View Article Online / Journal Homepage / Table of Contents for this issue

PCCP

Correlated *ab initio* study of nucleic acid bases and their tautomers in the gas phase, in a microhydrated environment and in aqueous solution†

Part 1. Cytosine

Semen A. Trygubenko,[a] Tetyana V. Bogdan,[a] Manuel Rueda,[b] Modesto Orozco,[b] F. Javier Luque,[b] Jiří Šponer,[c] Petr Slavíček[c] and Pavel Hobza*[c]

[a] *Department for Physical and Mathematical Sciences, The National University of Kiev-Mohyla Academy, Kyiv 04070, Ukraine*

[b] *Departament de Bioquímica i Biologia Molecular, Facultat de Química, Universitat de Barcelona, Martí i Franqués 1, Barcelona 08028, Spain and Departament de Fisicoquímica, Facultat de Farmàcia, Universitat de Barcelona, Avgda. Diagonal s/n, 08028, Barcelona, Spain*

[c] *J. Heyrovský Institute of Physical Chemistry, Academy of Sciences of the Czech Republic and Center for Complex Molecular Systems and Biomolecules, 182 23, Prague 8, Czech Republic. E-mail: hobza@indy.jh-inst.cas.cz*

Received 1st March 2002, Accepted 7th June 2002
First published as an Advance Article on the web 25th July 2002

Published on 25 July 2002. Downloaded by Pennsylvania State University on 18/09/2016 22:05:31.

Canonical, enol and imino tautomers of cytosine were studied theoretically in the gas phase, in a microhydrated environment (1 and 2 waters) and in bulk water. The structures of isolated, mono- and dihydrated tautomers were determined at the RI-MP2 level with the TZVPP basis set. The relative energies of isolated tautomers were calculated up to the CCSD(T) level using the cc-pVTZ basis set and at the MP2 level using the aug-cc-pVQZ basis set. For the MP2 and CCSD(T) predictions, complete basis set estimates were obtained using various extrapolation techniques. One of the enol forms is the global minimum at all theoretical levels in the gas phase while the canonical form represents the first local minimum. Already two water molecules reverse the relative stability of these two tautomers making the canonical form the global minimum. The effect of bulk solvent on the relative stability of cytosine tautomers was examined from self-consistent reaction field, Monte Carlo and molecular dynamics free energy calculations. Bulk solvent calculations unambiguously favored the canonical tautomer over the enol forms, in agreement with the trends found for the mono- and dihydrated cluster model. However, the bulk solvent results for relative energy changes differ from those of the cluster model. While the enol structure is predicted to be the least stable species in the bulk solvent, the microhydration model predicts it to be the first local minimum with a rather small energy difference (~1 kcal mol^{-1}) with respect to the global minimum.

I Introduction

The maintenance of the genetic code relies on the specific hydrogen-bonding recognition between nucleic acid (NA) bases.[1] According to the Watson–Crick model,[2] the adenine–thymine (AT) and guanine–cytosine (GC) base pairs are stabilized by two and three hydrogen bonds. Specific patterns of hydrogen-bonding donor/acceptor groups are provided by the AT and GC base pairs in both major and minor grooves of DNA duplexes. This structural feature is important since it enables one to read a DNA sequence without opening the base pairs, and is responsible for the specificity of recognition and binding of other molecules to the DNA.[3,4] The complex network of hydrogen-bonding interactions that modulates the structure and function of DNA is based on the predominance of the canonical tautomeric forms of NA bases. The eventual importance of the prototropic tautomerism was, nonetheless, recognized already by Watson and Crick.[2] Several models of spontaneous mutation in DNA involve the role of minor tautomeric forms of the bases.[5,6] Minor tautomeric forms can also be involved in the stabilization of certain anomalous DNA structures.[7] Recent studies have also shown the potential role of metal cation-nucleobases binding in the stabilization of rare imino tautomers, an effect that may be related to mutagenic effects caused by certain metals.[8,9]

High-level computational methods play a crucial role in determining the relative stability of tautomers in the gas phase where, for a number of nucleobases, several tautomers are known to co-exist. It is considerably more difficult to apply both theoretical and experimental methods to the study of nucleobase tautomerism when the bases are not isolated. In most cases, only the most stable tautomers of each base are detectable in the polar environment experiments, except for the, irrelevant to DNA, N7–N9 tautomerism of purine bases.[3,10] There have been numerous computational studies on the tautomeric equilibrium of bases, particularly of cytosine.[11–22] Based on these studies, the lowest energy tautomers of cytosine in the gas phase have been unambiguously identified (see Fig. 1). In addition to the canonical form **1**, two enol (**2a**, **2b**) and two imino (**3a**, **3b**) forms have been found to lie in a relatively narrow energy range. Nevertheless, since the

† Dedicated to the memory of our friend and distinguished scientist Professor Peter Kollman.

 DOI: 10.1039/b202156k

This journal is © The Owner Societies 2002

View Article Online

Published on 25 July 2002. Downloaded by Pennsylvania State University on 18/09/2016 22:05:31.

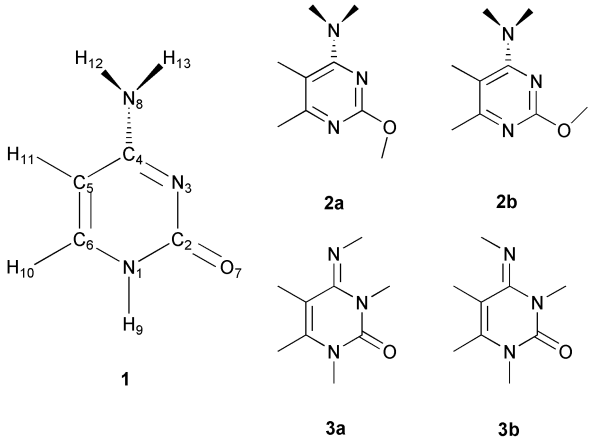

Fig. 1 The five most stable cytosine tautomers. Standard numbering and adopted nomenclature are presented.

minima are close in energy, the relative stability of these tautomers is very sensitive to the level of theory. Thus, relative to the canonical form 1, the energy difference of C2a and C3b are predicted to be −1.5 and +1.8 (MP2/6-31++G(d,p)),[16] +0.4 and +1.9 (B3LYP/6-31++G(d,p)),[16] −1.1 and +0.9 (MP2/6-311++G(d,p) supplemented by corrections up to the fourth order of the perturbational treatment with the 6–31G(d) basis),[20] and −1.5 and +0.3 (CCSD(T)/cc-pVTZ-(-f))[15] kcal mol^{-1}.

Since 1994 we have systematically studied theoretically[11a] all five canonical NA bases and more than 100 H-bonded and stacked NA base pairs using a consistent computational level of theory. The geometries of H-bonded pairs were optimized at the HF and sometimes MP2 levels with medium sized basis sets of atomic orbitals, while the stabilization energies of all NA base pairs were determined mainly at the MP2/6-31G(d;0.25) level. (Standard d-functions on non-hydrogen atoms were replaced by more diffuse ones with exponents 0.25). The MP2/6-31G(d;0.25) method has been extensively used to evaluate the base stacking. Isolated nucleobases were investigated up to the MP2/6-311G(2df,p) level, mainly with respect to the amino-group pyramidalization effects. Although these computations were entirely sufficient to provide a correct qualitative picture of nucleobase interactions, it is evident that further tuning of the computational procedures is needed to obtain quantitatively correct data.

The need to obtain more accurate estimates of the geometrical and energetic properties of isolated nucleobases and of their hydrogen-bonded and stacking interactions leads us to seek better computational techniques. Here we report the results of the investigation of the tautomerism of cytosine, a part of the study, which will be continued for other nucleobases and their molecular complexes. Gas phase calculations of the relative stability of the different tautomers of cytosine involve RI-MP2[23–27] geometry optimizations and single-point calculations at the MP2 and CCSD(T) levels of theory using a variety of very flexible basis sets, including the evaluations considering f-polarization functions. The systematic calculations allow us to use extrapolation techniques to determine the best estimates of the relative stability of the tautomers in the gas phase. However, it is necessary to supplement the gas phase evaluations by reliable estimates of solvent effects. Thus, the influence of solvation has been explored by using two different approaches. First, a molecular dynamics/quenching technique combined with subsequent *ab initio* calculations has been used to analyze the most stable complexes of cytosine tautomers with one and two water molecules. Second, *ab initio* self-consistent reaction field calculations and classical free energy calculations coupled to molecular dynamics and Monte Carlo techniques have been used to determine the differential free energy of hydration of the tautomers at room temperature and pressure. From these studies, a detailed picture of the tautomeric preference of cytosine in the gas phase, in a water cluster and in aqueous solution is obtained.

II Methods

II.1 General strategy

The gas-phase geometry of the cytosine tautomers was determined using the resolution of identity (RI) method combined with the MP2 procedure (RI-MP2)[23–27] implemented in TURBOMOLE.[28] The RI-MP2 method provides an accurate description of H-bonded and stacked DNA base pairs.[23] Because the method is computationally very effective (approximately by one order of magnitude if compared with the exact MP2 method) it is possible to use it with extended AO basis sets. This allows one to perform calculations that were up to now impractical. For the purpose of comparison, geometry optimization of the canonical tautomer of cytosine was also performed using the exact MP2 method. The double-polarized valence triple-ζ (TZVPP) basis set [5s3p2d1f/3s2p1d] and the default auxiliary basis set were systematically used.[29,30] The relative energies of the tautomers were subsequently determined at the RI-MP2 and CCSD(T) levels with systematic expansion of Dunning's correlation consistent[31] basis sets in order to extrapolate the energy difference to the basis set limit.[32–34] Calculations were carried out using the Gaussian-98[35] and Molpro[36] programs. All RI-MP2 calculations were done with the TURBOMOLE code and CCSD(T) with Molpro, the remaining calculations were performed with the Gaussian-98 code. A molecular dynamics/quenching technique with Cornell *et al*. AMBER empirical potential[37] was used to explore the potential energy surface of complexes of the tautomers of cytosine with one and two water molecules. The energetically most stable structures were reinvestigated at the RI-MP2 level to determine the relative energies and stabilization energies of cytosine···water and cytosine···$(water)_2$ clusters. Finally, the effect of hydration on the relative stability of cytosine tautomers was examined from self-consistent reaction field (SCRF), as well as free energy calculations coupled to molecular dynamics (MD) and Monte Carlo (MC) methods.[38–40]

II.2 Extrapolation technique

In order to correct the computed energies for the deficiencies due to incompleteness of the basis set expansion and electron correlation coverage, two different extrapolation techniques were used. In the first case, one extrapolates the HF, MP2 and CCSD(T) energies, and in the second - CCSD(T) energies on the basis of accurate MP2 energies and the extrapolated (MP2 − CCSD(T)) energy difference. With the results determined from at least three of the cc-pVNZ and aug-cc-pVNZ basis sets, it is possible to obtain high quality estimates of complete basis set limit energies.[32–34] However, since the cc-pVTZ basis set often approaches the edge of computational feasibility for moderately-sized systems, Truhlar[32] developed an extrapolation technique requiring only cc-pVDZ and cc-pVTZ energies. In accordance with this technique, HF and correlation energies are extrapolated separately, assuming a power dependence of energy on N (eqn. (1)).

$$E_X = E_\infty + AN^{-\alpha} \quad (1)$$

where N is 2 for cc-pVDZ, 3 for cc-pVTZ and E_∞, A and α are fitting parameters.

Provided the power coefficients are constants of the electronic structure method (*i.e.* different for the HF, MP2 or CCSD(T) methods, but independent of the molecular system), the resulting expression for the infinite basis set limit energy

View Article Online

Published on 25 July 2002. Downloaded by Pennsylvania State University on 18/09/2016 22:05:31.

estimate is given by

$$E_\infty = \frac{3^\alpha}{3^\alpha - 2^\alpha} E_3^{\mathrm{HF}} - \frac{2^\alpha}{3^\alpha - 2^\alpha} E_2^{\mathrm{HF}} + \frac{3^\beta}{3^\beta - 2^\beta} E_3^{\mathrm{corr}} - \frac{3^\beta}{3^\beta - 2^\beta} E_2^{\mathrm{corr}}, \quad (2)$$

where α and β denote power coefficients for the HF method and the correlated method, respectively, E_2 denotes the energy for the cc-pVDZ basis and E_3 the energy for the cc-pVTZ basis.

The same treatment can be applied to the aug-cc-pVNZ basis sets. Coefficients α and β were fitted by Truhlar[32] yielding $\alpha = 3.4$, $\beta_{\mathrm{MP2}} = 2.2$, $\beta_{\mathrm{CCSD}} = \beta_{\mathrm{CCSD(T)}} = 2.4$, and subsequently refined for a larger number of molecules to $\alpha = 3.39$, $\beta_{\mathrm{MP2}} = 2.2$, $\beta_{\mathrm{CCSD}} = 1.94$, $\beta_{\mathrm{CCSD(T)}} = 2.02$. This latter parametrization has been used here. Since there is a certain ambiguity in the choice of the α and β coefficients, we checked these coefficients on the triazine molecule, which resembles cytosine and allows us to perform CCSD(T) calculations with the cc-pVQZ basis set. Note also that this extrapolation technique is to a certain extent flexible, allowing its application to rather extended non-covalent complexes.[41]

The other extrapolation technique is based on the combination of accurate MP2 energies and the extrapolated (MP2 – CCSD(T)) energy difference (eqn. (3)). Recently this approach was used to calculate the benzene dimer interaction energy.[42] On the condition of the validity of eqn. (1), the correction for inclusion of higher electron correlation contributions can be determined from eqn. (4), where $\Delta_{\mathrm{CCSD(T)}}|_{\mathrm{inf}}$ is $\Delta_{\mathrm{CCSD(T)}}$ evaluated with the infinite basis set. Since the last two terms are of similar absolute value, they approximately cancel each other, even for small basis sets and, presumably, $\Delta_{\mathrm{CCSD(T)}}$ converges to the exact value faster than the energy itself. The basis set dependence of the $\Delta_{\mathrm{CCSD(T)}}$ within the cc-pVNZ and aug-cc-pVNZ basis set series has been investigated here for both the triazine and the cytosine tautomers.

$$\Delta_{\mathrm{CCSD(T)}} = E_{\mathrm{MP2}} - E_{\mathrm{CCSD(T)}} \quad (3)$$

$$\Delta_{\mathrm{CCSD(T)}} = \Delta_{\mathrm{CCSD(T)}}|_{\mathrm{inf}} + AN^{-\beta_{\mathrm{MP2}}} - BN^{-\beta_{\mathrm{CCSD(T)}}} \quad (4)$$

II.3 Molecular dynamics/quenching technique

Due to the complexity of the potential energy surface (PES) of molecular complexes involving nucleic acid bases,[43–45] we have used the molecular dynamics simulations combined with the quenching (Q) technique[46] to investigate clusters of cytosine tautomers with one and two water molecules (see ref. 45 for details). MD/Q simulations were carried out in the *NVE* microcanonical ensemble (*N*, *V*, and *E* mean number of particles, volume and total energy) within a quaternion formalism using the AMBER empirical potential,[37] which gives results in agreement with the *ab initio* data for NA base pairs. Simulations in the NVE ensemble give the properties of a cluster that does not interact with its surroundings. The respective code[47] uses a fifth-order predictor–corrector algorithm with a 0.5 integration step. MD simulations were performed at a constant total energy of -107.4 kcal mol^{-1} (corresponding average temperature is 298 K) which is high enough to allow crossing over relatively high energy barriers and thus to sample the whole PES. Every 1 ps the MD run was interrupted, the kinetic energy was removed and the structure of the cluster was fully optimized using the conjugate gradient method, finally storing both the geometry and the energy of the optimized structure. Then, the MD run takes off from the point it was interrupted. A total simulation time of 250 ns was completed.

The library of canonical cytosine in the AMBER force field was modified to parametrize the non-canonical tautomers from quantum mechanical calculations. Geometrical coordinates most strongly affected by tautomerization (bond lengths, valence and the dihedral angles of the hydroxy group of **2a,b** and the imino group of **3a,b**) were parametrized more thoroughly.[48] For new dihedrals careful parametrization using a differential fitting procedure according to Hopfinger and Pearlstein[49] was implemented. The atomic charges of the tautomers **2a**, **2b**, **3a** and **3b** were generated with a two-stage electrostatic potential fitting procedure (RESP)[50] at the HF/6-31G** level. For the dihedrals of the hydroxy group a set of thirteen optimized constrained geometries was calculated at the HF/6-31G* level. Then we calculated the molecular mechanical torsion curves for the same values of dihedrals, as used in the quantum chemical calculations, with zero dihedral force constants. The desired differential curve was obtained as a difference between these two curves. The differential curve was fitted with the least squares method implemented in *Mathematica* (Wolfram Research, Inc.) to the series of cosines in the form $\sum\{K_n(1+\cos(n\chi-\varphi))\}$, n = 1,2,3,4 and 6. As the cos(χ) and cos(2χ) appeared to be the main contributors to the given potential, only these two terms were used in the final fit.

The MD/Q procedure usually provided 4 to 8 structures within the suggested energy limit (~5 kcal mol^{-1}), which were then optimized at the HF/6-31G** level. The nature of the stationary point was verified on the basis of harmonic vibrational analysis at the same level of theory and the zero-point energy (ZPE) was computed. The harmonic approximation was supposed to be accurate enough, since Clary *et al.*[14] showed that anharmonic contributions to the ZPE for canonical cytosine with water complexes are less than 1 kJ mol^{-1}. Full geometry optimization of several of the most stable cytosine complexes with one and two water molecules was also performed at the RI-MP2 level with the TZVPP basis set employing the Turbomole 5.3 program suite.[28] The interaction energies were corrected for the basis set superposition effects using the Boys and Bernardi procedure.[51]

II.4 Statistical thermodynamic treatment of equilibrium geometries

Thermodynamic functions were determined from partition functions computed from the HF/6-31G(d,p) characteristics (geometry, vibrational frequencies). The rigid rotor–harmonic oscillator–ideal gas approximation was adopted.

II.5 Hydration

In order to determine the free energy of tautomerization in water, the relative free energies of hydration were computed by using both self-consistent reaction field and free energy calculations.[38–40]

SCRF calculations were performed using the recently reparametrized[52] *ab initio* HF/6-31G(d) optimized version of the MST[53–55] continuum model (also known as the polarizable continuum model[56,57]). Calculations were carried out using the gas phase geometries optimized at the RI-MP2 level, since small geometrical changes are expected to occur upon solvation of rigid molecules like those considered here.[58] MST calculations were performed using a locally modified version of Monstergauss.[59]

Monte Carlo-free energy perturbation (MC-FEP) simulations were performed to estimate the relative free energy of hydration associated with the mutation between the tautomers of cytosine. The solute was placed in a cubic box of nearly 9000 Å^3 containing approximately 270 TIP3P[60] water molecules. Simulations were performed in the isothermal–isobaric (NPT, 1 atm, 298 K) ensemble. Periodic boundary conditions were used in conjunction with preferential sampling in simulations. A residue-based non-bonded cut-off of 9 Å was used to evaluate intermolecular interactions. The mutation was carried out in 41 doublewidth sampling windows. In each window 2×10^6 configurations were used for equilibration and

View Article Online

Published on 25 July 2002. Downloaded by Pennsylvania State University on 18/09/2016 22:05:31.

3×10^6 configurations for averaging. The geometrical parameters of the mutated solutes were taken from the gas phase optimized geometry and were kept fixed during the simulations, which allows us to compare directly MST and MC-FEP relative free energies of hydration. Restrained electrostatic potential-derived atomic (RESP) charges computed at the RHF/6-31G(d) level were used with standard OPLS[61] Lennard-Jones parameters for the solute. The rotations and translations of the solute were adjusted to obtain around 40% acceptance. MC simulations were carried out with the BOSS4.2 program.[62]

Molecular dynamics-thermodynamic integration (MD-TI) simulations were also performed to estimate the relative hydration free energies between the tautomers of cytosine. The solute was placed in a cubic box containing approximately 420 TIP3P water molecules. Simulations were performed in the isothermal-isobaric (NPT, 1 atm, 298 K) ensemble using periodic boundary conditions, a non-bonded cut-off of 9 Å, SHAKE, and an integration time step of 1 fs. The mutation was carried out in 41 windows, each consisting of 10 ps of equilibration and 10 ps of averaging, leading to a total MD simulation of 820 ps for each mutation. HF/6-31G(d) RESP atomic charges were used in conjunction with van der Waals parameters taken from the AMBER force field.

III Results and discussion

III.1 Gas-phase tautomers

Based on the results of preceding studies of cytosine tautomers,[15,16,20] we have examined the relative stability in the gas phase of the canonical form (**1**), two enol forms (**2a**, **2b**) and two imino forms (**3a**, **3b**) (see Fig. 1). The geometries, rotational constants and dipole moments of all the tautomers are shown in Table 1, while their relative energies are summarized in Table 2.

Accuracy of the RI-MP2 method. The RI-MP2 and MP2 geometries of canonical cytosine were determined with the same basis set (TZVPP) to evaluate the precision of the RI-MP2 method in describing the geometrical characteristics. Both geometries were practically identical: the largest discrepancies in bond length, valence and dihedral angles amounted to 0.0002 Å, 0.01° and 0.02°, respectively.

Comparison of theoretical geometries obtained at various levels of theory for the canonical tautomer of cytosine is feasible owing to the existence of experimental rotational constants.[63] The HF/6-31G(d,p) geometry exhibits a rather large difference in the rotational constants (A, B and C parameters deviate from the experimental values by +2.9%, +3.3%, +3.1%, respectively). Such a difference in the rotational constants largely decreases for the MP2/cc-pVDZ[15] geometry (deviations from experimental values are −0.7%, −0.1%, and −0.4%). Addition of f-functions on non-hydrogen atoms and d-functions on hydrogens (present RI-MP2/TZVPP calculations) leads to a further improvement (deviations from experimental values are −0.6%, −0.1%, and −0.3%). These results indicate the importance of the inclusion of electron correlation effects and the use of flexible basis sets. We believe that obtaining even closer agreement between theoretical and experimental rotational constants would require passing from equilibrium geometry to vibrationally averaged geometry. The energies of interconversion between planar and non-planar structures, evaluated at the RI-MP2/TZVPP level, amount to 0.15, 0.31 and 0.37 kcal for **1**, **2a** and **2b**, respectively.

Basis set extrapolation. Energies corresponding to the basis set saturation limit (complete basis set–CBS) were estimated using Truhlar's extrapolation technique (see Methods section). Since the original procedure was based on calculations of atomization energies for small molecules, we were concerned about its suitability for the present systems. The extrapolation coefficients were, therefore, also derived from calculations performed with the cc-pVDZ, cc-pVTZ and cc-pVQZ basis for triazine as a reduced model system of cytosine. The dependence of the MP2 and CCSD(T) correlation energies on the cc-pVNZ (N = 2,3,4) basis sets for triazine is shown in Fig. 2, and the following coefficients were obtained: $\alpha = 2.77$, $\beta_{MP2} = 1.92$, $\beta_{CCSD(T)} = 2.23$. The MP2 calculations for cytosine were performed with up to the cc-pVQZ basis set, yielding the following coefficients: $\alpha = 2.84$, $\beta_{MP2} = 1.91$. There is a reasonable agreement between these values and those reported by Truhlar, suggesting that the extrapolation technique is quite robust and insensitive to the molecular system considered.

Relative energies of cytosine tautomers. Table 2 reports the relative gas phase energies for the cytosine tautomers. The enol form **2a** is the global energy minimum for all methods

Table 1 Geometries of cytosine tautomers. Optimization performed at the RIMP2/TZVPP level of theory

		Atom[a]													
Tautomer[a]		1	2	3	4	5	6	7	8	9	10	11	12	13	μ[d]
1	A[b]	3.8941 x[c]	0.0000	1.2340	2.3245	2.3365	1.1229	−1.0667	3.5132	−0.9133	0.9949	3.2489	4.3520	3.4701	6.30
	B	2.0269 y	0.0000	0.0000	0.0109	0.0123	0.0053	−0.0053	0.0650	−0.0027	0.0071	0.0282	−0.2044	−0.0985	
	C	1.3337 z	1.4130	2.0163	1.2853	−0.1470	−0.7510	2.0005	1.9468	−0.4259	−1.8237	−0.7198	1.4678	2.9384	
2a		3.9696	0.0000	1.0419	2.2425	2.3853	1.2202	−1.2013	3.3221	−1.8456	1.2494	3.3541	4.2093	3.1289	3.25
		2.0118	0.0000	0.0000	0.0082	0.0083	0.0017	−0.0056	0.0654	−0.0016	−0.0011	0.0172	−0.2392	−0.1616	
		1.3361	1.3300	2.1542	1.5737	0.1775	−0.5584	1.9370	2.4106	1.2172	−1.6405	−0.2979	2.0533	3.3715	
2b		3.9060	0.0000	1.0468	2.2531	2.3922	1.2237	−1.2149	3.3318	−1.0474	1.2569	3.3606	4.2170	3.1419	4.50
		2.0298	0.0000	0.0000	0.0085	0.0079	0.0011	−0.0061	0.0687	−0.0024	−0.0027	0.0163	−0.2386	−0.1786	
		1.3367	1.3248	2.1517	1.5799	0.1862	−0.5482	1.9045	2.4203	2.8552	−1.6304	−0.2900	2.0584	3.3768	
3a		3.8900	0.0000	1.2667	2.4991	2.3706	1.1514	−1.0177	3.6478	−0.9098	1.0075	3.2691	3.5541	1.2688	2.36
		2.0107	0.0000	0.0000	0.0009	−0.0000	−0.0004	−0.0000	0.0024	−0.0002	−0.0010	−0.0002	0.0030	0.0001	
		1.3255	1.3793	1.9274	1.2563	−0.1878	−0.7561	2.0462	1.8278	−0.4283	−1.8258	−0.7807	2.8428	2.9368	
3b		3.8705	0.0000	1.2608	2.4884	2.3677	1.1467	−1.0248	3.5492	−0.9104	1.0044	3.2572	4.3673	1.3013	4.63
		2.0278	0.0000	0.0000	0.0005	0.0001	−0.0002	−0.0000	0.0013	−0.0001	−0.0005	−0.0001	0.0015	0.0002	
		1.3306	1.3862	1.9294	1.2616	−0.1851	−0.7542	2.0413	1.9842	−0.4270	−1.8246	−0.7934	1.3818	2.9388	

[a] For atom numbering and tautomer classification refer to Fig. 1. Atom 1 is placed at the origin. [b] Rotational constants are given in GHz. [c] Coordinates are in Å. [d] Total dipole moment in Debye at RIMP2/TZVPP-auxTZVPP.

View Article Online

Published on 25 July 2002. Downloaded by Pennsylvania State University on 18/09/2016 22:05:31.

Table 2 Relative energies (in kcal mol^{-1}) of selected tautomers of cytosine in the gas phase. Geometries were obtained at the RIMP2/TZVPP level of theory. Our final gas-phase estimate (without ZPE) is shown on the last line

	Tautomer[a]				
Method	**1**	**2a**	**2b**	**3a**	**3b**
RIMP2/TZVPP	1.90	0.00	0.72	4.91	3.21
HF/6-31G(d,p) ZPE[b]	−0.27	0.00	−0.01	0.50	0.62
HF/6-31G(d,p) G[c]	−0.81	0.00	0.02	0.42	0.55
MP2/cc-pVDZ	2.57	0.00	0.67	4.62	2.96
MP2/cc-pVTZ	2.09	0.00	0.70	4.84	3.20
MP2/cc-pVQZ	1.65	0.00	0.72	4.74	3.01
MP2/aug-cc-pVDZ	1.60	0.00	0.69	4.70	2.96
MP2/aug-cc-pVTZ	1.58	0.00	0.71	4.76	2.99
MP2/aug-cc-pVQZ	1.46	0.00	–	–	2.94
CCSD(T)/cc-pVDZ	2.16	0.00	0.64	3.24	1.69
CCSD(T)/cc-pVTZ	1.61	0.00	0.67	3.44	1.87
CCSD(T)/cc-pVQZ[d]	1.23	0.00	0.69	3.36	1.73
CCSD(T)/aug-cc-pVDZ	1.22	0.00	0.66	3.34	1.70
E_{corr}(MP2/inf)[e]	1.39	0.00	0.00	1.34	1.61
E_{corr}(CCSD(T)/inf)[f]	0.77	0.00	−0.03	−0.20	0.11
E_{corr}(CCSD(T)/inf)[g]	0.80	0.00	−0.03	−0.13	0.19
E_{hf}(HF/inf)[e]	0.68	0.00	0.71	3.88	2.00
E_{hf}(HF/inf)[g]	0.80	0.00	0.71	3.94	2.07
E_{tot}(MP2/inf)[e]	2.06	0.00	0.71	5.22	3.61
E_{tot}(CCSD(T)/inf)[f]	1.44	0.00	0.69	3.68	2.11

[a] For atom numbering and tautomer classification refer to Fig. 1. [b] Zero-point energy corrections determined at the HF/6-31G(d,p) level. [c] Zero-point energy, thermal and entropy corrections determined at the HF/6-31G(d,p) level. [d] Based on extrapolating $\Delta_{CCSD(T)}$. [e] Extrapolation with coefficients fitted from extrapolating cytosine energies; corr, hf and tot mean correlation energy component, HF component and total relative energy. [f] CCSD(T) energy extrapolated employing exponents derived by extrapolating energies of triazine as a model system for cytosine. [g] Basis-set limit energies extrapolation with Truhlar coefficients.

including electron correlation, and the other enol tautomer (**2b**) represents the first local minimum. The energy difference between the enol forms is almost uniform at various levels of theory, and amounts to 0.6–0.7 kcal mol^{-1}. The canonical form **1** is at most level energetically less stable by 1.2–2.6 kcal mol^{-1}. Both imino forms, especially **3a**, are even less stable. A schematic representation of the dependence of the relative energy of the five isomers on the level of calculation is shown in Fig. 3. Within one computational method, consistent improvement of basis set affects only moderately the relative stability though it is to be noted that the cc-pVDZ basis set shows rather poor performance. However, transition from MP2 to CCSD(T) levels changes the relative energies of the tautomers, especially for the **3a** and **3b** species. Thus, while the imino tautomer **3b** is energetically well separated from the canonical structure **1** at the MP2 level, both structures are within 2 kcal mol^{-1} at the CCSD(T) level. To summarize, care should be paid to the choice of method and only methods including a large portion of correlation energy can be used for accurate prediction of tautomerization equilibria (*cf.* Table 2).

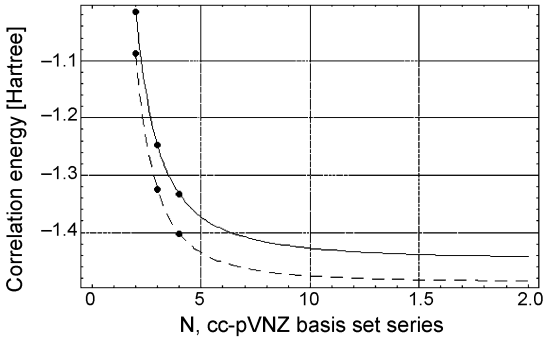

Fig. 2 MP2 and CCSD(T) correlation energy dependence within c-pVNZ series for triazine. Dashed line: CCSD(T), solid line: MP2.

Passing to the extrapolated MP2 energies (*i.e.* energies evaluated with the infinite AO basis set), we find only small differences with respect to the RIMP2/TZVPP energies. The same is true for the CCSD(T) extrapolated energies: here the differences with respect to the CCSD(T)/aug-cc-pVDZ level are not large. It follows from the evidence presented that already the non-extrapolated energies are close to saturation. Fig. 4 depicts the dependence of $\Delta_{CCSD(T)}$ as a function of N for the cc-pVNZ basis set for the triazine model system. The non-monotonic course of this dependence follows also from

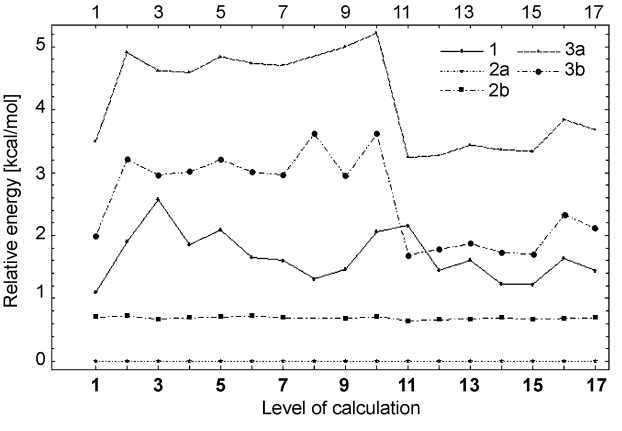

Fig. 3 Dependence of relative energy on method/basis set for cytosine tautomers. 1,[20] MP4/6-311++G(d,p); 2, RIMP2/TZVPP; 3, MP2/cc-pVDZ; 4,[15] MP2/cc-pVTZ(-f); 5, MP2/cc-pVTZ; 6, MP2/cc-pVQZ; 7, MP2/aug-cc-pVDZ; 8, MP2/aug-cc-pVTZ; 9, MP2/aug-cc-pVQZ; 10, MP2/inf; 11, CCSD(T)/cc-pVDZ; 12,[15] CCSD(T)/cc-pVTZ(-f); 13, CCSD(T)/cc-pVTZ; 14, CCSD(T)/cc-pVQZ; 15, CCSD(T)/aug-cc-pVDZ; 16, MP2/inf + $\Delta_{CCSD(T)}$/cc-pVDZ; 17, CCSD(T)/inf. (For single-point evaluation in 1, 4 and 12, respectively, MP2/6-31G(d), MP2/cc-pVTZ(-f) and MP2/cc-pVTZ(-f) geometry was used.)

View Article Online

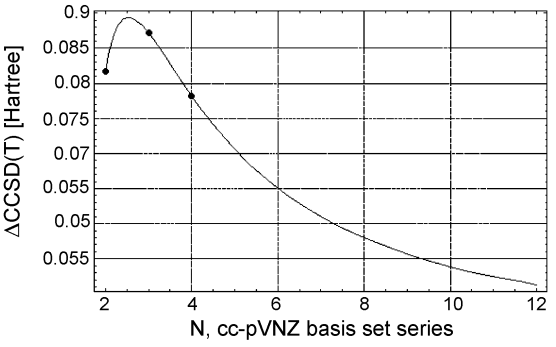

Fig. 4 $\Delta_{CCSD(T)}$ behavior as a function of basis set quality within the cc-pVNZ series.

the form of eqn. (4). Though the error over the whole range of N remains reasonable, even for the smallest cc-pVDZ basis, this approach does not yield a better estimate with a larger basis set. Instead, it suggests a crude estimate of higher order correlation contribution, which for **2a** amounts to $\Delta_{CCSD(T)}|_3 = 54.21$, $\Delta_{CCSD(T)}|_2 = 50.80$ and $\Delta_{CCSD(T)}|_{aug2} = 52.86$ kcal mol^{-1} for cc-pVNZ ($N = 3,2$) and aug-cc-pVNZ ($N = 2$), respectively. The corresponding values for other tautomers could be obtained from the entries in Table 2. The extrapolated MP2 relative energy for tautomer **1** is 2.06, *i.e.* slightly larger than the MP2/aug-cc-pVQZ value. It must again be recalled that the extrapolated MP2 relative energy is given as the sum of the extrapolated HF and the extrapolated MP2 correlation energy contributions. While the HF relative energy has monotonic dependence on the basis set, this is not true for the MP2 correlation energy contribution. In general, there is no reason to expect monotonic behavior for relative energy that is obtained from extrapolated data. The same arguments are valid also for the extrapolated CCSD(T) relative energies.

The most accurate and reliable energy data (CCSD(T)/inf) are presented in the last line of Table 2. In order to compare these values with experimental data it is necessary to include the ZPE energy (see the second line in Table 2). Evidently the ZPE corrections are small and do not bring changes in the relative stability of the cytosine tautomers. The final relative enthalpies (in kcal mol^{-1}) at 0 K are: **1** 1.17, **2a** 0.00, **2b** 0.68, **3a** 4.18, **3b** 2.73. If experiments are performed at non-zero temperatures, it is further necessary to include thermal corrections to the enthalpy and entropy terms (3rd line in Table 2, $T = 298.15$ K). These corrections also do not change the relative stability of the cytosine tautomers. The final relative free energies ($T = 298.15$ K, in kcal mol^{-1}) are thus: **1** 0.63, **2a** 0.00, **2b** 0.71, **3a** 4.10, **3b** 2.66. Energetically, exclusion of the f-functions (previous reference data by Kobayashi,[15] included for comparison in Fig. 3 under item 4) leads to a slight underestimation of the energy difference for high-energy tautomers.

Comparison of theoretical values with experimental data is difficult because there is no fully conclusive evidence yet. Thus, the canonical and enol forms have been detected by matrix isolation techniques, the former being destabilized by around 0.4 kcal mol^{-1},[64] which agrees with our theoretical estimate. Further, microwave experiments[63] have identified the presence of imino species. Though the experimental data are not very precise, the imino form should be destabilized by at least 1.4 kcal mol^{-1}.

Nonplanarity of the cytosine amino group. It is well established that amino groups of isolated nucleic acid bases are nonplanar due to a partial sp^3 pyramidalization of their amino group nitrogen atoms.[11a,65] The pyramidalization effects are neglected by presently available molecular mechanical force fields, but they are essential to stabilize a number of nonclassical molecular interactions seen in high-resolution DNA crystal structures.[66] Table 3 summarizes the RIMP2/TZVPP data for tautomers **1**, **2a** and **2b** (**3a** and **3b** species are planar), which we consider as new reference values. The amino group of the cytosine canonical form **1** is substantially non-planar. Non-equivalence of the two amino hydrogen dihedral angles (pyramidal-rotated geometry, see ref. 65*b*), which is caused by electrostatic repulsion between one of the amino group hydrogen atoms and the proximal H5 (C5) ring hydrogen atom[65] is worthy of notice. Note that the partial rotation of the amino group further weakens the double-bond contribution to the C4–N4 bond and pyramidal-rotated arrangements of nucleobase amino groups are often predicted to stabilize noncanonical contacts in nucleic acids.[66] The RIMP2/TZVPP data are in very good agreement with our preceding MP2/6-311G(2df,p) results.[65c] Formation of tautomers **2a** and **2b** slightly increases the nonplanarity since the shift of one ring hydrogen atom to an exocyclic position bolsters the formation of the amino nitrogen lone electron pair.

Table 3 Nonplanarity of the cytosine tautomer samino group (IUPAC atomic numbering). ΣXN4H is the sum of the three amino group valence angles. Inversion barriers amount to 0.15, 0.31 and 0.37 kcal for **1**, **2a** and **2b**, respectively

Tautomer	C5C4N4H	N3C4N4H	C2N3C4N4	ΣXN4H
1	−20.6	12.2	177.4	352.6
2a	−23.0	16.4	177.3	349.2
2b	−23.1	17.7	177.1	348.2

III.2 Microhydrated tautomers

Relative stability and binding patterns. In order to characterize the environmental effects on tautomerism, we have carried out extended classical MD/Q simulations of mono- and dihydrated cytosine tautomers. MD/Q simulations usually yielded 4 to 6 stable structures with one water molecule and 10 to 25 structures with two water molecules. The most stable structures were then re-investigated using QM methods. Fig. 5 shows three energetically most stable conformations of monohydrated structures arranged so that the relative energy of each tautomer, based on RI-MP2/TZVPP energies with HF/6-31G(d,p) ZPE correction, decreases from left to right. The relative energies and dipole moments of the depicted structures are presented in the Table 4. Fig. 5 shows that water-binding patterns are rather similar for the different tautomers. In the global minimum, oxygen and nitrogen ring atoms participate in double hydrogen bonding with water, while in the first local minimum (with the only exception of **3a**) water is bound through the amino and imino group nitrogens. Inspection of the results in Table 4 indicates that the qualitative ordering of relative energies is generally maintained at all levels of calculation. Nevertheless, inclusion of electron correlation modulates the magnitude of the energy differences, as noted in the case of structure **2** of the imino tautomers.

The four most stable structures of dihydrated cytosine are shown in Fig. 6 and their relative energies and dipole moments are summarized in Table 4. It is evident that the water dimer motif is energetically very stable since it is systematically included in all global minima. It is worth mentioning that the binding site of the water dimer coincides with the binding site of single water in the most stable structure of all five tautomers. For the canonical tautomer **1**, the order of relative stability of dihydrated structures fully agrees with that found in ref. 13. The importance of electron correlation effects is clearly shown in the change in the relative energy of the structures for tautomer **2a**.

Published on 25 July 2002. Downloaded by Pennsylvania State University on 18/09/2016 22:05:31.

View Article Online

Published on 25 July 2002. Downloaded by Pennsylvania State University on 18/09/2016 22:05:31.

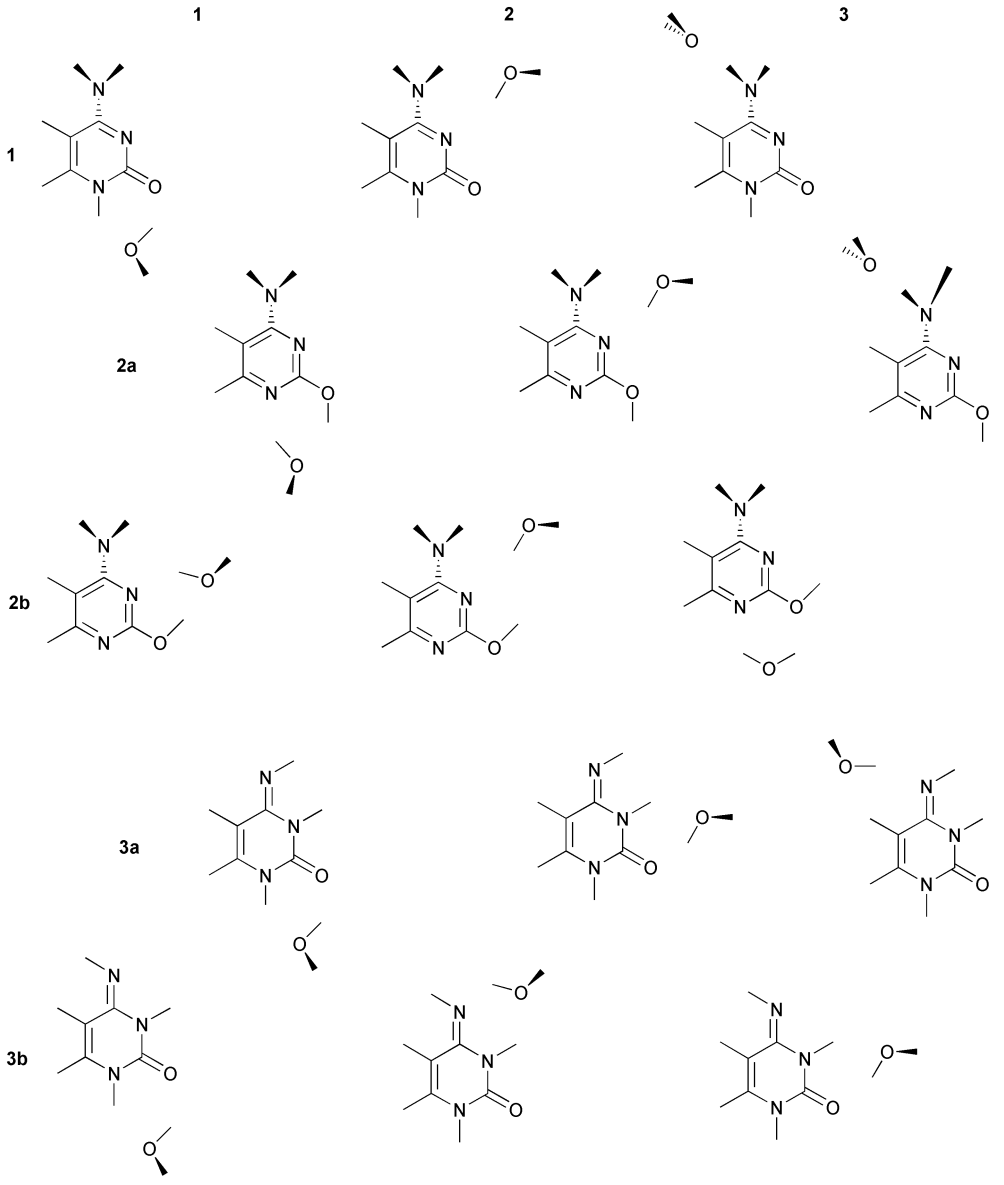

Fig. 5 The 15 most stable structures of cytosine tautomers with one water molecule optimized at the RIMP2/TZVPP level of theory. Stability is decreasing from left to right.

Bearing in mind the widespread use of the AMBER force field in contemporary large-scale MD simulations of nucleic acids, it is of particular interest to compare the empirical potential data with the RIMP2/ZPE values. For monohydrated complexes eight out of ten calculated relative energies agree within 1 kcal mol^{-1} while in the two remaining cases the difference is within 1–2 kcal mol^{-1}. In the case of the dihydrated complexes, three calculated relative energies show a difference between TIP3P/AMBER and QM data within 1 kcal mol^{-1} and six structures within the range 1–2 kcal mol^{-1}. Two **3b** structures show energy deviations higher than 2 kcal mol^{-1}. Taking into consideration the approximations inherent to the force field treatment we found these differences admissible, though not negligible. The HF/ZPE data show a somewhat better agreement with the reference RIMP2/ZPE points.

Table 5 contains the final values of the relative energies of the isolated tautomers as well as of the mono- and dihydrated complexes. At the RI-MP2 level (with the inclusion of ZPE) the canonical form **1** is 1.63 kcal mol^{-1} less stable in the gas phase than **2a**. It is evident that already the presence of one water molecule changes the relative energy difference between single tautomers, since it reduces the difference between **1** and **2a** to 0.14 kcal mol^{-1}, while that for **2a** and **2b** is increased to 1.97 kcal mol^{-1}. Addition of two water molecules changes the relative stability of tautomers **1** and **2a**, the former species being now about 1 kcal mol^{-1} more stable for the dihydrated complex. The relative stability of the second enol tautomer decreased upon microhydration and became the second local minimum. Both imino tautomers are less stable and their relative energy is changed only slightly upon addition of water molecules.

Interaction energies. RI-MP2/TZVPP interaction energies for all global minima and for several selected structures with binding patterns of particular interest are shown in Table 5. The interaction energies are generally within 9–12 and 18–22 kcal mol^{-1} for mono- and dihydrated tautomers. These are strong interactions, comparable to the stabilization energies of base pairs and substantially stronger than the water dimer. Deformation energy ranges typically between 0.5 and 1.5 kcal mol^{-1} for mono- and dihydrated tautomers, respectively (dihydrated structures **2a2** and **2b1** appear to be the most deformed (around 2.4 kcal mol^{-1}) in the series). It is also instructive to evaluate the many-body contributions to the interactions, showing the cooperativity of the water binding. For trimers, the three-body component of the interaction energy is not

View Article Online

Published on 25 July 2002. Downloaded by Pennsylvania State University on 18/09/2016 22:05:31.

Table 4 Relative energies (in kcal mol^{-1}) for tautomers of cytosine with one and two water molecules evaluated at different levels of theory[a]

Structure	MD/Q[b]	HF[c]	HF_{ZPE}[d]	RIMP2[e]	$RIMP2_{ZPE}$[f]	μ[g]
1–H_2O						
1	0.00	0.00	0.00	0.00	0.00	5.29
2	0.47	0.85	1.06	0.78	0.99	5.85
3	4.21	5.48	4.59	6.39	5.51	9.24
2a–H_2O						
1	0.00	0.00	0.00	0.00	0.00	4.20
2	−0.31	0.07	0.14	0.58	0.65	2.50
3	3.46	3.83	2.74	5.24	4.15	6.45
2b–H_2O						
1	0.00	0.00	0.00	0.00	0.00	5.30
2	1.37	1.72	1.22	1.32	0.82	3.45
3	1.55	2.01	0.95	2.92	1.85	7.15
3a–H_2O						
1	0.00	0.00	0.00	0.00	0.00	2.02
2	0.80	1.27	1.32	0.64	0.69	3.11
3	2.14	2.41	2.22	2.39	2.21	2.10
3b–H_2O						
1	0.00	0.00	0.00	0.00	0.00	3.52
2	2.06	1.28	1.33	0.42	0.48	5.36
3	2.76	2.34	2.15	2.29	2.10	4.96
1–$(H_2O)_2$						
1	0.00	0.00	0.00	0.00	0.00	4.72
2	2.24	1.63	1.47	0.36	0.20	4.83
3	0.19	0.92	1.10	0.85	1.02	5.57
4	1.23	1.23	1.11	1.81	1.69	5.18
2a–$(H_2O)_2$						
1	0.00	0.00	0.00	0.00	0.00	3.79
2	−0.95	−1.43	−1.25	−0.03	0.15	3.91
3	−1.17	−1.47	−1.16	0.28	0.59	1.98
4	−0.29	0.09	−0.27	—	—	—
2b–$(H_2O)_2$						
1	0.00	0.00	0.00	0.00	0.00	4.92
2	1.63	2.06	1.06	3.97	2.98	4.86
3	3.09	3.54	2.36	4.40	3.22	7.94
4	4.11	5.06	3.71	6.02	4.67	6.15
3a–$(H_2O)_2$						
1	0.00	0.00	0.00	0.00	0.00	1.82
2	0.82	0.65	0.63	0.81	0.79	3.35
3	3.62	3.31	3.27	1.72	1.68	3.61
4	4.13	3.50	3.12	3.19	2.80	3.31
3b–$(H_2O)_2$						
1	0.00	0.00	0.00	0.00	0.00	3.06
2	3.48	1.61	1.61	0.68	0.67	4.52
3	5.39	2.76	2.76	1.35	1.35	4.79
4	4.82	3.57	3.31	3.25	2.99	4.33

[a] Order of relative stability was referred to the most energetically preferable hydrated structure within complexes of one tautomer. [b] Modified Cornell *et al.* force field. [c] HF/6-31G(d,p). [d] HF/6-31G(d,p) corrected with ZPE evaluated at HF/6-31G(d,p). [e] RIMP2/TZVPP. [f] RIMP2/TZVPP corrected with ZPE evaluated at HF/6-31G(d,p). [g] Total dipole moment in Debye at RIMP2/TZVPP.

uniform and varies from −4.9 kcal mol^{-1} for **1**–$(H_2O)_2$, which contains a hydrogen-bonded water dimer, to −0.1 kcal mol^{-1} for **2a**–$(H_2O)_2$, where the two water molecules are not connected through a H-bond. Generally, the three-body term contributes up to 20% to the total interaction energy, which is comparable to the binding energy of a water dimer. Thus, the relative ordering of closely-lying minima, determined only on the basis of pairwise-additive empirical potential calculations, might not be well reproduced.

III.3 Hydrated tautomers

The differences in the free energy of hydration for cytosine tautomers are given in Table 6, which also reports the free energy of tautomerization in aqueous solution, as determined by adding the best estimate of the gas-phase free energy difference to the relative hydration free energy. There is very close agreement between the relative free energies of hydration determined from the MST, MC-FEP and MD-TI results, which gives confidence in the computed values. The canonical tautomer **1** is predicted to be better hydrated than the enol forms by around 7.0 (**2a**) and 5.7 (**2b**) kcal mol^{-1}, and than the imino tautomers by 3.8 (**3a**) and 3.9 (**3b**) kcal mol^{-1} (averaged values from MST, MC-FEP and MD-TI estimates). Based on the best estimates of the free energy differences in the gas phase (see above), the tautomer **1** is predicted to be the most stable form in dilute aqueous solution, the free energy of tautomerization relative to any other form being larger than 5.8 kcal mol^{-1}. Both enol and imino forms are predicted to be similarly destabilized in water. This is in accord with the fact that, compared to the tautomer **1**, the solvent-induced destabilization of the imino forms is smaller than that of the enol tautomers, which compensates for the greater stability of the latter species in the gas phase (see above). The predicted free energy difference of

View Article Online

Published on 25 July 2002. Downloaded by Pennsylvania State University on 18/09/2016 22:05:31.

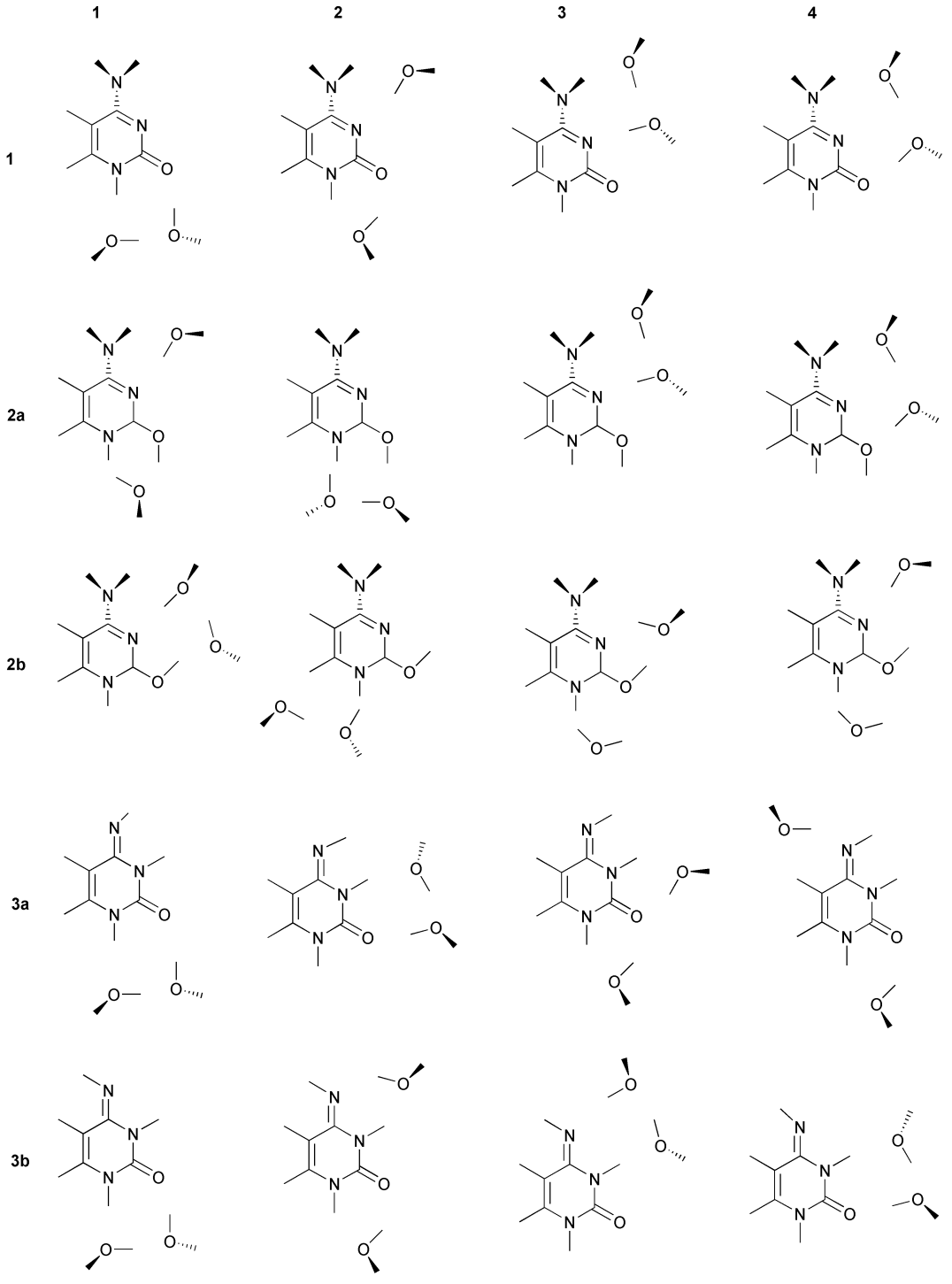

Fig. 6 The 20 most stable structures of cytosine tautomers with two water molecules optimized at the RIMP2/TZVPP level of theory. Stability is decreasing from left to right.

the imino species relative to tautomer **1** in water (+5.9 kcal mol^{-1}) is in agreement with the range of values determined experimentally (from 5.5 to 6.9 kcal mol^{-1}).[67]

III.4 Final assessment of the relative energies of cytosine tautomers in the gas phase, in a microhydrated environment and in water

MST, MC-FEP and MD-TI calculations unambiguously favored the canonical tautomer **1** over the enol form **2a**, in full agreement with the trends mentioned above for the mono- and dihydrated cluster models. The bulk solvent results differ, however, from the cluster model data in relative energy changes. While the former model predicts the enol structure **2a** to be largely destabilized (+6.4 kcal mol^{-1}), the latter model predicts it to be the first local minimum with a rather small energy difference (~1 kcal mol^{-1}) with respect to the global minimum. In the case of DNA base pairs we have shown[68] that microsolvation changes dramatically the structure of the pair and, further, the effect of a few waters (2–4) is comparable to that of a much larger set (64 waters). Both effects were explained by the very specific (and strong) interaction of the first few waters with the molecules of a pair. In the case of the isolated tautomer it is difficult to decide whether interaction of the first waters is also so decisive. We believe that the current results support the existence of three different environments: gas phase, small cluster and bulk solvent. It should be noted, nevertheless, that the microsolvation results would suggest

View Article Online

Published on 25 July 2002. Downloaded by Pennsylvania State University on 18/09/2016 22:05:31.

Table 5 Relative and interaction energies (in kcal mol^{-1}) of cytosine tautomers in the gas phase and mono- and dihydrated environment (global minima for each tautomer–water complex are presented)

	Relative energies[a b]				Interaction energies[c d]	
Structure	HF	HF_{ZPE}	RIMP2	$RIMP2_{ZPE}$	RIMP2[e]	$RIMP2_{TOT}$
1	1.71	1.43	1.90	1.63	—	—
2a	0.00	0.00	0.00	0.00	—	—
2b	0.74	0.73	0.72	0.71	—	—
3a	4.03	4.53	4.91	5.41	—	—
3b	2.31	2.94	3.21	3.83	—	—
1–H_2O	−0.38	−0.48	0.24	0.14	−12.21	−11.62
2a–H_2O	0.00	0.00	0.00	0.00	−10.59	−9.87
2b–H_2O	1.88	1.99	1.86	1.97	−9.44	−8.74
3a–H_2O	3.33	3.84	4.75	5.25	−10.58	−10.20
3b–H_2O	1.31	1.96	2.81	3.46	−10.83	−10.43
1–$(H_2O)_2$	−2.47	−2.62	−1.12	−0.96	−23.42 (−4.91)	−21.71
2a–$(H_2O)_2$	0.00	0.00	0.00	0.00	−20.14 (−0.14)	−18.66
2b–$(H_2O)_2$	1.25	1.56	1.05	1.67	−20.01 (−2.72)	−17.87
3a–$(H_2O)_2$	1.58	2.09	3.63	4.45	−21.54 (−4.12)	−20.18
3b–$(H_2O)_2$	−0.21	0.38	1.69	2.28	−21.80 (−4.35)	−20.36

[a] Order of relative stability of each tautomer is given with respect to tautomer **2a**. [b] For description of abbreviations used for methods *cf.* notes to Table 4. [c] Interaction energies were evaluated with augTZVPP basis set. [d] Total complexation energy RIMP2TOT is defined as the sum of the interaction energy RIMP2 and the deformation energies of the monomers.[11] [e] For trimers 3-body term components of the interaction energy are given (in parentheses).

some co-existence of tautomeric forms **1** and **2**, which is not supported by the prevailing view that only canonical forms of the main nucleobases are present in the polar solvent.[3] It is also instructive to compare the net hydration contributions to the relative tautomeric equilibria caused by the microhydration (Table 5) and continuous solvent approximation (Table 6). A single water molecule (RI-MP2+ZPE level) changes the tautomeric equilibrium with respect to the canonical form **1** by +1.49, +2.54, +1.33 and +1.12 kcal mol^{-1} for **2a**, **2b**, **3a**, and **3b**, respectively, which represents 20–42 % of the bulk solvent values (Table 6). Two explicit water molecules change the tautomeric equilibrium with respect to the canonical form **1** by +2.59, +3.34, +1.63 and +1.04 kcal mol^{-1} for **2a**, **2b**, **3a**, and **3b**, respectively, that is, 24–56% of the predicted bulk solvent effects. Thus, the microsolvated data are in-between the gas phase and bulk solvent predictions.

III.5 Biological relevance

In conclusion, let us briefly comment on the biological relevance of the cytosine tautomers. Obviously, the enol tautomers cannot be formed in nucleic acids, as they involve deprotonation of the N1 position of cytosine, where the sugar moiety is attached in nucleic acids. In contrast, the imino tautomers can be embedded into nucleic acids. Since their hydrogen bonding patterns differ from the canonical form **1** they could be associated with the formation of distinct noncanonical base pairs. The present results, in line with preceding studies, show that the imino tautomers are substantially destabilized compared to the canonical form in the polar solvent. This means that under normal conditions such tautomers are not present. Obviously, high rates of their formation would likely cause high mutation rates, thus the genetic material should be protected from their formation. It is nevertheless important to note that the imino tautomers are destabilized by environmental effects while the intrinsic energy terms would lead to a rather significant probability of imino tautomer formation.

Nevertheless, the destabilization of cytosine rare tautomers is not sufficient to exclude their involvement under certain circumstances and in certain environments, allowing expression of the intrinsic gas-phase trends. For example, the formation of rare tautomers can be boosted by favorable molecular interactions improving their free energy compared to the canonical amino form. In fact, such involvement of imino tautomers has already been suggested for several unusual nucleic acid structures: the pyrimidine–purine–pyrimidine triplex,[7a,e] the four-stranded intercalated i-DNA[7b,c] and the parallel stranded DNA duplex.[7d,f] Although the available experimental data as well as the computed results so far do not suggest that the tautomers are the dominant species in these molecules, their transient formation cannot be ruled out. The imino tautomer would certainly be involved in the case of a double-proton transfer mechanism of point mutations[69] and this process is likely to be substantially sensitive to the energy difference separating the cytosine tautomers. The formation of species formally identical to the imino tautomers is caused by metalation of the cytosine amino group,[8,9] although we have argued that the electronic structure of these metal-assisted imino tautomers is more similar to N3-protonated cytosine.[9] After the amino-metalation, the metal-assisted imino tautomer is the dominating species, which leads to mispairing of cytosine. Although no atomic resolution structure of amino-metalated oligonucleotide is available, the metal-assisted tautomers have been extensively studied for smaller systems. It should also be pointed out that the probability of formation of cytosine tautomers can be substantially enhanced by reducing the cytosine exposure to the solvent, which would shift the tautomeric equilibrium towards the above micro-hydration equilibria or even the gas-phase values. This may occur for example in biomolecular complexes of nucleic acids with proteins. Note that in the case of canonical Watson–Crick base pairing, the interaction

Table 6 MST, MC–FEP and MD–TI free energies of hydration (in kcal mol^{-1}) and free energy differences (ΔG_{sol}[a]) in aqueous solution for cytosine tautomers

Structure	MST	MC-FEP	MD-TI	Average	ΔG_{sol}
1	0.0	0.0	0.0	0.0	0.0
2a	+7.1	+7.5	+6.5	+7.0	+6.4
2b	+6.5	+5.5	+5.0	+5.7	+5.8
3a	+4.1	+3.5	+3.7	+3.8	+7.3
3b	+4.6	+4.0	+3.2	+3.9	+5.9

[a] ΔG_{sol} is obtained by combining the gas-phase relative free energies and the average hydration free energy difference.

View Article Online

with guanine base is another factor stabilizing the cytosine amino form, more important than the solvent screening. Once the cytosine is, for example, flipped out of the double helix into some rather hydrophobic environment, or involved in noncanonical contacts and base pairs, its imino-tautomerisation could be considerably easier.

Acknowledgements

This project, LN00A032 (Center for Complex Molecular Systems and Biomolecules) was supported by the Ministry of Education of the Czech Republic. We are grateful to Prof. J. Tomasi for providing us with his original code of the PCM model, which was modified to carry out the MST calculations. Financial support from the Dirección General de Investigación Científica y Técnica (grants PB98–1222 and PM99–0046) is also acknowledged. Some computations were carried out in the Supercomputer Center, Brno.

Published on 25 July 2002. Downloaded by Pennsylvania State University on 18/09/2016 22:05:31.

References

1 G. A. Jefrey, W. Saenger, *Hydrogen Bonding in Biological Structures*, Springer, Berlin, 1991.
2 J. D. Watson and F. H. C. Crick, *Nature*, 1953, **171**, 737.
3 W. Saenger, *Principles of Nucleic Acid Structure*, Springer Verlag, New York, 1984.
4 B. H. Geirstanger and D. E. Wemmer, *Annu. Rev. Biophys. Biomol. Struct.*, 1995, **24**, 463.
5 J. S. Kwiatkowski and B. Pullman, *Adv. Heterocycl. Chem.*, 1975, **18**, 199.
6 M. D. Topal and J. R. Fresco, *Nature*, 1976, **260**, 285.
7 (*a*) R. Soliva, C. A. Laughton and F. J. Luque, *Chem. Soc.*, 1998, **120**, 6147; (*b*) J. Šponer, J. Leszczynski, V. Vetterl and P. Hobza, *J. Biomol. Struct. Dyn.*, 1996, **13**, 695; (*c*) N. Špackova, I. Berger, M. Egli and J. Šponer, *J. Am. Chem. Soc.*, 1998, **120**, 6147; (*d*) N. U. Zhanpeisov, J. Šponer and J. Leszczynski, *J. Phys. Chem. A*, 1998, **102**, 10 374; (*e*) R. Soliva, F. J. Luque and M. Orozco, *Nucleic Acids Res.*, 1999, **27**, 2248; (*f*) D. Barsky and M. E. Colvin, *J. Phys. Chem. A*, 2000, **104**, 8570.
8 (*a*) B. Lippert, H. Schollhorn and U. Thewalt, *J. Am. Chem. Soc.*, 1986, **108**, 6616; (*b*) F. Zamora, M. Kunsman, M. Sabat and B. Lippert, *Inorg. Chem.*, 1997, **36**, 1583.
9 J. Šponer, J. E. Šponer, L. Gorb, J. Leszczynski and B. Lippert, *J. Phys. Chem. A*, 1999, **103**, 11 406.
10 P. Beak, *Acc. Chem. Res.*, 1997, **10**.
11 (*a*) P. Hobza and J. Šponer, *Chem. Rev.*, 1999, **99**, 3247; (*b*) J. Šponer and P. Hobza, *J. Phys. Chem.*.
12 (*a*) G. Fogarasi, *J. Mol. Struct.*, 1997, **413**, 271; (*b*) C. Aleman, *Chem. Phys.*, 2000, **253**, 13; (*c*) A. Les, L. Adamowicz and R. J. Bartlett, *J. Phys. Chem.*, 1989, **93**, 4001; (*d*) D. A. Estrin, L. Paglieri and G. Corongiu, *J. Phys. Chem.*, 1994, **98**, 5653; (*e*) T. K. Ha, H. J. Keller, R. Gunde and H. H. Gunthard, *J. Phys. Chem. A*, 1999, **103**, 6612; (*f*) N. Russo, M. Toscano and A. Grand, *J. Phys. Chem. B*, 2001, **105**, 4735; (*g*) N. Russo, M. Toscano and A. Grand, *J. Am. Chem. Soc.*, 2001, **23**, 10 272; (*h*) M. J. Nowak, L. Lapinski and J. Fulara, *Spectrochim. Acta Part A*, 1989, **45**, 229; (*i*) G. Fogarasi, *J. Phys. Chem. A*, 2002, **106**, 1381
13 T. van Mourik, D. M. Benoit, S. L. Price and D. C. Clary, *Phys. Chem. Chem. Phys.*, 2000, **2**, 1281.
14 D. C. Clary, D. M. Benoit and T. van Mourik, *Acc. Chem. Res.*, 2000, **33**, 441.
15 R. Kobayashi, *J. Phys. Chem. A*, 1998, **102**, 10 813.
16 J. R. Sambrano, A. R. Souza, J. J. Queralt and J. Andrés, *Chem. Phys. Lett.*, 2000, **317**, 437.
17 E. S. Kryachko, M. T. Nguyen and T. Zeegers-Huyskens, *J. Phys. Chem. A*, 2001, **105**, 1288.
18 E. S. Kryachko, M. T. Nguyen and T. Zeegers-Huyskens, *J. Phys. Chem. A*, 2001, **105**, 1934.
19 A. K. Chandra, M. T. Nguyen and T. Zeegers-Huyskens, *J. Mol. Struct.*, 2000, **519**, 1.
20 C. Colominas, F. J. Luque and M. Orozco, *J. Am. Chem. Soc.*, 1996, **118**, 6811.
21 L. Gorb and J. Leszczynski, *Int. J. Quantum Chem.*, 1998, **70**, 4.
22 L. Gorb and J. Leszczynski, *Int. J. Quantum Chem.*, 1997, **65**, 996.
23 P. Jurečka, P. Nachtigall and P. Hobza, *Phys. Chem. Chem. Phys.*, 2001, **3**, 4578.
24 R. A. Kendall and H. A. Früchtl, *Theor. Chem. Acc.*, 1997, **97**, 158.
25 M. Feyereisen, G. Fitzgerald and A. Komornicki, *Chem. Phys. Lett.*, 1993, **208**, 359.
26 O. Vahtras, J. Almlöf and M. W. Feyereisen, *Chem. Phys. Lett.*, 1993, **213**, 514.
27 D. E. Bernholdt and R. J. Harrison, *Chem. Phys. Lett.*, 1996, **250**, 470.
28 R. Ahlrichs, M. Bär, M. Häser, H. Horn and C. Kölmel, *Chem. Phys. Lett.*, 1989, **162**, 165.
29 A. Schaefer, H. Horn and R. Ahlrichs, *J. Chem. Phys.*, 1992, **97**, 2571.
30 A. Schaefer, C. Huber and R. Ahlrichs, *J. Chem. Phys.*, 1992, **97**, 2571.
31 T. H. Dunning, Jr., *J. Phys. Chem. A*, 2000, **104**, 9062.
32 D. G. Truhlar, *Chem. Phys. Lett.*, 1998, **294**, 45.
33 P. L. Fast, M. L. Sánchez and D. G. Truhlar, *J. Phys. Chem. A*, 1999, **103**, 5129.
34 A. J. C. Varandas, *J. Chem. Phys.*, 2000, **113**, 8881.
35 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P. M. W. Gill, B. G. Johnson, W. Chen, M. W. Wong, J. L. Andres, M. Head-Gordon, E. S. Replogle and J. A. Pople, *Gaussian 98* (Revision A.7), Gaussian, Inc., Pittsburgh PA, 1998.
36 MOLPRO is a package of *ab initio* programs written by H.-J. Werner and P. J. Knowles, with contributions from J. Almlof, R. D. Amos, A. Berning, P. Celani, D. L. Cooper, M. J. O. Deegan, A. J. Dobbyn, F. Eckert, S. T. Elbert, C. Hampel, G. Hetzer, R. Lindh, A. W. Lloyd, W. Meyer, A. Nicklass, T. Korona, K. Peterson, R. Pitzer, G. Rauhut, A. J. Stone, P. R. Taylor, M. E. Mura, P. Pulay, M. Schutz, H. Stoll and T. Thorsteinsson.
37 W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. E. Caldwell and P. Kollman, *J. Am. Chem. Soc.*, 1995, **117**, 5179.
38 J. Tomasi and M. Persico, *Chem. Rev.*, 1994, **94**, 2027.
39 C. J. Cramer and D. G. Truhlar, *Chem. Rev.*, 1999, **99**, 2161.
40 M. Orozco and F. J. Luque, *Chem. Rev.*, 2000, **100**, 4187.
41 P. R. Butler and A. M. Ellis, *Mol. Phys.*, 2001, **99**, 525.
42 S. Tsuzuki, T. Uchimaru, K. Matsumura, M. Mikami and K. Tanabe, *Chem. Phys. Lett.*, 2000, **319**, 547.
43 M. Kabeláč, F. Ryjáček and P. Hobza, *Phys. Chem. Chem. Phys.*, 2000, **2**, 4906.
44 M. Kabeláč and P. Hobza, *J. Phys. Chem. B*, 2001, **105**, 5804.
45 M. Kratochvíl, O. Engkvist, J. J. Šponer, P. Jungwirth and P. Hobza, *J. Phys. Chem. A*, 1998, **102**, 6921.
46 F. G. Amar and S. Berry, *J. Chem. Phys.*, 1986, **85**, 5943.
47 A. Heindenreich and J. Jortner, Package of MD programs for molecular clusters, 1992.
48 T. Fox and P. A. Kollman, *J. Phys. Chem. B*, 1998, **102**, 8070.
49 A. J. Hopfinger and R. A. Pearlstein, *J. Comput. Chem.*, 1984, **5**, 486.
50 P. Cieplak, W. D. Cornell, C. I. Bayly and P. A. Kollman, *J. Mol. Biol.*, 1967, **24**, 345.
51 S. F. Boys and F. Bernardi, *Mol. Phys.*, 1989, **214**, 15.
52 C. Curutchet, M. Orozco and F. J. Luque, *J. Comput. Chem.*, 2001, **21**, 1180.
53 M. Orozco, M. Bachs and F. J. Luque, *J. Comput. Chem.*, 1995, **16**, 563.
54 F. J. Luque, Y. Zhang, C. Aleman, M. Bachs, J. Gao and M. Orozco, *J. Phys. Chem.*, 1996, **100**, 4269.
55 F. J. Luque, C. Aleman, M. Bachs and M. Orozco, *J. Comput. Chem.*, 1996, **17**, 806.
56 S. Miertus and J. Tomasi, *Chem. Phys.*, 1982, **65**, 239.
57 S. Miertus, E. Scrocco and J. Tomasi, *Chem. Phys.*, 1981, **55**, 117.
58 F. J. Luque, J. M. López-Bes, J. Cemelí, M. Aroztegui and M. Orozco, *Theor. Chem. Acc.*, 1997, **96**, 105.
59 M. Peterson and R. Poirier, *MonsterGauss*, Department of Biochemistry. Univ. of Toronto. Canada. Version modified by

View Article Online

R. Cammi and J. Tomasi, 1987 and by F. J. Luque and M. Orozco 2000.
60 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926.
61 J. Pranata, S. G. Wierschke and W. L. Jorgensen, *J. Am. Chem. Soc.*, 1991, **113**, 2810.
62 W. L. Jorgensen, *BOSS*, Version 4.2. Yale University, New Haven, CT, 1999.
63 R. D. Brown, P. D. Godfrey, D. McNaughton and A. P. Pierlot, *J. Am. Chem. Soc.*, 1989, **111**, 2308.
64 (*a*) E. D. Radchenko, G. G. Sheina, N. A. Smorygo and Y. P. Blagoi, *J. Mol. Struct.*, 1984, **116**, 387; (*b*) K. Kuczera, M. Szczesniak and K. Szczepaniak, *J. Mol. Struct.*, 1988, **172**, 101; (*c*) K. Szczepaniak, M. Szczesniak and W. B. Person, *J. Am. Chem. Soc.*, 1988, **110**, 8319.
65 (*a*) J. Šponer and P. Hobza, *J. Mol. Struct. (THEOCHEM)*, 1994, **305**, 35; (*b*) J. Šponer and P. Hobza, *J. Phys. Chem.*, 1994, **98**, 3161; (*c*) O. Bludsky, J. Šponer, J. Leszczynski, V. Špirko and P. Hobza, *J. Chem. Phys.*, 1996, **105**, 11 042.
66 (*a*) J. Šponer and P. Hobza, *J. Am. Chem. Soc.*, 1994, **116**, 709; (*b*) J. Šponer, J. Florian, J. Leszczynski and P. Hobza, *J. Biomol. Struct. Dyn.*, 1996, **13**, 827; (*c*) B. Luisi, M. Orozco, J. Šponer, F. J. Luque and Z. Shakked, *J. Mol. Biol.*, 1998, **279**, 1123; (*d*) D. Vlieghe, J. Šponer and L. van Meervelt, *Biochemistry*, 1999, **38**, 16 443.
67 M. Dreyfus, O. Bensaude, G. Dodin and J. E. Dubois, *J. Am. Chem. Soc.*, 1976, **98**, 583.
68 M. Kabeláč and P. Hobza, *Chem. Eur. J.*, 2001, **10**, 2067.
69 J. Florian and J. Leszczynski, *J. Am. Chem. Soc.*, 1996, **118**, 3010.

Published on 25 July 2002. Downloaded by Pennsylvania State University on 18/09/2016 22:05:31.