Structural and electronic changes of cytosine upon addition of atomic sulfur

P. Mollinedo-Rosado, R. Santamaria *

Instituto de Física, UNAM, A.P. 20-364, D.F. 01000, Mexico

Received 10 August 2006; received in revised form 18 October 2006; accepted 19 October 2006

Available online 6 November 2006

Abstract

When sulfur becomes a contaminant it may affect important biophysical and biochemical processes. In this work, we investigate the interaction of cytosine with atomic sulfur. We find two equally probable places where sulfur may link cytosine. The changes in energetics and variations in the vibrational spectra, dipole moment, charge populations and electron density with respect to isolated cytosine are quantified. The computations are performed according to the Kohn–Sham version of density functional theory. We conclude on the possibility that sulfur may locally interfere in the formation of the Watson–Crick pairs and in the stacking of contiguous nucleic acid bases of the DNA/RNA helix.

© 2006 Elsevier B.V. All rights reserved.

Keywords: Cytosine; Sulfur contamination; Electronic structure; Vibrational spectrum; Density functional theory

1. Introduction

The nucleic acids RNA and DNA are the central molecules in the storage and transmission of the genetic information [1]. This has been one of the main reasons to investigate the nucleic acids at the quantum mechanical level [2–6]. The RNA and DNA are polymers that share some interesting similarities, for instance, it is well known that such polymers have three nucleic acid bases (guanine, adenine and cytosine) in common, being the smallest the cytosine molecule and the largest the guanine molecule [7].

On the other hand, sulfur, as a single species or taking part in a compound, is found in a relatively high abundance in the biosphere. Sulfur is characterized as a very reactive element [8] and it has been used, for instance, in chemical warfare agents [9], where the damage on the DNA is significant [10,11]. There is also evidence of sulfur compounds interfering in the biochemical activity and structural organization of higher plants [12–15] and macroinvertebrates [16], specially located in urban environments. The migration of sulfur compounds is also considered a threat to the ecological equilibrium of marine environments [17]. Taking the reactive nature of sulfur into consideration, one immediately wonders of the possibility that it might strongly interact with the RNA and DNA chain molecules, and the consequences that such possible interaction brings in the storage and transmission of the genetic information. Since the genetic information is based on the sequence of the nucleic acid bases, the strategic places where sulfur may damage the genetic code are precisely such bases. In this work, we are particularly interested in investigating the structural, electronic and energetic changes of cytosine due to the presence of sulfur in its simplest form, namely, its atomic form. Most of the investigations on sulfured compounds of cytosine have focused in studying thiocytosine [18,19], derivatives of the thiocytosine [20,21] and other compounds of larger molecular weight [22]. The importance of analyzing this system depends on the position that sulfur links the cytosine because, according to the position, the sulfur could prevent the formation
of Watson–Crick pairs or the piling of the nucleic acid bases in the RNA and DNA molecules.

In order to assess the changes of cytosine due to the presence of atomic sulfur, we carry computations and compare with the basic properties of cytosine and sulfur, infinitely separated from each other, assuming both in vacuum, and in their ground-state configurations. In this regard, an ab initio method that provides reliable results is required. However, from the computationally point of view, it should be sufficiently flexible to perform all the necessary calculations. The method which meets this requirements is the Kohn–Sham version of density functional theory (KS–DFT), which is described in the following section.

2. Method

The KS–DFT level of theory allows to obtain a variety of molecular properties. These properties like, for example, the energy, are constructed in terms of the electron density and the molecular orbitals. It is with the gaussian basis set DZVP (double-zeta valence plus polarization [23]) that the molecular orbitals of the electrons are built. The electron density originates from such molecular orbitals. Then, the coulomb, exchange and correlation potentials are calculated from the electron density. The functionals chosen to construct the exchange and correlation energy are Becke – 88 (B88) and Lee–Yang–Parr (LYP), respectively [24,25]. In the KS–DFT context an effective potential that rules the behavior of each electron is derived and adopts the general form:

$$ v_{\text{eff}}(\mathbf{r}) = v_{\text{nuc}}(\mathbf{r}) + \frac{\delta J[\rho]}{\delta \rho(\mathbf{r})} + v_{\text{B88–LYP}}(\mathbf{r}) $$

where $v_{\text{nuc}}(\mathbf{r})$ is the nuclear potential, $J[\rho]$ is the coulomb energy term and $v_{\text{B88–LYP}}(\mathbf{r})$ is the exchange-correlation potential [26]. In the KS–DFT scheme we must solve $n$ integro-differential equations of the type:

$$ \left[ -\frac{1}{2} \nabla^2 + v_{\text{eff}}(\mathbf{r}) \right] \phi_i = e_i \phi_i \quad i = 1, \ldots, n $$

where $n$ is the number of electrons of the system and $\nabla^2$ is the kinetic energy operator corresponding to the non-interacting system of electrons. This equation has to be solved iteratively and self-consistently until the convergence in the electron density and energy is achieved. The selected convergence thresholds in both cases are $1 \times 10^{-2}$ au. Once the molecular orbitals are known, the others quantities like the electron density, potentials, energy, and more, can be calculated. In order to find the geometrical structures with minimum energy we have chosen the Quasi-Newton algorithm with linear searches and approximate updates of the Hessian matrix like in Ref. [27]. The computations have been performed in a PC cluster using the NWChem software package [28].

3. Results

3.1. Method validation

In the literature one finds that the level of theory used here is acceptable to study cytosine, because the theoretical results compare well with experimental data [29–31]. However, our goal is to investigate not isolated cytosine, but cytosine interacting with sulfur. In order to validate the method when these two compounds interact, we have calculated the structures and vibrational spectra of organic molecules containing sulfur such as 2,4-imidazolidinedithione,5,5-dimethyl and 1,2-epithiopropane [32]. Our aim is to compare bond lengths, bond angles and their vibrational spectra with other higher-level methods and experimental measurements, whenever possible, to indirectly assess the reliability of the DFT method for the special case of cytosine interacting with sulfur.

Table 1 gives, according to the DFT method and experimental measurements, the bond lengths and angles of the compound 2,4-imidazolidinedithione,5,5-dimethyl. Except for the $C–S$ bonds (for which the length is overestimated by approximately 0.1 Å) most DFT bond lengths are in agreement with the experimental data [32]. The bond angles showing the worst deviations are of the type CNC, NCN and CCN, where the DFT deviations are between 5° and 9°. In the case of the 1,2-epithiopropane compound, the bond lengths and angles compare better with higher-level methods like the MP2, QCISD and CCD (the experimental results are not presented as they substantially deviate from all the theoretical methods) [33]. The largest deviation between the DFT and the other methods occurs for bond lengths of the type $C–S$, they are overestimated by 0.06 Å. The angles show fluctuations around $1^\circ$.

The computation of the vibrational spectra also constitutes a reasonable test of the method because second derivatives of the energy are calculated, which exhibit the fine behavior of the energy as a function of the nuclear coordinates, rather than a bare number solely. In the low

<table>
<thead>
<tr>
<th>Bond</th>
<th>Exp.</th>
<th>DFT</th>
</tr>
</thead>
<tbody>
<tr>
<td>N1–C2</td>
<td>1.419</td>
<td>1.377 (0.042)</td>
</tr>
<tr>
<td>C2–S5</td>
<td>1.564</td>
<td>1.660 (−0.096)</td>
</tr>
<tr>
<td>C3–N1</td>
<td>1.411</td>
<td>1.416 (−0.005)</td>
</tr>
<tr>
<td>C3–S9</td>
<td>1.564</td>
<td>1.671 (−0.106)</td>
</tr>
<tr>
<td>C4–C2</td>
<td>1.512</td>
<td>1.556 (−0.044)</td>
</tr>
<tr>
<td>C4–C7</td>
<td>1.539</td>
<td>1.553 (−0.014)</td>
</tr>
<tr>
<td>C4–C8</td>
<td>1.539</td>
<td>1.553 (−0.014)</td>
</tr>
<tr>
<td>N6–C4</td>
<td>1.488</td>
<td>1.487 (0.001)</td>
</tr>
<tr>
<td>N6–C3</td>
<td>1.429</td>
<td>1.367 (0.062)</td>
</tr>
</tbody>
</table>

* The dimensions are Å and degrees. Numbers in parenthesis represent deviations with respect to the experiment (Exp) [32].
frequency region, the frequencies of 1,2-epithiopropane are similar to the experimental ones, except for a small shift. For example, two DFT peaks of significant intensity appear at 562 and 597 cm\(^{-1}\), whereas the equivalent experimental peaks occur at 600 and 620 cm\(^{-1}\) [32]. In the interval from 840 to 1475 cm\(^{-1}\) the theoretical and experimental data show a better overlap. The main peaks of this region occur at 849, 890, 985, 1025, 1054, 1155, 1342, 1381, 1449 cm\(^{-1}\) and the experimental peaks approximately appear at 880, 900, 1000, 1026, 1050, 1160, 1340, 1380 and 1430 cm\(^{-1}\). At higher frequencies it is observed a greater shift when the theoretical values of 2955, 3043 and 3075 cm\(^{-1}\) are compared with the experimental measurements 2850, 3000 and 3080 cm\(^{-1}\). We have a similar situation for the compound 2,4-imidazolidinedithione,5,5-dimethyl [32]. The conclusion in this Section is that the level of theory looks appropriate to study the structural and electronic changes of cytosine produced by its interaction with atomic sulfur, except for an overestimation of 0.1 Å in bond lengths of the type C–S.

### 3.2. Geometrical structures and energetics

After optimizing the system cytosine plus sulfur, several geometries are obtained, they differ by the way sulfur links cytosine. However, our discussion will be only focused on the most energetically stable structures (Fig. 1). We distinguish them with the symbols: S1 for the structure where sulfur stays in the molecular plane and attached to N3, S2 for the structure where sulfur is found out of the plane and attached to the bond C5–C6, and S3 for the structure where sulfur stays in the molecular plane but attached to oxygen. The geometrical characterization of the three compounds are given in Tables 2 and 3.

In the case of the S1 compound, the bonds C2–N3 and N3–C4 appear slightly elongated by 0.05 Å with respect to the bonds of isolated cytosine. For S2, the C4–C5 and C5–C6 bonds show similar behavior. In contrast, for the S3 compound, only the N1–C2 exhibits the major change by reducing the bond length from 1.450 to 1.388 Å. In the case of angles (Table 3), the angles C5C4N7, N3C4C5 and N1C2O8 of S1 change by 3.5°, −3.1° and 3.0° (a positive number indicates a wider angle with respect to the corresponding angle in cytosine without sulfur, while a negative number means the opposite). For S2, the angles C5C4N7 and N1C6C5 get smaller by 3.2° and 3.0°. For the S3 compound the angle N1C2N3 gets bigger by 6.1°, while the angle N3C2O8 becomes smaller by approximately the same quantity.

The electronic energy (\(E_{\text{elec}}\)) plus the zero point vibrational energy (\(E_{\text{zpe}}\)), their sum denoted by \(E^{\text{S1}}\) (with \(i = 1–3\)), is given in Table 4. The \(E^{\text{S1}}\) energies take as a reference point the energy of isolated cytosine plus the energy of isolated sulfur. For example, for the S1 compound we have \(E^{\text{S1}} = (E^{\text{S1}}_{\text{elec}} + E^{\text{S1}}_{\text{zpe}}) - (E^{\text{isol}}_{\text{elec}} + E^{\text{isol}}_{\text{zpe}})\).

The \(E^{\text{S1}}\) energy indicates that sulfur linked to oxygen in cytosine is less stable than sulfur linked to nitrogen N3 or the carbon bond C5–C6. On the other hand, the difference between the \(E^{\text{S1}}\) and \(E^{\text{S2}}\) energies is smaller than the resolution of the method and it is not possible to decide on the most stable structure. By repeating the calculations with an extended basis set like the TZVP (triple-zeta valence plus polarization) [23], the difference between the energies \(E^{\text{S1}}\) and \(E^{\text{S2}}\) becomes slightly greater. Still, we are unable to decide on the most stable compound from an energetic point of view. Under the consideration of a possible basis-set superposition error the calculations have been repeated taking into consideration the counterpoise correction. Again, the difference between the compounds S1 and S2 is lower than the method resolution. Such a fact points out the

---

**Table 2**

<table>
<thead>
<tr>
<th>Bond</th>
<th>Cytosine</th>
<th>S1</th>
<th>S2</th>
<th>S3</th>
</tr>
</thead>
<tbody>
<tr>
<td>N1–C2</td>
<td>1.450</td>
<td>1.414 (−0.035)</td>
<td>1.431 (−0.019)</td>
<td>1.388 (−0.062)</td>
</tr>
<tr>
<td>C2–N3</td>
<td>1.386</td>
<td>1.440 (0.054)</td>
<td>1.405 (0.019)</td>
<td>1.348 (−0.038)</td>
</tr>
<tr>
<td>N3–C4</td>
<td>1.335</td>
<td>1.387 (0.052)</td>
<td>1.313 (−0.022)</td>
<td>1.346 (0.011)</td>
</tr>
<tr>
<td>C4–C5</td>
<td>1.450</td>
<td>1.453 (−0.017)</td>
<td>1.507 (0.057)</td>
<td>1.448 (−0.002)</td>
</tr>
<tr>
<td>C5–C6</td>
<td>1.376</td>
<td>1.375 (−0.001)</td>
<td>1.485 (0.109)</td>
<td>1.377 (0.001)</td>
</tr>
<tr>
<td>C6–N1</td>
<td>1.369</td>
<td>1.377 (0.009)</td>
<td>1.403 (0.034)</td>
<td>1.369 (0.000)</td>
</tr>
<tr>
<td>N7–C4</td>
<td>1.387</td>
<td>1.359 (−0.028)</td>
<td>1.377 (−0.010)</td>
<td>1.378 (−0.009)</td>
</tr>
<tr>
<td>O8–C2</td>
<td>1.238</td>
<td>1.227 (−0.011)</td>
<td>1.235 (−0.003)</td>
<td>1.312 (0.074)</td>
</tr>
<tr>
<td>H9–N7</td>
<td>1.022</td>
<td>1.038 (0.016)</td>
<td>1.023 (0.001)</td>
<td>1.022 (0.000)</td>
</tr>
<tr>
<td>H10–N7</td>
<td>1.020</td>
<td>1.019 (−0.001)</td>
<td>1.020 (0.000)</td>
<td>1.019 (−0.001)</td>
</tr>
<tr>
<td>H11–C5</td>
<td>1.092</td>
<td>1.091 (−0.001)</td>
<td>1.096 (0.004)</td>
<td>1.092 (0.000)</td>
</tr>
<tr>
<td>H12–C6</td>
<td>1.094</td>
<td>1.092 (−0.002)</td>
<td>1.093 (−0.001)</td>
<td>1.094 (0.000)</td>
</tr>
<tr>
<td>H13–N1</td>
<td>1.023</td>
<td>1.022 (−0.001)</td>
<td>1.023 (0.000)</td>
<td>1.063 (0.040)</td>
</tr>
</tbody>
</table>

---

\(a\) The dimensions are Å. Numbers in parenthesis represent deviations with respect to the numbers of isolated cytosine. The last rows show bond lengths between atoms of cytosine and sulfur.

---

**Fig. 1.** The most energetically-stable structures of (a) cytosine and compounds (b) S1, (c) S2 and (d) S3. The dipole moment orientations and magnitudes are also shown.
DZVP CPU low computational cost (by using the
ESi energy (\(\text{ES}_i\)) and deformation energy (\(\text{def}\), in kcal/mol, for the S1, S2 and S3 compounds.

degeneracy between the S1 and S2 compounds and, in addition, that DZVP is an acceptable basis set. In evaluating the following properties we will use the DZVP basis set for its low computational cost (by using the TZVP basis set the CPU times scale, under the best conditions, by a factor of 1.7 with respect to the times using the DZVP basis set).

Table 4 also shows the atomization energy \(E_{\text{atom}}\) of S1, S2 and S3. They are calculated, for example for S1, in the form \(E_{\text{atom}}^{S1} = (E_{\text{elec}}^{S1} - \sum_{i=1}^{14} E_{\text{elec}}^{i}) - (E_{\text{elec}}^{S1} - \sum_{i=1}^{3} E_{\text{elec}}^{i})\), where the terms in the second brackets represent the reference point with respect to the atomization energy of cytosine. The purpose of taking such a reference point is to avoid working with large energy values. To our surprise, the energy degeneration between S1 and S2 persists. On the other hand, the atomization energy of S3 indicates that the binding of atoms is weaker than in the S1 and S2 cases. Table 4 additionally shows deformation energies, defined as the energy of cytosine with the structure it has in the sulfur compound minus the electronic energy of isolated cytosine. For example, for the compound S1 we have \(E_{\text{def}}^{S1} = E_{\text{elec}}^{S1-\text{sulfur}} - E_{\text{elec}}^{S1}\). The deformation energy breaks the degeneracy between S1 and S2. The sulfur linked to the carbon bond C5–C6 leads to a greater deformation in the cytosine structure than in the case the sulfur connects to nitrogen N3 or to oxygen.

### 3.3. Vibrational spectroscopy

The cytosine plus sulfur compound constitutes a system of 14 atoms, leading to 36 normal modes of vibration. In Fig. 2 the frequencies of the systems S1, S2 and S3 are compared with the frequencies of cytosine. In the S1 case, most peaks are similar to these of cytosine (Fig. 2a).

Table 4

<table>
<thead>
<tr>
<th>Energy</th>
<th>Base</th>
<th>Geometry</th>
<th>S1</th>
<th>S2</th>
<th>S3</th>
</tr>
</thead>
<tbody>
<tr>
<td>(E_{\text{elec}}^{S1})</td>
<td>DZVP</td>
<td>–38.142</td>
<td>–36.857</td>
<td>–21.256</td>
<td></td>
</tr>
<tr>
<td>(E_{\text{elec}}^{S1})</td>
<td>TZVP</td>
<td>–38.647</td>
<td>–35.132</td>
<td>–21.873</td>
<td></td>
</tr>
<tr>
<td>(E_{\text{atom}}^{S1})</td>
<td>DZVP</td>
<td>–38.882</td>
<td>–38.122</td>
<td>–21.771</td>
<td></td>
</tr>
<tr>
<td>(E_{\text{def}}^{S1})</td>
<td>DZVP</td>
<td>4.231</td>
<td>18.904</td>
<td>7.391</td>
<td></td>
</tr>
</tbody>
</table>

\(E_{\text{elec}}^{S1}\), atomization energy (\(E_{\text{atom}}^{S1}\)) and deformation energy (\(E_{\text{def}}^{S1}\)), in kcal/mol, for the S1, S2 and S3 compounds.

### 3.4. Atomic population analysis

In order to assess the charge redistribution caused by sulfur bonded to cytosine, we analyze the Mulliken partial populations of cytosine, S1, S2 and S3, because they can give us information of the charge concentration effects on each nucleus (see Table 5).

The atoms of compound S1 presenting the largest charge variations with respect to the cytosine charge values are C2, N3 and C4, which are precisely the closest atoms of cytosine to sulfur. For the S2 compound the atoms are C5 and C6, as expected, while for the S3 compound there is only one atom showing the largest fluctuation, and that is C2, against what it is expected as it is oxygen the atom that is linked to sulfur. In the S1 and S3 cases, sulfur is shown slightly more electron-negative than cytosine because a small charge transfer of 0.21 and 0.23 e from cytosine to sulfur is observed, respectively. In the S2 case, sulfur and cytosine are practically neutral, there is essentially no charge transfer (which does not imply that an internal charge redistribution has not taken place in cytosine and in sulfur).

### 3.5. Dipole moments

We have also calculated the dipole moment \(\mathbf{d}\) to investigate charge redistributions (see Fig. 1). The cytosine and
the S1, S2 and S3 compounds exhibit dipole moments of 6.41 D (6.56 D in Ref. [29]), 8.32, 5.51 and 9.83 D, respectively. The dipole of S1 points towards the sulfur atom because this last one is more electronegative than cytosine. In cytosine the dipole moment points towards the oxygen. The dipole of S2 lies outside the molecular plane, but its projection onto the molecular plane is almost parallel to the oxygen bond. The dipole of S3 practically lies in the molecular plane pointing towards the sulfur atom.

Note that by using the Mulliken population charges it is possible to calculate the dipole moments of the molecules and compare with the dipole moments computed according to the KS–DFT method. Then, if we calculate the dipole moment from the Mulliken charges (Table 5) in a purely classical way $\mathbf{P}_{iq} = q_i \mathbf{r}_{iq}$, where $q_i$ is the atomic partial charge, and the center of mass is taken as the origin, the dipoles calculated classically for cytosine, S1, S2 and S3 are, in that order, 6.09, 8.17, 5.18 and 9.50 D, which are close to those obtained from the KS–DFT computation (6.41, 8.32, 5.51 and 9.83 D). Therefore, the Mulliken populations can be considered sufficiently acceptable in our case.

### 3.6. The electron density

The Mulliken scheme and the dipole moment condense the information of a spatial charge distribution to point charges and a polarization vector. However, the spatial charge distribution is a volumetric quantity, represented by the electron density $\rho$, that can be analyzed to study local and global charge distributions. This is the reason we calculate the difference between the electron densities of S1, S2 and S3 with respect to this of cytosine, in other words, we compute $\delta \rho = \rho_{S_i}(\mathbf{r}) - \rho_{cyt}(\mathbf{r})$ ($i = 1-3$) at each spatial point. In Fig. 3, $\delta \rho$ is exhibited and projected onto the structure of cytosine, as it is the molecule of reference. In the S1 case, the regions with greater electron density variation appear along the C2–O8 bond and N3 of the amino group. The largest and darker lobe is assigned to the sulfur atom. In the S2 case, the regions of cytosine affected by the presence of sulfur appear more dispersed and along the bonds C2–O8, H9–N7–H10 and C5–S14–C6 (Fig. 3c). The largest lobe, as in the S1 case, is associated to the sulfur atom, but appears out of the molecular plane of cytosine. In the S3 case, the region with a greater electron density variation appears along the C2–O8 bond.

### 4. Conclusions

When sulfur reaches the innermost places of the cell, as the cell nucleus, where abundant DNA segments exist,
the sulfur shows the ability to link cytosine, altering its geometric and electronic structure. In this work, we have quantified the changes that cytosine suffers due to the presence of sulfur. A stability analysis based on energetics indicates the existence of several positions where sulfur may link cytosine. Only two compounds are shown to be the most stable. The vibrational spectra of the sulfur-ated cytosines exhibit clear differences, as the presence of sulfur in cytosine produces the appearance of new normal modes and frequency shifts (with respect to isolated cytosine) that strongly depend on the sulfur position in cytosine. The charge distribution of cytosine is also affected by the position of sulfur. According to an analysis within the Mulliken scheme, the centers showing significant variations are the immediate atomic neighbors of sulfur, except for the less stable sulfur-ated compound. A minimum charge transfer occurs from cytosine to the sulfur atom, specially in the S2 case. Still, the net charge of cytosine differs between the sulfur-ated compounds. The Mulliken analysis is an effective tool to visualize charge changes in atomic centers, but renders little information of the spatial charge redistribution of the interacting species. In this regard, the electric dipole moment becomes helpful since it presents different orientations and magnitudes in response to the sulfur position. The sulfur atom affects considerably the electron density. Important charge redistributions are observed like, for instance, along the C2–O8 bond of cytosine. The structural and electronic changes of cytosine, produced by atomic sulfur, may prevent the formation of Watson–Crick pairs as sulfur can stay in the middle of hydrogen bridges. The piling of nucleic acids in DNA or RNA helices could be also affected, because sulfur may stay in between contiguous nucleic acid bases. This work should be helpful to understand some of the risks of sulfur contamination and the possible consequences on the DNA/RNA chains.

Acknowledgement

The authors want to acknowledge the allotment of CPU time in the ultra-f PC cluster of IFUNAM.

References