J Mol Model (2013) 19:3447–3461
DOI 10.1007/s00894-013-1850-8

ORIGINAL PAPER

Stability and isomerization of complexes formed by metal ions and cytosine isomers in aqueous phase

Hongqi Ai · Jingjing Liu · Kwaichow Chan

Received: 10 January 2013 / Accepted: 4 April 2013 / Published online: 25 May 2013

Abstract We present a systematic study of the stability of the formation of complexes produced by four metal ions ($M^{+/2+}$) and 14 cytosine isomers (C_n). This work predicts theoretically that predominant product complexes are associated with higher-energy $C_4M^{+/2+}$ and $C_5M^{+/2+}$ rather than the most stable $C_1M^{+/2+}$. The prediction resolves successfully several experimental facts puzzling two research groups. Meanwhile, in-depth studies further reveal that direct isomerization of $C_1 \leftrightarrow C_4$ is almost impossible, and also that the isomerization induced by either metalation or hydration, or by a combination of the two unfavorable. It is the single water molecule locating between the H1(−N1) and O2 of the cytosine that plays the dual roles of being a bridge and an activator that consequently improves the isomerization greatly. Moreover, the cooperation of divalent metal ion and such a monohydration actually leads to an energy-free $C_1 \leftarrow C_4$ isomerization in the gas phase. Henceforth, we are able to propose schemes inhibiting the free $C_1 \leftarrow C_4$ isomerization, based purely on extended hydration at the divalent metal ion.

Keyword Complex stability · Cytosine · Isomerization · Metal ions

Electronic supplementary material The online version of this article (doi:10.1007/s00894-013-1850-8) contains supplementary material, which is available to authorized users.

H. Ai (✉) · J. Liu
Shandong Provincial Key Laboratory of Fluorine Chemistry and Chemical Materials, School of Chemistry and Chemical Engineering, University of Jinan, Jinan City 250022, People's Republic of China
e-mail: chm_aihq@ujn.edu.cn

K. Chan
Department of Natural Science, Albany State University, Albany, GA 31705, USA

Introduction

Deoxyribonucleic acid (DNA) is the main carrier of human genetic information [1]. Yet cations Na^+, K^+, Ca^{2+}, and Mg^{2+} can trigger proton transfer between bases of the nucleic acid, changing its configuration and functions. Many gene mutations and organism lesions derive from such changes [2]. At higher concentrations, interactions between metal ions and nucleic acid bases could break the hydrogen bond(s) of the base pairs, thus damaging the integrity of nucleic acid polymer structure [3]. Hence it is of great significance to keep the configuration of nucleic acid intact and to realize the normal physiological functions by reducing or preventing gene mutation caused by cations [4, 5]. As a part of the subject in maintaining normal physiological functions of the DNA, stability investigation on these bases and their metalated complexes have been receiving more and more attention [6–12].

As far as the isolated cytosine isomers were concerned, Kobayashi [6] predicted the stability ordering for thhe first five most stable isomers in the gas phase was $C_4 > C_5 > C_1 > C_2 > C_3$ obtained at the CCSD(T)/cc-pvtz(−f) level (see Fig. S1 in Support information (SI)), and cautioned that density functional theory (DFT) would offer an unreliable stability ordering for these isomers. Even so, both DFT and *ab initio* methods are employed routinely to investigate various properties of complexes produced by the interaction between cytosine isomer and metal ion in gas phase. For example, Vázquez and Martínez [7] probed the structures of both C_1 and C_4 and their interactions with divalent metal ions Ca^{2+}, Cd^{2+}, and Zn^{2+} in gas phase at the B3LYP/LANL2DZ level. They suggested that the most stable metalated complex (C_1M^{2+}) is derived from the third most stable cytosine isomer C_1. Russo et al. also confirmed that C_1 produced the most stable complex when it is bound by either a monovalent alkali metal ion (Na^+or K^+) [8] or a divalent alkali earth metal ion (Mg^{2+}or Ca^{2+}) using the B3LYP/6-311+G(2df,2p) method. Kabeláč and Hobza's study [9] at the RI-MP2/TZVPP level

predicted the strongest stability of $C_1M^{+/2+}$($M^{+/2+}$ = Na^+, Mg^{2+}, or Zn^{2+}) with the preferred bidentate O2⋯$M^{+/2+}$⋯N3 form and reported that their stability energy was 10 % stronger than other complexes. Šponer et al. [10] studied the interactions between Mg^{2+} and both N3 and N4 sites of C_1 in gas phase at the MP2/6-31G(d)//HF/6-31G(d) level. Vázquez and Martínez [11] probed the interaction between cytosine isomer and Al, Cu and Ag using the B3LYP and argued that the most stable anionic complexes were also in C_1M^- (M=Al, Cu, or Ag) form. Tureček and Yao [12] investigated hydrogen atom addition to both cytosine and cytosine-water (1:1 water adducts and 1:2 water adducts) at the B3-MP2/6-311++G(3df,2p) level, where the isomer C_1 of cytosine was employed. Summarily C_1 is the optimal isomer for metal ion to bind to no matter what valence (−1, 0, +1,+2) or type the metal is in the gas phase [7–12].

To our best knowledge few studies reported previously on the interaction between cytosine and metal ions in aqueous phase, it is the aqueous-phase surroundings, not the gas-phase, that approximate more closely the real physiological surroundings *in vivo*. Moreover, most of the derivatives of cytosine widely used experimentally are indeed produced in solution phase. For example, theoretical and experimental results have verified that the effects of both microsolvation and bulk solvation were substantial not only on the electronic structure and ionization of DNA bases in water [13] , but also on the stability ordering of the cytosine isomers [14, 15]. Thus our intention is to undertake a comprehensive study on complexes formed by cytosine and various metal ions ($M^{+/2+}$) Na^+, K^+, Ca^{2+}, and Mg^{2+} in aqueous phase, including the microhydrated environment and bulk solution. In this study, we will focus on not only the stability of these metal complexes but also the isomerization kinetics between C_1 and C_4 so that some experimental confusion can be understood and accounted for reasonably well. We hope that the proposed schemes, as presented here, have the potential to help regulate the gene mutation induced by base isomerization, surrounding factors, or their variant combinations.

Computational methods

Polarized continuum model (PCM) [16] in combination with default radius UA0 are adopted to simulate bulk aqueous environment, in which the dielectric constant is set to ε=78.4 due to its history of successful treatment on such systems [12–14]. All $C_nM_{I/II}^{+/2+}$ complexes in gas and aqueous phases are optimized separately and their frequencies are calculated at the B3LYP/6-31+G(d) basis set level [17]. Various energies of these $C_nM_{I/II}^{+/2+}$ complexes in both gas and aqueous phases are refined at the B3LYP/6-311++G(d,p) level. Some results sensitive to the B3LYP method for verification are further calculated at the SCS-MP2 [18] instead of the CCSD(T) level thanks to Grimme et al. [19] , who once pointed out that the results from SCS-MP2 calculations were very close to those from CCSD(T), but at only a tiny fraction of the computational cost. The corrections of zero point energies at 298 K are included for relative energies calculations. All electronic structure calculations are performed with the help of the Gaussian 03 suite of programs [20].

Results and discussion

Two different active sites on 13 cytosine isomers C_n (n=1, 2,..13) reported systematically by Kabeláč and Hobza [9] (Fig. S1) are chosen to interact with $M^{+/2+}$ respectively and the bound schemes are displayed in Fig. 1. For these $C_{n'}$ isomers (n′=2,..13), two stable forms of $C_nM_I^{+/2+}$ and $C_nM_{II}^{+/2+}$ can be generated (modes I and II). Herein, in principle only the more stable one of the two modes is displayed in the Table S1 of SI. For some cytosine isomers, such as C_2, …, C_5, one to two of their metalated complexes enjoy stability in gas-phase; if not in the gas phase, at least one will enjoy high degrees of stability in aqueous phase or vice verse. Thus 13+*m* complexes $C_nM_{I/II}^{+/2+}$ are discussed here. Their relative energies are listed in Table S1. Isomer C_x is also added in our discussion due to its lower stability in aqueous phase [15]. Thus metalated complexes presented in aqueous phase total to 14+*m*.

Choice of computational models and methods

First UA0, UAHF, BONDI, UFF, PAULING and KLAMT radii in combination with the PCM model are adopted to probe effects of different radii on the computational results. Then calculations employing both PCM model and default radius of UA0 are performed to compare effects of different models on the calculational results. In all these calculations, four complexes derived from the binding between C_1, C_2, C_3 or C_4 and Na^+ or Ca^{2+} are taken as models to test the applicability of selected computational schemes. The results are shown in Table S2. The comparisons reveal that the energy orders obtained from all above schemes for $C_nNa_{I/II}^+$ and $C_nCa_{I/II}^{2+}$ are C_1Na^+<$C_2Na_I^+$<$C_3Na_I^+$<$C_4Na_{II}^+$ and $C_1Ca_I^{2+}$< $C_2Ca_{II}^{2+}$< $C_3Ca_I^{2+}$< $C_4Ca_{II}^{2+}$, respectively, no matter what radius or solvent model are employed. By contrast, PCM/UA0-relative energies are the lowest in absolute value. Because PCM/UA0-relative energies of the most stable C_n isomers calculated are in excellent agreement with experimental results, see ΔEs in A1 column of Table 1, we report only PCM/UA0 results for the following discussions [15, 21].

The first six most stable cytosine C_n (n=1-5, *x*) isomers are employed to verify the applicability of B3LYP method in computing both the aqueous-phase stability of these cytosine isomers and their metalated complexes. Our B3LYP results in Table 1 show that the C_n stability order in gas phase

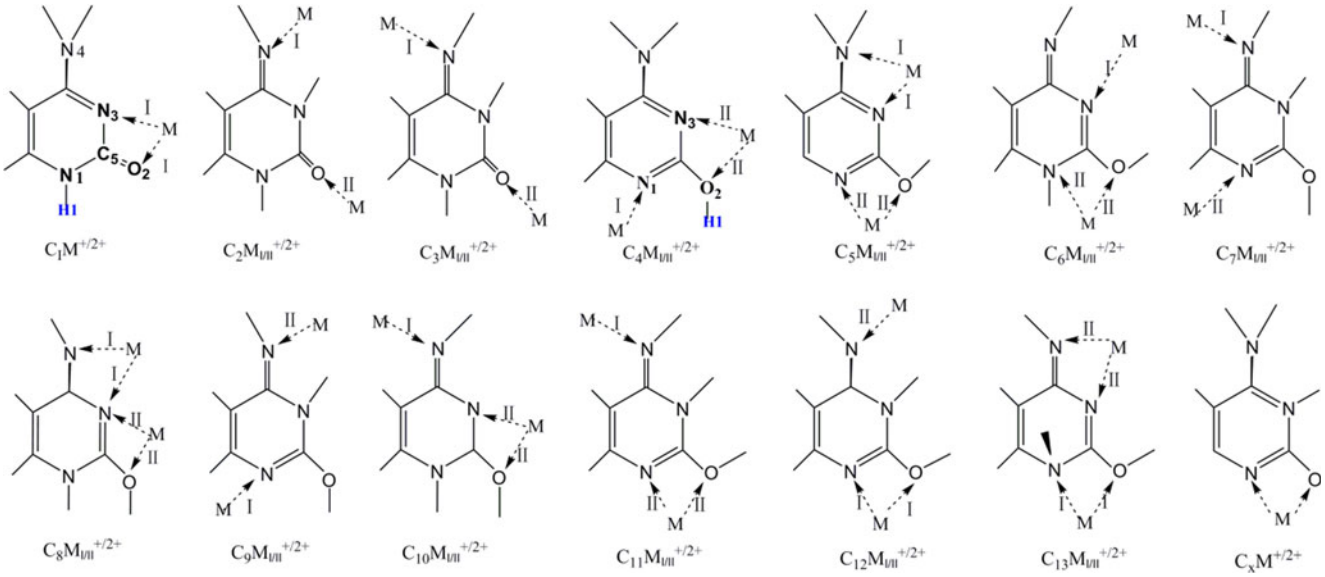

Fig. 1 Two binding modes (I/II) of metal ions in 14 C_n complexes $C_nM_{I/II}^{+/2+}$

is $C_1 > C_4 > C_5 > C_2 > C_3$, which is consistent with what were predicted by Russo et al. [8] at the B3LYP/6-311+G(2df,2p) level. Obviously, the order is not in line with the experimental observation, nor with the CCSD(T) results obtained by Kobayashi in 1998 [6]. Our gas-phase SCS-MP2 ordering $C_4 > C_5 > C_1 > C_2 > C_3$, no matter with and without ZPVE corrections, is indeed in good agreement with experimental observation and theoretical predictions by Kobayashi [6] and Trygubenko et al. [14] obtained at the CCSD(T) level. Here the SCS-MP2 results are obtained by using the cc-pVTZ basis set in order to compare with that of the CCSD (T) found in ref [6], in which a similar basis set (cc-pVTZ(−f)) was employed. Moreover, Kobayashi's study [6] revealed that, with the same basis set, the MP2 method could predict correct stability ordering of cytosine isomers in the gas phase, while DFT method could not. For the ordering in the aqueous phase, our B3LYP prediction is $C_1 > C_x > C_2 > C_3 > C_4 > C_5$, in excellent agreement with Sambrano et al.'s B3LYP-result [22] (column C) and Dreyfus et al.'s experiment [21] (column E). MP4 method with large basis set can also predict perfectly well [14, 23] the stability of C_2 relative to C_1 (columns D1, D2). In comparison with experimental values, the MP4-prediction for the C_x relative stability (23.4 [23] vs experimental 15.0 kJ mol^{-1} [21]) seems to have a larger discrepancy. Moreover, two MP4 results offer a reverse prediction for the stability ordering of C_4 and C_5 (columns D1, D2). Sambrano et al.'s MP2/6-31++G(d,p) results [22] indicated that C_x was less stable than the C_2 and C_4 in aqueous phase, which was also inconsistent with experimental observation [21]. Thus based on an overall consideration of various factors, the DFT method should do

Table 1 Relative energy (kJ mol^{-1}) of C_n in gas (ΔE_g) and aqueous (ΔE_s) phases

ΔE_g					ΔE_s					
	A1	A2	B	C2	A1	B	C	D1	D2	E
C_1	0	0	12.1(11.5)	5.2(5.2)	0	0(0)	0	0	0	0
C_2	8.7	9.6	21.1(22.3)	8.3(8.3)	22.3	27.4(28.5)	24.6	24.7	25.5	24.0
C_3	15.6		28.9(29.6)	14.1(14.1)	25.3	32.9(33.8)	26.9	30.5	30.1	
C_4	4.4	5.4	0(0)	0(0)	32.1	17.0(16.0)		26.8	28.5	27.5*
C_5	7.8	8.8	2.9(2.9)	2.7(2.7)	32.7	17.5(16.6)		24.3	29.7	
C_x	27.8				16.0		18.3		23.4	15.0

Note: Computations are performed at the following levels: A1: B3LYP/6-311++G(d,p)//B3LYP/6-31+G(d); B: SCS-MP2//CC-PVTZ//B3LYP/6-31+G(d); C: B3LY/6-31++G(d,p) from ref [22]. D1: MP4SDQ/CC-PVTZ//B3LYP/6-31+G(d)from ref [14]. A2, C2, D1, D2, and E were from ref [8] obtained at B3LYP/6-311+G(2df,2p) level, ref [6] at CCSD(T)//MP2/CC-PVTZ(−f) level, ref [14] by combining the gas-phase relative free energies and the average (three methods) hydration free energy difference, and ref [23] at MP4/6-311++G(d,p)//MP2/6-31G(d) level, as well as experimental determination[21], respectively. *Value is from ref [15] obtained at MP2/6-31G(d,p)//3-21G level in combination with SCRF model. Data in parenthesis are ZPVE-corrected results

better than the MP in predicting the aqueous-phase stability of the cytosine isomers. Due to our focus on the stability of complexes in aqueous phase, the above results encourage us to employ the B3LYP method to deal with metal complexes both from high accuracy in prediction and low computation cost points of view.

As for the complexes of the metalated cytosine isomers in the gas phase, both B3LYP and MP2 methods are frequently employed to predict their stability and the metal-cytosine interactions [7–11]. A more systematic study on the stability ordering of a series of metalated cytosine complexes was carried out by Kabeláč and Hobza [9] in 2006. As extensions of Kabeláč and Hobza's results, an up-to-date study on the gas-phase stability ordering of metalated cytosine complexes was performed by Kobayashi [24] , in which a previously not studied stable complex $C_xM^{+/2+}$ in the gas phase was predicted at a higher computational level (CCSD(T)) and an energetic inconsistency between the predictions of B3LYP and CCSD(T) was also argued. In such a situation, our gas-phase B3LYP results should be compared to the post-SCF [24] results with caution so that the aqueous-phase stability ordering of the metalated complexes can be credited correctly at the B3LYP level.

The results of the comparison are listed in Table 2, in which Russo and Toscano's data [8] are also included. The present stability order of $C_nNa_{I/II}^+$ in gas phase is $C_1Na^+ > C_xNa^+ > C_5Na_{II}^+ > C_4Na_{II}^+ > C_2Na_{II}^+ > C_2Na_I^+ > C_3Na_I^+ > C_{13}Na_{II}^+ > C_3Na_{II}^+ > C_8Na_I^+ > C_6Na_{II}^+ > C_5Na_I^+ > C_7Na_I^+ > C_9Na_{II}^+ > C_{10}Na_I^+ > C_{12}Na_I^+ > C_{11}Na_I^+ > C_{13}Na_{II}^+$ obtained at the B3LYP//6-311++G(d,p) level. The order is obviously in good agreement with those predictions of Russo and Toscano [8], obtained at the B3LYP/6-311+G(2df,2p) level, in which only the stability orderings of C_1Na^+/K^+, $C_2Na_I^+/K_I^+$, $C_4Na_{II}^+/K_{II}^+$, and $C_5Na_I^+/K_I^+$ were considered; and with that of Kobayashi, obtained at the B3LYP//cc-pvtz level [24], in which 25 stable structures were covered. In comparison with the post-SCF results of Kobayashi [24], our results agree satisfactorily both in magnitude and in the ordering in most cases, provided differences in basis set (Kobayashi's cc-pvtz vs 6-311++G(d,p)) and methodology (CCSD(T) vs B3LYP) are taken into account. Several minor discrepancies in the ordering do occur, a consequence of being disturbed by the stability changes of $C_8Na_I^+$ and $C_{13}Na_I^+$. The discrepancies derived from the CCSD(T) underestimates the energy (about 10 kJ mol^{-1}) of the two complexes (present and Kobayashi's own [24]); while B3LYP results overestimate the two values. It is interesting that other complexes are in good agreement in the ordering at both B3LYP and CCSD(T) levels. It is important to note that these two complexes ($C_8Na_I^+$ and $C_{13}Na_I^+$) hold poorer

Table 2 A comparison for relative energies (kJ mol^{-1}) of $C_nM^{+/2+}$ and C_n in gas phase obtained at the present B3LYP/6-311++G(d,p) level with available theoretical results

$M^{+/2+}$	Na^+				Ca^{2+}	Mg^{2+}	C_n	
	This work		Ref [24] CCSD(T)	B3LYP	ΔE_g	ΔE_g	ΔE_a	ΔE_g
$C_1M^{+/2+}$	0	cwcm	0	0	0	0	0	0
$C_xM^{+/2+}$	16.7	ct2b	20.5	17.6	−1.1,0.4[b]	−1.8,-0.4[b]	16.0	27.8
$C_2M_{II}^{+/2+}$	74.5	ct3ab	77.8	77.8	152.2,159.8[b]	-	22.3	8.7
$C_2M_I^{+/2+}$	81.1	ct3at	86.2	82.0	166.6	179.8,209.6[b]	-	9.6[a]
$C_3M_I^{+/2+}$	82.4	ct3bt	87.0	88.3	162.4,173.2[b]	172.6,200.4[b]	25.3	15.6
$C_3M_{II}^{+/2+}$	89.9	ct3bb	93.3	93.7	175.9			
$C_4M_{II}^{+/2+}$	55.1,59.8[a]	ct1bm	50.6	61.1	121.9,120.9[b]	126.7,133.5[b]	32.1	4.4,5.4[a]
$C_5M_{II}^{+/2+}$	37.2, 41.0[a]	ct1ab	32.6	41.4	76.1,73.6[b]	81.6, 86.6[b]	32.7	7.8[a]
$C_5M_I^{+/2+}$	122.2							8.8[a]
$C_6M_{II}^{+/2+}$	111.4	ct5ct	110.5	112.1	148.8	151.4	85.6	78.4
$C_7M_I^{+/2+}$	125.1	ct4dt	124.7	126.4	190.9	199.6	83.6	70.6
$C_8M_I^{+/2+}$	102.8	ct5bt	92.9	103.3	72.3,65.3[b]	50.9, 51.9[b]	100.5	134.9
$C_9M_{II}^{+/2+}$	127.5	ct4bt	127.6	128.9	202.8	213.6	79.5	59.0
$C_{10}M_I^{+/2+}$	146.7	ct5dt	145.6	147.3	178.8	181.2	96.8	116.9
$C_{11}M_I^{+/2+}$	166.1	ct4ct	165.7	166.5	236.7	245.4	94.4	106.3
$C_{12}M_I^{+/2+}$	155.6	ct4ab	146.9	160.7	232.3	246.1	90.2	91.2
$C_{13}M_{II}^{+/2+}$	87.4	ct5at	77.0	87.4	82.4,77.0[b]	59.9,62.3[b]	88.5	93.0
$C_{13}M_I^{+/2+}$	252.5							

[a] from ref [8] at the B3LYP/6-311+G(2df,2p) level. [b] from ref [24] at the CCSD(T)/cc-pvtz level. Data in red refer to those that disturb the ordering of $C_nM^{+/2+}$.

stability in the aqueous phase (Table S1) and will therefore not be included in our later discussions, thus their inconsistencies are irrelevant.

Likewise, the present stability orderings of selected $C_{n'}Mg^{2+}$ and $C_{n'}Ca^{2+}$ complexes (n′=1-5, 8, and 13) are in good agreement with those of CCSD(T) results, except the C_xMg^{2+}. To be chosen as a complex in the stability ordering, the complex must be in the top five in either the gas phase (<100 kJ mol^{-1}) or the aqueous phase (see the detailed stability ordering in the aqueous phase in Table S1). The selection finally covers seven dominant complexes in the stability ladder. These complexes are our targeted interests and hence discussions pervail in the following sections (listed in Table 3). As Kobayashi [24] once mentioned, B3LYP does underestimate the relative energy (−1.1 kJ mol^{-1}) of the C_xCa^{2+}; however, the underestimation does not affect the stability ordering of the C_xCa^{2+} in the aqueous phase since the energy difference between C_xCa^{2+} and C_1Ca^{2+} in aquous phase has extended to 14.0 kJ mol^{-1}. The presentation of B3LYP here is analogous to the predicted stability orderings of C_1 and C_4 in gas and aqueous phases mentioned above [6, 14, 21], which further supports the validity of the B3LYP method in predicting the stability of these $C_nM^{+/2+}$ complexes in aqueous phase.

Stability order of C_nM^+ complexes in aqueous phase

Now backtrack to our concerns regarding aqueous-phase stability of $C_nM^{+/2+}$ complexes. A detailed comparson of stability ordering in Table S1 reveals that all the top most stable metalated complexes in aqueous phase are derived from the following six cytosine isomers, C_1-C_5 and C_x. The only exceptions are C_8 and C_{13} bound by Mg^{2+}. Both $C_8Mg_I^{2+}$ and $C_{13}Mg_{II}^{2+}$ have competitive stability ordering among their isomers in aquoues phase and they even rank ahead of $C_4Mg_I^{2+}$. Thus Table 3 only covers the most stable $C_nM^{+/2+}$ complexes derived from C_1-C_5 and C_x, plus C_8, C_{13} (only for Mg^{2+} binding) in aqueous phase. The corresponding structures of $C_nM^{+/2+}$ complexes listed in Table 3 are dispayed in Fig. 2; data in the gas phase are also listed for the convenience of comparison. Results show that the stability ordering for these two kinds of monovalent metal-cytosine complexes is $C_1M^+ > C_xM^+ > C_2M_I^+ > C_3M_I^+ > C_5M_{II}^+ > C_4M_{II}^+$. The ordering only contains the more stable complex of two modes I & II. The order will alter and comprise new candidates based on the kind of metal ion involved.

As far as C_nNa^+ complexes, the new ordering follows $C_1Na^+ > C_xNa^+ > C_2Na_I^+ > C_3Na_I^+ > C_3Na_{II}^+ > C_2Na_{II}^+ > C_5Na_{II}^+ > C_4Na_{II}^+ > C_4Na_I^+$. It is interesting to note that both ordering and even magnitude of the first four most stable complexes are firmly in agreement with that of each of the corresponding cytosine reactants, C_1(0.0) > C_x (16.0) > C_2 (22.3)> C_3(25.3 kJ mol^{-1}). This result indicates that Na^+ binding hardly affects the ordering stability of the most stable cytosine isomers in aqueous phase and that Na^+ prefers binding amino nitrogen (N4) to carboxyl oxygen (O2) in reactants C_2 and C_3.

By contrast, there are a lot of changes in stability ordering for the last two metalated C_4 and C_5 complexes in aqueous phase. Firstly, C_4 and C_5 are first and second most stable isomers (Table 1), whereas C_1 only occupies third place in the stability ladder in gas phase [6, 14]. Solvent effect overturns the stability ordering and turns the order into C_1> C_4 > C_5 due to the larger dipole moment of C_1 in gas phase (6.8D in Table S3) [14, 22]. The favorable stability of C_4 and C_5 in gas phase is conducive to the stability of their metalated product complexes. Thus Table 3 shows their complexes do have stronger stability in aqueous phase. Different from C_1 and C_x, which offer bidentate coordination sites for metal ions and can produce only a single complex, C_4 or C_5 can generate two different complexes, denoted as modes I and II above.

Table 3 Relative energies (kJ mol^{-1}) of $C_nM^{+/2+}$ and C_n in aqueous (ΔE_a) and gas (ΔE_g) phases at the B3LYP/6-311++G(d,p) level

$M^{+/2+}$	Na^+		K^+		Ca^{2+}		Mg^{2+}		C	
	ΔE_a	ΔE_g	ΔE_a	ΔE_g	ΔE_a	ΔE_g	ΔE_a	ΔE_g	ΔE_a	ΔE_g
$C_1M^{+/2+}$	0	0	0	0	0	0	0	0	0	0
$C_xM^{+/2+}$	15.8	16.7	14.9	17.2	14.0	−1.1	13.2	−1.8	16.0	27.8
$C_2M_I^{+/2+}$	22.7	81.1	24.1	74.2	22.8	166.6	16.8	179.8	22.3	8.7
$C_2M_{II}^{+/2+}$	*30.3*	*74.5*	*24.4*	*62.3*	*24.6*	*152.2*	*36.5*	*202.8*		
$C_3M_I^{+/2+}$	25.4	82.4	27.2	76.9	25.6	162.4	22.1	172.6	25.3	15.6
$C_3M_{II}^{+/2+}$	*28.8*	*89.9*	*36.7*	*77.7*	*28.7*	*175.9*	*40.5*	*224.2*		
$C_4M_{II}^{+/2+}$	*39.3*	*55.1*	*35.6*	*50.8*	*55.1*	*121.9*	*98.6*	*126.7*	32.1	4.4
$C_4M_I^{+/2+}$	50.4	109.2	44.2	98.4	64.1	225.3	70.1	-		
$C_5M_{II}^{+/2+}$	*35.3*	*37.2*	*33.8*	*33.9*	*46.9*	*76.1*	*51.8*	*81.6*	32.7	7.8
$C_5M_I^{+/2+}$	55.7	122.2	-	-	-		-	-	-	
$C_8M_I^{+/2+}$					86.9	72.3	73.3	50.9	100.5	134.9
$C_{13}M_{II}^{+/2+}$					*78.6*	*82.4*	*65.8*	*59.9*	88.5	93.0

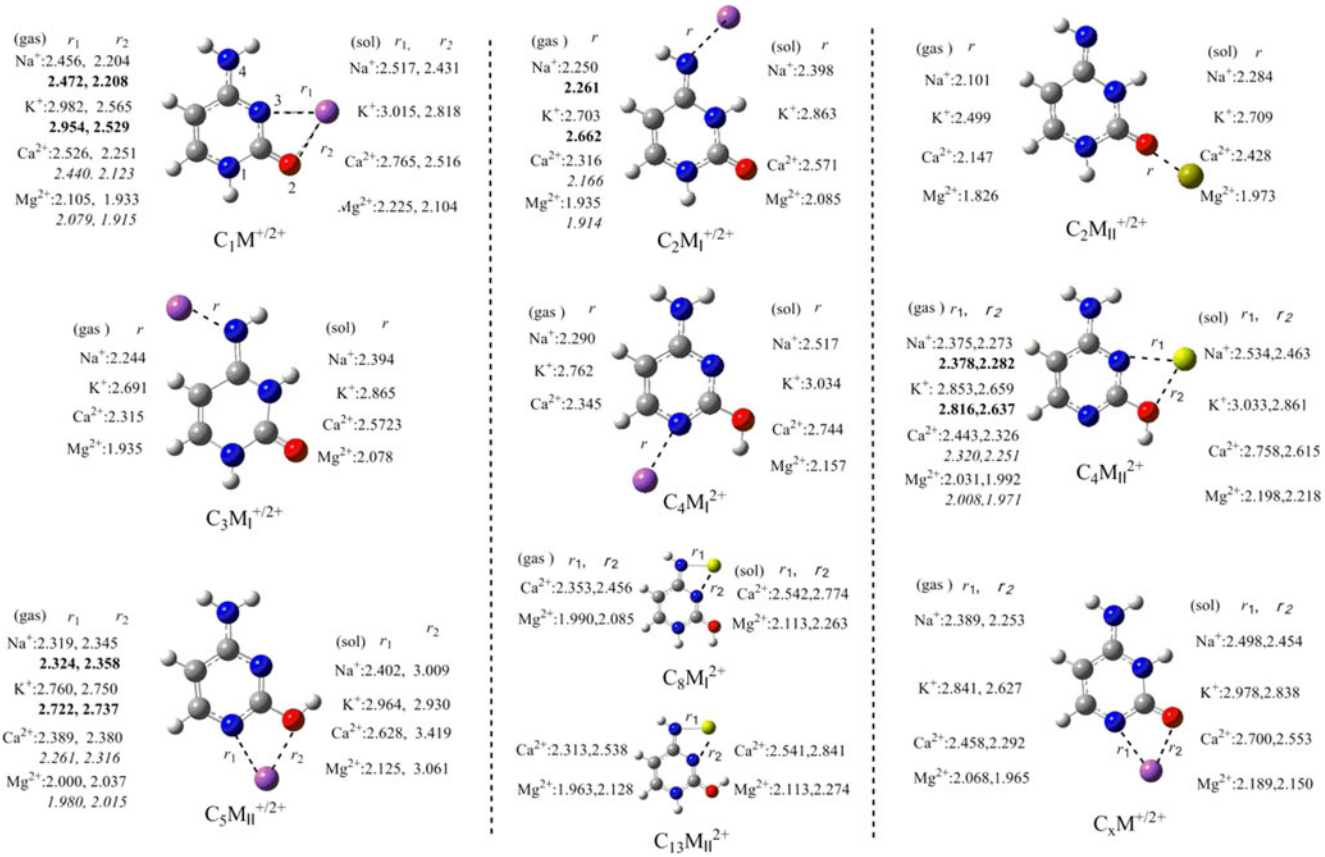

Fig. 2 Optimized geometries and selected bonding distances, r, r_1 and r_2 of $C_{n'}M^{+/2+}$ (n'=1-5, 8, 13). Distance in Å. Data in black from ref [8], in italic from ref [29]

The structural difference between C_4 and C_5 lies in the orientation of the hydroxyl hydrogen. For C_5, Na^+ can approach it from either I or II direction (see $C_5M_{I/II}^{+/2+}$ in Fig. 1) and create the bidentately coordinated $C_5Na_{II}^+$ (Fig. 2) or $C_5Na_I^+$ (not shown due to poorer stability). Table S1 shows that $C_5Na_I^+$ (122.2 kJ mol^{-1}) has over three times more energy than $C_5Na_{II}^+$, and over two times more than $C_4Na_{II}^+$ in gas phase; hence the solvent effect can greatly increase the stability of the complex in aqueous phase, due to its larger dipole moment (10.5D). Even so, the relative energy of $C_5Na_I^+$ is still the highest (55.7 kJ mol^{-1}) among these $C_nNa_{I/II}^+$ complexes (n=1-5, x) in aqueous phase. An observation for $C_5Na_I^+$ geometry (not shown) reveals that the amino hydrogens of C_5 isomer twist and deform after the metal ion approaches the amino group that leads to poorer stability of this complex [9, 24] in both phases. Because of their poor stability, $C_5M_I^{+/2+}$ complexes will not be discussed any further. For C_4, Na^+ can also bind to it from two different directions (see $C_4M_{I/II}^{+/2+}$ in Fig. 1) and generate three different end products. They are a bidentate $C_4Na_{II}^+$ from the II direction, and two $C_4Na_I^+$ complexes from the I direction. Figure 2 shows that one $C_4Na_I^+$ complex holds monodentate geometry (still nominated as $C_4Na_I^+$), yet another degenerates into bidentate $C_5Na_{II}^+$. Similar to the case of $C_5Na_I^+$, the monodentate $C_4Na_I^+$ has higher relative energy (109.2 kJ mol^{-1}, in Table 3), but also larger dipole moment (10.4D in Table S3) in gas phase, and as a result, it has a stronger stability (50.4 kJ mol^{-1}) in aqueous phase. As before, the relative energy of $C_4M_I^+$ with other ions involved in gas phase and in aqueous phase is also listed in Table 3. The table shows that the stability ordering is $C_5M_{II}^{+/2+}$ > $C_4Na_{II}^+$> $C_4Na_I^+$ in both gas and aqueous phases, indicating that a metal ion prefers binding in the II orientation of C_5 isomer to that of C_4. That is, the bidentate coordination at the I orientation of C_4 is more favorable to the stability of products than either monodentate binding in the same orientation or the bidentate coordination at the II orientation of C_4 in light of thermodynamics.

The above observations, especially the degenerated phenomenon of $C_4Na_I^+$, can be employed to clarify the inconsistency between the theoretical prediction and experimental metal ion affinity (MIA). Experimental determination from Cerda and Wesdemioti [25] revealed that MIA of Na^+-cytosine was 177.0 kJ mol^{-1}. Russo et al.'s computation [8] implied that the MIA mainly derived from Na^+-C_5($C_5Na_{II}^+$, 176.6), instead of Na^+-C_4($C_4Na_{II}^+$, 163.2), Na^+-C_1(C_1Na^+, 209.6), or Na^+-C_2(C_2Na^+, 135.1). (Note distinguishing the different nomenclatures for identical complexes presented in Fig. 1

and in ref [8].) Normally, the experimental MIA derives mainly from the most favorable complex C_1Na^+ thermodynamically. Facing the discrepancy, Russo et al. met with the challenge to rationalize it and pinned their hopes on the supposal of the presence of the C_4 isomer in the experimental environment. In fact, both theoretical [6, 14] and experimental [21] measurements revealed that C_4 was not only indeed present in the experimental environment but also was the most stable isomer, for which Gorb et al. [26] estimated the relative concentration of cytosine C_4 to account for more than half the overall population of cytosine isomers (65.7 %). Lately, a more accurate estimate from Bazsó et al. in 2011 [27] was as follows: 0.22: 0.26 : 0.44 : 0.08 for the mole ratio of C_1 : C_5 : C_4 : C_2. More than ten of theoretical and experimental measurements for the ratio from other research groups were compared and assessed in full detail in the Bazsó et al.'s paper. A conclusion can be drawn that C_4 and C_5 isomers are the first and second abundant complexes that interact with other reagents, such as metal ions. Now this invites a more interesting question. Why is it that the MIA of the C_5 complex $C_5Na_{II}^+$ [176.6 kJ mol^{-1}] is much more closer to the experimental measure (177.0) than that of C_4 complex $C_4Na_{II}^+$ [163.2]? A plausible explanation is offered by Russo et al. [8]that overcoming the 65.3 kJ mol^{-1} energetic barrier for Na^+-O-H group rotation is the key to facilitate the conversion between $C_4Na_{II}^+$ and $C_5Na_{II}^+$. The process to facilitate the conversion is feasible only at high temperature, however. Our gas-phase $C_4Na_I^+$ optimization processes (Fig. 3) also advocates the existence of other more competitive options. First, as the most stable isomer, C_4 has the most population density interacting with Na^+. If Na^+ approaches C_4 in the I direction, then what the final product is, $C_4Na_I^+$ or $C_5Na_{II}^+$, depends on the Na^+ location. Figure 3A exhibits the optimization process of $C_4Na_I^+$, in which Na^+ is positioned on the same plane as the C_4 ring and pointing to N1 (refer to the atomic labels in Fig. 1), and the final monodentate *Cs* structure is formed. From the aspect of reaction kinetics, Fig. 3A indicates that the generation process of $C_4Na_I^+$ is energy-free. Thermodynamically, the energy of $C_4Na_I^+$ is 109.2 kJ mol^{-1} more than the most stable $C_1Na_I^+$, and 95.0 kJ mol^{-1} more than $C_5Na_{II}^+$, thus rejecting the product as stability favorable and implying the existence of other options. If Na^+ is positioned at the same orientation but staying directly over or below the plane of the C_4 ring (note such possibility is far higher than above planar, see the first geometry in Fig. 3B), then the final product will degenerate into bidentate $C_5Na_{II}^+$, in which a hydroxyl hydrogen rotates to the opposite direction along with the approaching metal ion. Figure 3B clearly delineates that the potential energy surface is smooth and that the ceaseless adjustment of the metal ion is energy free along with the rotation of the hydroxyl hydrogen, except for four high-energy states. A detailed analysis discerns that the first high-energy state (step 16) derives mainly from the compression of the angle ∠N3C5O2, i.e,. from 114.6° (in step 15) to 111.5° (step 16); whereas the latter three states (steps 21, 34 and 39) originate the compressions of the angle ∠N1C5O2, from 116.5° (front step 20) to 111.5° (step 21), from 112.4° (step 33) to 107.6° (step 34) and from 113.7° (step 38) to 109.2° (step 39), respectively. The highest energy difference lies between step 16 and step 15 with about 0.018 a.u. Likewise, two similar processes of mode I can also be observed in the generation of $C_4K_I^+$ and the conversion of $C_4K_I^+ \rightarrow C_5K_{II}^+$. In contrast, the conversion of $C_4K_I^+ \rightarrow C_5K_{II}^+$ experiences a huge energetic barrier of about 0.185 au, which rises about ten-fold of that in $C_4Na_I^+ \rightarrow C_5Na_{II}^+$ conversion. The barrier is also mainly a result from the compression of the angle ∠N1C5O2 (step 31 in Fig. 3C). The generation of both $C_4Na_{II}^+$ and $C_4K_{II}^+$ enjoys barrier free processes, confirmed by the $C_4K_{II}^+$ example in Fig. 3D. The results signify that the Na^+-cytosine product with the greatest abundance should be in $C_5Na_{II}^+$ form, which covers the products of accessible conversion of $C_4Na_I^+ \rightarrow C_5Na_{II}^+$ and of natural reaction of C_5+Na^+ with mode II. K^+-cytosine product, however, should be in $C_4K_{II}^+$ form due both to the greatest abundance of reactant C_4 [6, 14, 21, 27] and the difficulty in $C_4K_I^+ \rightarrow C_5K_{II}^+$ conversion kinetics, although the product $C_5K_{II}^+$ is fairly stable. It is worth noting that the optimization process strongly depends on the structures of initial reactants and their relative positions. Anyway, the process of compressing the angle ∠N1C5O2 must occur during Na^+/K^+-cytosine interaction because the angle in the free C_4 (116.7°) is larger than those in the $C_5Na_{II}^+/K_{II}^+$ (112.5°/113.7°), indicating that the energetic barrier originated from the compression has to be overcome for the reaction to happen. The above analysis manifests why experimental MIA [25] of cytosine-Na^+ complex corresponds to that of $C_5Na_{II}^+$, whereas that of cytosine-K^+ complex corresponds to $C_4K_{II}^+$ instead of $C_5K_{II}^+$ [8].

Our present results can account for why the experimental measure for the MIA of K^+-cytosine complex from Cerda and Wesdemiotis [25] is 110.kJ mol^{-1}, about 25 kJ mol^{-1} less than that from Yang and Rodgers [28] , which throw Yang and Rodgers at a loss. As designed by Yang and Rodgers, two types of transition states (TSs) for dissociation $C_5K^+ \rightarrow C_4+K^+$ were considered (corresponding to $M^+(C_3) \rightarrow M^+ + C_2$ in ref [28]). One is a phase space limit (PSL) for which they assumed that the TSs were loose and product-like because the interaction between the alkali metal cation and cytosine was largely electrostatic. Another is called the tight transition state (TTS) for which they applied the parameters of TTSs for unimolecular tautomerization of the M^+-cytosine complexes determined from theoretical calculations. Then they measured two MIAs for the process with 113.6±4.3 (TTS) and 136.0± 3.6 kJ mol^{-1} (PSL), respectively. Obviously, the TTS result, not the PSL, coincides well with 110.kJ mol^{-1} determined by Cerda and Wesdemiotis [25] experimentally. As displayed in Fig. 3D, TTS corresponds to the orientation II (product: $C_4K_{II}^+$)

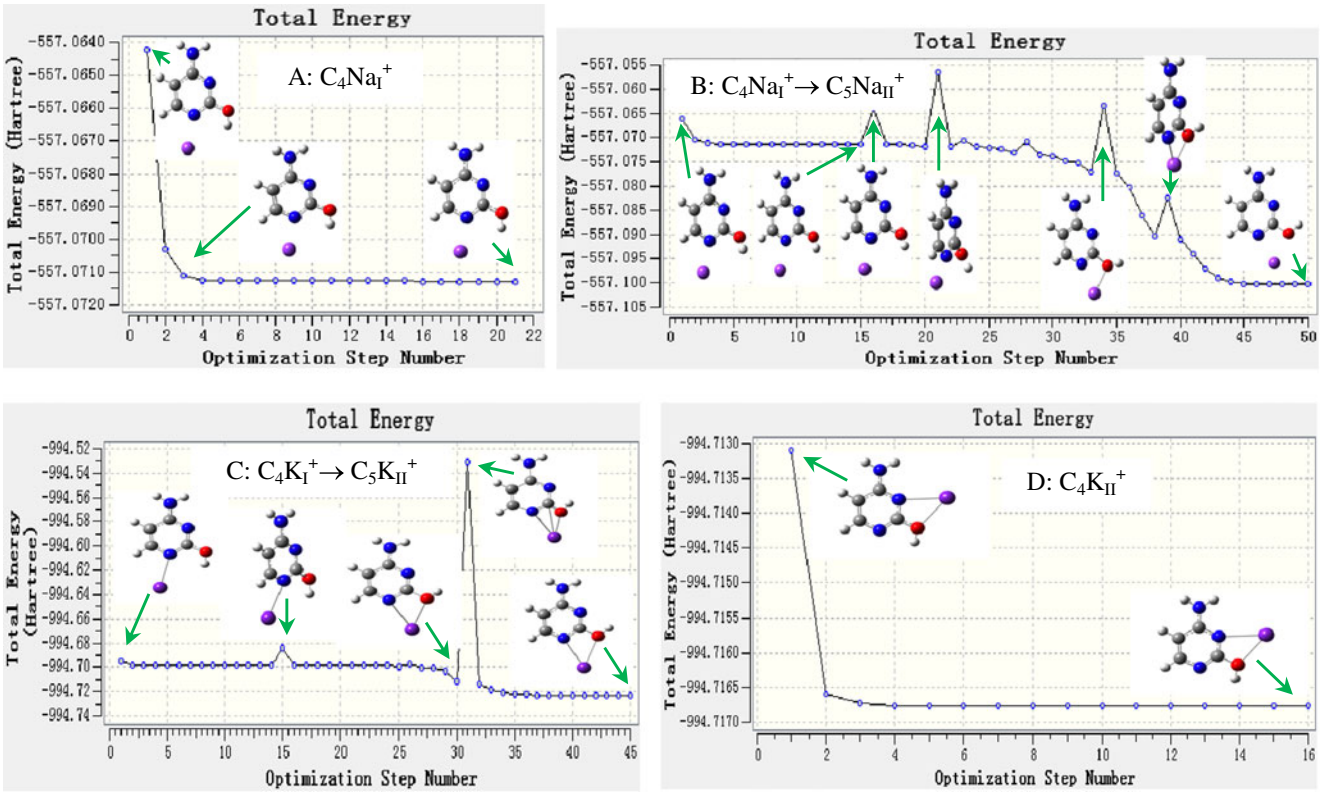

Fig. 3 Optimization processes and selected geometrical structures of $C_4Na_I^+$(A & B), $C_4K_I^+$ (C), and $C_4K_{II}^+$ (D) obtained at the B3LYP/6-31+G(d) level. Metal ion in pink, nitrogen in blue, carbon in gray, hydrogen in white and oxygen in red

mentioned above, in which the structural parameters of C_4 should alter inconspicuously during the association and dissociation of $C_4K_{II}^+$, i.e., $C_4K_{II}^+ \leftrightarrow C_4+K^+$ (1). Their PSL results however should correspond to the orientation II with the dissociation process of $C_5K^+ \rightarrow C_4+K^+$(2) and MIA=136.0±3.6 kJ mol^{-1}, in which C_4 transforms into C_5 during the course when metal ion approaches C_4 following orientation II, as shown in Fig. 3C. Then a problem arises. Which MIA will represent the real reaction? Yang and Rodgers [28] argued that process (2) should play a dominant role in the cytosine-K^+ association or dissociation process, which is inconsistent with the report of Cerda and Wesdemiotis [25]. There exists a counter-example in Yang and Rodgers' [28] measurement, which can be employed to support the experimental result of Cerda and Wesdemiotis. Yang and Rodgers also examined their experimental MIA result of $C_4K_{II}^+ \rightarrow C_4+K^+$ process with theoretical method and showed that their PSL-experimental MIA was 133.1± 3.5 kJ mol^{-1}, higher than their theoretical prediction of 115.1 kJ mol^{-1}. Their theoretical result is obviously more credible, as they are confirmed by both prior theoretical [8] treatment (108.8) and experimental [25] (110.0) reports. Thus their TTS result (113.6±4.3 kJ mol^{-1}) rather than the PSL in process (2) should be more plausible in this case. Anyway, an acceptable explanation for the two experimental phenomena mentioned above is based only on the hypothesis that the metalated C_4 product is predominant.

Here the common $C_4M_I^{+/2+}$and $C_5M_{II}^{+/2+}$ presentations are employed because besides Na^+ and K^+, Ca^{2+} cation also generates other similar products and bestows comparable reaction process (not shown) when it is bound to C_4. The selected geometrical parameters are also shown in Fig. 2. Only Mg^{2+}-bound C_4 in mode I creates $C_5Mg_{II}^{2+}$ directly, possibly due to stronger charge effect of Mg^{2+}. Russo et al. [8] also reported the gas-phase distances between Na^+/K^+ and the coordination sites of C_1, C_4 and C_5, respectively, shown in black in Fig. 2, are in excellent agreement with our results. The counterpart distances in aqueous phase, compared to those in the gas phase, is extended somewhat, indicating that the solvent effect not only alters the stability of these complexes but also their geometrical structures.

For the stability of complexes involving K^+ ions, a comparison reveals that the order becomes $C_1K^+ > C_xK^+ > C_2K_I^+ > C_2K_{II}^+ > C_3K_I^+ > C_5K_{II}^+ > C_4K_{II}^+ > C_3K_{II}^+ > C_4K_I^+$ in aqueous phase. The ordering still keeps $C_1K^+ > C_xK^+ > C_2K_I^+ > C_3K_I^+ > C_5K_{II}^+ > C_4K_{II}^+$ if the high-energy mode of each complex is excluded. The energy difference between $C_4K_I^+$ and $C_5K_{II}^+$ is 64.6 kJ mol^{-1}, 20.4 kJ mol^{-1} less than the difference between $C_4Na_I^+$ and $C_5Na_{II}^+$ in gas phase, implying the disadvantage of the $C_4K_I^+ \rightarrow C_5K_{II}^+$ conversion

thermodynamically, relative to the $C_4Na_I^+ \rightarrow C_5Na_{II}^+$ conversion. Taken in this sense, the contribution to the product $C_5K_{II}^+$ from the reaction C_4+ K^+ in mode I should be less, which also partly accounts for why $C_4K_{II}^+$ is the predominant product in the experimental reaction of cytosine+ K^+, whereas the $C_5Na_{II}^+$ is the predominate product of reaction of cytosine + Na^+[8,25]. The present results favors the viewpoint that the main product of reaction cytosine + M^+ corresponds to $C_5M_{II}^+$ (for M=Na) or $C_4M_{II}^+$ (for M=K), instead of the mixture of any two or more stable complexes, which is deduced from the full agreement with experimental measure for the MIA of the product of cytosine + M^+ with that calculated for $C_5M_{II}^+$ (M=Na) or $C_4M_{II}^+$ (M=K).

Stability order of C_nM^{2+} complexes in aqueous phase

Similar to those mono-valent metalated cytosine complexes, the stability ordering of Ca^{2+}, Mg^{2+}-involved cytosine isomers also preserves $C_1M^{2+} > C_xM^{2+} > C_2M_I^{2+} > C_3M_I^{2+} > C_5M_{II}^{2+} > C_4M_{I/II}^{2+}$. The last complex $C_4M_{I/II}^{2+}$ denotes $C_4M_I^{2+}$ (only M=Mg) and $C_4M_{II}^{2+}$ (only M=Ca) respectively. As for the univalent $C_4Na_I^+/K_I^+$, C_4 bound by Ca^{2+} in mode I in gas phase also holds monodentate $C_4Ca_I^+$ form. Thus it can be argued that Ca^{2+} involved cytosine complexes should have properties comparable to the univalent $C_4M_I^+$ both in stability ordering and geometric structure. In contrast, Mg^{2+} involved cytosine complexes indeed possesses several diverse features.

First, C_xMg^{2+} instead of C_1Mg^{2+} becomes the most stable complex in the gas phase. As well as for the C_xCa^{2+}. In light of the CCSD(T) result, $C_xCa^{2+}$with lower stability than C_1Ca^{2+} maybe derived from insufficient calculation of the B3LYP [25]. From this perspective, Mg^{2+} is thus the only one that can induce the stability of C_x over C_1. In aqueous phase, all C_1 complexes bound by the four different ions become the most stable ones, so is the reactant C_1 in the aqueous phase, indicating that the solvation effect can greatly "trowel" the difference originated from the distinction of ionic charge or ionic kind. Second, only C_8 and C_{13} bound by Mg^{2+} become the third and fourth most stable complexes in the gas phase, and the stability of the two complexes C_8Mg^{2+} and $C_{13}Mg^{2+}$ is also better than the C_4Mg^{2+} in the aqueous phase. Third, only Mg^{2+} can cause the C_4 in the I orientation ($C_4Mg_I^{2+}$) degenerating into a C_5 (i.e., $C_5Mg_{II}^{2+}$) spontaneously in gas phase. The $C_4M_I^{+/2+}$ complexes involving the other three ions can still remain in the monodentate structures (see $C_4M_I^{+/2+}$ in Fig. 2), however. In aqueous phase, $C_4Mg_I^{2+}$ recovers to the monodentate state, as three other $C_4M_I^{+/2+}$ complexes do. We would attribute the special features of these Mg^{2+} involved complexes to Mg^{2+}'s stronger binding ability to the ligand. This viewpoint can be strongly supported by shortening Mg^{2+}-cytosine distances than the corresponding counterparts of $M^{+/2+}$-cytosine (M=Na, K, Ca) in Fig. 2, obtained both from present calculations and from Russo et al. [29] . Moreover, Russo et al. had reported theoretically that the Mg^{2+}-cytosine binding was stronger by about 30 % than that of the corresponding Ca^{2+}-cytosine binding. For example, the MIAs of C_1Mg^{2+}, $C_4Mg_{II}^{2+}$and $C_5Mg_{II}^{2+}$ were predicted to be 187.4, 156.1 and 167.6 kcal mol^{-1}, respectively, far stronger than the corresponding counterparts of cytosine-Ca^{2+} (140.1, 110.1 and 122.1) [29].

Isomerization of $C_1 \leftrightarrow C_4$ induced by metal ion

For the most stable isomer of cytosine in gas phase, C_4 [6, 14, 21, 27], and the most stable isomer of cytosine in aqueous phase, C_1 [15, 21], their structural distinction lies at the different locations of hydrogen. If the hydrogen locates at the N1 site, then it is the isomer C_1; if it transfers to the neighboring O2 site, then C_1 becomes C_4. Matrix isolation infrared studies revealed that isolated cytosine placed in an inert environment exhibited an interesting $C_1 \leftrightarrow C_4$ tautomerism. On the other hand, it is well established that cytosine adopts predominantly the amino-oxo form C_1 in aqueous phase or in crystal [15, 21]. A small contribution from the imino-oxo and imino-hydroxyl tautomers to cytosine population had also been suggested in aqueous phase [27]. Another factor encouraging us to consider $C_1 \leftrightarrow C_4$ isomerization is that the C_1M complex is the most stable one in both gas [7–12] and aqueous phases, and that C_4M_{II} is the fourth most stable structure in aqueous phase. Due to the largest ratio of the C_4 [27, 28] in the gas phase, its metalated products $C_4M_{II}^{+/2+}$ and $C_5M_{II}^{+/2+}$ had also been confirmed to be predominant experimentally [8] although the stability of these complexes was not the most stable. One of our aims is to investigate the potential of point mutation induced by surrounding factors such as metal ion binding and aqueous effect. Thus the cooperation of both the metal ion binding and aqueous effect could be a significant factor to boost the $C_1 \leftrightarrow C_4$ tautomerism.

Fig.S2 displays the isomerization between C_1 and C_4($C_1 \leftrightarrow C_4$) induced by four different ions and in both gas and aqueous phases. Results show that the direct isomerization between C_1 and C_4 is hardly performed due to higher activation energy barrier (143.0 kJ mol^{-1} in $C_1 \rightarrow C_4$ process (i) and 138.6 kJ mol^{-1} in $C_1 \leftarrow C_4$ process (ii)) in the gas phase. Kosenkov et al. [30] and Gorb [26] also confirmed the result and once predicted the activation energies of reaction (i) were 142.8 and 138.1 kJ mol^{-1}, respectively. Moreover, Kosenkov et al. [30] observed that increased temperature did not improve the reaction in kinetics. Our investigations find that aqueous effect also aggravates such an already unfavorable isomerization process. Metalation at both N3 and O2 sites of the cytosine isomers C_1 and C_4 leads to the isomerization processes of $C_1M \leftrightarrow C_4M$ more arduously, confirmed by the increasing activation energies. Moreover, the more charges of the metal ion, the greater are the increases in activation

energies. The conclusion can be manifested from two aspects. First activation energies in divalent $C_1M \leftrightarrow C_4M$ processes (e.g., $C_1Mg^{2+} \leftrightarrow C_4Mg^{2+}$) are indeed higher than that in the monovalent$C_1M \leftrightarrow C_4M$ processes (e.g., $C_1Na^{+} \leftrightarrow C_4Na^{+}$). Another is the fact that metal ion with identical valence but a different radius induces different activation energies. The larger the ion radius, the lower the charge density. Consequently, the process with larger ion involved holds lower activation energy. Obviously, the phenomenon occurs in both gas and aqueous phases; however, the hydration disfavors the isomerization process.

Bridge effect of monohydration on the isomerization of $C_1M \leftrightarrow C_4M$

To investigate the hydration effect on the isomerization of $C_1M \leftrightarrow C_4M$ process, one water molecule (W) is situated between H1(−N1) and O2. The water molecule positioned here is expected to play a bridge role for transferring H1 between N1 and O2, and then the product will switch between C_1MW and C_4MW. Figure 4 displays these isomerization processes between C_1MW and C_4MW. Results show that the monohydration greatly reduces the activation energies of the $C_1 \leftrightarrow C_4$ and $C_1M \leftrightarrow C_4M$ processes, arguably a great improvement of the isomerization kinetics. For example, in $C_1 \rightarrow C_4$ process the improvements of activation energy are 95.1 and 151.3 kJ mol^{-1}, respectively, in gas phase and aqueous phase. For $C_1M \rightarrow C_4M$ process, the improvements are 113.0 and 187.2 kJ mol^{-1}, respectively, in gas phase and aqueous phase for the $C_1Na^{+} \rightarrow C_4Na^{+}$ process, whereas the counterparts for $C_1Ca^{2+} \rightarrow C_4Ca^{2+}$ are 138.0 and 273.0 kJ mol^{-1}, respectively. In the reverse process, the improvement becomes more remarkable. For instance, the activation energy in aqueous-phase $C_1Ca^{2+} \leftarrow C_4Ca^{2+}$ process is 283.7 kJ mol^{-1}, but in $C_1Ca^{2+}W \leftarrow C_4Ca^{2+}W$ process it decreases greatly to 21.5 kcal mol^{-1}. The above results manifest that the improvement effect derived from the monohydration is better in the processes involving monovalent metal ions over non-metal ions, and in divalent metal ion over monovalent metal ions, as well as in aqueous-phase over the gas-phase. And interestingly, gas-phase $C_1Ca^{2+}W \leftarrow C_4Ca^{2+}W$ process is energy free if the ZPVE is included, indicating a strong tunnel effect. Likewise, $C_1Mg^{2+}W \leftarrow C_4Mg^{2+}W$ process is also an energy free process with or without ZPVE, implying the monohydration transforms the C_4Mg^{2+} into C_1Mg^{2+} spontaneously (shown in Fig. 5). In bulk aqueous phase, the activation energy is only 17.0 kJ mol^{-1}, a rather minor barrier.

The role of the water molecule in $C_1 \leftrightarrow C_4$ and $C_1M \leftrightarrow C_4M$ processes can be analyzed by careful observation. Herein we prefer choosing the structures of TSs to discuss the probability of $C_1 \leftarrow C_4$ isomerization induced by the M-W cooperative factors. In TSs of $C_1M \leftrightarrow C_4M$ processes, the proton H1 occupies a medial position between N1 and O2 and an analysis on TS vibration mode shows that the proton with the virtual frequency mode does move between N1 and O2, as a pendulum swings back and forth. In the TSs of $C_1MW \leftrightarrow C_4MW$ processes including that of $C_1W \leftrightarrow C_4W$, H1 transfer depends on the carrier-water molecule to complete $C_1 \leftrightarrow C_4$ isomerization. Figure 4 clearly illustrates that both H1 and one hydrogen of the water in all but Ca^{2+}-involved gas-phase TSs delineate a process with concerted proton transfer. Such processes experience lower activation energy. In contrast, H1(−O2) in $C_4Ca^{2+}W$ will transfer and approach first to the water molecule, and then generate a H_3O^{+}. Finally, the H_3O^{+} releases a proton to N1 site and produces $C_1Ca^{2+}W$. Obviously, in the $C_1Ca^{2+}W \leftarrow C_4Ca^{2+}W$ process, the transfer of both H1 and one hydrogen atom of the H_3O^{+} is ordinal instead of concerted. The 0.996 Å O-H distance of water molecule of TS in $C_1Ca^{2+}W \leftarrow C_4Ca^{2+}W$ process expands on that the water molecule in such case almost keeps its structure and confirms the above prediction. In fact, the sequential proton transfer process can be observed more obviously from the gas-phase isomerization of $C_1Mg^{2+}W \leftarrow C_4Mg^{2+}W$ in Fig. 5. Moreover, the two divalent metal ion-involved $C_1MW \leftarrow C_4MW$ processes are energy free ones, highlighting the significance of the divalent charge.

In aqueous phase, both H1 and one hydrogen of the water in all these TSs are in active states, signifying that the transfers of two protons are concerted. The phenomenon manifests that the solvent effect also plays a key role in such an isomerization process. Figure 6 shows that it is more ready for the isomerization trend of $C_1MW \leftarrow C_4MW$ process along with the increase of charge density over the metal ion in gas phase. The trend doesn't change but becomes gentler in aqueous phase.

In addition, the applicability of B3LYP to these monohydrates has been assessed by comparing with the SCS-MP2 results. As a comparable method to the CCSD(T) qualitatively and very close to the later in predicting the isomerization energies of the organic molecules [19], SCS-MP2 is employed again here to perform single-point calculations with the 6-311++G(d,p) basis set on the basis of the B3LYP/6-31+G(d) geometries for some selected complexes. The relative energies are also listed in black in Fig. 4.

The C_1W and C_4W complexes are selected for both gas and bulk aqueous phases; and their corresponding metalated complexes $C_1Na^{+}W$ and $C_4Na^{+}W$ occur only in gas phase. The SCS-MP2 result shows that the monohydrated $C_4(C_4W)$ is still (−2.6 kJ mol^{-1}) more stable than monohydrated $C_1(C_1W)$ in gas phase; but in the bulk aqueous phase, the B3LYP-result reveals that C_4W is less stable than the C_1W. Gorb et al.'s relative energies ΔE [26] ($\Delta E = E_{C4W} - E_{C1W}$) were −2.1 and 7.9 kJ mol^{-1}, obtained at MP2/6-311++G(d,p)//MP2/6-31G(d) and MP4(SDQ)/6-31+G(d,p)//MP2/6-31G(d) levels, respectively. Obviously it is also a contradicting prediction. Their −2.1 kJ mol^{-1} result is consistent with our SCS-MP2 due

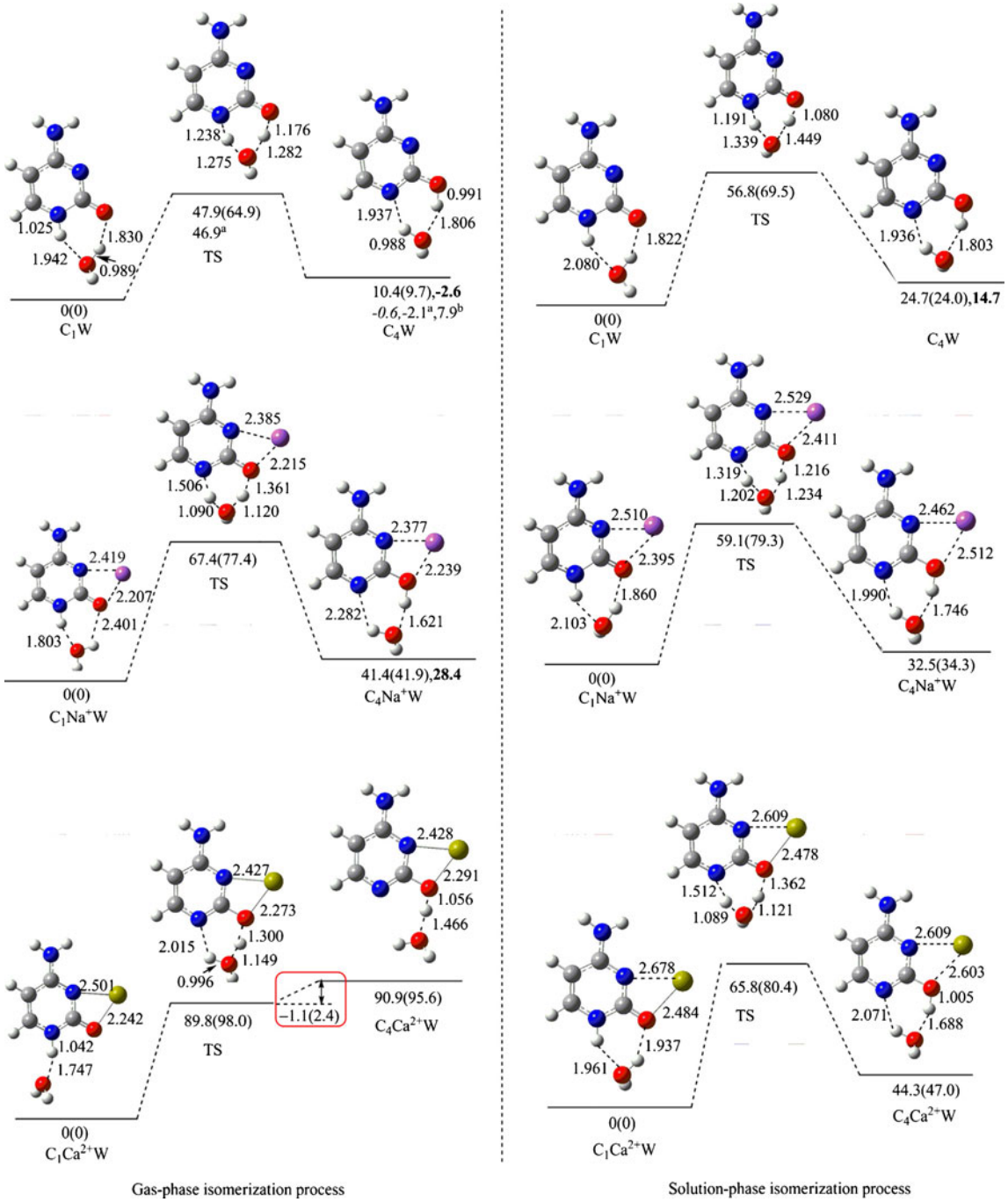

Fig. 4 Structures and activation energies of isomerization processes of $C_1W \leftrightarrow C_4W$ and $C_1M^{+/2+}W \leftrightarrow C_4M^{+/2+}W$. Data in parentheses do not include the ZPVE corrections, in black denotes the present SCS-MP2/6-311++G(d,p)//B3LYP/6-31G(d) results, in italic are from ref [14] obtained at the RIMP2+ZPVE level. [a,b] are from ref [26] obtained at MP2/6-311++G(d,p)//MP2/6-31G(d) and MP4(SDQ)/6-31+G(d,p)//MP2/6-31G(d) levels, respectively. Energy in kJ mol^{-1} and distance in Å

to similar level and basis set being used. The MP4(SDQ) result evidently overestimates the energy of C_4W, indicating that a larger basis set is a requisite for computing such a sensitive system correctly [6, 24], although MP4(SDQ) level is better than MP2 in accuracy. Trygubenko et al.'s prediction for the value [14] at the RIMP2+ZPVE level is −0.6 kJ mol^{-1}, in rational agreement with our SCS-MP2 result. Even in such an agreeable case, the B3LYP still overestimates the monohydrated C_4 energy, as it has done so for the isolated C_4. In bulk aqueous phase, both B3LYP and SCS-MP2 make a concordant prediction for the stability of C_1W and C_4W . For the ΔE prediction of C_1Na^+W and C_4Na^+W in gas phase, both methods uncover that

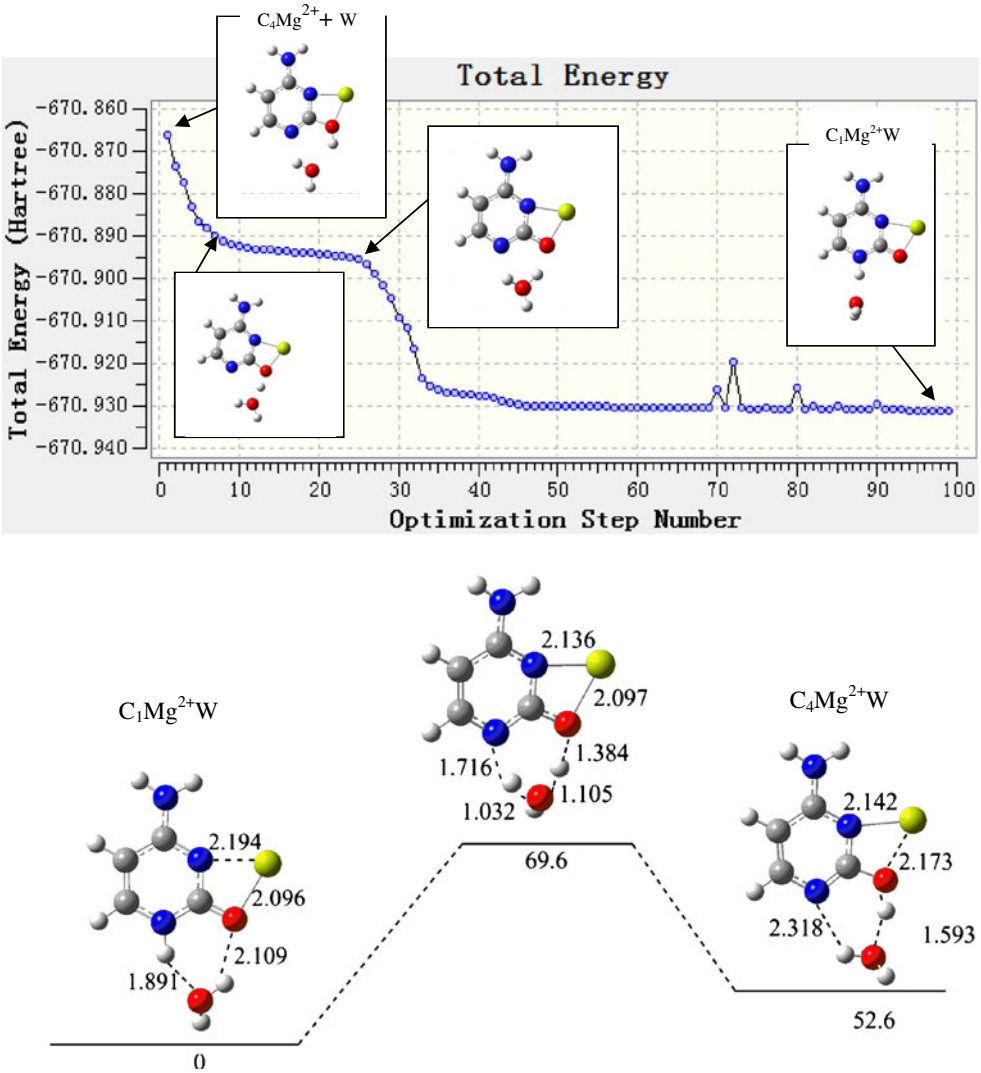

Fig. 5 $C_1 \leftrightarrow C_4$ isomerization processes assisted by the cooperation of both Mg^{2+} binding and monohydration (i.e., $C_1Mg^{2+}W \leftrightarrow C_4Mg^{2+}W$) in gas (top) and aqueous (bottom) phases. Energy in kJ mol^{-1} and distance in Å

C_1Na^+W is more stable, indicating that B3LYP is suitable for the prediction of aqueous-phase C_n isomer as well as their metalated complexes in both gas and aqueous phases.

Regulation effect of microhydration on the isomerization of $C_1M \leftarrow C_4M$

Divalent metal ions are usually multi-coordinated in aqueous phase [31] , which implies that both activation energy-free $C_1Mg^{2+}W \leftarrow C_4Mg^{2+}W$ and $C_1Ca^{2+}W \leftarrow C_4Ca^{2+}W$ processes could be artificial. Thus we further extend our investigation on the microhydration at the position of metal ion on the basis of both $C_1Mg^{2+}W \leftarrow C_4Mg^{2+}W$ and $C_1Ca^{2+}W \leftarrow C_4Ca^{2+}W$ processes. We aim to obtain the explicit number of water molecules bound at the metal ion, which can inhibit the two energy-free isomerization processes.

Figure 7 displays two further hydrated processes, $C_1Mg^{2+}2/3\ W \leftarrow C_4Mg^{2+}2/3\ W$ and $C_1Ca^{2+}2/3\ W \leftarrow C_4Ca^{2+}2/3\ W$. Results reveal that monohydration at the Ca^{2+} site is sufficient to keep $C_4Ca^{2+}2$ W stable; and the activation energy of the $C_1Ca^{2+}2\ W \leftarrow C_4Ca^{2+}2\ W$ process becomes 4.2 kJ mol^{-1} after ZPVE correction, a positive value. Dihydration at the metal ion can further increase the energy by 8.5 kJ mol^{-1}, signaling that the more the water molecules are bound, the less favorable the isomerization process of $C_1Ca^{2+}nW \leftarrow C_4Ca^{2+}nW$ will be. This trend can be verified indirectly by the higher activation energy of 21.5 kJ mol^{-1} obtained in the $C_1Ca^{2+}W \leftarrow C_4Ca^{2+}W$ process in the bulk aqueous phase (PCM). From the thermodynamic point of view, further hydration at the Ca^{2+} means improving the stability of the structure of $C_4Ca^{2+}W$ while decreasing the energy of $C_4Ca^{2+}W$ relative to $C_1Ca^{2+}W$. For example, the relative energy of $C_4Ca^{2+}3$ W to $C_1Ca^{2+}3$ W is 81.0 kJ mol^{-1}, lower than that (85.3 kJ mol^{-1}) of $C_4Ca^{2+}2$ W relative to $C_1Ca^{2+}2$ W. In contrast to the relative energy (55.1 kJ mol^{-1}) of $C_4Ca^{2+}W$ to $C_1Ca^{2+}W$ in aqueous phase, there is still room

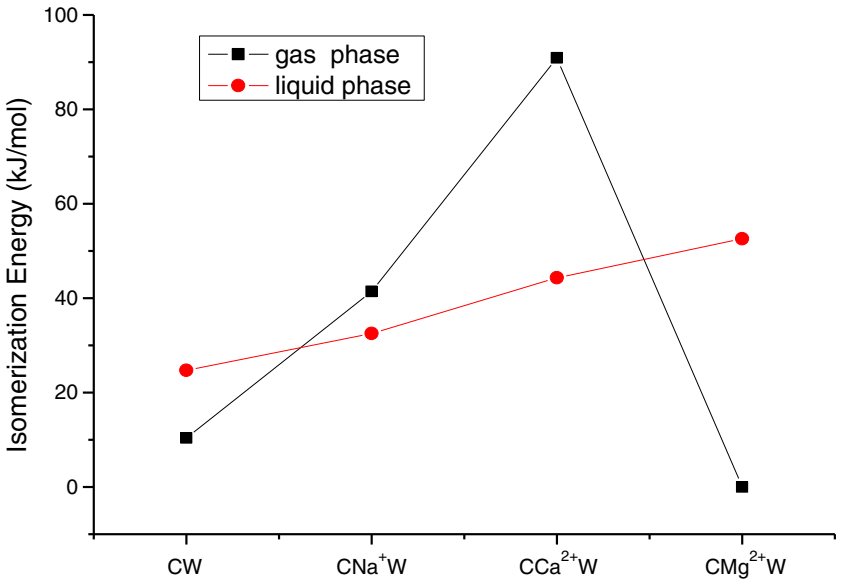

Fig. 6 Isomerization energies (ΔE) of $C_4W \rightarrow C_1W+\Delta E$ and $C_4MW \rightarrow C_1MW+\Delta E$ ($M=Na^+$, Ca^{2+} and Mg^{2+}, respectively) processes. ΔE in either gas phase (♦) or liquid phase (•) denotes the energy difference between C_1W and C_4W (corresponding to CW on abscissa axis) or between C_1MW and C_4MW (corresponding to CNa^+W, $CCa^{2+}W$, and $CMg^{2+}W$)

to decrease the relative energy, i.e., further hydration at the Ca^{2+} would continue to improve the stability of the $C_4Ca^{2+}W$ more. Mg^{2+}has smaller ionic radius than Ca^{2+}, thus the number of water molecules required to inhibit the activation energy-free $C_1Mg^{2+}W \leftarrow C_4Mg^{2+}W$ process are different. Calculations reveal the number is two, i.e., not less than two water molecules bound at the Mg^{2+} is required to keep the $C_4Mg^{2+}W$ (product $C_4Mg^{2+}3$ W) stable. Although monohydration at the metal ion can also improve the stability of the $C_4Mg^{2+}W$, the energy of the product $C_4Mg^{2+}2$ W is still higher by 1.4 kJ mol^{-1} than that of the TS after the ZPVE is included, however. According to this trend, microhydration at the metal ion would further improve the stability of the product of $C_4Mg^{2+}nW$. The phenomenon can help one gain insight into inhibiting and even regulating the isomerization process of $C_1 \leftarrow C_4$, induced by both metal binding and hydration.

For divalent $C_1MnW \rightarrow C_4MnW$ processes ($n=1, 2, 3$) in gas phase, the activation energy is raised as the hydration number (n) increases, indicating the impossibility of the $C_1M \rightarrow C_4M$ isomerization. From the geometry point of view, the angle ∠N1H1O(W) becomes smaller and smaller as the n increases, where O(W) of the angle stands for the oxygen of water attached at the H1 site. This implies that the water molecule, initially playing the role of a bridge for the transfer of H1, is staying away from the O2 site as the number of water molecules attached at the metal ion position (see the angle in the C_1MnW in Fig. 7) increases. Obviously this trend disfavors the $C_1M \rightarrow C_4M$ isomerization; that accounts for why divalent $C_1MnW \rightarrow C_4MnW$ isomerization process becomes more and more difficult along with the increasing n. The result further confirms that C_1M is the most stable form in both gas and aqueous phase.

Conclusions

The stability orders of $C_nNa_{I/II}^+$, $C_nK_{I/II}^+$, $C_nCa_{I/II}^{2+}$, $C_nMg_{I/II}^{2+}$ (n=1-13,x) in aqueous phase have been determined and their most stable complexes predicted in $C_1M^{+/2+}$ form in bulk aqueous phase. Like the stability ordering of the first six most stable isomers of cytosine in aqueous phase, $C_1 > C_x > C_2 > C_3 > C_4 > C_5$, the metalated complexes also maintain the ordering in the case of C_4 product degenerated into C_5, i.e., $C_1M^{+/2+} > C_xM^{+/2+} > C_2M^{+/2+} > C_3M^{+/2+} > C_5M^{+/2+} \geq C_4M^{+/2+}$. The stability ordering of these metalated complexes in aqueous phase differs greatly from that in the gas phase, indicating strong aqueous effect on the stability of these complexes.

Nonetheless, more attention should be paid to, both $C_5M^{+/2+}$ and $C_4M^{+/2+}$ since they are both derived from C_4, which is the most stable cytosine isomer in the gas phase. How much the ratio of a metalated cytosine complex is depends on the ratio of the reactant cytosine rather than on the stability of the metalated complexes themselves, from which one gains insight into why experimentally produced metalated complexes correlate to the higher-energy $C_{4/5}Na^+$ and $C_{4/5}K^+$ instead of the most stable C_1Na^+/K^+. The present study urges one to attach greater importance to the derivatives generated from the ground-state base structure; although the derivative may not always be the most stable one. Otherwise, rational explanations for a lot of experimental observations can hardly be obtainable.

As the most stable isomers C_1 in aqueous phase and C_4 in gas phase, both $C_1 \leftrightarrow C_4$ conversion and the conversion conditions, such as gas phase, microhydration, or bulk aqueous phase, are of special importance and significance for understanding the majority of the constituents of the cytosine isomers or their metalated derivatives, and as a result of surroundings.

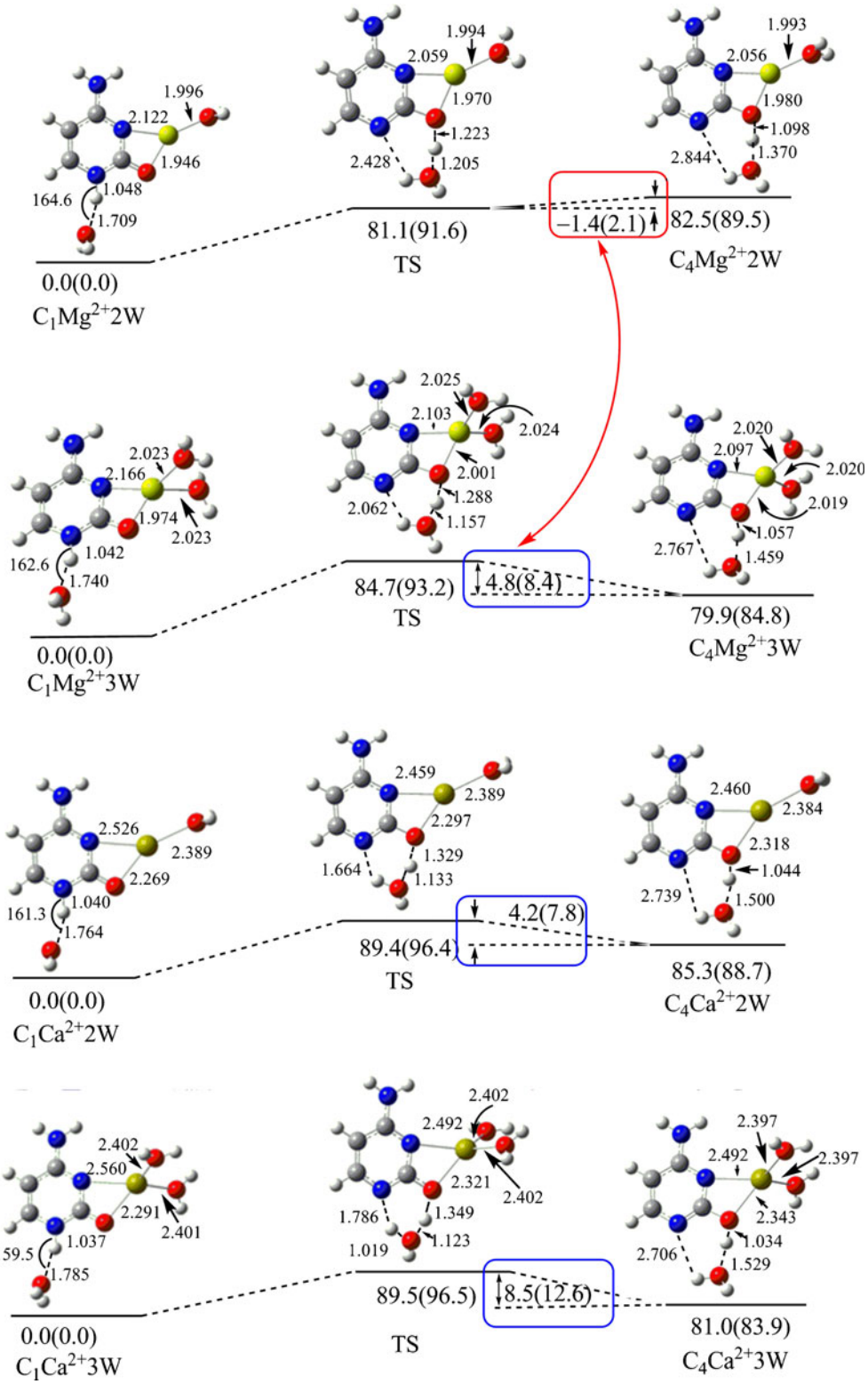

Fig. 7 Structures, relative energies and activation energies (in the box) of isomerization processes of $C_1M^{2+}nW \leftrightarrow C_4M^{2+}nW$ (n=2, 3; M=Ca, Mg). Further hydration at the Mg ion can turn the spontaneous (activation energy free) $C_1Mg^{2+}1W \leftarrow C_4Mg^{2+}1W$ process into an energy-required one. The activation energies are compared and highlighted in the red and blue boxes linked by a red double-headed arrow. Values in and outside parentheses denote those energies without and with ZPVE corrections. Energy in kJ mol^{-1}, distance in Å and angle in degree

The present work offers systematic and qualitative schemes to master such conversions, through which other DNA/RNA bases or biomolecules with different ground-state structures in gas and aqueous phases can refer to and draw lessons from.

That the stability orderings of both cytosine isomers and its metalated complexes are sensitive to the choice of computational method in gas phase and B3LYP may be insufficient in predicting the stability of some isomers or complexes correctly.

However, in aqueous phase, the B3LYP method has been verified to be reliable and recommended in the viewpoint of both accuracy and minimizing computational costs.

There exists a volume of studies on the gas-phase properties of DNA bases and their various derivatives; yet less attention has been paid to the properties of aqueous-phase derivatives. Since most derivatives of these bases and other biomolecules occur in aqueous phase, it is of more significant interest to study the properties in the aqueous phase and the relevance of the properties in both gas and aqueous phases. We hope the present paper serves as the first swallow of the spring, and others will follow.

Acknowledgments This work is supported by National Natural Science Foundation of China (NSFC)(Nos.20973084, 21003162), NSFC-National Research Foundation of Korea (No.21211140340) and NSF (No.Y2008B56) of Shandong Province.

References

1. Portugal FH, Cohen JS (1977) A century of DNA: a history of the discovery of the structure and function of the genetic substance. MIT Press, Cambridge
2. Goddard JD, Mezey PG, Csizmadia IG (1975) A note on a non-empirical molecular orbital study of some cytosine and thymine tautomers. Theor Chem Accounts 39:1–6
3. Kaim W, Schwederski B (1994) Bioinorganic chemistry: inorganic elements in the chemistry of life. John Wiley & Sons, Chichester
4. Saenger W, Cantor CR (1984) Principles of nucleic acid structure. Springer, New York
5. Zhou RY, Han YC, Jiang W, Yang H, Liu C (2007) Divalent metal ion-activated DNA cleavage activity of copper superoxide dismutase. Chem J Chin Univ 28:212–216
6. Kobayashi R (1998) A CCSD (T) study of the relative stabilities of cytosine tautomers. J Phys Chem A 102:10813–10817
7. Vázquez MV, Martínez A (2007) Ca, Cd, Zn, and their ions interacting with cytosine: a theoretical study. J Phys Chem A 111:9931–9939
8. Russo N, Toscano M, Grand A (2001) Bond energies and attachments sites of sodium and potassium cations to DNA and RNA nucleic acid bases in the gas phase. J Am Chem Soc 123:10272–10279
9. Kabeláč M, Hobza P (2006) Na^+, Mg^{2+}, and Zn^{2+} binding to all tautomers of adenine, cytosine, and thymine and the eight most stable keto/enol tautomers of guanine: a correlated ab initio quantum chemical study. J Phys Chem B 110:14515–14523
10. Šponer J, Burda JV, Sabat M (1998) Interaction between the guanine-cytosine Watson- Crick DNA base pair and hydrated group IIa (Mg^{2+}, Ca^{2+}, Sr^{2+}, Ba^{2+}) and Group IIb (Zn^{2+}, Cd^{2+}, Hg^{2+}) metal cations. J Phys Chem A 102:5951–5957
11. Vázquez MV, Martínez A (2008) Theoretical study of cytosine-Al, cytosine-Cu and cytosine-Ag (neutral, anionic and cationic). J Phys Chem A 112:1033–1039
12. Tureček F, Yao C (2003) Hydrogen atom addition to cytosine, 1-methylcytosine, and cytosine-water complexes: a computational study of a mechanistic dichotomy. J Phys Chem A 107:9221–9231
13. Slavíček P, Winter B, Faubel M, Bradforth SE, Jungwirth P (2009) Ionization energies of aqueous nucleic acids: Photoelectron spectroscopy of pyrimidine nucleosides and ab initio calculations. J Am Chem Soc 131:6460–6467
14. Trygubenko SA, Bogdan TV, Rueda M, Orozco M, Luque FJ, Šponer J, Slavíček P, Hobza P (2002) Correlated ab initio study of nucleic acid bases and their tautomers in the gas phase, in a microhydrated environment and in aqueous phase. Phys Chem Chem Phys 4:4192–4203
15. Gould IR, Green DVS, Young P, Hillier IH (1992) A theoretical study using ab initio methods of tautomerism in cytosine in the gas phase and in water. J Org Chem 57:4434–4437
16. Miertus S, Scrocco E, Tomasi J (1981) Perspective on "Electrostatic interactions of a solute with a continuum. A direct utilization of ab initio molecular potentials for the prevision of solvent effects". Chem Phys 55:117–129
17. Lee C, Yang W, Parr RG (1998) Development of the colle-salvetti correlation-energy formula into a functional of the electron density. Phys Rev B 37:785–789
18. Møller C, Plesset MS (1934) Note on an approximation treatment for many-electron systems. Phys Rev 46:618–622
19. Grimme S, Steinmetz M, Korth M (2007) How to compute isomerization energies of organic molecules with quantum chemical methods. J Org Chem 72:2118–2126
20. Frisch MJ, Trucks GW, Schlegel HB (2004) Gaussian 03, Revision C.02. Gaussian Inc, Wallingford
21. Dreyfus M, Bemaude O, Dodin G, Dubois JE (1976) Tautomerism in cytosine and 3-methylcytosine. A thermodynamic and kinetic study. J Am Chem Soc 98:6338–6349
22. Sambrano JR, de Souza AR, Queralt JJ, Andrés J (2000) A theoretical study on cytosine tautomers in aqueous media by using continuum models. Chem Phys Lett 317:437–443
23. Colominas C, Luque FJ, Orozco M (1996) Tautomerism and protonation of guanine and cytosine: implications in the formation of hydrogen-bonded complexes. J Am Chem Soc 118:6811–6821
24. Kobayashi R (2012) Correlated ab initio quantum chemical study of the interaction of the Na^+, Mg^{2+}, Ca^{2+}, and Zn^{2+}ions with the tautomers of cytosine . J Phys Chem A 116:4987–4994
25. Cerda BA, Wesdemiotis C (1996) Li^+, Na^+, and K^+binding to the DNA and RNA nucleobases. bond energies and attachment sites from the dissociation of metal ion-bound heterodimers. J Am Chem Soc 118:11884–11892
26. Gorb L, Podolyan Y, Leszczynski J (1999) A theoretical investigation of tautomeric equilibria and proton transfer in isolated and monohydrated cytosine and isocytosine molecules. J Mol Struct (THEOCHEM) 487:47–55
27. Bazsó G, Tarczay G, Fogarasi G, Szalay PG (2011) Tautomers of cytosine and their excited electronic states: a matrix isolation spectroscopic and quantum chemical study. Phys Chem Chem Phys 13:6799–6807
28. Yang Z, Rodgers MT (2012) Tautomerization in the formation and collision-induced dissociation of alkali metal cation-cytosine complexes. Phys Chem Chem Phys 14:4517–4526
29. Russo N, Toscano M, Grand A (2003) Gas-phase absolute $Ca^{2+}$and Mg^{2+} affinity for nucleic acidbases. A theoretical determination. J Phys Chem A 107:11533–11538
30. Kosenkov D, Kholod Y, Gorb L, Shishkin O, Hovorun DM, Mons M, Leszczynski J (2009) Ab initio kinetic simulation of gas-phase experiments: tautomerization of cytosine and guanine. J Phys Chem B 113:6140–6150
31. Wakisaka A, Watanabe Y (2002) Relation of hydrophobic effect with salt effect: on the viewpoint of cluster structure. J Phys Chem B 106:899–901