Computational and Theoretical Chemistry 1076 (2016) 101–108

Contents lists available at ScienceDirect

Computational and Theoretical Chemistry

journal homepage: www.elsevier.com/locate/comptc

# The effect of cytosine methylation on its halogen-bonding properties

Marta Marín-Luna *, Ibon Alkorta, José Elguero

*Instituto de Química Médica, CSIC, Juan de la Cierva, 3, E-28006 Madrid, Spain*

ARTICLE INFO

*Article history:*
Received 27 November 2015
Received in revised form 9 December 2015
Accepted 9 December 2015
Available online 17 December 2015

*Keywords:*
DNA methylation
DNA-bases tautomerism
Halogen bond
Molecular electrostatic potential
MP2 calculations

ABSTRACT

This study shows the influence of a 5-methyl substituent on the interaction between 1-methylcytosine, 1,5-dimethylcytosine, 1-methyluracil and 1-methylthymine and dichloride ($Cl_2$) and fluorine chloride (ClF). The methyl derivatives were selected because of the important role played by a methyl group at position 5 of the pyrimidine ring on the biological properties of nucleobases. Besides binding energies obtained at the MP2/aug-cc-pVDZ level, a variety of theoretical methods were used to analyze the structures of the different minima obtained. A total of 116 complexes have been studied.



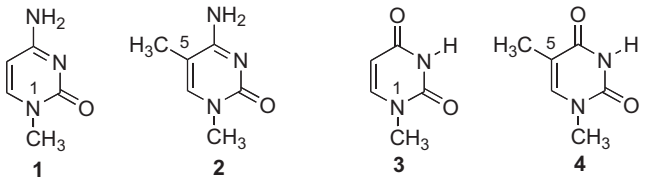
## 1. Introduction

This work is related to two topics, 5-methylation of cytosine and halogen bonds and to their interaction. Besides cytosine and 5-methylcytosine, uracil, thymine and more specifically their 1-methyl derivatives will be discussed.

Transcription of genes is regulated by chemical marks in particular by addition of a methyl group to the DNA helix [1,2]. The introduction of a methyl group at position 5 has profound consequences in the biological behavior of cytosine. Cytosine DNA methylation sequence consists in two main steps: initially the UHRF1 protein distinguishes hemi-methylated from unmethylated DNA [3] and then a methyl group is transferred from S-adenosyl-L-methionine (SAM) to the C(5) position of cytosine by a family of cytosine (DNA-5) methyltransferases (DNMT's) [4–6]. The methylation mechanism involves an $S_N2$-type reaction where the methyl group is transferred as a cation, according to quantum mechanical calculations [7,8]. DNA methylation occurs almost exclusively at CpG islands and has an important contributing role in the regulation/inhibition of gene expression [9,10]. This turn is decisive in epigenetic mechanisms [11], oncogenesis, hematological malignancies (MDS), the structure of the Holliday junction, and has lead to the design of inhibitors of DNA methylation, like 5-aza-2′-deoxycytidine [12–15]. On the other hand, when cytosine is replaced with 5-methylcytosine the i-motif still forms without any decrease in stability [16] (see, however [17]).

Tautomers are organic compounds that are interconvertible by an intramolecular proton transfer reaction [18,19]. Nucleobases are potential compounds to undergo tautomeric modifications that may lead to mutations in the DNA helix. This phenomenon now known as the "rare tautomer hypothesis" was originally postulated by Watson in 1953 [20].

The tautomerism of cytosine and 5-methylcytosine and the photoconversion between tautomers was studied at the QCISD and CCSD levels [21]. The effect of discrete water molecules on the tautomerism of cytosine has been calculated at the MP2 level [22]. Fragile X metal retardation is due to hypermethylation of cytosine residues; theoretical MP2(full)/6-31G(d) calculations were carried out to study the effect of 5-methylation of cytosine on the base pairing energies. Calculated proton affinities of **1** (964.7 kJ $mol^{-1}$) and **2** (977.9 kJ $mol^{-1}$) were reported (protonation on N3 and **a** tautomers as represented) [23].




There is abundant information about the fact that coordination with metals strongly modifies the tautomeric composition of cytosine, 5-methylcytosine, uracil, thymine and more specifically of their 1-methyl derivatives (**1**–**4**) [24–26]. The differences in the effect of the 5-methyl group in cytosines (DNA, associated with

* Corresponding author. Tel.: +34 91 562 29 00; fax: +34 91 564 48 53.
*E-mail address:* martamarin@iqm.csic.es (M. Marín-Luna).
*URL:* http://are.iqm.csic.es/ (M. Marín-Luna).

http://dx.doi.org/10.1016/j.comptc.2015.12.007
2210-271X/

gene silencing [27]) and in uracil/thymine (RNA *vs.* DNA) are sufficiently important and have been associated with the very different role they play in DNA and RNA.

Other molecules can also affect to the stability order of the nucleobase tautomers. Thus we have studied the influence of intermolecular halogen bonds on the tautomerism of guanine [28]. There, we have shown that in most cases the complexes formed by 9-methylguanine with dihalogen molecules ($Br_2$, BrCl, BrF, $Cl_2$ and ClF) are able to alter the order of stability of the different tautomers. Papers reporting tautomeric control by halogen bonds are still very rare [29–31]. We have devoted several papers to halogen bonds, for instance [32–35].

According to the IUPAC "A halogen bond occurs when there is evidence of a net attractive interaction between an electrophilic region associated with a halogen atom in a molecular entity and a nucleophilic region in another, or the same, molecular entity" [36]. The halogen bond [37] is intimately relied to the notion of σ-hole, "a region of positive electrostatic potential, on the outermost portion of the halogen's surface, centered on the *R–X* axis" [38,39].

In the present article, we want to explore the effect on the tautomerism due to the complexation of two representative dihalogen molecules (ClF and ClCl) with four nucleobases, two unsubstituted in position 5 and two with a methyl group in such position: cytosine, 5-methylcytosine, uracil and thymine (5-methyluracil). These nucleobases have been computed as 1-methyl derivatives (1-methylcytosine (**1**), 1,5-dimethylcytosine (**2**), 1-methyluracil (**3**) and 1-methylthymine (**4**), respectively), as simple models of those found in DNA and RNA. In addition, the influence of a methyl group at C(5) on the XB interaction and the ability of the dihalogen molecules to distinguish between methylated nucleobases and non-methylated ones have been analyzed in detail.

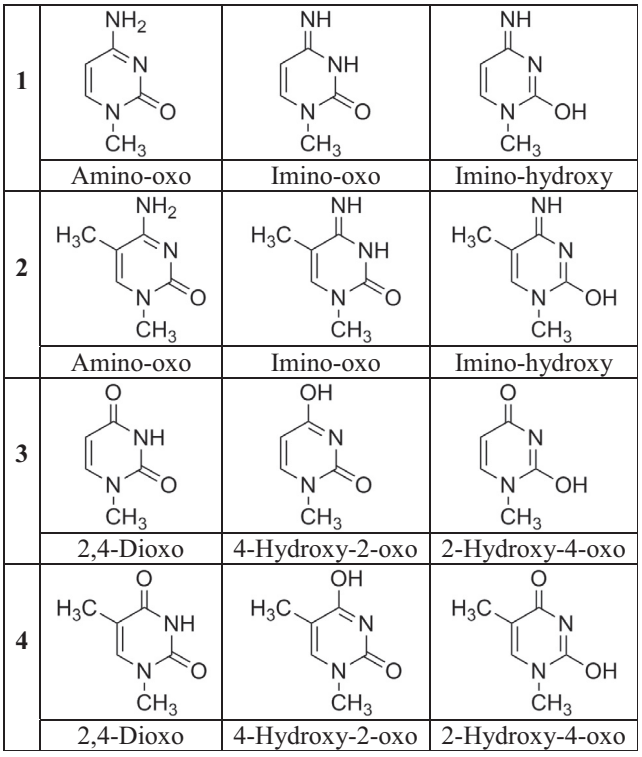



**Fig. 1.** Tautomers of 1-methylcytosine (**1**), 1,5-dimethylcytosine (**2**), 1-methyluracil (**3**) and 1-methylthymine (**4**) generated by the AMBIT-tautomer program.

## 2. Computational methods

The potential tautomeric forms of each isolated molecule have been generated with the AMBIT-Tautomer program [40]. The geometries of the systems have been optimized at MP2 [41]/aug-cc-pVDZ computational level [42]. This theoretical level has been shown to be appropriate on previous studies about weak interactions and offers a reasonable compromise of computation time for the large number of compounds under study [43–46]. Frequency analyses have been applied to confirm the minimal nature of the optimized structures. The binding energies are defined as the differences in energy between the complexes and the sum of the energies of isolated tautomers in their minimal energy geometry. All calculations have been carried out with the Gaussian-09 program [47]. The molecular electrostatic properties of the monomers have been calculated using the Gaussian 09 program and plotted with the Jmol program [48]. The electron density has been analyzed by means of the Atom in Molecules (AIM) methodology [49–51]. The location and properties of the bond critical points (BCP) and the molecular graphs have been calculated and plotted using the AIMAll program [52].

## 3. Results and discussion

The Ambit program of automatic generation of tautomers [40], proposes three tautomers for each compound (Fig. 1). It should be noted that this program does not differentiate between the *E*/*Z* forms of the imino group or the *sE*/*sZ* of the hydroxy ones.

All the tautomers of Fig. 1 have been calculated at the MP2/aug-cc-pVDZ theoretical level. The *E*/*Z* orientation of the imino and the *sE*/*sZ* of the hydroxy groups have been considered in all appropriate cases. In the case of 1-methylcytosine (**1**) and 1,5-dimethylcytosine (**2**), only the oxo (amino-oxo and imino-oxo) tautomers have been taken in account in this study due to the reported high relative energies of the imino-hydroxy tautomers respect to the most stable amino-oxo tautomer (70–105 kJ $mol^{-1}$) [53–55].

The relative energies of the tautomers, shown in ascending order of energy of unsubstituted tautomers at C(5) position, are depicted in Fig. 2. In the case of 1-methylcytosine (**1**) and its 5-methyl derivative (**2**), the amino-oxo tautomers are the most stable while tautomers with two carbonyl groups, **3a** and **4a**, are the preferred ones in the cases of 1-methyluracil (**3**) and 1-methylthymine (**4**), which agrees with the literature (see above). For cytosine derivatives (**2**) the stability order of the tautomers is the same as those of **1** (**a** > **b** > **c**). In the case of **3** and **4**, the ranking of stability is the same save for the **c**/**d** pairs. In the case of 1-methyluracil (**3**), the 4-hydroxy tautomer **3c**, in which hydroxy group presents an *anti* (*sE*) disposition, is more stable than the 2-*anti*-hydroxy tautomer **3d** (*sE*) by 9.5 kJ $mol^{-1}$. However, in **4c** and **4d** the most stable tautomer is the 2-*anti*-hydroxy, **4d**, by 6.6 kJ $mol^{-1}$.

It is noticeable that the presence of a methyl substituent on the pyrimidine ring plays an important role in the relative energies of tautomers in comparison to when the methyl group is absent. Thus, while the relative energies of the tautomers of **2** are lower than **1** (**b** –3.1, **c** –5.7 kJ·$mol^{-1}$) these are greater in the case of **4** compared with **3** (**b** +13.1, **c** +19.0, **d** +4.9, **e** +8.2 kJ·$mol^{-1}$). These differences can be due to the fact that in **2a** there is a repulsion between the methyl and amino group that is reduced in **2b** and **2c**. In contrast, in **4b** and **4c** there is a repulsion between the OH and methyl groups but not in **4a** and in the tautomers of **3**.

In the literature, calculations concerning all these 1-methyl derivatives were carried out with the semi-empirical AM1 method and agree qualitatively with the values of Fig. 2. There are calculations concerning some of the molecules we have studied, e.g. **1a** (0.0), **1b** (5.4) and **1c** (11.2 kJ $mol^{-1}$) at the CCSD(T)/AVDZ//MP2/AVDZ