Homodimers of Cytosine and 1-MethylCytosine. A DFT study of geometry, relative stability and H-NMR shifts in gas-phase and selected solvents

ARTICLE *in* JOURNAL OF MOLECULAR MODELING · MARCH 2014

Impact Factor: 1.74 · DOI: 10.1007/s00894-014-2115-x · Source: PubMed

6 AUTHORS, INCLUDING:

Guvanchmyrat Paytakov
Jackson State University
4 PUBLICATIONS **20** CITATIONS

SEE PROFILE

Dmytro Mykolayovych Hovorun
National Academy of Sciences of Ukraine
379 PUBLICATIONS **2,053** CITATIONS

SEE PROFILE

J Mol Model (2014) 20:2115
DOI 10.1007/s00894-014-2115-x

ORIGINAL PAPER

Homodimers of Cytosine and 1-MethylCytosine. A DFT study of geometry, relative stability and H-NMR shifts in gas-phase and selected solvents

Guvanchmyrat Paytakov · Leonid Gorb · Andriy Stepanyugin · Svitlana Samiylenko · Dmytro Hovorun · Jerzy Leszczynski

Received: 16 October 2013 / Accepted: 13 December 2013

Abstract Dimers of cytosine and its N^1-methylated counterpart were investigated in gas-phase and in various solvents including chloroform, dimethylsulfoxide, and water. The studies were performed at DFT/M06-2X/6-31+G(d,p) level of theory. Relative stabilities of tautomers of cytosine solvated explicitly by a small number of solvent molecules were evaluated. Further solvation effect calculations for homodimers were carried out with conductor-like polarizable continuum model (CPCM). H-NMR shifts and IR frequencies for optimized structures were calculated and compared with available experimental data.

Keywords Chloroform · CPCM · Cytosine dimer · DFT · DMSO · H-NMR · Hydrogen bonding · IR frequency · 1-methylcytosine dimer · M06-2X · π-π stacking · Water

Introduction

It is a well known fact that hydrogen bonding and stacking represent two main types of the interactions that are mainly responsible for the occurrence of numerous DNA and RNA forms [1, 2]. Traditionally, at quantum-chemical level, the nature of those contacts is studied by the modeling of the interaction of small number of isolated DNA bases in gas-phase. Those calculations are performed using different basis sets and different quantum-chemical approximations [3, 4]. In particular, Arshadi and coworkers reported that a cytosine-cytosine dimer having centrosymmetrical type of association (**CS**) has the strongest hydrogen bonds among all regular and irregular DNA base pairs [5]. To study the interaction of DNA bases, the benchmark calculations utilize Møller-Plesset perturbation theory of the second order, interpolated to complete basis results. These results are augmented by the contribution from coupled-cluster (CCSD(T)) level calculations performed with relatively lower basis sets [6–9]. It is assumed that such a combined approach provides virtually better accuracy than the one reached at experimental studies, since it is expected that in the case of gas-phase experimental conditions a mixture of different base pairs is obtained [10]. One of the goals for such simulations has been to compare the energy contributions from different base pair steps during the stacking interactions, with those characterizing the hydrogen bonding interactions that are present in different forms of DNA and RNA. To date the most accurate results of this type are collected in reference [11]. Based on those data one may conclude that energy of hydrogen bonding is always significantly larger (for instance, interaction energy of hydrogen bonding in the case of cytosine-cytosine dimer is c.a. 20 kcal mol^{-1} and for stacking is c.a. 10 kcal mol^{-1}) than the energy of stacking interaction. Nevertheless, the stacking contribution plays a significant role too [12]. In spite of real scientific significance of the gas-phase data, they cannot be attributed directly to the interaction in the strand of DNA or RNA because of several reasons. One of them is that in DNA or RNA the bases are hydrated, or interacting with surrounding enzymes. It is also known that DNA is hydrated very selectively [13, 14], especially in the

Electronic supplementary material The online version of this article (doi:10.1007/s00894-014-2115-x) contains supplementary material, which is available to authorized users.

G. Paytakov · J. Leszczynski (✉)
Interdisciplinary Center for Nanotoxicity, Department of Chemistry and Biochemistry, Jackson State University, Jackson, MS 39217, USA
e-mail: jerzy@icnanotox.org

L. Gorb · A. Stepanyugin · S. Samiylenko · D. Hovorun
Department of Molecular Biophysics, Institute of Molecular Biology and Genetics, National Academy of Sciences of Ukraine, Key State Laboratory in Molecular and Cell Biology, 150 Vul. Zabolotnogo, Kyiv 03143, Ukraine

Published online: 20 February 2014

leaving cell. However, the size and shape of a hydration shell is rather unknown. Because of that another important area of DNA and RNA bases study is represented by investigation of their behavior in the solutions with different dielectric permittivity [see refs. 15, 16 and the references therein]. Such models help to overcome such experimentally observed difficulties as very low solubility of DNA bases, problems with the experimental studies of stacked and H-bonded DNA base pairs, and the fact that H-bonded water can absorb in the same IR region as nucleic acid bases. In addition, in the case of the application of NMR techniques the protons of the solvent can rapidly exchange with amino protons of the bases. Also, denaturation agents of DNA, such as urea, formamide, methanol (CH_3OH), and dimethylsulfoxide (DMSO), are used to lower DNA duplex thermostability during electrophoresis or polymerase chain reaction. Nevertheless, the quantum-chemical data that characterize the interaction in these solvents cannot be used as the source of the interaction energies in DNA and RNA strands; they rather represent the interaction energies for the extreme case when the base is completely hydrated (solvated). It also should be mentioned that the number of studies devoted to interaction of DNA and RNA bases in solvated state (both experimental and computational) is even more limited than the one related to the gas-phase investigations. Most of them are devoted to hydration of single DNA base molecule or its tautomers [17, 18].

The limitations of the data on the interaction of DNA and RNA bases in solvated state challenge us to study such a phenomenon using the methods of computational chemistry. The current publication is devoted to investigation of the interaction of two cytosine (**C**) species and its derivative 1-methylcytosine (**1mC**) in gas-phase and in such solvents as chloroform, DMSO, and water. Available H-NMR experimental data in solvents like chloroform and DMSO led us to consider performing calculations in these media in order to validate our results. Actual dielectric permittivity around the DNA strands is less than the permittivity originated from aqueous media since strands are surrounded by proteins. In addition, chloroform and DMSO are the examples of nonpolar and polar solvents respectively, where one can test correlation between polarity of the solvent and dimer interaction strength. The study includes traditional computational chemistry analysis of the geometry, interaction energy and also the prediction of NMR chemical shifts and IR-frequencies.

Finally we would like to mention the following. To describe correctly both stacking and hydrogen bonding, most of the studies use Møller-Plesset perturbation theory and coupled-cluster approximations. Current investigations apply another computational approximation–density functional theory (DFT). The successful application of DFT to describe these interactions (especially stacking one) became possible only recently due to the appearance of the so-called family of hybrid meta functionals [19]. In addition, incorporation of the dispersion term into the functionals such as BLYP and B3LYP [20, 21] provides new tools to study stacking interactions. Since a DFT level is significantly less time consuming that the methods based on a Møller-Plesset perturbation theory, it allows to study the systems of much bigger size [22]. Therefore, further careful comparison between wave function and density functional based methods would facilitate computational investigations of larger DNA fragments using the latter approaches.

Computational details

Gaussian 09 program package [23] was applied to perform all calculations. Hybrid meta-GGA exchange-correlation density functional M06-2X [24] with 6-31+G(d,p) basis set was used to fully optimize selected structures, calculate interaction energies, vibrational frequencies and NMR shifts. As stated in reference [25], it is necessary to include diffuse and polarization functions when stacking interactions are calculated. Basis set superposition errors (BSSE) for the gas-phase dimers were corrected using counterpoise (CP) method proposed by Boys and Bernardi [26]. Solvent media calculations have been performed with conductor-like polarizable continuum model (CPCM) and the universal force field (UFF) radii was used to define cavities [27]. NMR spectra were calculated at M06-2X/6-31+G(d,p) level using gauge-independent atomic orbitals (GIAO) [28], the NMR computational formalism, and the CPCM solvation model.

Although several tautomeric forms (amino-oxo, amino-hydroxy and imino-hydroxy) coexist for cytosine we have considered the amino-oxo form only, since both previous theoretical and experimental studies reported that this form dominates in higher concentration in solvent [17, 29–31].

Results and discussion

Although there are controversial results reported on the geometry of cytosine in solvent, the data of the majority of the publications suggest that cytosine retains its canonical form [17, 29–33]. We compared the binding capability of water to various tautomers of cytosine in aqueous media. In the case of water solution our results show that among amino-hydroxy, imino-oxo and amino-oxo tautomers, the canonical form (amino-oxo tautomer) is the most stable (4.9–6.1 kcal mol^{-1}) [29, 34, 35]. In addition, water binds to amino-oxo form more strongly (0.49–1.34 kcal mol^{-1}), comparing to other tautomers (see Table S1 and Fig. S3).

The considered geometries of dimers are presented in Fig. S1. To select them, we took into account the results of computational investigations [8, 36, 37] which suggest the most stable structures among hydrogen bonded and stacked

dimers. The effect of N^1 methylation was introduced by replacement of N^1-H hydrogen by CH_3-group.

Geometry Figure 1 depicts optimized structures of selected dimers in gas-phase. We found that application of M06-2X DFT functional results in the gas-phase structures of cytosine dimers that are very similar to the ones which were obtained at the MP2 level (see Table 1) [11]. As expected, solvation changes the intra- and intermolecular parameters of dimer geometry to raise dimer dipole moment. These changes occur in parallel to the increased value of polarity of a solvent. A noticeable elongation of intermolecular H-bonding distances for all dimers (except stacking dimers and H···N in **A** and **MA**) results in weaker interaction energies. Optimized geometries and bond lengths of dimers are very similar in such solvents as DMSO and water. We also found that stacked dimers do not retain their perfect antiparallel structure upon optimization. The structure as shown in Fig. 1 is characterized by slide and twist of about 40°, and tilt of stacked monomers. It is in accordance with previously reported data from investigations of analogous dimers [38–40]. Here, amino hydrogen bond ($N\text{-}H^a$) is elongated by 0.15–0.25 Å and N^1-H bond is longer by 0.29–0.32 Å when it contributes to formation of H-bonding.

Since the presence of methyl group results in significant distortion of the geometry, we display in Table 1 the interatomic distances ($O\cdots H^aN$, $N^3\cdots H^aN$ and O···H, in the case of **MA**, **MB**, and **MCS** dimmers, respectively) which do characterize such a change. As follows from the data presented in Table 1 the gas-phase hydrogen bonded dimers formed by 1-methylcytosine retain the structure of similar dimers formed by cytosine. However, the structures MA and MCS are distorted, due to involvement of the CH_3 group in hydrogen bonding (see Electronic supplementary material for distorted geometries). Replacement upon methylation of more positively charged N^1-H hydrogen by almost inert methyl hydrogen results in a visible change of the total size of hydrogen bonds in **MA** and **MCS** (for **A**:**MA**, $N^1H\cdots N^3$= 1.777:$CH\cdots N^3$=2.628 and for **CS**:**MCS** NH···O=1.687: CH···O=2.601 (in Å) see Table 1). General trend is decrease of the $N^3\cdots H$ distance as polarity of solvent increases, but this is not the case in **MA**. We predicted slight shortening in C^4-N bond when $N\text{-}H^a$ influence increases. Unusual trend of $N^3\cdots H$ hydrogen bonding distances observed in **MA** dimer is due to the loss of planarity in polar solvents.

Energy To the best of our knowledge, the data on relative stability of different cytosine dimers in gas-phase and water solution are slightly controversial and not systematic [41]. Experimentally, the dimers of **A** and **CS** type have been observed in gas-phase by double resonance laser spectroscopy [42]. The computer simulation [43] performed using MPWB1K DFT functional suggests a low concentration of dimers in gas-phase. Formation of **CS** type of dimer was revealed and this species has negative Gibbs free energy at 490 K. The very accurate MP2 simulations augmented by CCSD(T) contribution [11] suggest structure of **B** type to be the most stable. However, the values of energy differences among the other hydrogen bonded types of dimers are not provided in this study. Finally, the study [44] reveals the

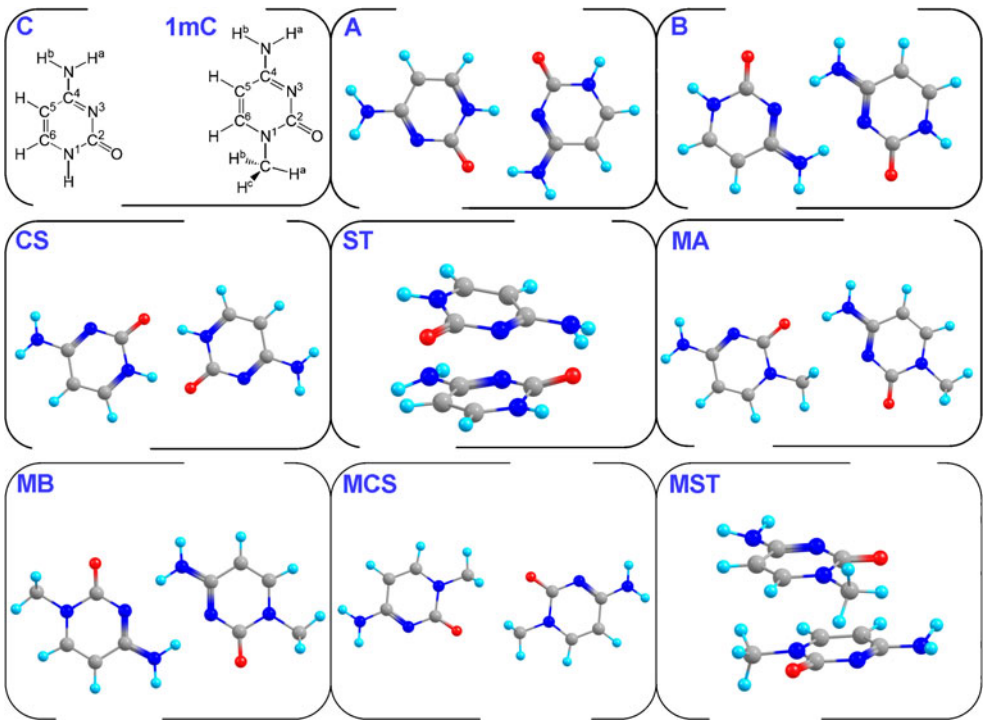

Fig. 1 Two dimensional structures of cytosine and 1-methylcytosine and optimized structures of cytosine and 1-methylcytosine dimers at M06-2X/6-31+G(d,p) level

Table 1 Selected bond lengths and intermolecular distances (in Å) and dipole moments (D) for optimized monomer and dimer structures at M06-2X/6-31+G(d,p) level of theory

Structure	Parameter	Gas-phase	In $CHCl_3$	In DMSO	In water
C	N^1-H (Å)	1.011 (1.002*)	1.012	1.012	1.012
	C^2-O (Å)	1.216 (1.219*)	1.229	1.234	1.235
	C^4-N (Å)	1.355 (1.367*)	1.347	1.344	1.343
	N-H (Å) [a]	1.008 (1.005*)	1.009	1.009	1.009
	μ (D)	6.8	8.8	9.64	9.7
A	C^4-N (Å)	1.334	1.335	1.335	1.335
	N-H^a (Å)	1.033	1.026	1.024	1.024
	H^a…O (Å)	1.797	1.876	1.896	1.903
	C^2-O (Å)	1.237	1.243	1.247	1.247
	N^1-H (Å)	1.04	1.042	1.044	1.044
	H…N (Å)[3]	1.777	1.776	1.775	1.772
	μ (D)	4.3	5.2	5.6	5.7
B	C^4-N (Å)	1.334 (1.334†)	1.336	1.336	1.336
	N-H^a (Å)	1.030 (1.029†)	1.028	1.027	1.027
	H^a…N^3 (Å)	1.905 (1.861†)	1.943	1.955	1.956
	μ (D)	0.1	0.1	1.15	1.2
CS	C^2-O (Å)	1.239	1.246	1.248	1.248
	O…H (Å)	1.687	1.741	1.758	1.760
	N^1-H (Å)	1.043	1.037	1.036	1.035
	μ (D)	0.4	0.1	0.0	1.4
ST	C^4-N (Å)	1.356	1.35	1.348	1.348
	N-H^b (Å)	1.014	1.014	1.015	1.015
	H^b…O (Å)	2.286	2.224	2.147	2.145
	C^2-O (Å)	1.225	1.234	1.240	1.240
	μ (D)	5.1	7.9	9.8	9.8
1mC	C^4-N (Å)	1.356	1.349	1.346	1.346
	N-H^a (Å)	1.008	1.009	1.009	1.009
	N^1-C (Å)	1.457	1.460	1.461	1.461
	C^2-O (Å)	1.219	1.230	1.235	1.235
	μ (D)	6.4	7.9	9.24	9.3
MA	C^4-N (Å)	1.344	1.343	1.342	1.342
	N-H^a (Å)	1.021	1.02	1.021	1.021
	H^a…O (Å)	1.890	1.922	1.888	1.889
	C^2-O (Å)	1.228	1.238	1.241	1.242
	N^1-C (Å)	1.463	1.461	1.463	1.463
	H…N^3 (Å)	2.628(51)	2.755(60)	2.459	2.466
	μ (D)	1.2	2.6	8.01	8.1
MB	C^4-N (Å)	1.336	1.337	1.338	1.338
	N-H^a (Å)	1.029	1.027	1.027	1.027
	H^a…N^3 (Å)	1.910	1.945	1.957	1.959
	μ (D)	0.0	0.0	0.58	0.6
MCS	N^1-C (Å)	1.461	1.462	1.462	1.463
	C-H^a (Å)	1.090	1.090(92)	1.092(91)	1.092
	H^a…O (Å)	2.601(11)	2.372(684)	2.413(736)	2.364
	C^2-O (Å)	1.223	1.232(3)	1.236(35)	1.236
	μ (D)	0.9	0.4	1.05	0.0
MST	C^4-N (Å)	1.362(72)	1.357(53)	1.346(48)	1.348(50)
		1.011(10)	1.010(08)	1.009(07)	1.009(08)

Table 1 (continued)

Structure	Parameter	Gas-phase	In $CHCl_3$	In DMSO	In water
	N-H[a,b] (Å)	1.011(09)	1.010(09)	1.009(08)	1.009(08)
	C^2-O (Å)	1.227(23)	1.232(33)	1.236(35)	1.236(36)
	N^1-C (Å)	1.459(64)	1.461(61)	1.461(61)	1.462(60)
		1.088(95)(93)			
	C-H[a,b] (Å) [c]	1.087(92)(94)	1.091(91)(90)	1.090(92)(90)	1.088(93)(90)
		1.	1.088(93)(90)	1.092(90)(89)	1.091(91)(89)
	μ (D)	5.0	6.8	5.28	7.8

*Parameters calculated at MP2/cc-pVTZ(-f) level of theory [1]. † Parameters calculated at MP2/cc-pVTZ level of theory [2]

interaction enthalpies and entropies only for the most stable stacked and hydrogen bonded structures in water solution. Nevertheless, general trends which we revealed based on the application of M06-2X functional are in line with described literature data. Namely, our gas-phase calculations predict very similar stability for hydrogen bonded dimers as described in literature, with slight preference for the structures of **A** and **CS** types. The stacking structure is found to be significantly less stable. In addition, in the case of gas-phase data the value of interaction energy calculated at M06-2X level is in excellent correspondence with the one calculated at the MP2 level, augmented by CCSD(T) contribution (see Table 2). Here it is worthy to note that M06-2X level is much less computer time demanding than the MP2 calculations.

Interaction energies of dimers in water are significantly lower compared to gas-phase data. This raises the question whether dimerization could occur in highly polar solvents. Comparing interaction energies of monohydrated cytosine cluster (~7 kcal mol^{-1}) and cytosine dimers (~11 kcal mol^{-1}) in aqueous solution we do expect that dimerization process would occur. The data describing the process of dimerization in solvents demonstrate the dependence of interaction energy on the polarity of solvent in the following way: the dimers are always less stable in solvent compared with gas-phase. The stacking structure is also bonded, however it is always the least stable. Interestingly, the value of interaction energy in stacking dimer agrees well with the interaction enthalpy calculated in the approximation of Langevin dipoles. However, in the case of the value of the interaction energy characterizing hydrogen bonding we do not have perfect correspondence with the data obtained using approximation of Langevin dipoles. (see Table 2). Therefore, the pattern of relative stability looks as follows:

CS≈**A** > **B** >> **ST**.

The predicted tendency in the stability of hydrogen bonded and stacking dimers is completely different in the case of dimerization of 1-methyl-cytosine:

MB > **MST** > **MA** > **MCS**.

One may expect the predominance of **MB** type of the dimer structure over the structures of **MA**, **MC**, and **MST** types. This tendency can be explained by the fact that hydrogen attached to N1 is replaced by methyl group. Since methyl group involvement in interatomic interaction(s) makes them weaker, it is very clear why the **MB** dimer is the most stable one. The described above tendency remains the same in solvents, however

Table 2 Calculated interaction energies (kcal mol^{-1}) for selected structures of cytosine (**C**) dimers and 1-methylcytosine (**1mC**) dimers

Structures	Gas-phase	In $CHCl_3$	In DMSO	In water	
	IE	CP IE	IE	IE	IE
C					
A	−23.13	−22.21	−15.29	−11.88	−11.93
B	−21.05	−20.23 (−20.8[a])	−14.91	−9.99	−10.12
CS	−23.05 (−22.85[b])	−22.20 (−21.73[b])	−16.67	−11.96	−12.06
ST	−13.89	−12.46 (−10.7[d])	−8.69	−6.66	−6.60 (−3.7[c])
1mC					
MA	−12.49	−11.92	−8.97	−6.03	−6.18
MB	−20.58	−19.83	−14.76	−10.16	−10.22
MCS	−7.43	−7.03	−5.26	−1.93	−1.91
MST	−15.14	−13.26	−11.46	−7.64	−7.83

[a] Values calculated at MP2/cc-pVTZ level, ref. [2] [b] this study, calculated at MP2/6-311++G(d,p) level [c] calculated at "frozen core" MP2/6-31G(d-0.25), ref. [3] [d] calculated at MP2/aug-cc-pVTZ level, ref. [4]

in contrast to the case involving dimers of cytosine the stacking structure of **MST** is the second species in the stability pattern. Not only are geometries of dimers very similar when compared in DMSO and aqueous solutions but also interaction energies of dimers in both solvents are comparable. This indicates that not the shape of solvent, but magnitude of dielectric permittivity plays a crucial role in determining geometries and stabilities of dimers.

H-NMR shifts and IR-frequencies As we already mentioned there is no experimental data describing the interaction of cytosine and 1-methylcytosine in solutions. This is due to a very low solubility of those species. To the best of our knowledge, there is only one NMR study which reveals the change of hydrogen NMR shifts during the association [45]. The data presented in this study suggest the very slight dependence of chemical shifts on the concentration of the components during the association.

To evaluate accuracy of the M06-2X level predicted NMR chemical shifts we verified its accuracy for different NMR experimental data related to the association of cytosine and protonated cytosine (hemiprotonated dimer) in chloroform. The experimental and calculated chemical shifts are collected in Fig. S2. One may see a satisfactory correspondence between experimental and calculated data. Therefore, one may expect that M06-2X approximation is reliable to analyze and predict NMR data for the species considered here.

The H-NMR shifts for monomers and dimers are tabulated in Table 3. The tremendous change in proton NMR shifts could be observed for interacting hydrogens, compared to those predicted for its monomeric structure. Dimer species type **A** could be distinguished from the other type of dimers because only this type shows two chemical shifts around 15.2–16 ppm and 9.8–11.5 ppm, depending on polarity of the solution. On the other hand, **B** type dimer shows only one signal with chemical shift in the range of 11.0–11.9 ppm. **CS** dimer reveals one signal with chemical shift in the range of 14.7 to 13.0 ppm for gas-phase, and in water, respectively. In addition, at least one strongly deshielded chemical shift for the hydrogen(s) N^1H, NH^a, are observed for all types of cytosine dimers, except **ST**.

Table 3 Selected H-NMR shifts (ppm) calculated at M06-2X/6-31+G(d,p) level of theory for cytosine (**C**) and 1-methylcytosine (**1mC**) homodimers in gas-phase and selected solutions

	Gas-phase					Chloroform					DMSO						Water				
	C	**A**	**B**	**CS**	**ST**	**C**	**A**	**B**	**CS**	**ST**	**C**	**A**	**B**	**CS**	**ST**	*Exp*[40]	**C**	**A**	**B**	**CS**	**ST**
C^5H	5.6	5.9	5.7	5.9	5.6	6	6.4	6.1	6.3	5.9	6.2	6.5	6.2	6.4	6	*5.6*	6.2	6.4	6.2	6.5	6
C^6H	7.5	8.3	7.3	7.8	7.5	7.8	8.4	7.6	7.9	7.5	7.9	8.4	7.8	7.9	7.5	*7.3*	7.9	8.4	7.8	8	7.5
N^4H^a	5	5	11.9	5	4.8	5.3	5.3	11.4	5.5	5.3	5.4	5.4	11.2	5.6	5.5	*7*	5.4	5.5	11.2	5.4	5.5
N^4H^b	4.3	4.3	5	4.3	4.1	5	5	5.5	5.2	5	5.3	5.3	5.6	5.5	5.4	*6.9*	5.3	5.4	5.6	5.3	5.4
N^1H	7.1	15.2	6.8	14.7	7.3	7.7	15.5	7.5	13.8	7.6	7.9	15.6	7.7	13.5	7.8	*10.4*	7.9	15.5	7.7	13.5	7.8
N^1H		7	6.8	14.7	6.8		7.5	7.5	13.8	5.8		7.7	7.7	13.5	5.9			7.8	7.7	13.5	5.9
C^6H		7.5	7.3	7.8	7.2		7.6	7.6	7.9	7.5		7.7	7.8	7.9	7.6			7.9	7.8	8	7.6
C^5H		5.8	5.7	5.9	5.7		6.3	6.1	6.3	5.1		6.4	6.2	6.4	5.2			6.3	6.2	6.5	5.2
N^4H^a		11.5	11.9	5	4.8		10.2	11.4	5.5	6.3		9.9	11.2	5.6	6.7			9.8	11.2	5.4	6.7
N^4H^b		4.8	5	4.3	5.8		5.3	5.5	5.2	7.4		5.5	5.6	5.5	7.6			5.5	5.6	5.3	7.6
	1mC	**MA**	**MB**	**MCS**	**MST**	**1mC**	**MA**	**MB**	**MCS**	**MST**	**1mC**	**MA**	**MB**	**MCS**	**MST**		**1mC**	**MA**	**MB**	**MCS**	**MST**
C^5H	5.6	5.6	5.8	5.7	6	6.1	6.1	6.2	6	6.1	6.2	6.3	6.3	6.1	6.1		6.3	6.3	6.4	6.2	6.2
C^6H	7.4	7.2	7.2	7.5	7.4	7.7	7.6	7.5	7.9	7.7	7.8	7.9	7.7	8	7.5		7.9	7.9	7.7	8	7.8
NH^a	5	9.3	11.8	4.8	4.6	5.3	8.8	11.3	5.1	5	5.4	9.2	11.1	5.3	5.4		5.4	9.2	11.1	5.4	5.2
NH^b	4.2	4.5	4.9	4.1	4.2	4.9	5.1	5.3	4.9	4.8	5.2	5.2	5.4	5.2	5.1		5.2	5.2	5.4	5.2	5
CH	2.5	2.3	2.5	7.5	7.8	2.9	2.8	2.9	7.9	7.8	3	3	3.1	7.8	7.7		3	3	3.1	8	7.8
CH	3.6	3.6	3.6	5.7	5.6	3.5	3.5	3.5	6	5.9	3.5	3.4	3.5	6.3	6.2		3.5	3.4	3.5	6.2	6.1
CH	3.6	3.5	3.6	4.8	4.5	3.5	3.5	3.5	5.1	4.9	3.5	3.4	3.5	5.3	5.3		3.5	3.4	3.5	5.4	5.2
C^5H		6	5.8	4.1	3.8		6.4	6.2	4.9	4.7		6.3	6.3	5.2	5			6.3	6.4	5.2	5
C^6H		7.7	7.2	4.2	3.6		7.9	7.5	4.4	2.7		8	7.7	3.9	2.8			8	7.7	4.5	2.8
NH^a		5	11.8	3.1	4.2		5.3	11.3	3.1	4.1		5.5	11.1	3.2	3.6			5.5	11.1	2.9	3.9
NH^b		4.1	4.9	4.2	2.7		5	5.3	3.6	3.3		5.3	5.4	3.7	3.2			5.3	5.4	3.3	3.3
CH		3.3	2.5	3.1	2.7		3.4	2.9	3.1	2.9		2.9	3.1	3.5	3.1			2.9	3.1	2.9	2.9
CH		4.5	3.6	4.2	4.1		4.3	3.5	4.4	3.7		4.7	3.5	3	3.4			4.6	3.5	4.5	3.8
CH		4.5	3.6	4.2	2.5		4.1	3.5	3.6	2.6		3.6	3.5	4.8	2.8			3.6	3.5	3.3	2.7

Table 4 Selected IR frequencies (cm^{-1}) and intensities (presented in parenthesis) (km mol^{-1}) for cytosine (**C**)and 1-methylcytosine (**1mC**)monomers and dimers calculated at M06-2X/6-31+G(d,p) level

	Frequency type									Frequency type								
	Dimer C^{exp}	NH2(s)[b]	NH2(s)[f] *3450*	N1H[b]	N1H[f] *3470*	NH2(a)[b]	NH2(a)[f] *3572*	CO[b]	CO[f]	Dimer $1mC^{exp}$	NH2(s)[b]	NH2(s)[f] *3450*	N1H[b]	N1H[f]	NH2(a)[b]	NH2(a)[f] *3570*	CO[b]	CO[f]
Gas-phase	**C**		3644 (137)		3656 (84)		3791 (71)		1834 (850)	1mC		3648 (117)				3795 (69)		1803 (789)
	A^{exp}		*3450*		*3475*	*3525*	*3572*											
	A	3224 (2152)	3643 (99)	3077 (1593)	3647 (114)	3734 (120)	3785 (60)	1766 (451)	1811 (1331)	MA	3460 (1107)	3642 (115)	3747 (118)	3787 (65)	3103 (11)	3086 (42)	1772 (615)	1794 (1087)
	B	3286 (2856)			3667 (226)	3733 (203)			1810 (1861)	MB	3293 (3007)		3736 (196)			3090 (93)		1789 (1930)
	CS		3643 (222)	3045 (5245)			3790 (136)		1769 (2159)	MCS		3639 (219)		3782 (124)	3102 (37)		1799 (1821)	
	ST	3572 (115)	3623 (108)	3644 (72)	3655 (92)	3714 (83)	3764 (58)	1798 (455)	1809 (1141)	MST	3603 (76)	3610 (72)	3727 (47)	3743 (46)	3077 (46)	3071 (47)	1769 (282)	1793 (1132)
In chloroform	**C**		3632 (251)		3640 (99)		3774 (114)		1759 (1088)	1mC		3634 (189)		3772 (110)		3088 (40)		1738 (1180)
	A	3359 (1365)	3634 (166)	2999 (2902)	3651 (180)	3721 (197)	3775 (105)	1728 (874)	1755 (1555)	MA	3484 (1156)	3636 (193)	3725 (178)	3777 (115)	3103 (28)	3093 (46)	1724 (636)	1733 (1391)
	B	3299 (3088)			3650 (353)	3714 (324)			1740 (2051)	MB	3314 (3170)		3720 (317)			3096 (97)		1725 (3389)
	CS		3631 (349)	3153 (5223)			3776 (230)		1722 (1007)	MCS		3634 (376)		3775 (211)	3093(102) (21(31))		1734(14) (1679(357))	
	ST	3578 (211)	3626 (201)	3650 (128)	3646 (154)	3723 (138)	3767 (106)	1743 (1244)	1747 (1121)	MST	3626 (145)	3617 (155)	3756 (90)	3751 (94)	3091 (44)	3088 (52)	1740 (2115)	1725 (493)
In dimethylsulfoxide	**C**		3627 (336)		3633 (74)		3767 (131)		1728 (894)	1mC		3628 (219)		3765(127)		3089 (42)		1705 (391)
	A	3392 (1251)	3625 (196)	2978 (3215)	3637 (214)	3715 (231)	3763 (127)	1715 (1171)	1732 (1271)	MA	3444 (1228)	3626 (223)	3717 (204)	3764 (131)			1692 (540)	1710 (293)
	B	3309 (3130)			3648 (396)	3710 (367)			1724 (1318)	MB	3321 (3194)		3716 (363)			3097 (98)		1704 (1636)
	CS		3627 (397)	3179 (5159)			37689 (264)		1696 (3061)	MCS		3632(4) (216(1))		3773(5) (128(9))	3090(102) (22(38))		1716 (2141)	
	ST	3556 (303)	3634 (282)	3647 (131)	3645 (196)	3713 (167)	3772 (131)	1717 (1479)	1727 (894)	MST	3624 (233)	3629 (201)	3765 (125)	3767 (125)			1722 (2547)	1704 (616))
In water	**C**		3627 (338)		3633 (75)		3767 (131)		1728 (882)	1mC		3628 (221)		3764 (128)		3089 (42)		1705 (398)
	A	3367 (1271)	3629 (193)	3000 (3225)	3625 (210)	3716 (218)	3771 (128)	1717 (1239)	1734 (1220)	MA	3445 (1222)	3627 (224)	3717 (205)	3765 (132)	3085 (27)	3092 (46)	1692 (543)	1710 (294)
	B	3310 (3126)			3647 (399)	3710 (370)			1723 (1286)	MB	3322 (3185)		3716 (365)			3096 (98)		1704 (1694)
	CS		3627 (408)	3188 (5104)			3765 (267)		1694 (2962)	MCS		3634 (435)		3773 (256)	3090 (54)		1707 (892)	
	ST	3555 (305)	3634 (282)	3647 (132)	3644 (196)	3713 (168)	3771 (131)	1717 (1500)	1727 (880)	MST	3625 (189)	3622 (220)	3760 (118)	3759 (121)	3097 (48)	3088 (54)	1722 (2406)	1708 (478)

Interestingly, among **1mC** homodimers only in **MA** the H-NMR chemical shift larger than 11 ppm is observed. **MB** type 1-metylcytosine dimer is distinctive, with its unique chemical shift located in the range 11–11.8 ppm. **MA** shows H^a shift around 9±0.3 ppm and deshielded methyl hydrogens reveal shifts around 3.6–4.7 ppm. From NMR shifts it is also clear that in more polar solvents such as water and dimethylsulfoxide, **MCS** does not adopt symmetric structure. These specific chemical shits could help to differentiate and analyze different types of dimers present and dominant, under various conditions.

We also made a comparison of observed experimental data and our calculated results in case of DMSO. First of all, we noticed that the general correspondence between calculated and experimental data is worse than in the case of dimerization of protonated and non-protonated cytosine described above. Nevertheless, presented data clearly suggest that the structure observed experimentally has C_s symmetry since, in agreement with experimental data, there is no cleavage of the N^4H_2 protons which is observed in the case of the **A**-type structure. Unfortunately, the experimental data presented in [45] is not sufficient to make a distinctive conclusion about the structure of 1-methylcytosine dimer.

Finally, since the frequency of vibrations in IR region is a fundamental characteristic which characterizes the intermolecular interaction, we predicted IR frequencies for all considered species and compared them with available experimental data [42]. The non-scaled IR frequencies calculated at M06-2X level are tabulated in Table 4. The comparison of IR spectrum of isolated cytosine species with the spectrum of A and **ST** type of dimers suggests that there are two types of carbonyl vibrational frequencies which are considerably red-shifted. Also extremely intense peaks are observed for the $NH2(s)^b$, $N1H^b$, and CO^f vibrations of cytosine dimers and intensity of $NH_2(s)^b$ stretching decreases whereas intensities of $N1H^b$, CO^f, and CO^b vibrational stretching frequencies increase considerably in polar solvents. In order to distinguish among the various forms of cytosine dimers, it is enough to look at missing symmetric, antisymmetric, bound, and free NH, NH_2 or CO stretching vibrations. For instance for **B** type dimer one may notice that $NH2(s)^f$, N^1H^b, $NH_2(a)^f$, and CO^b vibrational frequencies are missing. In the case of **CS** type dimer $NH_2(s)^b$, N^1H^f, $NH_2(a)^b$, and CO^b vibrational frequencies are not observed. We also would like to mention that obviously, all derivatives of 1-methyl-cytosine lack N^1H vibrational stretching frequency.

Conclusions

This study focuses on the interaction of cytosine and 1-methylcytosine homodimers in gas-phase, chloroform, DMSO, and in aqueous solutions. We found the following pattern of relative stabilities of dimers.

$$\text{In case of cytosine : } \mathbf{CS} \approx \mathbf{A} > \mathbf{B} > \mathbf{ST}$$
$$\text{In case of } 1-\text{methylcytosine : } \mathbf{MB} > \mathbf{MST} >> \mathbf{MA} > \mathbf{MCS}.$$

Attachment of methyl group to the N^1 atom of cytosine changes the stability order due to significant distortion of the geometry, compared to cytosine dimers. Strong polar solvents like water and DMSO decrease interaction energy between dimer components (up to 50 %), when compared to gas-phase. The comparison of experimental H-NMR shifts with the calculated ones suggests that the **CS** structure (see Fig. 1) is the one observed experimentally during the dimerization in DMSO.

Acknowledgments This work has been supported by the NSF CREST Interdisciplinary Center for Nanotoxicity, Grant # HRD-0833178. Extreme Science and Engineering Discovery Environment (XSEDE) computational facilities, supported by National Science Foundation grant number OCI-1053575 have been used for this project. Mississippi Center for Supercomputing Research (MSCR) is acknowledged for providing generous supercomputer facilities.

References

1. Mitra S, Hazra TK, Izumi T (2003) Encyclopedia of physical science and technology (Third Edition). Academic, New York, pp 853–876
2. Demeunynck M, Bailly C, Wilson WD (2003) DNA and RNA Binders, From Small Molecules to Drugs, vol 1, 2.Wiley-VCH, Weinheim
3. Hobza P, Šponer J (1999) Chem Rev 99:3247–3276
4. Gu J, Leszczynski J, Schaefer HF (2012) Chem Rev 112:5603–5640
5. Arshadi S, Bekhradnia AR, Ebrahimnejad A (2011) Can J Chem 89: 1403–1409
6. Kobayashi R (1998) J Phys Chem A 102:10813–10817
7. Šponer J, Riley KE, Hobza P (2008) Phys Chem Chem Phys 10: 2595–2610
8. Šponer J, Leszczyński J, Hobza P (1996) J Phys Chem 100:5590–5596
9. Cerny J, Hobza P (2007) Phys Chem Chem Phys 9:5291–5303
10. Jurečka P, Hobza P (2003) J Am Chem Soc 125:15608–15613
11. Jurecka P, Šponer J, Hobza P (2004) J Phys Chem B 108:5466–5471
12. Šponer J, Leszczynski J, Vetterl V, Hobza P (1996) J Biomol Struct Dyn 13:695–706
13. Fang Y, Bai C, Wei Y, Lin SB, Kan L-S (1995) J Biomol Struct Dyn 13:471–482
14. Pal SK, Zhao L, Xia T, Zewail AH (2003) Proc Natl Acad Sci U S A 100:13746–13751
15. Zendlová L, Hobza P, Kabeláč M (2007) J Phys Chem B 111:2591–2609
16. Fazaeli R, Monajjemi M, Ataherian F, Zare K (2002) J Mol Struct THEOCHEM 581:51–58
17. Sambrano JR, Souza ARD, Queralt JJ, Andrés J (2000) Chem Phys Lett 317:437–443
18. Sathyabama V, Anandan K, Kanagaraju R (2009) J Mol Struct 897: 106–110
19. Zhao Y, Truhlar DG (2005) Phys Chem Chem Phys 7:2701–2705
20. Grimme S, Antony J, Ehrlich S, Krieg HJ (2010) Chem Phys 132: 15410–15419

21. Grimme S, Ehrlich S, Goerigk L (2011) J Comput Chem 32:1456–1465
22. Zubatiuk TA, Shishkin OV, Gorb L, Hovorun DM, Leszczynski J (2013) Phys Chem Chem Phys 15:18155–18166
23. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Mennucci B, Petersson GA, Nakatsuji H, Caricato M, Li X, Hratchian HP, Izmaylov AF, Bloino J, Zheng G, Sonnenberg JL, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Vreven T, Montgomery Jr. JA, Peralta JE, Ogliaro F, Bearpark M, Heyd JJ, Brothers E, Kudin KN, Staroverov VN, Kobayashi R, Normand J, Raghavachari K, Rendell A, Burant JC, Iyengar SS, Tomasi J, Cossi M, Rega N, Millam JM, Klene M, Knox JE, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Martin RL, Morokuma K, Zakrzewski VG, Voth GA, Salvador P, Dannenberg JJ, Dapprich S, Daniels AD, Farkas Ö, Foresman JB, Ortiz JV, Cioslowski J, Fox DJ (2010) Gaussian 09, Revision C.01. Gaussian, Inc, Wallingford
24. Zhao Y, Truhlar DG (2008) Theor Chem Acc 119:525–525
25. Hobza P, Šponer J (2002) J Am Chem Soc 124:11802–11808
26. Boys SF, Bernardi F (1970) Mol Phys 19:553–566
27. Takano Y, Houk KN (2005) J Chem Theory Comput 1:70–77
28. Wolinski K, Hinton JF, Pulay P (1990) J Am Chem Soc 112:8251–8260
29. Shishkin OV, Gorb L, Leszczynski J (2000) J Phys Chem B 104: 5357–5361
30. Fogarasi G, Szalay PG (2002) Chem Phys Lett 356:383–390
31. Bazso G, Tarczay G, Fogarasi G, Szalay PG (2011) Phys Chem Chem Phys 13:6799–6807
32. Baklagina YG, Milevskaya IS, Minaev BF, Eisner YE (1968) Mol Bio 2:303–307
33. Baryshnikov GV, Minaev BF, Minaeva VA, Podgornaya AT (2012) Chem Heterocycl Compd 47:1268–1279
34. Shukla MK, Leszczynski J (2002) J Phys Chem A 106:11338–11346
35. Gorb L, Podolyan Y, Leszczynski J (1999) J Mol Struct THEOCHEM 487:47–55
36. Šponer J, Hobza P (1996) Chem Phys 204:365–372
37. Kelly REA, Lee YJ, Kantorovich LN (2005) J Phys Chem B 109: 22045–22052
38. Czyznikowska Z, Zalesny R (2009) Biophys Chem 139:137–143
39. Hobza P, Šponer J, Polasek M (1995) J Am Chem Soc 117:792–798
40. Amutha R, Subramanian V, Nair BU (2002) Theor Chem Acc 107: 343–350
41. Kabeláč M, Hobza P (2001) J Phys Chem B 105:5804–5817
42. Nir E, Huenig I, Kleinermanns K, Vries MSD (2003) Phys Chem Chem Phys 5:4780–4785
43. Kosenkov D, Kholod Y, Gorb L, Shishkin O, Hovorun DM, Mons M, Leszczynski J (2009) J Phys Chem B 113:6140–6150
44. Florián J, Šponer J, Warshel A (1999) J Phys Chem B 103:884–892
45. Potyahaylo AL, Samijlenko SP, Stepanyugin AV, Hovorun DM (2004) Proc SPIE 5507:190–194