153.8	152.3	161.4
eiC1*bd**	161.5	157.6	173.5
eiC1*ac* (H^+ on N1)	263.6	264.0	261.0
eiC1*bc* (H^+ on N1)	249.3	249.6	245.6
C2*c*	0.0	0.0	0.0
C2*d**	36.9	35.2	39.9
C2*c* (H^+ on NH_2)	172.3	173.3	168.0
C2*d* (H^+ on NH_2)	151.1	151.9	147.7
C2'*c*	27.9	28.5	26.2
C2'*d*	28.0	28.7	27.1
iC2*ac*	91.2	91.6	87.8
iC2*ad*	97.8	98.6	90.8
iC2*bc*	109.8	110.6	103.1
iC2*bd*	110.9	111.7	105.5
iC2*ac* (H^+ on N1)	269.5	270.2	265.0
iC2*bc* (H^+ on N1**)**	256.9	257.6	252.9
iC2'*ad* (H^+ on N3)	238.4	234.6	247.4
iC2'*bd** (H^+ on N3)	261.7	260.7	260.3
C3	14.5	14.9	15.6
eC3*c*	107.0	107.9	107.2
eC3*d*	51.7	52.6	51.6
eiC3*ad*	120.9	122.1	120.8
eiC3*bc**	163.3	162.0	166.4
eiC3*bd**	157.2	156.2	163.4
iC3*b*	185.6	186.3	182.9

iC3*a* (H^+ on N1)	215.7	216.8	210.8
iC3*b* (H^+ on N1)	236.6	237.6	231.7
eiC3'*bc*	295.0	297.0	289.8
eiC3'*bd*	250.6	251.4	247.0
C4*a*	145.8	143.7	146.7
C4'	193.1	193.6	189.7
iC4*a*	39.6	36.9	47.9
iC4*b*	38.0	37.6	31.4
eiC4*ac*	54.5	53.9	54.7
eiC4*ad*	89.6	89.9	87.7
eiC4*bc*	46.9	46.2	48.5
eiC4*bd*	77.4	77.6	76.1
eC4'*ac*	119.1	119.8	114.5
eC4'*ad*	77.4	74.5	86.2
eC4'*bc*	122.2	122.8	119.4
eC4'*bd*	78.2	77.7	77.1
iC4*a** (H^+ on N1)	188.3	190.8	182.8
iC4*b** (H^+ on N1)	177.9	180.3	172.7
iC4*a** (H^+ on N3)	194.0	193.9	197.6
iC4*b** (H^+ on N3)	114.9	116.2	116.4
iC4*a* (H^+ on N1, deprot. N4)	278.4	279.2	277.1
iC4*b* (H^+ on N1, deprot. N4)	282.2	283.1	278.7
eiC4*ad* (deprot. N4)	167.6	172.3	156.9
eiC4*ac* (H^+ on N1, deprot. N4)	370.3	370.9	369.6
eiC4*ad** (H^+ on N1, deprot. N4)	423.2	423.5	419.9
eiC4*bc* (H^+ on N1, deprot. N4)	352.7	353.1	351.9
eiC4*bd** (H^+ on N1, deprot. N4)	400.6	400.6	399.8
C5 (H^+ on C5)	140.2	141.2	138.8
C5* (H^+ on NH_2)	225.0	223.4	232.4
C5*	89.0	86.9	98.1
eC5*c*	90.4	90.4	90.1
eC5*d**	124.4	122.6	133.3
eC5'*c*	166.5	165.8	174.6
eC5'*d*	124.8	122.9	134.1
eiC5*ac*	204.8	205.4	202.3
eiC5*ad*	209.5	210.4	207.5
eiC5*bc*	213.8	214.8	211.5
eiC5*bd*	214.1	215.1	213.6
eiC5*ac* (H^+ on N1)	374.3	374.8	372.2
eiC5*bc* (H^+ on N1)	347.8	348.3	346.3
eiC5'*bd* (H^+ on N1)	347.8	348.3	346.1
eiC5'*ac* (H^+ on N3)	388.6	390.4	385.3
eiC5'*ad* (H^+ on N3)	345.9	346.5	339.8

eiC5'*bd* (H^+ on N3)	361.9	363.3	360.1
C5' (H^+ on NH_2)	312.4	313.8	310.3
C5* (H^+ on N1)	250.2	247.2	261.2
C5'* (H^+ on N3)	186.5	191.6	176.4
eC5*c* (H^+ on C5)	132.5	132.9	131.9
eC5*d* (H^+ on C5)	130.0	130.4	128.2
C5 (H^+ on C5, H^+ on NH_2)	311.0	312.4	308.3
eiC5*ac* (H^+ on C5)	231.8	232.4	228.7
eiC5*ad* (H^+ on C5)	276.1	277.9	273.3
eiC5'*ac* (H^+ on C5)	215.7	216.2	214.6
eiC5'*ad* (H^+ on C5)	184.2	184.3	183.4
eiC5'*bc* (H^+ on C5)	223.5	224.3	219.7
eiC5'*bd* (H^+ on C5)	187.1	187.5	185.4
C6 (H^+ on NH_2)	215.2	213.7	222.3
C6*	77.2	75.2	86.3
eC6*c**	78.7	76.3	88.4
eC6*d**	113.8	112.0	122.9
eC6'*c**	168.0	167.3	176.1
eC6'*d**	126.4	124.6	135.7
eiC6*ac*	199.7	200.2	197.9
eiC6*ad*	205.2	206.0	202.4
eiC6*bc*	218.5	219.3	218.2
eiC6*bd*	219.7	220.6	219.3
eiC6*ac* (H^+ on N1)	367.7	368.3	363.0
eiC6*ad** (H^+ on N1)	367.7	365.8	374.0
eiC6*bc* (H^+ on N1)	353.2	351.3	360.2
eiC6*bd* (H^+ on N1)	353.6	351.7	360.6
eiC6'*ac* (H^+ on N3)	398.1	397.4	404.7
eiC6'*ad* (H^+ on N3)	355.2	355.8	351.7
eiC6'*bd* (H^+ on N3)	379.9	381.0	377.5
C6' (H^+ on NH_2)	320.3	322.0	317.6
C6* (H^+ on N1)	239.3	238.7	242.4
C6'* (H^+ on N3)	155.1	159.3	149.6

Table S3. Experimental and computed IR vibrational bands for the $[CH_3Hg(\mathbf{C})]^+$ ion.

C2*c*

Wavenumbers (cm^{-1}) Exp.	Calc.[a]	DFT-computed intensities (km/mol)	Vibrational mode
1210	1202	56	δ CH + δ NH
	1215	60	δ CH_3 umbrella
1295	1316	32	δ C6H
1480	1486	162	ν C4N
	1500	78	ν C4C5
1580	1588	987	ν C2O
1630	1626	309	δ NH_2 scissoring
	1647	593	ν C5C6

a) Scaled by a factor of 0.97

C2*d*

Wavenumbers (cm^{-1}) Exp.	Calc.[a]	DFT-computed intensities (km/mol)	Vibrational mode
1210	1200	67	δ CH + δ NH
	1218	97	δ CH_3 umbrella
1295	1323	119	δ C6H
1480	1486	235	ν C4N
	1498	110	ν C4C5
1580	1586	1305	ν C2O
1630	1628	373	δ NH_2 scissoring
	1652	682	ν C5C6

a) Scaled by a factor of 0.97

C3

Wavenumbers (cm^{-1}) Exp.	Calc.[a]	DFT-computed intensities (km/mol)	Vibrational mode
1210	1206	78	δ CH + δ NH
	1213	44	δ CH_3 umbrella
1295	–		
1480	1490	172	ν C4N
	1521	191	ν C4C5+ ν N1C6
1580	–		
1630	1639	720	ν C4N

a) Scaled by a factor of 0.97

C2'*d*

Wavenumbers (cm^{-1})		DFT-computed intensities (km/mol)	Vibrational mode
Exp.	Calc.[a]		
1210	1212	93	δ CH_3 umbrella
1295	1299	379	δ N3H
1405	–		
1480	1474	139	ring breathing
1580	1596	108	δ NH_2 scissoring
1630	1609	1496	ν C2O+ ν N3C4
	1647	501	ν C4N

a) Scaled by a factor of 0.97

Table S4. $[(\mathbf{C})CH_3]^+$ energies of the different isomers at the B3LYP/6-31++G(d,p) level of theory.

Notation **Cn** (n = 1-6) indicates the position of the methyl group in the ring. Labels *a*, *b*, *c*, *d* are relative orientations. Prefix **e**, **i** denotes enolic and imino forms. Additional commas are used to note that N3 instead of N1 is protonated if both sites are available; otherwise, the available N1 and/or N3 are protonated unless explicitly indicated. Some structures are omitted for different reasons: (i) converged to already listed structures; (ii) not converged. Symbol (*) is used for structures corresponding to first order or second order saddle points. See some examples in Table **S1**.

CMe^+	Δ (E+ZPE)	ΔH	ΔG
C1 (H^+ on N1)	203.4	203.7	202.0
C1 (H^+ on NH_2)	173.4	173.6	172.5
C1	39.5	39.3	40.0
eC1*c*	40.8	40.1	41.9
eC1*d*	77.1	76.7	78.0
eiC1*ac*	162.2	162.4	160.5
eiC1*ad*	166.9	166.9	166.3
eiC1*bc*	180.1	180.7	177.4
eiC1*bd*	180.4	180.6	179.5
eiC1*ac* (H^+ on N1)	327.1	326.1	327.6
eiC1*ac* (H^+ on N1)	312.5	311.4	313.1
C2*c*	62.0	61.4	62.6
C2*d*	94.0	93.4	94.7
C2*c* (H^+ on NH_2)	209.2	209.6	206.0
C2*d* (H^+ on NH_2)	198.0	198.2	196.5
C2'*c*	136.0	136.1	136.2
C2'*d*	95.8	95.7	96.0
iC2*ac*	171.7	171.2	171.8
iC2*ad*	178.2	177.9	178.0
iC2*bc*	190.8	190.5	190.5
iC2*bd*	191.9	191.7	191.5
iC2*ac* (H^+ on N1)	348.8	348.8	346.6
iC2*ad** (H^+ on N1)	410.6	409.3	409.1
iC2*bc* (H^+ on N1)	334.2	334.1	332.5
iC2*bd** (H^+ on N1)	392.0	390.5	392.1
iC2'*ac** (H^+ on N3)	370.1	366.6	373.5
iC2'*ad** (H^+ on N3)	317.5	315.6	319.1
iC2'*bc** (H^+ on N3)	395.9	393.0	399.0
iC2'*bd** (H^+ on N3)	341.0	339.4	342.5
C3	42.9	41.9	44.5
C3 (H^+ on NH_2)	272.5	273.2	269.1
eC3*c**	131.3	129.9	133.8
eC3*d*	80.2	79.5	81.7
eiC3*ac**	162.7	160.4	165.6
eiC3*ad*	167.2	167.5	166.1

eiC3*bc**	185.6	183.5	188.5
eiC3*bd*	184.4	184.5	183.5
iC3*a*	237.1	236.2	236.8
iC3*b*	257.4	256.6	257.9
iC3*a* (H^+ on N1)	256.6	256.6	254.3
iC3*b* (H^+ on N1)	264.5	264.3	262.3
eiC3'*ac* (H^+ on N3)	353.2	353.3	352.0
eiC3'*ad* (H^+ on N3)	295.4	294.4	296.4
eiC3'*bd* (H^+ on N3)	332.9	332.3	333.0
C4*a*	169.9	169.2	167.9
C4*b**	174.2	171.2	176.7
C4'	259.0	258.9	257.1
iC4*a*	49.4	48.9	49.5
iC4*b*	44.6	43.9	45.0
eiC4*ac*	49.1	48.3	49.6
eiC4*ad*	83.4	83.4	83.3
eiC4*bc*	48.0	47.4	47.9
eiC4*bd*	81.6	81.7	80.5
eC4'*ac*	127.0	127.5	126.5
eC4'*ad*	84.2	83.7	84.7
eC4'*bc*	124.6	124.8	124.3
eC4'*bd*	81.3	80.6	82.1
iC4*a** (H^+ on N1)	197.7	199.8	193.2
iC4*b** (H^+ on N1)	190.0	192.3	184.6
iC4*a* (H^+ on N3)	122.4	125.7	116.2
iC4*b** (H^+ on N3)	203.4	205.8	197.4
iC4*a** (H^+ on N3, deprot. N4)	238.9	235.3	242.2
iC4*b** (H^+ on N3, deprot. N4)	256.5	253.4	259.2
iC4*a* (H^+ on N1, deprot. N4)	256.1	256.4	254.2
iC4*b* (H^+ on N1, deprot. N4)	259.7	260.2	256.3
eiC4ac* (deprot. N4)	169.1	167.5	170.7
eiC4ad* (deprot. N4)	174.3	173.1	175.6
eiC4bc* (deprot. N4)	182.2	180.9	183.5
eiC4bd (deprot. N4)	183.6	184.7	178.6
eiC4'*ac** (H^+ on N3, deprot. N4)	352.3	348.3	356.2
eiC4'*ad** (H^+ on N3, deprot. N4)	307.0	303.3	310.8
eiC4'*bc** (H^+ on N3, deprot. N4)	322.0	321.3	321.0
eiC4'*bd** (H^+ on N3, deprot. N4)	324.8	321.8	327.6
eiC4*ac** (H^+ on N1, deprot. N4)	340.6	338.8	341.1
eiC4*ad** (H^+ on N1, deprot. N4)	393.3	391.2	393.8
eiC4*bc** (H^+ on N1, deprot. N4)	322.5	320.6	323.3
eiC4*bd** (H^+ on N1, deprot. N4)	371.0	368.7	371.9
C5 (H^+ on C5)	160.8	160.9	159.9

C5 (H^+ on NH_2)	160.4	160.9	158.9
C5	22.6	22.0	23.9
eC5*c*	18.0	17.1	19.9
eC5*d*	52.0	51.7	53.3
eC5'*c*	95.9	96.7	96.2
eC5'*d*	53.6	53.3	55.2
eiC5*ac*	133.3	133.1	134.3
eiC5*ad*	138.2	138.3	138.9
eiC5*bc*	148.8	149.0	149.2
eiC5*bd*	149.4	149.6	149.9
eiC5*ac* (H^+ on N1)	307.0	306.6	306.3
eiC5*ad** (H^+ on N1)	359.9	359.1	358.9
eiC5*bc* (H^+ on N1)	288.4	288.1	288.0
eiC5*bd** (H^+ on N1)	337.6	336.8	337.2
eiC5'*ac* (H^+ on N3)	325.4	327.0	321.7
eiC5'*ad* (H^+ on N3)	280.6	280.6	279.9
eiC5'*bd* (H^+ on N3)	302.4	302.8	300.8
C5'* (H^+ on NH_2)	255.1	252.9	256.7
C5 (H^+ on N1)	179.9	181.2	176.1
C5' (H^+ on N3)	97.9	100.7	93.7
eC5*c* (H^+ on C5)	152.2	151.5	151.7
eC5*d* (H^+ on C5)	149.7	148.9	149.2
C5 (H^+ on C5, H^+ on NH_2)	307.2	308.0	304.6
eiC5*ac* (H^+ on C5)	269.1	268.4	269.5
eiC5*ad* (H^+ on C5)	316.1	316.5	315.5
eiC5*bc* (H^+ on C5)	248.9	248.2	248.8
eiC5*bd* (H^+ on C5)	294.1	294.5	293.0
eiC5'*ac* (H^+ on C5)	228.3	227.7	227.5
eiC5'*ad* (H^+ on C5)	198.5	197.7	197.9
eiC5'*bc* (H^+ on C5)	243.1	242.7	242.0
eiC5'*bd* (H^+ on C5)	208.5	207.8	207.7
C6 (H^+ on C5 migr. from C6)	111.3	112.3	107.0
C6 (H^+ on NH_2)	136.5	137.2	133.9
C6	0.0	0.0	0.0
eC6*c*	0.9	0.5	1.7
eC6*d*	36.3	36.5	36.5
eC6'*c*	75.3	76.5	74.7
eC6'*d*	33.8	33.9	34.3
eiC6*ac*	123.0	122.9	123.4
eiC6*ad*	128.8	129.0	128.9
eiC6*bc*	142.0	142.3	142.0
eiC6*bd*	143.6	144.0	143.5
eiC6*ac* (H^+ on N1)	297.4	297.2	296.5

eiC6*ad** (H^+ on N1)	350.9	350.3	349.9
eiC6*bc* (H^+ on N1)	282.7	282.5	282.2
eiC6*bd** (H^+ on N1)	332.6	332.0	332.3
eiC6'*ac* (H^+ on N3)	308.5	309.6	306.6
eiC6'*ad* (H^+ on N3)	264.6	264.5	263.8
eiC6'*bd* (H^+ on N3)	288.7	289.1	287.2
C6'* (H^+ on NH_2)	230.0	231.3	227.4
C6* (H^+ on N1)	84.5	88.7	75.2
C6'* (H^+ on N3)	73.9	77.3	68.6
eC6*c* (H^+ on C5 migr. from C6)	107.5	107.9	103.2
eC6*d* (H^+ on C5 migr. from C6)	105.3	105.8	99.6
C6 (H^+ on C6, H^+ on NH_2)	506.7	507.9	504.3
eiC6*ac* (H^+ on C6)	389.1	389.3	387.9
eiC6*ad* (H^+ on C6)	374.0	372.7	376.4
eiC6*bc* (H^+ on C6)	359.7	360.2	357.3
eiC6*bd* (H^+ on C6)	397.7	399.0	394.9

Table S5. $[(\mathbf{C})C_2H_5]^+$ energies of the different isomers at the B3LYP/6-31++G(d,p) level of theory.

Notation **Cn** (n = 1-6) indicates the position of the ethyl group in the ring. Labels *a*, *b*, *c*, *d* are relative orientations. Prefix **e**, **i** denotes enolic and imino forms. Additional apostrophes are used to note that N3 instead of N1 is protonated if both sites are available; otherwise, the available N1 and/or N3 are protonated unless explicitly indicated. Some structures are omitted for different reasons: (i) converged to already listed structures; (ii) not converged. Symbol (*) is used for structures corresponding to first order or second order saddle points. See some examples in Table **S1**.

Note. For each isomer, different conformations of the ethyl group were considered and calculated. For the sake of brevity, only the most stable minima are included in the list.

CEt$^+$	Δ (E+ZPE)	ΔH	ΔG
C1 (H^+ on N1)	190.7	190.9	190.5
C1 (H^+ on NH_2)	164.1	164.2	164.1
C1	28.7	28.3	30.2
eC1*c*	30.0	29.1	32.1
eC1*d*	66.7	66.3	68.4
eiC1*ac*	150.2	149.9	150.8
eiC1*ad*	155.0	154.7	155.7
eiC1*bc*	168.0	168.0	168.2
eiC1*bd*	168.3	168.1	168.8
eiC1*ac* (H^+ on N1)	310.5	309.5	312.0
eiC1*bc* (H^+ on N1)	295.8	294.8	297.5
eiC1*bd* (H^+ on N1)	345.2	346.1	341.3
C2*c*	42.4	41.9	43.0
C2*d*	72.5	72.1	73.3
C2*c* (H^+ on NH_2)	190.9	191.5	188.3
C2*d* (H^+ on NH_2)	180.3	180.7	179.1
C2'*c*	113.8	114.2	114.1
C2'*d*	76.0	76.2	76.3
iC2*ac*	147.9	147.6	148.0
iC2*ad*	154.0	154.0	153.9
iC2*bc*	166.7	166.6	166.5
iC2*bd*	167.7	167.8	167.4
iC2*ac* (H^+ on N1)	327.6	327.8	325.2
iC2*ad** (H^+ on N1)	386.4	385.3	384.8
iC2*bc* (H^+ on N1)	312.7	312.8	310.9
iC2*bd** (H^+ on N1)	367.8	366.4	368.0
iC2'*ac* (H^+ on N3)	348.0	348.6	344.3
iC2'*ad* (H^+ on N3)	298.4	298.4	297.1
iC2'*bc** (H^+ on N3)	371.9	369.0	375.2
iC2'*bd* (H^+ on N3)	322.4	322.8	320.6
C3	35.5	34.6	37.9
C3 (H^+ on NH_2)	261.1	261.1	261.5

eC3*c**	140.5	138.1	146.2
eC3*d*	71.8	71.1	74.5
eiC3*ac*	149.8	149.2	151.6
eiC3*ad*	158.0	157.8	159.4
eiC3*bc*	174.1	173.7	175.7
eiC3*bd*	176.3	176.1	177.7
iC3*a*	227.6	226.9	228.5
iC3*b*	248.2	247.6	249.7
iC3*a* (H^+ on N1)	247.1	246.9	246.1
iC3*b* (H^+ on N1)	256.0	255.8	255.2
eiC3'*ac* (H^+ on N3)	325.6	325.7	326.7
eiC3'*ad* (H^+ on N3)	282.3	281.5	284.0
eiC3'*bc* (H^+ on N3)	308.2	307.6	309.9
eiC3'*bd* (H^+ on N3)	325.1	324.6	326.0
C4*b*	149.9	149.4	148.4
C4'	235.9	235.9	234.2
iC4*a*	36.6	36.2	36.6
iC4*b*	31.5	30.9	31.9
eiC4*ac*	38.2	37.4	38.7
eiC4*ad*	72.5	72.5	72.4
eiC4*bc*	37.5	36.8	37.4
eiC4*bd*	71.2	71.2	70.1
eC4'*ac*	114.6	115.2	114.1
eC4'*ad*	71.9	71.5	72.5
eC4'*bc*	111.6	112.0	111.4
eC4'*bd*	68.6	67.9	69.4
iC4*a* (H^+ on N1)	113.6	117.1	107.4
iC4*b* (H^+ on N1)	114.1	117.7	108.1
iC4*a* (H^+ on N3)	110.4	113.8	104.2
iC4*b* (H^+ on N3)	104.6	107.7	99.1
iC4*a* (H^+ on N3, deprot. N4)	228.6	224.8	233.9
iC4*b* (H^+ on N3, deprot. N4)	184.9	189.1	174.2
iC4*a* (H^+ on N1, deprot. N4)	262.4	262.3	261.3
iC4*b* (H^+ on N1, deprot. N4)	250.2	248.4	251.6
eiC4ac* (deprot. N4)	160.3	158.4	163.6
eiC4ad* (deprot. N4)	179.5	179.9	179.4
eiC4bc* (deprot. N4)	172.1	172.8	169.5
eiC4bd (deprot. N4)	173.6	174.4	171.0
eiC4'*ac** (H^+ on N3, deprot. N4)	341.9	337.6	347.7
eiC4'*ad* (H^+ on N3, deprot. N4)	296.7	292.8	302.4
eiC4'*bc* (H^+ on N3, deprot. N4)	202.7	208.0	190.1
eiC4'*bd* (H^+ on N3, deprot. N4)	205.2	210.0	194.1
eiC4*ac* (H^+ on N1, deprot. N4)	345.0	344.7	344.7

eiC4*ad* (H^+ on N1, deprot. N4)	383.5	381.2	385.7
eiC4*bc* (H^+ on N1, deprot. N4)	310.9	310.7	309.7
eiC4*bd* (H^+ on N1, deprot. N4)	362.0	359.5	364.6
C5 (H^+ on C5)	157.4	157.5	157.3
C5* (H^+ on NH_2)	160.8	161.2	160.6
C5	24.2	23.6	26.6
eC5*c*	19.5	18.6	22.4
eC5*d*	53.4	53.1	55.7
eC5'*c*	98.9	99.6	100.2
eC5'*d*	56.8	56.3	59.2
eiC5*ac*	134.4	134.1	136.3
eiC5*ad*	139.3	139.2	140.9
eiC5*bc*	149.8	149.8	151.2
eiC5*bd*	150.3	150.4	151.8
eiC5*ac* (H^+ on N1)	308.0	307.5	308.0
eiC5*bc* (H^+ on N1)	289.1	288.7	289.5
eiC5*bd* (H^+ on N1)	338.1	337.3	338.5
eiC5*ac* (H^+ on N3)	327.0	328.5	323.9
eiC5*ad* (H^+ on N3)	282.4	282.3	282.8
eiC5*bd* (H^+ on N3)	304.3	304.6	304.0
C5' (H^+ on NH_2)	249.3	250.2	247.9
C5 (H^+ on N1)	181.9	182.9	177.2
C5' (H^+ on N3)	102.1	105.3	98.3
eC5*c* (H^+ on C5)	149.5	148.8	149.9
eC5*d* (H^+ on C5)	146.9	146.2	147.4
C5 (H^+ on C5, H^+ on NH_2)	303.7	304.1	303.3
eiC5*ac* (H^+ on C5)	263.7	263.0	265.1
eiC5*ad* (H^+ on C5)	310.0	310.5	310.3
eiC5*bc* (H^+ on C5)	245.1	244.6	245.8
eiC5*bd* (H^+ on C5)	289.4	290.0	288.9
eiC5'*ac* (H^+ on C5)	225.8	225.2	226.8
eiC5'*ad* (H^+ on C5)	196.1	195.3	196.3
eiC5'*bc* (H^+ on C5)	241.2	240.9	240.7
eiC5'*bd* (H^+ on C5)	206.5	205.9	206.4
C6 (H^+ on NH_2)	137.4	137.9	136.2
C6	0.0	0.0	0.0
eC6*c*	1.1	0.5	2.9
eC6*d*	36.4	36.4	37.6
eC6'*c*	75.6	76.7	76.2
eC6'*d*	34.3	34.2	36.1
eiC6*ac*	122.4	122.3	123.7
eiC6*ad*	128.0	128.2	129.1
eiC6*bc*	141.3	141.5	142.2

eiC6*bd*	142.7	143.0	143.5
eiC6*ac* (H^+ on N1)	295.3	295.2	294.6
eiC6*ad* (H^+ on N1)	295.3	295.2	294.6
eiC6*bc* (H^+ on N1)	280.7	280.5	280.9
eiC6'*ac* (H^+ on N3)	307.3	308.3	306.7
eiC6'*ad* (H^+ on N3)	263.5	263.4	263.8
eiC6'*bc* (H^+ on N3)	287.7	288.1	287.3
eiC6'*bd** (H^+ on N3)	287.6	285.8	290.9
C6' (H^+ on NH_2)	228.4	229.7	227.0
C6 (H^+ on N1)	85.3	89.2	78.2
C6' (H^+ on N3)	74.1	77.4	69.9
eC6*c* (H^+ on C5 migr. from C6)	105.5	105.7	103.9
eC6*d* (H^+ on C5 migr. from C6)	103.8	104.0	102.1
C6 (H^+ on C6, H^+ on NH_2)	501.4	503.2	498.1
eiC6*ac* (H^+ on C6)	383.4	384.0	382.3
eiC6*ad* (H^+ on C6)	422.2	423.7	419.9
eiC6*bc* (H^+ on C6)	245.7	245.3	245.7
eiC6*bd* (H^+ on C6)	391.3	393.0	388.3

Table S6. Experimental and computed IR vibrational bands for the $[(\mathbf{C})CH_3]^+$ ion.

C2*c*

Wavenumbers (cm^{-1}) Exp.	Calc.[a]	DFT-computed intensities (km/mol)	Vibrational mode
1335	(1315)	103	δ C5H + δ C6H + δ N1H
1500	1495	105	ν C4N
	1508	75	ν C4C5+ν N1C6
1525	–		
1610	1594	511	ν C2O
1650	1637	187	δ NH_2 scissoring
	1654	735	ν N3C2 + ν C5C6
1800	–		

a) Scaled by a factor of 0.97

C2*d*

Wavenumbers (cm^{-1}) Exp.	Calc.[a]	DFT-computed intensities (km/mol)	Vibrational mode
1335	(1355)	24	δ C5H + δ C6H + δ N1H
1500	1490	74	ν N3C4
	1498	132	ν C4N
1525	–		
1610	1592	645	ν C2O
1650	1633	100	δ NH_2 scissoring
	1661	734	ν N3C2 + ν C5C6
1800	–		

a) Scaled by a factor of 0.97

C3

Wavenumbers (cm^{-1}) Exp.	Calc.[a]	DFT-computed intensities (km/mol)	Vibrational mode
1335	–		
1500	–		
1525	1516	121	ring breathing
1610	–		
1650	1646	727	ν C4N
1800	1791	525	ν C2O

a) Scaled by a factor of 0.97

C6

Wavenumbers (cm^{-1})		DFT-computed intensities (km/mol)	Vibrational mode
Exp.	Calc.[a]		
1335	1325	37	δ C5H + δ N1H
1500	–		
1525	1540	147	ν C4C5
	1546	99	ring breathing
1610	1615	163	ν C5C6+ δ NH_2 scissoring
1650	1650	791	ν C4N
1800	1810	711	ν C2O

a) Scaled by a factor of 0.97

C1

Wavenumbers (cm^{-1})		DFT-computed intensities (km/mol)	Vibrational mode
Exp.	Calc.[a]		
1335	1324	134	δ C5H + δ N1H
1500	–		
1525	1522		ν N3C4
1610	1618	89	ν C5C6+ δ NH_2 scissoring
1650	1654	863	ν C4N
1800	1790	580	ν C2O

a) Scaled by a factor of 0.97

Figure S6. Computed vibrational spectra of **C2*d*** and **C1** forms of $[(\mathbf{C})CH_3]^+$ cation. The experimental IRMPD spectrum of methylated cytosine is overlayed in grey.

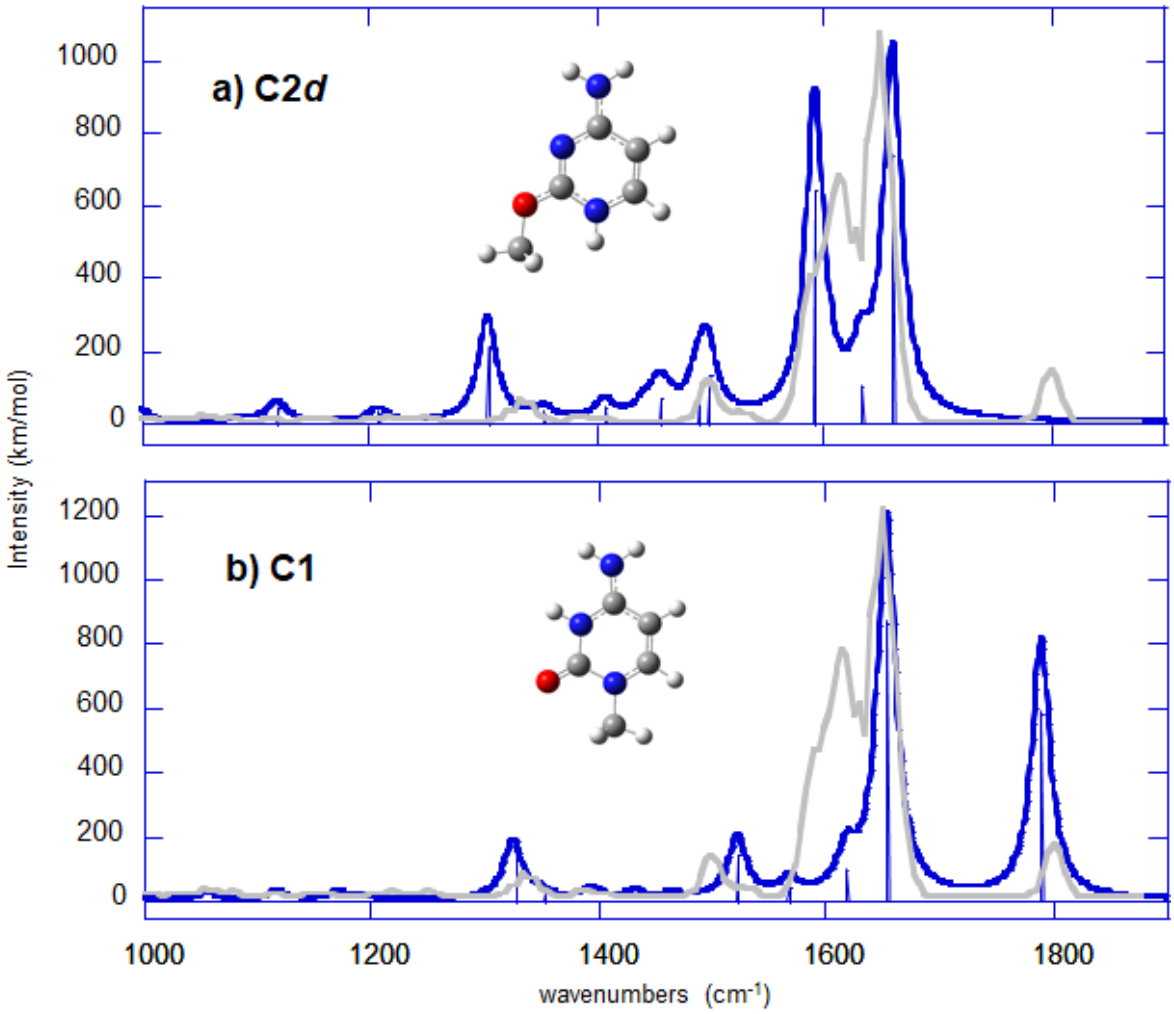


Table S7. Scan of the C2-O-Hg angle through the molecular plane from rotamer **C2*d*** to rotamer **C2*c*** (global minimum) with their corresponding electronic energy values.

C-O-Hg angle	Electronic energy (a.u.)
215.60 (**C2*d***)	-588.20185
206.03	-588.20152
196.46	-588.20102
186.89	-588.20099
177.32	-588.20197
167.75	-588.20304
158.18	-588.20348
148.61	-588.20241
139.04	-588.21381
129.47	-588.21621
119.90 (**C2*c***)	-588.21722

The scan was carried out fixing the C-O-Hg angle and freezing all coordinates but those of the methyl group attached to Hg. Considering the **C2*d*** as the starting point (0.0 kJ/mol), the highest point of the scan (+2.3 kJ/mol) is located at an angle of 186.89º, as shown in Figure **S7**.

Figure S7. Plot of electronic energy values associated to the scan of the C2-O-Hg angle through the molecular plane from rotamer **C2*d*** to rotamer **C2*c*** (global minimum).