Charles University in Prague
Faculty of Science

Theoretical Study of Nucleic Acid Bases: Tautomerism, Stability and Properties

Jaroslav Rejnek

Ph.D. Thesis

Advisor: Prof. Ing. Pavel Hobza, DrSc.

Prague 2006

Declaration of the author

I declare that I have worked out this thesis by myself using the cited references.

Prague, July 14, 2006

Jaroslav Rejnek

Acknowledgements

I must first express my gratitude towards my advisor, Professor Pavel Hobza. His leadership, support, attention to detail, hard work, and scholarship have set an example I hope to match some day.

I would like to thank the many friends I have worked with in Pavel Hobza's group.

Finally, I thank my parents for instilling in me confidence and a drive for pursuing my Ph.D. And for the one who was always by my side understanding, tolerant and ready to listen, Martina.

Contents

Appendix A: M. Hanus, M. Kabeláč, **J. Rejnek**, F. Ryjáček, P. Hobza: Correlated *ab initio* Study of Nucleic Acid Bases and Their Tautomers in the Gas Phase, in a Microhydrated Environment, and in Aquaeous Solution. Part 3. Adenine. *J. Phys. Chem. B*, **2004**,108, 2087.

Appendix B: **J. Rejnek,** M. Hanus, M. Kabeláč, F. Ryjáček, P. Hobza: Correlated *ab initio* Study of Nucleic Acid Bases and Their Tautomers in the Gas Phase, in a Microhydrated Environment, and in Aquaeous Solution. Part 4. Uracil and Thymine. *Phys. Chem. Chem. Phys.*, **2005**, 7, 2006.

Appendix C: J. Fanfrlík, **J. Rejnek**, M. Hanus, P. Hobza: Hydration Gibbs Energies of Nucleic Acid Bases Determined by Gibbs Energy Perturbation, Continuous and Hybrid Approaches. *Collect. Czech. Chem. Commun.* **2005**, 70, 1756.

Appendix D: **J. Rejnek**, P. Hobza: H-bonded nucleic acid base pairs containing unusual base pairs tautomers: Complete Basis Set Calculations at the MP2 and CCSD(T) levels. in preparation

1 Introduction

1.1 Subject of the Thesis

In the beginning of computational chemistry, chemists were lucky just to describe for example a hydrogen atom. Since the time computational methods have evolved rapidly. Nowadays theoretical calculations are without a doubt one of the most powerful tools to explain not only the origin and nature of molecular interaction but they are also able to predict many important properties and behavior of studied systems. This unremitting development in computational chemistry enables us to study amazing systems such as proteins, DNA and many other biologically relevant systems. Computational chemistry offers now a more detailed insight into components of biological systems, their interactions and nature of their interactions. Contemporary computational chemistry aims to increase the complexity of studied systems and to use the most biologically relevant environments as possible.

The same is true for the thesis present here. Recent advances in quantum chemistry and molecular dynamics in the last decade have raised the interest of nucleic acid base pairs and their interactions. The structure of DNA and molecular interactions maintaining the structure and the stability of DNA are still one of the most topical questions in chemistry since Watson and Crick found its structure. According to Crick H-bonding could not provide necessary exact specificity since he believed that hydrogens of bases did not have fixed locations but rather they were instantly hopping between possible tautomeric positions.[1] This idea in a slightly different form have survived and is also a cornerstone of this thesis.

Present theoretical study is aimed to estimate the tautomeric equilibrium of nucleic acid bases. At the present time it is not possible to perform calculations only in the gas phase but it is inevitable to consider the effect of native environments. It is not always possible to take all these effects into account, therefore the best balance between accuracy and approximation must be chosen. This kind of studies provides then information not only about system which is studied but also about theoretical method used, their accuracy and comparison with other possible approximations.

This thesis presents a study of nucleic acid bases tautomers in the gas phase, microhydrated environment and in a bulk water environment. It estimates the tautomeric equilibrium of nucleic acid bases in the gas-phase and water environment, compares different

methods estimating solvation free energies and finally reveals accurate interaction energies of tautomeric nucleic acid base pairs.

1.2 Molecular interactions

Covalent interactions (bonds) hold the atoms of biopolymers and small molecules together. Covalent bond energies are on the order of 100 kcal/mol. Covalent bonds do not break when proteins fold or unfold, or when DNA anneals or melts. Those processes (protein folding/unfolding, *etc.*) are controlled by non-covalent interactions. Each given non-covalent interaction is rather weak, often with a free energy difference of less than RT. But their numbers are huge. In huge numbers, small non-covalent forces drive the spontaneous folding or unfolding of proteins and nucleic acids, recognition between complementary molecular surfaces, maintaining the structure of biomolecules of which a double-helical DNA structure represents a clear example.

The subject of our study - nucleic acid bases - are also involved in such interactions. The following interactions are especially important: (i) planar interactions between the bases leading to the formation of the hydrogen-bonded structures stabilized mainly by the electrostatic interactions, (ii) vertical interactions leading to the formation of the stacked structures stabilized mainly by the London dispersion interaction, and (iii) interactions with ions and water molecules.

1.3 Tautomerism

The maintenance of the genetic code relies on the specific hydrogen-bonding recognition between nucleic acid (NA) bases. According to Watson-Crick model, the adenine-thymine (AT) and guanine-cytosine (GC) base pairs are stabilized by two and three hydrogen bonds.[1] Specific patterns of hydrogen bonding donor/acceptor groups are provided by the AT and GC base pairs both in 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.[2,3] 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. Prototopic tautomerism of nucleic acid (NA) bases concerns keto- / enol- and amino- / imino- forms. The eventual importance of prototropic tautomerism was, nonetheless, recognized already by Watson and Crick.[4]

The situation is different with gas-phase experiments where various tautomers coexist. Passing from the gas-phase to bulk water requires a description of hydration effects. Providing the dipole moment of various tautomers differs considerably from that of the canonical form, water can dramatically change the relative stability of various tautomers. Many studies on the tautomeric equilibria of NA bases exist and they directed attention particularly to cytosine and guanine.[5-7] Rare tautomers may be involved in various biochemical processes including the point mutations.[8,9]

1.4 Solvation

It is well-known that the presence of solvent plays an important role in the stabilization of the three-dimensional structure of DNA. Water acts both as a proton acceptor and as a proton donor, and it can affect the structural features that are necessary for the biological function of NA. Depending on the humidity, ionic strength and ratio of bases presented, DNA adopts multiple conformational forms (e.g. B-DNA, A-DNA, and Z-DNA),[2,10] while the most relevant biological architecture is the antiparallel, right-handed B-DNA double helix. Each DNA structure is the result of a balance of a number of interactions of the individual DNA building blocks. Inseparable kind of these interactions are interactions with ions and water molecules, both existing in the natural medium of biomolecules.

A number of methods have been proposed for modeling these interaction computationally. Two general approaches are used: classical ensemble treatments and quantum mechanical continuum models.

2 Methods

2.1 Potential and Free Energy Calculations and Solvation

The theoretical description of NA bases tautomers is not easy and requires usage of correlated quantum-chemical methods and of the most advanced molecular-mechanical methods.

Since it is necessary to take the role of environment into account, it is inevitable to pass from the description of potential energy surface (PES) to the free energy, which means to cover the entropy. The description in terms of free energy is also important when comparing results with experiments since it enables to characterize the system for non-zero temperatures.

In the case of covalent interactions the bonding enthalpies are much higher compared to the entropic term and thus enthalpies determine the value of the free energy. For non-covalently bound complexes such as nucleic acid base pairs the situation is different. The enthalpy is comparable to the entropy due to restrictions of movement after association of subsystems.

2.1.1 Molecular Dynamics / Quenching Technique

Due to the complexity of the potential energy surfaces of molecular complexes involving nucleic acid bases, we have used the molecular dynamics (MD) simulations combined with the quenching (Q) technique to investigate clusters of nucleic acid tautomers with one and two water molecules.

It is, in fact, a combination of molecular dynamics and energy minimization. MD/Q simulations were carried out using the modified Cornell et al. potential[11] in the NVE microcanonical ensamble. The quenching technique should be performed at a temperature high enough (the kinetic energy should be higher than the highest energy barrier between two minima of the PES) so that all representative conformations are sampled. The basic procedure of quenching consists of stopping the MD simulation repeatedly after a limited number of steps, removing the kinetic energy term and performing a non restricted minimization using the conjugate gradient method. The energies and coordinates of the resulting minima are stored and subsequently the MD simulation continues from the point where it was stopped. By

using the quenching technique it is possible to localize not only the global minimum but all the local minima as well. The MD/Q procedure yields not only the PES but also the free energy surface (FES). In the latter case one determines the population of individual energy minima during a long MD/Q simulation. This population is directly proportional to the change of the free energy of a system. Long runs of the MD/Q simulations allow construction of the complete FES.

2.1.2 Molecular Mechanics Free Energy Calculations

Calculation of thermodynamic quantities from molecular simulation is based on the principles of statistical mechanics. The classical statistical mechanics provides well-known relation between the free energy and partition function, however, it is not suited for a practical calculation due to slow convergence of the free energy. Instead of such equation we can introduce the free energy difference between two states:

$$\Delta G = -RT \ln(e^{-\frac{\Delta V}{RT}}) \qquad (1)$$

In practice, a perturbation is applied to force the transition from one state to another. The desired quantity is then averaged through the simulated thermodynamic ensemble. The results are the same when averaging through the time or the space.

Then, statistical procedures are used to calculate the work done on the system by the perturbation. The free energy is a state function, which means that the free energy difference is only depending on the initial and final state, no matter what path is taken to go from one to the other. As a consequence, one can choose any nonphysical path to perform your calculations as long as they can be related through thermodynamic cycles (see chapter 2.1.2.2.) to the physical process you are interested in.

2.1.2.1 Molecular Dynamics - Thermodynamic Integration

The Hamiltonian is made dependent of a coupling parameter λ representative of the state of the system. The free energy difference is evaluated as:

$$\Delta G_{BA} = G(B) - G(A) = \int_{\lambda_A}^{\lambda_B} \frac{\partial G(\lambda)}{\partial \lambda} d\lambda = \int_{\lambda_A}^{\lambda_B} \left\langle \frac{\partial H(\lambda)}{\partial \lambda} \right\rangle_{\lambda'} d\lambda \quad (2)$$

The relative free energy between two states A and B is expressed as an integral from λ_A to λ_B over the ensemble average of the derivative of the Hamiltonian with respect to the coupling parameter λ. The integration can be performed continuously while slowly changing the coupling parameter λ from λ_A to λ_B during the simulation (the so-called slow growth method). However this approach can be problematic when the system lag behind the changing Hamiltonian and never equilibrate appropriately. A more controlled approach is to simulate the system at a number of fixed λ points and to evaluate the integral numerically. This way the convergence of the simulations at each λ point can be checked independently. Then the method is also based on the coupling parameter approach but it estimates the free energy from the probability of finding the system in either of the two states:

$$\Delta G_{BA} = -k_B T \ln \left\langle \exp^{-(H(\lambda_B) - H(\lambda_A))/k_B T} \right\rangle_{\lambda_A} \quad (3)$$

The free energy difference is calculated as an ensemble average over the state A but the equation can equally be written as an ensemble average over the state B:

$$\Delta G_{BA} = +k_B T \ln \left\langle \exp^{-(H(\lambda_A) - H(\lambda_B))/k_B T} \right\rangle_{\lambda_B} \quad (4)$$

In finite sampling, the ensemble average only converges to the correct answer if configurations sampled in state A also have a high probability in state B. The end state B must therefore not be too different from the reference state A.

The extent of the sampling reached during the simulation will be one of the two primarily limiting factors concerning the accuracy of the calculations. The other one being the underlying model, or force field, used to describe the system.

2.1.2.2 Thermodynamic cycle

A thermodynamic cycle refers to how a system can go through a number of changes and return to the same state finally. In general we calculate the free energy differences between two states or systems. The thermodynamic cycle is another demonstration of state behavior of free energy. It uses the fact that all possible paths changing one system to another are equivalent, and thus calculated energies equal. We can draw the thermodynamic cycle for many different systems such as enzyme – substrate binding or solvation. (Figure 1)

The resulting equation for the difference in the free energy changes is as follows:

$$\Delta\Delta G = \Delta G_A\,(solv.) - \Delta G_B\,(solv.) = \Delta G_{AB}\,(vac.) - \Delta G_{AB}\,(aq.) \qquad (5)$$

In our case the term $\Delta\Delta G$ refers to the change of free energy of solvation between two different NA bases tautomers ant thus it informs us about the change of equilibrium in water.

Figure 1. Thermodynamic cycle of the solvation (A and B are two different systems – chemically or geometrically; ΔG are individual changes of the free energy during conversion or solvation).

Let us have a closer look at our thermodynamic cycle. According to the equation *(5)* we have, in principle, two possibilities how to calculate the $\Delta\Delta G$. The first one is to calculate ΔG_A (solv.) and ΔG_B (solv.). That means to calculate in one single step the free energy difference of the change of tautomeric form and simultaneously the change of environment. This kind of calculation is very difficult, almost infeasible. The second possibility is much easier and includes calculation of ΔG_{AB} (vac.) and ΔG_{AB} (aq.). These simulations cover the free energy difference between tautomeric forms in vacuo (ΔG_{AB} (vac.)) and in water (ΔG_{AB}

(aq.)). The resulting $\Delta\Delta G$ is then calculated by substracting these two values according to the equation *(5)*.

2.1.2.3 Soft Core Method

The linear interpolation of the Lennard-Jones and Coulomb potentials gives problems when growing particles out of nothing or when making particles disappear (λ close to 0 or 1). The problem arises when states A and B are extremely different, and thus the range of transformation is very large and computationally demanding. To circumvent these problems, the singularities in the potentials need to be removed. This is done with soft-core potentials.[12]

2.1.3 Implicit Solvation Model

Solvent treatment within a computational model can range from a less-detailed description such as by a continuum model to a more explicit, atomistic description, such as is employed in a molecular dynamics simulation (discussed previously). A continuum model in computational molecular sciences can be defined as a model in which a number of the degrees of freedom of the constituent particles (a large number, indeed) are described in a continuous way, usually by means of a distribution function.

The solvation free energy (ΔG_{sol}) is the free energy change to transfer a molecule from vacuum to solvent. The solvation free energy can be considered to have three components:

$$\Delta G_{sol} = \Delta G_{elec} + \Delta G_{vdw} + \Delta G_{cav} \qquad (6)$$

where ΔG_{elec} is the electrostatic component. This contribution is particularly important for polar and charged solutes due to the polarisation of the solvent, which we model as a uniform medium of constant dielectric ε. ΔG_{vdw} is the van der Waals interaction between the solute and solvent. ΔG_{cav} is the free energy required to form the solute cavity within the solvent. This component is positive and comprises the entropic penalty associated with the reorganisation of the solvent molecules around the solute together with the work done against the solvent pressure in creating the cavity.

The electrostatic contribution (ΔG_{elec})

The electrostatic component of the free energy of solvation is usually derived by placing a charge or a dipole in a spherical or ellipsoidal cavity. The first contributions to the study of solvation effects were made by Born and Onsager.[13] The solute dipole within the cavity induces a dipole in the surrounding medium, which in turn induces an electric field within the cavity – the reaction field. The reaction field then interacts with the solute dipole, so providing additional stabilisation of the system. The magnitude of the reaction field was determined by Onsager to be:

$$\phi_{RF} = \frac{2(\varepsilon - 1)}{(2\varepsilon + 1)a^3}\mu \qquad (7)$$

where μ is the dipole moment of the solute; a and ε are the radius of the cavity and the dielectric constant of the medium. The energy of a dipole in an electric field ϕ_{RF} is $-\phi_{RF}\,\mu$, but for a polarisable dipole it is necessary to add an additional term which represents the work done assembling the charge distribution within the cavity $\phi_{RF}\,\mu/2$.

The reaction field model can be incorporated into quantum mechnics, where it is commonly reffered to as the self-consistent reaction field (SCRF) method. The reaction field is considered to be a perturbation of the Hamiltonian for an isolated molecule. The modified Hamiltonian of the system is then given by:

$$H_{tot} = H_0 + H_{RF} \qquad (8)$$

where H_0 is the Hamiltonian of the isolated molecule and H_{RF} is the perturbation:[14]

$$H_{RF} = -\hat{\mu}^T \frac{2(\varepsilon - 1)}{(2\varepsilon + 1)a^3}\langle\psi|\hat{\mu}|\psi\rangle \qquad (9)$$

where $\hat{\mu}$ is the dipole moment operator written in matrix form and $\hat{\mu}^T$ is its transpose.The electrostatic contribution to the solvation free energyis then given by:

$$\Delta G_{elec} = \langle \psi | H_{TOT} | \psi \rangle - \langle \psi_0 | H_0 | \psi_0 \rangle + \frac{2(\varepsilon - 1)}{(2\varepsilon + 1)a^3} \frac{1}{2} \mu^2 \qquad (10)$$

The third term in equation *(10)* is the correction factor corresponding to the work done in creating the charge distribution of the solute within the cavity in the dielectric medium. ψ_0 is the gas-phase wavefunction.

A drawback of the SCRF method is its use of a spherical cavity since molecules are rarely exactly spherical in the shape. The more realistic cavity shape is obtained from the van der Waals radii of the atoms of the solute. This approach called polarisable continuum method (PCM) has been implemented in a variety of *ab initio* and semi-empirical quantum mechanical programs. The cavity surface is divided into a large number of small surface elements, and there is a point charge assocoiated with each surface element. This system of point charges represents the polarisation of the solvent, and the magnitude of each surface charge is proportional to the electric field gradient at that point. The total electrostatic potential at each surface element equals the sum of the potential due to the solute and the potential due to the other surface charges:

$$\phi(r) = \phi_\rho(r) + \phi_\sigma(r) \qquad (11)$$

where $\phi_\rho(r)$ is the potential due to the solute and $\phi_\sigma(r)$ is the potential due to the surface charges.

First, the cavity surface is determined from the van der Waals radii of the atoms. That fraction of each atom`s van der Waals sphere which contributes to the cavity is then divided into a number of small surface elements of calculable surface area. An initial value of the point charge for each surface element is then calculated from the electric field gradient due to the solute alone:

$$q_i = -\left[\frac{\varepsilon - 1}{4\pi\varepsilon}\right] E_i \Delta S \qquad (12)$$

where ε is the dielectric constant of the medium, E_i is the electric field gradient and ΔS is the area of the surface element. The contribution $\phi_\sigma(r)$ due to the other point charges can then be calculated using Coulomb`s law. These charges are modified iteratively until they are

self-consistent. The potential $\phi_\sigma(r)$ from the final part of the charge is then added to the solute Hamiltonian ($H = H_0 + \phi_\sigma(r)$) and the SCF calculation initiated. After each SCF calculation new values of the surface charges are calculated from the current wavefunction to give a new value of $\phi_\sigma(r)$ which is used in the next iteration until the solute wavefunction and the surface charges are self-consistent.

The calculation of ΔG_{elec} is then similar to the previous case:

$$\Delta G_{elec} = \int \psi H \psi d\tau - \int \psi_0 H_0 \psi_0 d\tau - \frac{1}{2}\int \phi(r)\rho(r)dr \qquad \textit{(13)}$$

where $\rho(r)$ is the charge distribution of the surface elements.

In our studies we used a method very similar to PCM called conductor-like screening model (COSMO).[15] The cavity is considered to be embedded in a conductor with an infinite dielectric constant. The advantage of this is that screening effects in an infinitely strong dielectric (i.e. a conductor) are much easier to handle. On the surface of a conductor the potential due to the solute and due to the surfaces charges is set to zero, which gives rise to a convenient boundary condition when determining the surface charges. For an alternative dielectric these charges are scaled by a factor:

$$q` = q\frac{\varepsilon_r - 1}{\varepsilon_r + 0{,}5}. \qquad \textit{(14)}$$

2.2 Total Interaction Energies

The interaction energy ΔE of a dimer A...B is defined as the electronic energy difference between the energy of the dimer ($E^{A...B}$) and the sum of the energies isolated monomers (E_A, E_B).

To calculate interaction energies in this way leads to an overestimation of a true value. The discrepancy arises from the phenomenom called Basis Set Superposition Error (BSSE). As the atoms of interacting molecules (or of different parts of the same molecule) approach one another, their basis functions overlap. Each monomer "borrows" functions from other nearby components, effectively increasing its basis set and improving the calculation of derived properties such as energy. If the total energy is minimised as a function of the system geometry, the short-range energies from the mixed basis sets must be compared with the long-range energies from the unmixed sets, and this mismatch introduces an error. The BSSE would be expected to be particularly significant when small basis sets are used and decreases when larger sets are used.

In general, there exist two different possibilities how to get rid of the BSSE. One way to overcome the BSSE problem is via the counterpoise correction method of Boys and Bernardi,[16] in which the entire basis set is used for all calculation. The calculation of the energy of the subsystem A is performed in the presence of „ghost" orbitals of B (without the nuclei or electrons of B) and similar calculation is performed for B. The second – evident solution would be the use of extremely large basis sets (BSSE decreases with larger basis sets). Unfortunately, the convergence is very slow.

However, BSSE-free energies can be obtained in a similar way to use of large basis sets by extrapolating to the complete basis set limit (CBS). For the extrapolation to the CBS limit we need energies determined by systematically improved basis sets. We used the correlation-consistent polarized valence basis sets of Dunning[17] (cc-pVDZ and cc-pVTZ).

Correlation energy is known to converge far slower to its basis set limit than the HF energy. Therefore both contributions must be extrapolated separately, using different terms. In our work, we used two-point extraplation formulas introduced by Helgaker and co-workers[18,19]:

$$E_X^{HF} = E_{CBS}^{HF} + A\, e^{-\alpha X}, \quad E_X^{corr} = E_{CBS}^{corr} + BX^{-3} \qquad (15)$$

where E_X and E_{CBS} are energies for the basis set with the largest angular momentum X (X = 2 for DZ, 3 for TZ) and for the complete basis set respectively and α is the coefficient from the original work.

Although many scientists deals with the term Interaction Energy, the definition still differs from group to group. Sometimes the interaction energy is for example the same as binding energy, sometimes not. The main difference between these definitions is usualy related to deformation energy E_{Def}.[20,21]

The concept of deformation energy which is always repulsive is connected with the consideration of BSSE in geometry optimization. If the counterpoise-corrected gradient optimization is used (the BSSE is included in each geometry iteration) the deformation energy is not defined. In other case (the BSSE is included a posteriori at the end of the optimization) the deformation energy is defined. It is evaluated as the energy difference between the monomers adopting the final deformed geometry (as adjusted in the complex) and relaxed isolated monomers, all evaluated with the monomer basis set.[20,21] Then the total interaction energy is defined in the following way:

$$\Delta E^{A...B} = E^{A...B} - (E^{A} + E^{B}) + E_{Def} \qquad \textit{(16)}$$

In our study we constantly included the deformation energy into the interaction energy and resulting energy call the total interaction energy.

All calculations were carried out using following software packages:
TURBOMOLE 5.6, 5.7 and 5.8
GAUSSIAN 03
AMBER package (versions 6 and 7)
GROMACS 3.1.4
MOLPRO 2002.6

3 Systems Studied

3.1 Tautomers of Nucleic Acid Bases

Different tautomers of nucleic acid bases are obtained when considering different positions of hydrogen around the base. Altogether , 14 amino and imino tautomers of adenine (Figure 2) and 13 keto and enol tautomers of uracil (Figure 3) and thymine (figure is not presented since the uracil/thymine differ only in the presence of methyl group at position 5) were „created" in this way.

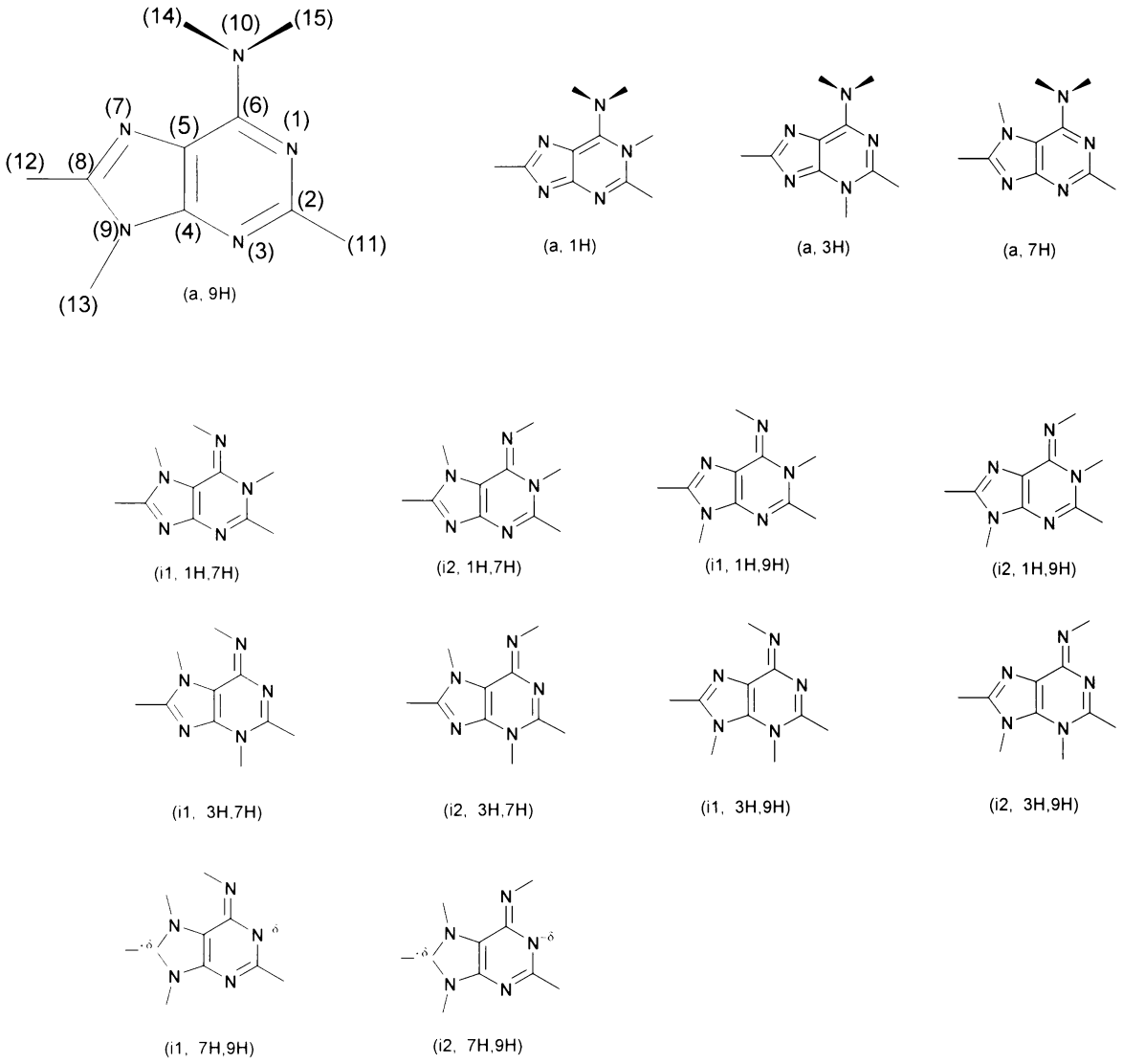

Figure 2. Fourteen adenine tautomers. Standard numbering and adopted nomenclature are presented. Single lines represent hydrogens.

We presented 12 "classical" and two unusual "zwitterions-like" structures of adenine: (i1, 7H,9H) and (i2, 7H,9H).

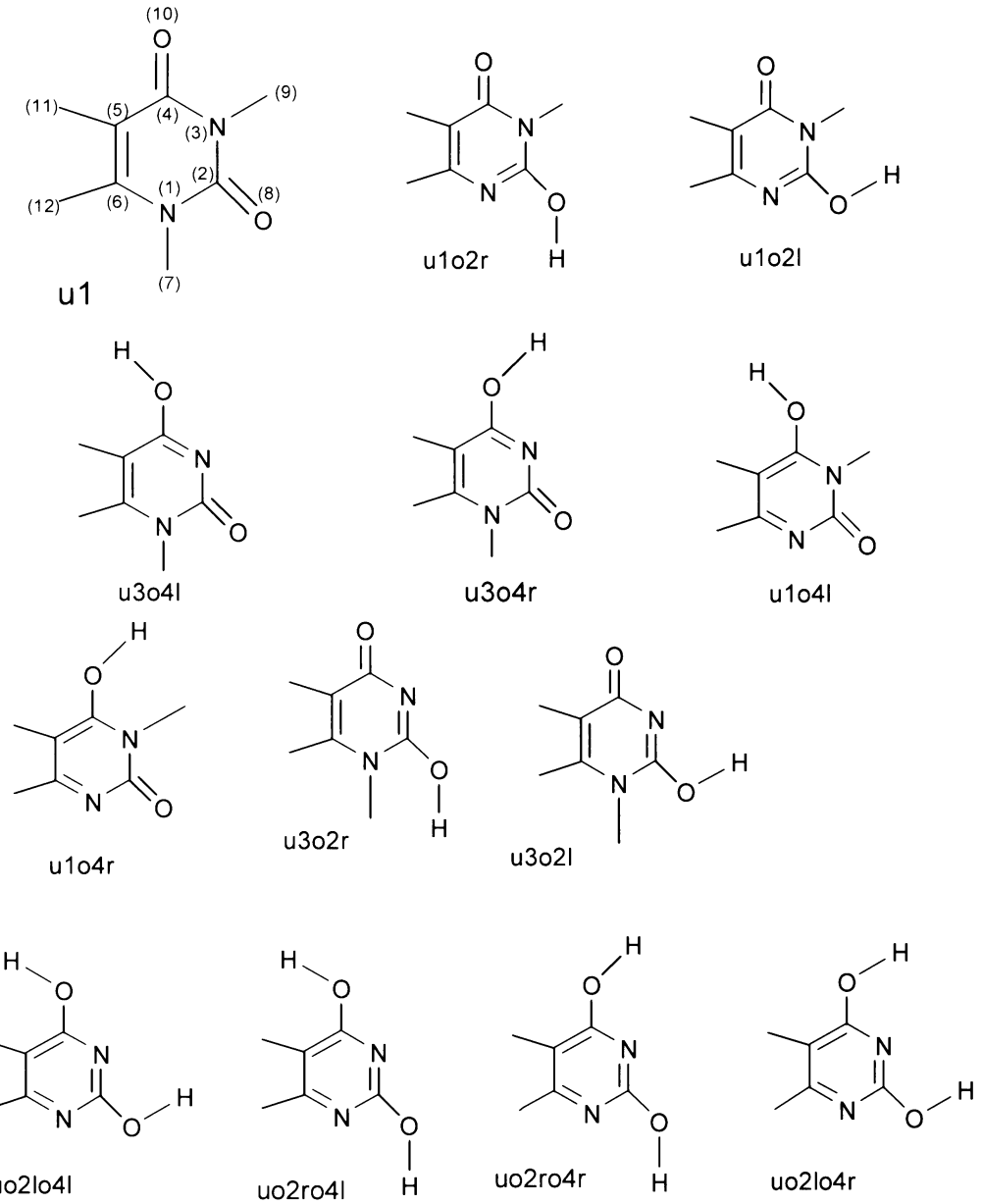

Figure 3. The thirteen uracil tautomers. Standard numbering and adopted nomenclature are presented. In the case of bases, single lines represent hydrogens, hydroxy group hydrogens of bases are depicted explicilty.

The interaction of all these tautomers (altogether 40: 14 adenine, 13 uracil and thymine tautomers) with one and two water molecules was then explored using MD/Q technique. Resulting structures are showed in the Results section (Figures 6, 7 and 8), where they are also discussed.

In the subsequent work we tried to impose our knowledge about calculations in water environment and tried to compare both methods (COSMO and MD-TI) used on a broad-spectrum of NA bases tautomers. Since it was a study of a so called "hybrid model" where microhydrated tautomers (tautomers with one water molecule) play a crucial role, structures of these clusters are presented. Microhydrated structures of adenine, uracil and thymine were also part of the previous study and structures of uracil are depicted in Figure 8 in the Result section. Microhydrated structures of cytosine, guanine and isoguanine are showed here in Figure 4.

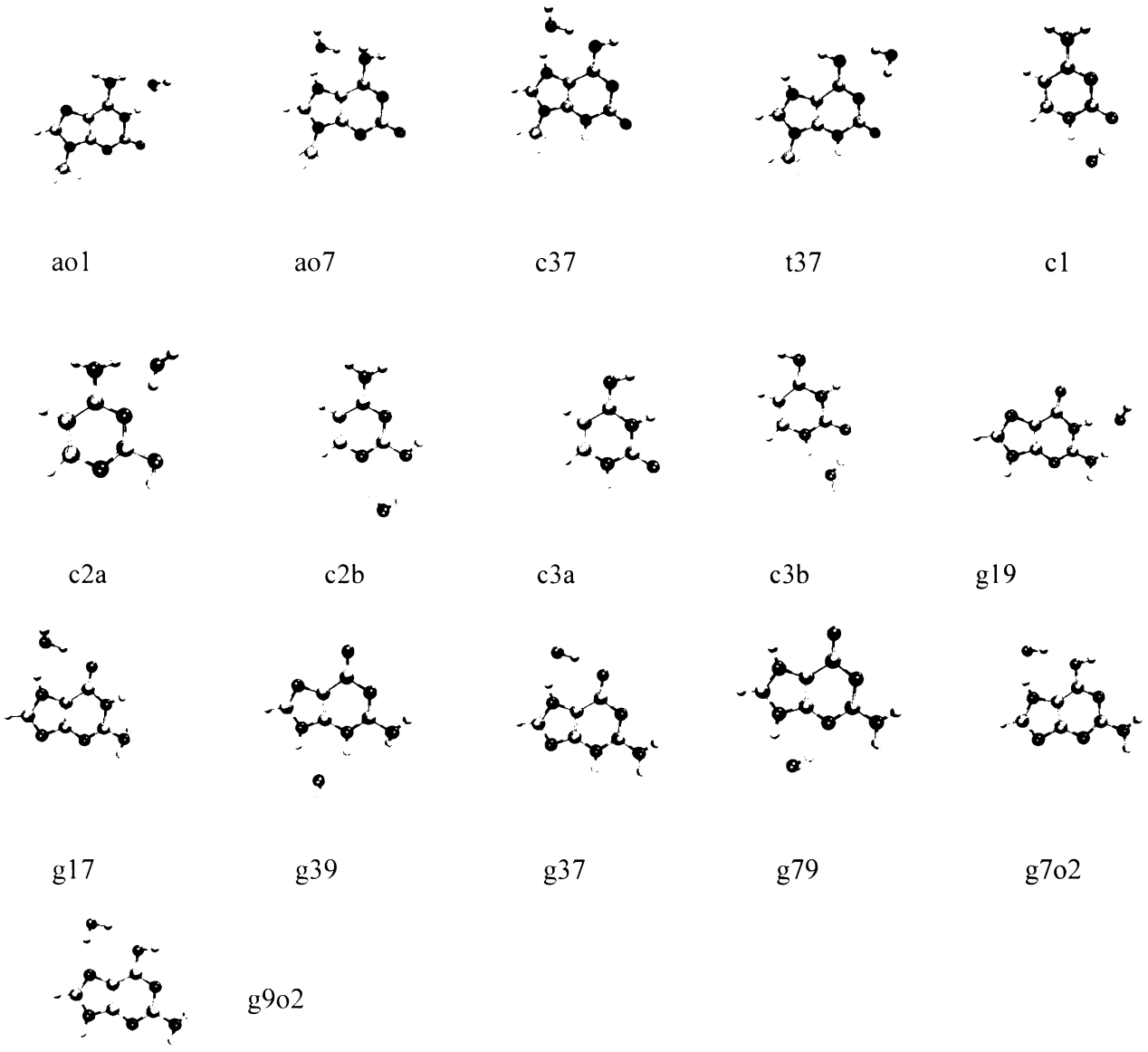

Figure 4. The most populated structures of each isoguanine, cytosine and guanine tautomers optimized at the MP2/cc-pVDZ level of theory.

3.2 Base Pairs of Nucleic Acid Tautomers

The last presented work concerns nucleic acid base pairs created from single tautomeric bases studied in our previous papers.[22,23] Base pairs of tautomeric nucleic acid bases were created using our chemical intuition with respect to geometry restriction. This procedure yielded altogether 15 base pairs: 7 adenine-thymine and 8 guanine-cytosine base pairs. These structures are depicted in the Figure 5. All of these structures could be incorporated to a DNA molecule because the position that binds to sugar remains free.

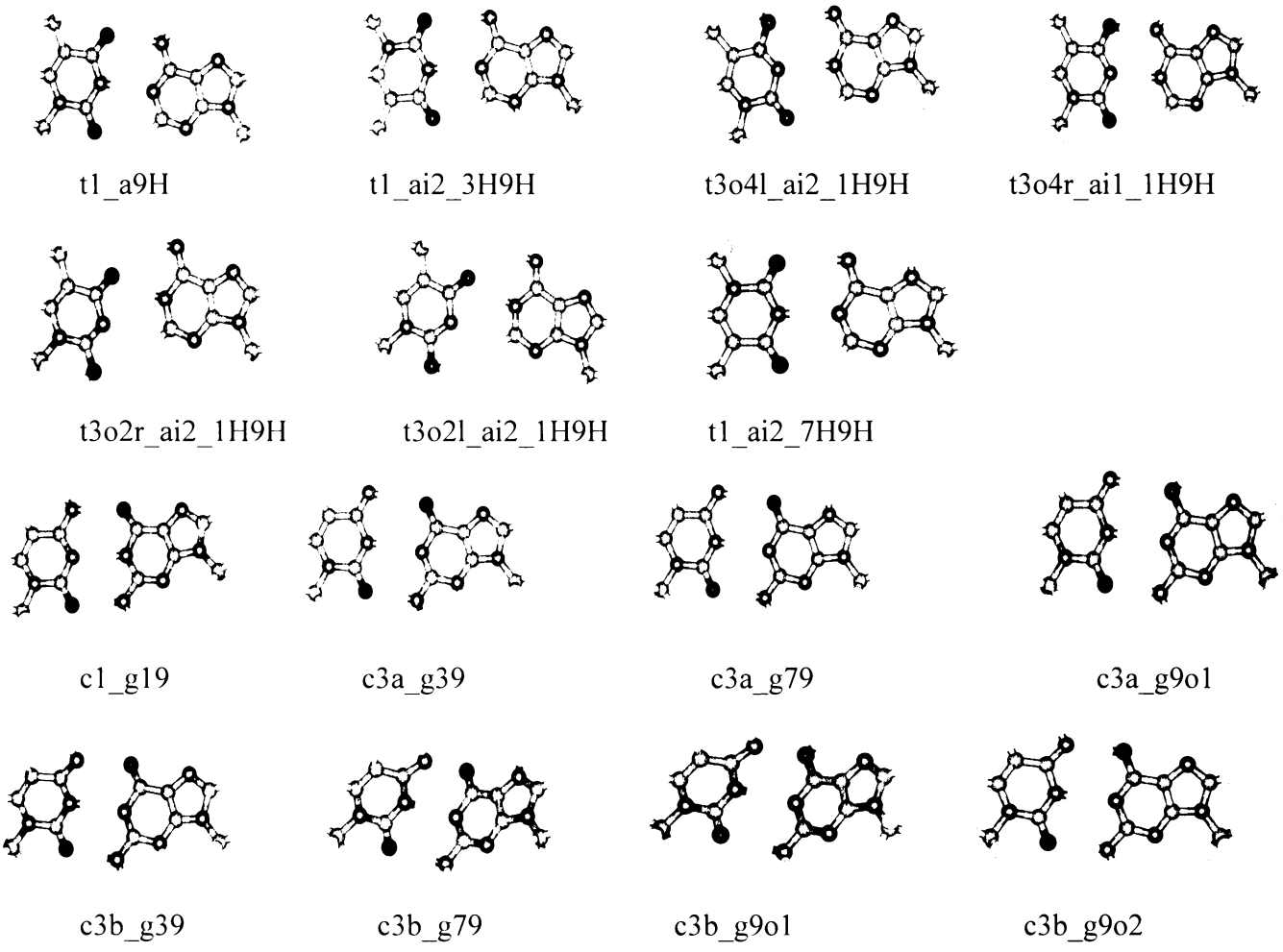

Figure 5. Nucleic acid base pairs containing tautomeric forms of nucleic acid bases.

4 Results

4.1 Equilibrium Properties of Nucleic Acid Tautomers

4.1.1 Adenine

The detailed study of adenine tautomers can be found in Appendix A.

Gas-phase tautomers. The relative stability of four amino forms having hydrogen at nitrogen N9 (canonical form; a, 9H), nitrogen N1 (a, 1H), nitrogen N3 (a, 3H) and nitrogen N7 (a, 7H), eight standard imino forms having hydrogens at nitrogens N7 and N1 (i1, 1H,7H; i2, 1H,7H), nitrogens N9 and N1 (i1, 1H,9H; i2, 1H,9H), nitrogens N7 and N3 (i1, 3H,7H; i2, 3H,7H), nitrogens N9 and N3 (i1, 3H,9H; i2, 3H,9H), and two unusual imino tautomers having hydrogens at nitrogens N7 and N9 (i1, 7H,9H; i2, 7H,9H) were examined (see Figure 2). Table 1 shows their (gas phase) relative energies, enthalpies and free energies.

Table 1. Relative energies (ΔE), zero-point vibration energies (ΔZPVE), and free energies (ΔG) of adenine tautomers in the gas phase. All energies are in kcal/mol.

Method Structure	ΔE(RI-MP2)[a]	ΔE(MP2)[b]	ΔZPVE[c]	$\Delta(G_0^{298}-E)$[c]	ΔG_0^{298} [c]
(a, 9H)	0.00	0.00	0.00	0.00	0.00
(a, 1H)	17.74	18.79	-0.26	-0.35	17.40
(a, 3H)	7.99	9.22	-0.56	-0.56	7.43
(a, 7H)	7.63	7.79	-0.12	-0.16	7.47
(i1, 1H,7H)	16.55	17.61	-0.30	-0.51	16.05
(i2, 1H,7H)	16.09	16.83	-0.25	-0.32	15.77
(i1, 1H,9H)	12.07	12.36	0.05	-0.03	12.05
(i2, 1H,9H)	18.53	19.08	-0.34	-0.44	18.09
(i1, 3H,7H)	24.29	25.84	-0.75	-1.22	23.07
(i2, 3H,7H)	17.47	18.24	-0.30	-0.54	16.93
(i1, 3H,9H)	31.56	32.55	-1.39	-1.68	29.89
(i2, 3H,9H)	31.96	32.56	-2.08	-1.77	30.19
(i1, 7H,9H)	44.96	47.96	-0.86	-1.17	43.79
(i2, 7H,9H)	35.54	37.56	-0.33	-0.44	35.10

[a] RI-MP2/TZVPP//RI-MP2/TZVPP. [b] MP2/aug-cc-pVDZ//RI-MP2/TZVPP. [c] MP2/6-31G**.

Investigating the energy characteristic we found that the canonical form clearly corresponds to the global minimum while the first and second local minima [(a, 7H) and (a, 3H) amino forms] are considerably less stable (by about 8 kcal/mol). The imino isomers (i1, 1H,9H), (i2, 1H,7H), (i1, 1H,7H), (i2, 3H,7H), (i2, 1H,9H) and amino isomer (a, 1H) are energetically less stable (by 12 – 19 kcal/mol). A very large energy difference of more than 30 kcal/mol was found for rare unusual imino tautomers (i, 7H,9H) and (i, 3H,9H). The above-mentioned results were obtained from the RI-MP2/TZVPP calculations.

Final relative energies (ΔE) and relative free energies at 298 K (ΔG_0^{298}) are also summarized in the Table 1. Relative free energies were determined from relative energies and $\Delta(G_0^{298} - E)$ terms (MP2/6-31G**). From the energy, enthalpy and free energy results we can conclude that only the canonical form can exist in a gas phase.

Microhydrated tautomers. The MD/Q simulations on monohydrated and dihydrated tautomers yielded about 4 stable structures (*cf.* Table 2 and Figures 6 and 7) with one water and about 20 stable structures with two water molecules for each tautomer. Four energetically most stable structures were then studied using *ab initio* methods; their structures are presented in Figures 6 and 7. The stability of these structures decreases from left to right.

The relative energies, their interaction energies and relative energies of global minima of these structures are presented in Table 2. Stabilization energies for adenine...water complexes are large, between 8.8 and 19.5 kcal/mol. Large stabilization energies are due to the favorable position of water which forms a bridge between adenine proton donor and acceptor positions.

Following expectation, the largest stabilization energy was found for imino (i, 7H,9H) tautomers possessing the largest dipole moment. Interestingly, however, the second largest stabilization energy belongs to the (i2, 3H,7H) tautomer having only small dipole moment of 3.2 D. In addition, stabilization energies for dihydrated adenine tautomers are large (18.1 – 34.1 kcal/mol) and also here it is due to a very favorable orientation between tautomers and water molecules, and water molecules themselves. Figure 7 shows that global minima of tautomers depicted contain mostly the water dimer motif. As in the previous case the largest stabilization energy was detected in the case of imino (i2, 7H,9H) tautomer which has very large dipole moment. Relative energies of isolated adenine tautomers are repeated in Table 2 and the question arises whether mono- and dihydration change this order.

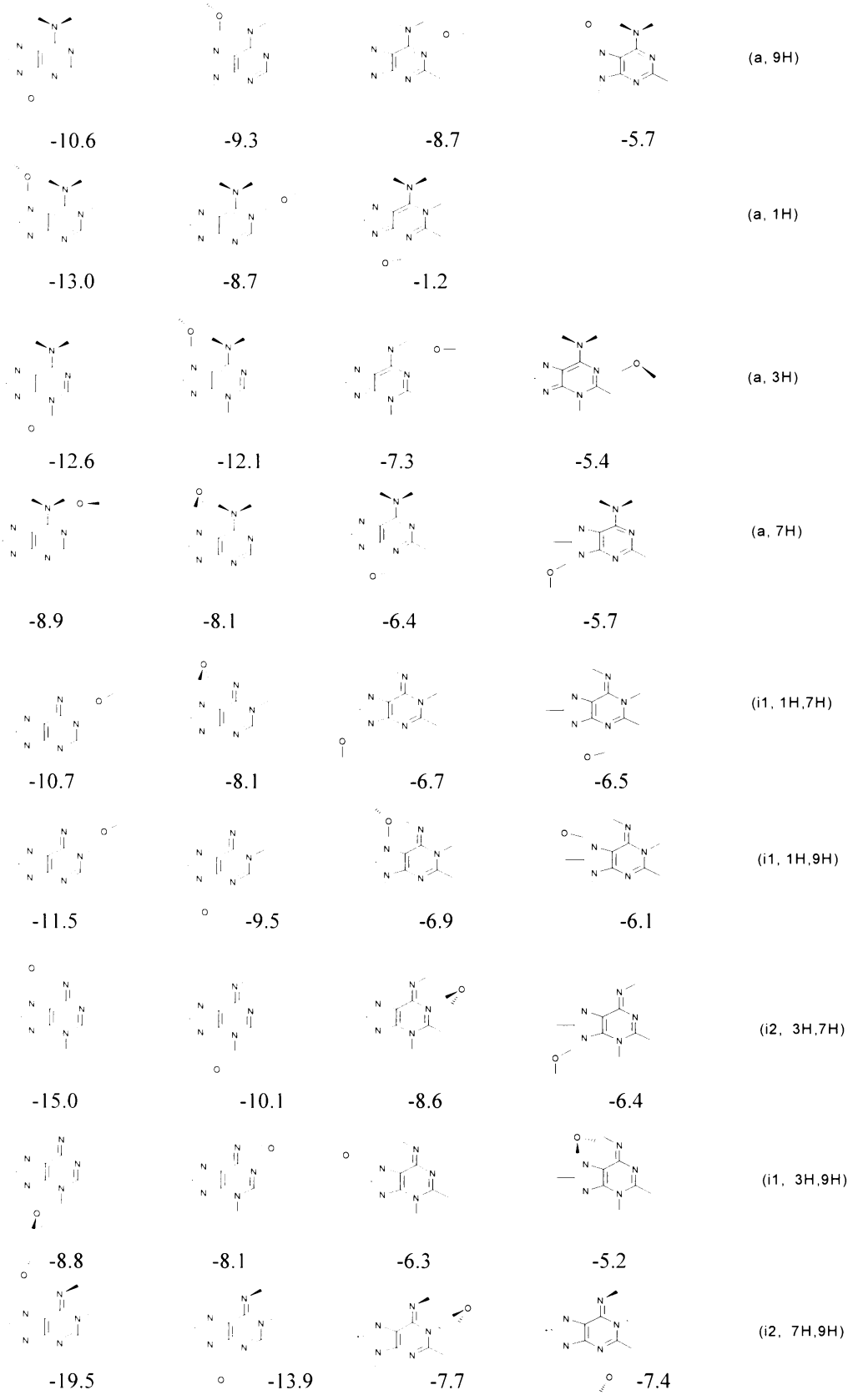

Figure 6. Nine of the most stable structures of adenine tautomers with one water molecule optimized at the RIMP2/TZVPP level of theory. The stability is decreasing from left to right. The interaction energies in kcal/mol (Cf. Table 2, last column) are presented below the structures.

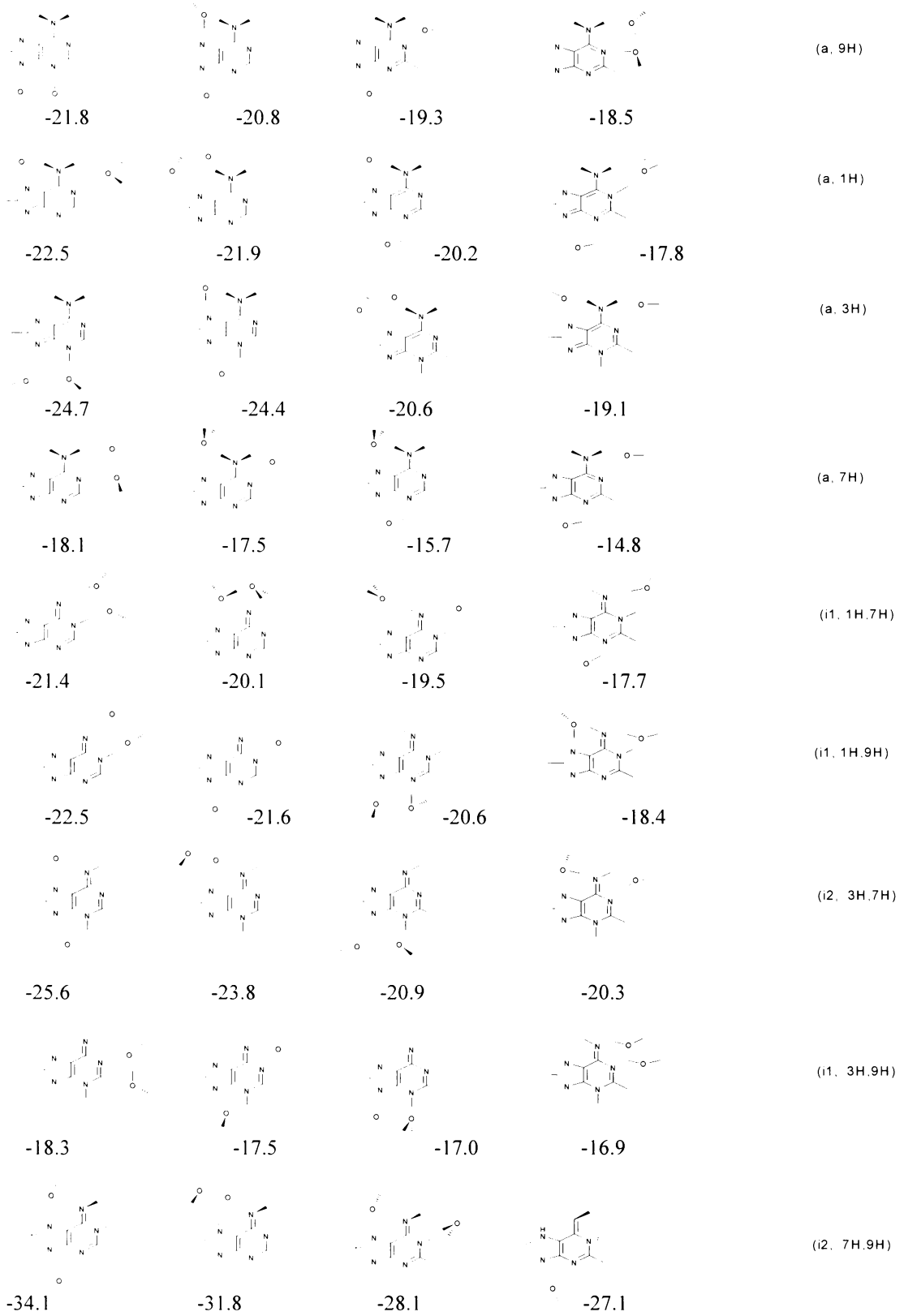

Figure 7. Nine of the most stable structures of adenine tautomers with two water molecules optimized at the RIMP2/TZVPP level of theory. The stability is decreasing from left to right. The interaction energies in kcal/mol (Cf. Table 2, last column) are presented below the structures.

Table 2. Relative and interaction energies (in kcal/mol) of adenine tautomers in the gas phase and mono- and dihydrated environment (global minima for each tautomer-water complex are presented).

Structure	Relative energies[a,b]		Interaction energies[c,d]	
	RI-MP2	RI-MP2$_{ZPVE}$	RI-MP2	RI-MP2$_{TOT}$
(a, 9H)	0.00	0.00	-	-
(a, 1H)	17.74	17.49	-	-
(a, 3H)	7.99	7.43	-	-
(a, 7H)	7.63	7.51	-	-
(i1, 1H,7H)	16.55	16.23	-	-
(i2, 1H,7H)	16.09	16.15	-	-
(i1, 1H,9H)	12.07	11.22	-	-
(i2, 1H,9H)	18.53	18.20	-	-
(i1, 3H,7H)	24.29	24.04	-	-
(i2, 3H,7H)	17.47	17.17	-	-
(i1, 3H,9H)	31.56	30.82	-	-
(i2, 3H,9H)	31.86	31.56	-	-
(i1, 7H,9H)	44.96	42.88	-	-
(i2, 7H,9H)	35.54	34.15	-	-
(a, 9H) – (H_2O)	0.00	-	-11.21	-10.56
(a, 1H) – (H_2O)	14.95	-	-15.01	-12.96
(a, 3H) – (H_2O)	5.81	-	-13.29	-12.61
(a, 7H) – (H_2O)	9.04	-	-9.39	-8.90
(i1, 1H,7H) – (H_2O)	16.28	-	-11.43	-10.73
(i1, 1H,9H) – (H_2O)	10.97	-	-12.33	-11.52
(i2, 3H,7H) – (H_2O)	12.32	-	-16.28	-14.95
(i1, 3H,9H) – (H_2O)	33.42	-	-9.12	-8.81
(i2, 7H,9H) – (H_2O)	25.23	-	-23.21	-19.53
(a, 9H) – ($H_2O)_2$	0.00	-	-23.36	-21.81
(a, 1H) – ($H_2O)_2$	17.48	-	-24.87	-22.54
(a, 3H) – ($H_2O)_2$	4.89	-	-26.60	-24.71
(a, 7H) – ($H_2O)_2$	11.43	-	-19.81	-18.06
(i1, 1H,7H) – ($H_2O)_2$	16.98	-	-23.17	-21.35
(i1, 1H,9H) – ($H_2O)_2$	11.32	-	-24.46	-22.46
(i2, 3H,7H) – ($H_2O)_2$	13.43	-	-27.50	-25.59
(i1, 3H,9H) – ($H_2O)_2$	35.78	-	-18.99	-18.29
(i2, 7H,9H) – ($H_2O)_2$	23.56	-	-36.95	-34.13

[a] Order of relative stability of each tautomer is given with respect to canonical tautomer. Total energies of tautomers, mono- and dihydrated tautomers are considered. RI-MP2$_{ZPVE}$ is defined as a sum of relative RI-MP2 energy and ΔZPVE; the former energy is evaluated with TZVPP basis set while the latter are at the MP2/6-31G** level. [b] For description of abbreviations used for methods cf. notes to Table 1. [c] Interaction energies were evaluated with TZVPP basis set. [d] Total complexation energy RI-MP2$_{TOT}$ is defined as a sum of the RI-MP2 interaction energy and deformation energies of the monomers.

The energy gap between non-hydrated global and local minima is too large which means that only the canonical form can exist in the gas phase. From the Table 2 it follows that stabilization energy of the (a, 3H) tautomer with one water is larger than that of the canonical form which means that the energy gap between these two tautomers (7.43 kcal/mol) is reduced by 2.05 kcal/mol when monohydration is considered. It is true that stabilization energies of other tautomers with one water molecule are larger or even considerably larger than that of the canonical form but the energy destabilization of these tautomers is also large. Considering microhydration results (and extrapolating dihydration to full bulk hydration) it must be stated that the energy gap between canonical and amino (a, 3H) tautomer is considerably reduced which suggests that the latter form might co-exist in a microhydrated environment with the canonical tautomer.

Hydrated tautomers. Relative hydration free energy for adenine tautomers are shown in Table 3, which also gives gas-phase free energy, hydration free energy and free energy of tautomerization in an aqueous solution determined by MD-TI method, C-PCM (COSMO) procedure and hybrid model.

MD-TI Method. Relative hydration free energies determined by the MD-TI method (the third column in Table 3) vary in the surprisingly *broad range of -4 to -21 kcal/mol* and the largest values were found for unusual rare imino (i, 7H,9H) tautomers that correspond with their largest dipole moments. Because, however, the energy destabilization of these two tautomers in the gas phase is in absolute value even larger, the resulting relative free energy of tautomerization (the fourth column in Table 3) for these tautomers is still highly positive which means they are (with respect to the canonical tautomer) destabilized. On the other hand, one of the lowest values of hydration free energies found for both (a, 3H) and (a, 7H) amino tautomers reduces the energy destabilization of these two forms in the gas phase. Consequently, their relative free energies of tautomerization in aqueous solution are less favorable than that of the canonical form by a relatively modest 2.5 and 2.8 kcal/mol, respectively. It should be mentioned that results concerning the (a, 3H) tautomer are supported by the microhydration results.

C-PCM (COSMO) model. When investigating the hydration free energies obtained from the COSMO model (fifth column in Table 3) we found that with exception of the (i2, 7H,9H) structure they agree reasonably with MD-TI data. The largest difference (~ 1.8 kcal/mol) was

found for the (i1,3H,9H) tautomer while for other tautomers it is less than 1.5 kcal/mol. It should be noted that both methods give the same relative trends and they both predicted the same structures, which are the most solvent stabilized structures [(a, 1H), (i1, 3H,9H) and (i2, 7H,9H)].

After performing the optimization at the B3LYP/6-31G* level, followed by a single point calculation at HF/6-31G*/ UAHF level (United Atoms radii optimized for HF/6-31G* level of theory), slightly different solvation free energies resulted. From the sixth column of Table 3 it becomes clear that the "optimized" C-PCM free energies are systematically (in absolute value) larger than the "non-optimized" values.

We can thus presently conclude that performing physically justified optimization resulted in no improvement but also no deterioration of "non-optimized" gas-phase data.

Hybrid model. At first, we included the most strongly interacting water molecule (*cf.* Table 2) to the solvated system. Summing the relative interaction energies of each tautomer and the respective C-PCM (COSMO) free energy we obtained the relative hydration free energies (the seventh column in Table 3). Comparing these values with the C-PCM determined for bare adenine tautomers we found that the present values are closer to the "optimized" results. The difference is mostly small and the only exception represent (i2, 3H,7H) and (i2, 7H,9H) tautomers where the monohydrated results are larger by more than 3 kcal/mol.

A question arises whether it is correct to consider the most strongly bound water from the gas phase microhydration. The other structures, being less stable in the gas-phase, can be favored in a continuum water environment. The eighth column of Table 3 shows that for canonical, (a, 3H) and (a, 7H) tautomers the first local minimum from the gas-phase calculation become the global minimum upon the inclusion of the continuum solvent. The "additional" stabilization range (from 1.1 to 1.7 kcal/mol) is definitely not negligible.

Similarly as in the case of bare adenine we performed the geometry optimization of monohydrated adenines in the presence of continuum solvent. From the Table 3 (the ninth column) it is clear that larger absolute values of hydration free energies resulted but the difference is not dramatic. Worth mentioning is the fact that hydration free energy of the (i2, 7H,9H) tautomer is now comparable (in absolute value even slightly larger) with the respective MD-TI value, which strongly favors the reliability of MD-TI procedure. We believe that the inclusion of a few specific waters is especially important if the solute dipole moment become large. In this case the continuum model is not efficient enough to describe

the solvation and consequently too small (in absolute scale) hydration free energies resulted. This is the case of (i2, 7H,9H) tautomer where C-PCM values were comparable to the MD-TI values only after consideration of explicit water(s). This led us later to test the hybrid model on a broad spectrum of different tautomers (see section 4.1.3.).

From Table 3 it becomes, however, clear that the combination of the specific hydration (estimated including the geometry optimization in the continuum solvent) and C-PCM model yields satisfactory results comparable with MD-TI data.

Consideration of two and three specific water molecules in the C-PCM model yields basically similar results as in the case of monohydration. Comparing the C-PCM results for adenine...$(H_2O)_n$ complexes, where n=0 (bare adenine), 1, 2 and 3, we can state that inclusion of water(s) do not deteriorate the hydration free energies obtained for bare adenines.

Table 3. Relative gas-phase free energies (ΔG_0^{298}), relative free energies of hydration, evaluated using MD-TI method (ΔG^{TI}), COSMO method ($\Delta G^{C\text{-}PCM}$) and hybrid model (ΔG^{HYB}), and relative free energies in aqueous solution ΔG^{298}(TI) of adenine tautomers.

Method Structure[a]	ΔG_0^{298} [b]	ΔG^{TI}	ΔG^{298}(TI)	$\Delta G^{C\text{-}PCM}$	$\Delta G^{C\text{-}PCM}$ opt [c]	ΔG^{HYB}(1w) vac min [d]	ΔG^{HYB}(1w) solv min [e]	ΔG^{HYB}(1w) opt [c, d]	ΔG^{HYB}(2w) vac min [d]	ΔG^{HYB}(2w) solv min [e]	ΔG^{HYB}(2w) opt [c, d]	ΔG^{HYB}(3w) sup [f]	ΔG^{HYB}(3w) opt [c, f]
(a, 9H)	0.00	0.00	0.00	0.00	0.00	0.00	-1.07	0.00	0.00	0.00	0.00	0.00	0.00
(a, 1H)	17.40	-11.49	5.91	-10.05	-14.36	-14.54	-14.54	-17.75	-11.34	-13.28	-13.90	-12.21	-12.30
(a, 3H)	7.43	-4.97	2.46	-3.49	-4.55	-4.46	-5.95	-4.87	-4.71	-4.71	-5.44	-4.82	-5.30
(a, 7H)	7.47	-4.68	2.79	-5.04	-6.88	-5.64	-7.31	-6.97	-4.18	-4.18	-5.53	-5.68	-7.33
(i1, 1H,7H)	16.05	-7.80	8.25	-7.07	-7.35	-7.41	-7.41	-7.69	-7.05	-7.05	-7.44	-4.99	-6.60
(i1, 1H,9H)	12.05	-4.04	8.01	-3.58	-3.70	-4.03	-4.03	-4.26	-3.67	-3.79	-4.11	-2.81	-3.74
(i2, 3H,7H)	16.93	-5.13	11.80	-3.64	-4.03	-7.75	-7.75	-8.82	-3.95	-5.33	-3.66	-5.38	-6.18
(i1, 3H,9H)	29.89	-12.18	17.71	-14.00	-14.83	-15.07	-15.07	-16.07	-10.43	-12.71	-10.07	-12.09	-14.53
(i2, 7H,9H)	35.10	-21.25	13.85	-14.16	-15.79	-18.94	-18.94	-22.44	-18.76	-18.76	-17.87	-17.72	-20.78

[a] Cf. Figure 2. [b] Cf. Table 1. [c] geometries optimized in the continuum solvent.. [d] global minimum in the gas phase.. [e] global minimum in the continuum solvent. [f] superimposition of monohydrated structures. All energies are in kcal/mol.

4.1.2 Uracil and Thymine

The study performed on uracil and thymine tautomers is very similar to the study of adenine and therefore the detailed description of results is omitted and just the general results and conclusions are presented. For the detailed description see Appendix B.

Gas-phase tautomers. *Uracil.* We examined the relative stability of one diketo form having hydrogen both at nitrogen N1 and at nitrogen N3 (canonical form), eight enol forms, and four dienol forms having no hydrogens on both N1 and N3 nitrogens (Cf. Figure 3). Looking at relative RI-MP2 energies in the second column of Table 4, we can state that all enol and dienol forms are energetically much less stable (by 9.4 – 28 kcal/mol) than the canonical tautomer.

RI-MP2 results obtained with extended AO basis set provided accurate relative energy characteristics. Columns 3 and 4 of Table 4 show that the inclusion of the CCSD(T) correction term does not change the relative energy values since these corrections are rather small (the largest one amounts to 0.7 kcal/mol) and thus are not critical. This finding is important since the CCSD(T) calculations are CPU time demanding. Column 5 of Table 4 shows that ΔG values at T=298 K are systematically larger than any energy values which means that the ΔG difference between canonical form and other tautomers is even higher.

Present gas phase results are in perfect agreement with previous theoretical and experimental results proving that only the canonical form can exist in the gas phase. U1o2r and uo2lo4r are the most stable rare tautomers, both being less stable by more than +11 kcal/mol.

Thymine. Investigated thymine tautomers are obviously very similar to that of uracil. Investigating various entries in Table 4, we can only state that absolute and relative energy and free energy values for various thymine tautomers are very similar to these found for uracil tautomers.

Table 4. Relative energies (ΔE), zero-point vibration energies (ΔZPVE), and free energies (ΔG) (all in kcal/mol) of uracil and thymine tautomers in the gas phase.

Tautomer	ΔE(MP2)[a] ΔE(MP2+ZPVE)[b]	ΔE(RI-MP2)[a]	CCSD(T)[c]	ΔE (RI-MP2+CCSD(T))	ΔG[b] (kcal/mol)	μ (Debye)
u1	0.00	0.00	0.00	0.00	0.00	5.0
u1o2l	19.88 (19.09)	17.71	-0.41	17.30	18.99	2.5
u1o2r	11.07 (10.88)	9.86	-0.32	9.54	11.06	3.7
u1o4l	22.28 (21.59)	20.56	-0.36	20.20	21.62	7.8
u1o4r	25.71 (24.66)	23.61	-0.35	23.26	24.49	6.2
u3o2l	19.78 (19.20)	17.73	-0.45	17.29	19.20	7.0
u3o2r	31.31 (29.75)	27.98	-0.72	27.26	29.71	10.0
u3o4l	20.32 (19.73)	17.46	-0.37	17.10	19.75	8.6
u3o4r	12.54 (12.33)	10.81	-0.21	10.60	12.46	5.5
uo2lo4l	17.43 (16.9)	14.54	-0.15	14.38	17.25	4.0
uo2lo4r	12.82 (12.55)	10.56	-0.07	10.49	12.85	2.5
uo2ro4l	17.44 (17.00)	14.45	-0.14	14.31	17.26	4.6
uo2ro4r	11.61 (11.42)	9.44	-0.03	9.41	11.73	1.6
Tautomer	ΔE(MP2)a ΔE(MP2+ZPVE)b	ΔE(RI-MP2)a	CCSD(T) c	ΔE (RI-MP2+CCSD(T))	ΔGb (kcal/mol)	μ (Debye)
t1	0	0.00	0.00	0.00	0.00	4.0
t1o2r	10.71 (10.47)	9.41	-0.13	9.28	10.65	3.2
t1o2l	19.54 (18.66)	17.23	-0.17	17.06	18.47	2.1
t1o4r	27.74 (26.44)	24.18	-0.11	24.07	26.53	6.4
t1o4l	25.18 (24.21)	21.68	-0.21	21.47	24.58	8.1
t3o2r	32.09 (30.40)	27.36	-0.24	27.11	30.79	9.6
t3o2l	19.18 (18.62)	17.25	-0.18	17.07	18.68	6.3
t3o4r	13.35 (13.14)	11.77	-0.34	11.43	13.28	5.8
t3o4l	21.76 (21.17)	19.61	-0.45	19.16	21.26	9.0
to2lo4l	18.49 (17.95)	15.49	-0.14	15.35	18.21	4.4
to2lo4r	14.34 (13.98)	10.55	0.01	10.56	13.13	2.7
to2ro4l	18.47 (17.94)	15.48	-0.14	15.35	18.22	5.0
to2ro4r	13.16 (12.74)	9.45	0.01	9.46	13.46	1.9

[a] RI-MP2/TZVPP//RI-MP2/TZVPP.

[b] MP2/6-31G**

[c] (CCSD(T)/6-31G** - MP2/6-31G**)

Microhydrated tautomers. The four energetically most stable mono- and dihydrated structures of uracil and thymine obtained from the MD/Q simulation were optimized using *ab initio* methods. Only structures of monohydrated uracil are depicted in Figure 8.

The relative and interaction energies of mono- and dihydrated uracil and thymine tautomers are presented in Table 5 where, for the sake of comparison, the relative RI-MP2 energies are presented as well.

Uracil. The stabilization energies of the structures of rare tautomeric forms with a new hydration motif (different from u1 tautomer hydration motif) are rather small (-2.2 and -2.9 kcal/mol) for tautomers u1o2r and u1o4l, where water is located between two CH group. They indicate that the CH site is active for hydrogen bonding. Investigating the entries in Figure 8 and Table 6, we find that the hydration energies of various tautomers are large and vary dramatically between tautomers. These energies correlate approximately to the dipole moment of tautomers and the fact that no tight correlation between these values exists suggest that other energy terms (other than electrostatic) are also responsible for the stabilization of the hydrated tautomer.

Investigating the most stable structures of dihydrated tautomers, we found that hydration sites agree with monohydrated ones. In five of seven cases the second water hydrated the same site and simultaneously made a bridge with the first water. These structures are more stable than structures hydrated at two different sites (without a close contact between water molecules). The unusual hydration motif found in monohydrated uracil (between two CH groups) is not found among the four most stable dihydrated structures, due to the fact that the number of uracil...$2H_2O$ configurations, where water interacts with N and O atoms, is much larger, eliminating low interaction energy sites (i.e. CH sites). Comparing the fourth mono- and dihydrated structures of u1, we found that the second water, which hydrates the CH site along with the first one, causes surprisingly large stabilization of 8.8 kcal/mol. This clearly tells us that the hydration of the CH group might be substantial and should be taken into consideration.

From the comparison of Table 4 and Table 5 it becomes clear that inclusion of the monohydrated structures brings some moderate modifications that cannot change the order of various tautomers. The approach of the second water does not yield any further change of relative energies, which indicates the important role of monohydration.

Thymine. Energies are similar to those of uracil tautomers and also changes in the relative energies of various thymine tautomers upon hydration are similarly small.

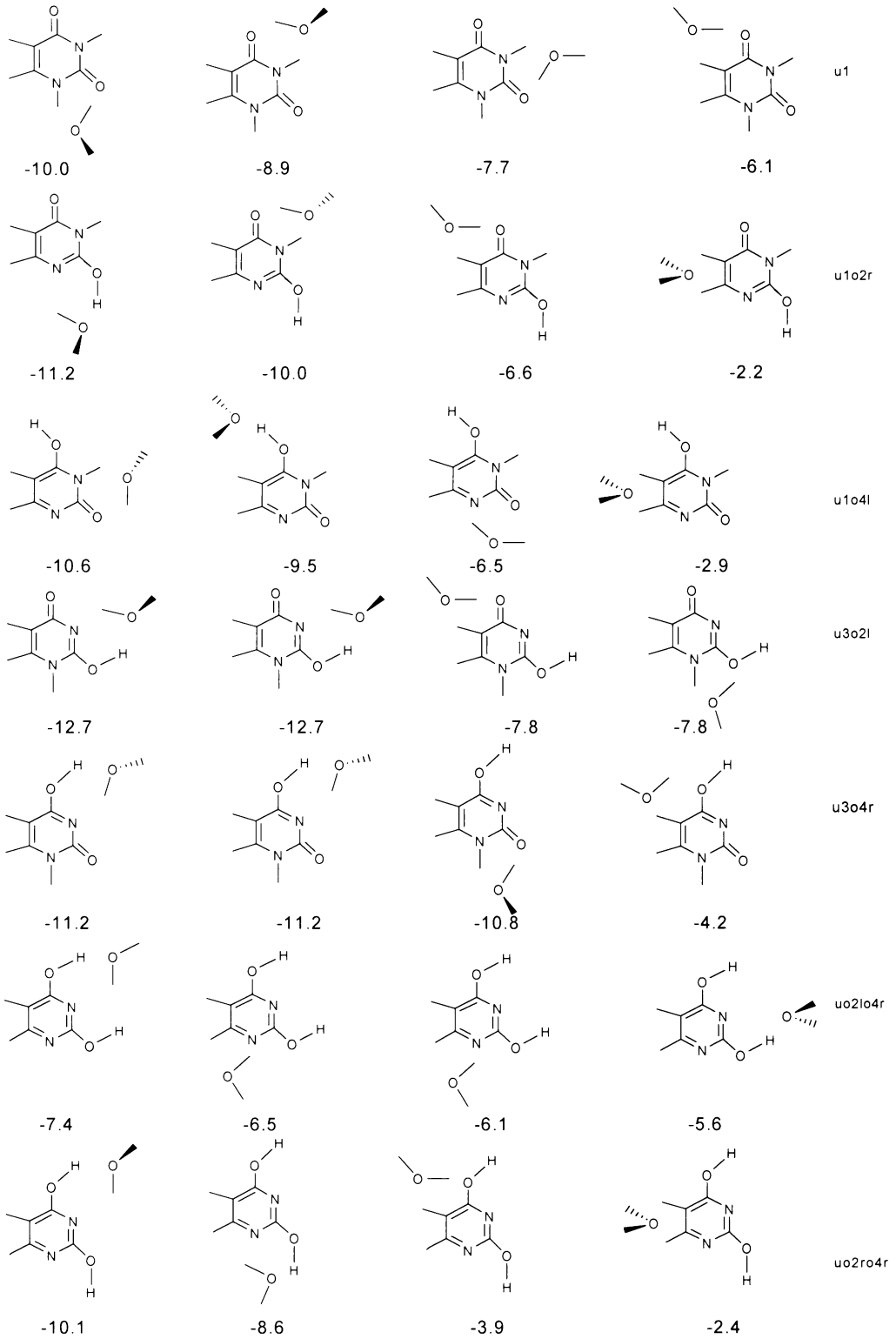

Figure 8. Seven of the most stable structures of uracil tautomers with one water molecule optimized at the RI-MP2/TZVPP level of theory. The stability decreases from left to right. The total interaction energies in kcal/mol are presented below the structures. Structures uo2lo4r (2 and 3) and u3o2l (1,2 and 3,4) were detected at empirical level but they belong to the same minimum structure at the *ab initio* level. In the case of bases and water molecules single lines represent hydrogens, hydroxy group hydrogens of bases are depicted explicilty.

Table 5. Relative and interaction energies (in kcal/mol) of uracil and thymine tautomers in mono- and dihydrated environment (global minima for each tautomer-water complex are presented).

	relative energies[a,b]	interaction energies[c,d]			relative energies[a,b]	interaction energies[c,d]	
structure	RI-MP2	RI-MP2	RI-MP2$_{TOT}$	structure	RI-MP2	RI-MP2	RI-MP2$_{TOT}$
u1-(H_2O)	0.00	-10.47	-9.96	**t1-(H_2O)**	0.00	-10.41	-9.89
u1o2r-(H_2O)	8.31	-12.27	-11.18	**t1o2r-(H_2O)**	7.94	-12.18	-11.13
u1o4l-(H_2O)	19.70	-11.38	-10.59	**t1o4l-(H_2O)**	21.21	-11.57	-10.77
u3o2l-(H_2O)	14.61	-13.92	-12.70	**t3o2l-(H_2O)**	16.68	-13.23	-10.31
u3o4r-(H_2O)	9.32	-12.16	-11.17	**t3o4r-(H_2O)**	10.02	-12.36	-11.34
uo2lo4r-(H_2O)	13.30	-8.36	-7.43	**to2ro4r-(H_2O)**	9.29	-11.07	-10.29
uo2ro4r-(H_2O)	9.11	-10.90	-10.14	**to2lo4r-(H_2O)**	13.35	-8.12	-7.39
u1-(H_2O)$_2$	0.00	-21.96	-20.57	**t1-(H_2O)$_2$**	0.00	-20.97	-19.73
u1o2r-(H_2O)$_2$	8.52	-24.49	-21.35	**t1o2r-(H_2O)$_2$**	7.21	-22.36	-20.77
u1o4l-(H_2O)$_2$	19.94	-22.69	-21.01	**t1o4l-(H_2O)$_2$**	15.91	-30.24	-24.61
u3o2l-(H_2O)$_2$	14.93	-25.99	-23.53	**t3o2l-(H_2O)$_2$**	11.12	-27.33	-24.36
u3o4r-(H_2O)$_2$	8.97	-23.74	-21.95	**t3o4r-(H_2O)$_2$**	8.05	-24.00	-22.14
uo2lo4r-(H_2O)$_2$	16.26	-15.82	-15.29	**to2ro4r-(H_2O)$_2$**	8.66	-22.14	-19.89
uo2ro4r-(H_2O)$_2$	9.81	-21.39	-19.74	**to2lo4r-(H_2O)$_2$**	15.04	-15.70	-15.23

[a] The order of the relative stability of each tautomer is given with respect to the canonical tautomer. Relative total energies of mono- and dihydrated tautomers are considered. RI-MP2$_{ZPVE}$ is defined as a sum of relative RI-MP2 energy and $\Delta ZPVE$; the former energy is evaluated with the TZVPP basis set, while the latter are at the MP2/6-31G** level. [b] For description of abbreviations used for methods, see notes to Table 1. [c] Interaction energies were evaluated with the TZVPP basis set. [d] Total complexation energy, RI-MP2$_{TOT}$ is defined as a sum of the RI-MP2 interaction energy and deformation energies of the monomers.

Hydrated tautomers. Gas phase relative free energies, relative hydration free energies determined by MD-TI, COSMO and hybrid approaches for uracil and thymine tautomers are shown in Table 6.

MD-TI relative hydration free energies (the second column of Table 6) of uracil and thymine are similar. In both cases the enol x1o4l and x3o4r tautomers (x = t or u) are the best hydrated and this finding is in accord with the microhydration results. The COSMO results (the fourth column of Table 6) agree reasonably well with the MD-TI values for all thymine and all uracil tautomers and the relative order of the tautomers is basically retained. Comparing relative ΔG for uracil and thymine tautomers, we can state that methylation at position 5 has a minor effect and ΔG remains about the same for both tautomers.

Table 6. Relative gas-phase free energies (ΔG_0^{298}), relative free energies of hydration, evaluated using MD-TI method (ΔG^{TI}), COSMO method ($\Delta G^{C\text{-}PCM}$) and hybrid model (ΔG^{HYB}), and relative free energies in aqueous solution ΔG^{298}(TI) of uracil and thymine tautomers.

Structure [a]	ΔG_0^{298} [b]	ΔG^{TI}	ΔG^{298}(TI)	$\Delta G^{C\text{-}PCM}$	$\Delta G^{C\text{-}PCM}$ opt [c]	ΔG^{HYB} (1w) vac min [d]	ΔG^{HYB} (1w) solv min [d]	ΔG^{HYB} (1w) opt vac [c, d]	ΔG^{HYB} (1w) opt solv [c, e]	ΔG^{HYB} (2w) vac min [d]	ΔG^{HYB} (2w) solv min [e]	ΔG^{HYB} (2w) opt vac [c, d]	ΔG^{HYB} (2w) opt solv [c, e]
u1	0.00	0.00	0.00	0.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
u1o2r	11.06	0.78	11.84	0.92	0.39	-0.70	-0.01	-0.85	-0.67	2.53	2.53	-1.24	-1.24
u1o4l	21.62	-6.48	15.14	-5.92	-7.34	-6.91	-7.77	-7.50	-9.05	-7.01	-7.01	-7.49	-7.49
u3o2l	19.20	-3.22	15.98	-4.64	-5.70	-6.12	-5.43	-6.56	-6.38	-5.12	-5.12	-6.89	-6.89
u3o4r	12.46	-3.65	8.81	-2.34	-3.11	-3.18	-2.49	-3.48	-3.29	-0.31	-2.89	-0.38	-3.17
uo2lo4r	12.85	2.31	15.16	1.63	1.23	2.19	1.31	0.76	0.44	1.87	1.87	1.46	1.46
uo2ro4r	11.73	3.08	14.81	2.48	2.08	1.92	2.62	1.79	1.98	4.16	3.39	-1.02	1.65

Structure [a]	ΔG_0^{298} [b]	ΔG^{TI}	ΔG^{298}(TI)	$\Delta G^{C\text{-}PCM}$	$\Delta G^{C\text{-}PCM}$ opt [c]	ΔG^{HYB} (1w) vac min [d]	ΔG^{HYB} (1w) solv min [d]	ΔG^{HYB} (1w) opt vac [c, d]	ΔG^{HYB} (1w) opt solv [c, e]	ΔG^{HYB} (2w) vac min [d]	ΔG^{HYB} (2w) solv min [e]	ΔG^{HYB} (2w) opt vac [c, d]	ΔG^{HYB} (2w) opt solv [c, e]
t1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
t1o2r	10.65	0.98	11.63	0.72	0.09	-0.73	-0.49	-0.69	-0.74	1.30	-0.07	2.44	0.64
t1o4l	24.58	-5.46	19.12	-4.62	-6.06	-5.21	-7.49	-5.36	-5.32	-6.57	-5.96	-4.06	-6.47
t3o2l	18.68	-2.50	16.18	-4.59	-5.72	-2.73	-5.26	-6.47	-6.07	-7.17	-6.50	-6.59	-7.31
t3o4r	13.28	-3.50	9.78	-3.04	-4.01	-3.72	-3.48	-4.01	-4.05	-1.93	-4.12	-1.16	-4.84
to2lo4r	13.13	3.01	16.14	0.96	0.11	1.06	1.29	0.33	0.57	1.42	2.09	1.86	1.14
to2ro4r	13.46	3.90	17.36	1.86	1.30	1.28	1.52	1.57	1.53	0.29	0.56	1.80	0.83

[a] Cf. Figure 1. [b] See Table 1. [c] Geometries optimized in the continuum solvent. See Methods. [d] Global minimum in the gas phase. [e] Global minimum in the continuum solvent. All energies are in kcal/mol.

The main conclusion from the MD-TI calculations is that preferential hydration of enol forms is fully confirmed at the COSMO level. Taking all these results into account we must state that the bare COSMO technique gives reasonable values of hydration free energies at much lower cost than computer-time demanding MD-TI calculations.

Optimization of tautomer structures in a solvent causes a systematic shift toward lower relative values. The absolute average error (AAE) values (relative to the MD-TI results) were enlarged for both uracil (1.1 kcal/mol) and thymine (1.8 kcal/mol) tautomers.

Results connected to the hybrid model are discussed in the next section.

4.1.3 Free energy perturbation and Continuous hybrid approaches

All results discussed below are summarized in Tables 7 and 8. Complete study is presented in Appendix C.

Adenine. The relative hydration order was closer to the MD-TI order when the hybrid model was used. Only the a7 and a19r tautomers were exchanged compared to he MD-TI method. The gap between MD-TI and COSMO hybrid results for the a79l tautomer also became smaller (2.7 kcal/mol) and the absolute average error slightly decreased, compared to the results of the bare COSMO. The hybrid model brought the COSMO values and the relative order of adenine tautomers closer to the TI values.

Cytosine. Both TI and COSMO results indicate that the canonical tautomer c1 is hydrated better than other cytosine tautomers. A continuum model gave similar relative hydration order of tautomers compared to TI results (c1 > c2a > c3a > c3b > c2b) and the only difference concerned the c2a tautomer, which was the second in the TI series and the last one in the COSMO series. However, the difference between the TI and COSMO results for this tautomer was less than 3 kcal/mol.

The monohydration was less specific and considerably weaker compared with adenine tautomers and thus consideration of the water dimer criteria was more useful. The absolute average error was smaller for the hybrid approach than for bare COSMO.

Guanine. Based on the TI results, guanine tautomers were hydrated in the following decreasing order: g79 > g39 > g37 > g17 > g19 > g7o2 > g9o2. The continuum model gave different results and the g19 and g17 tautomers were exchanged, which was also true for the g39 and g79 tautomers. In the case of guanine the worst agreement between COSMO and TI results was observed. The use of the hybrid model lowered the absolute average error from 9.1

to 8.3 (criterion a) or 8.2 (criterion b) and also changed the order of the g17 and g19 tautomers (g39 > g79 > g37 > g17 > g19 > g7o2 > g9o2).

Thymine. The continuum model gave slightly different order of hydration than TI method. The usage of hybrid model with water dimer criterion did not change the continuum model ordering. The hybrid model with criterion a) gave results which were usually closer to the TI results. The absolute average error was comparable for all continuum approaches.

Uracil. The continuum model gave similar relative results like TI and the only difference concerned the u3o4r and u3o2l tautomers, which were exchanged in the COSMO series (i.e. the order is u1o4l > u3o2l > u3o4r > u1> u1o2r > uo2lo4r > uo2ro4r). The absolute COSMO values were in a good agreement with the TI ones. The absolute average error of the COSMO method was only 1.1 kcal/mol. The hybrid model did not shift the COSMO results closer to the TI results and the absolute average error increased. Also the hydration ordering was worse.

Isoguanine. The TI results indicated that isoguanine tautomers are hydrated in the following decreasing order: t37 > c37 > ao7 > ao1. The ao7 and c37 tautomers were exchanged in the COSMO series. When using hybrid approaches, the order of tautomers remained the same but the absolute average error increased.

Table 7. The absolute average error (AAE) of the relative hydration Gibbs energies ($\Delta\Delta G_{HYD}$) obtained by COSMO and the hybrid model compared with the MD-TI in kcal/mol.

Base	COSMO	Hybrid model	
		a	*b*
Adenine	1.8	1.7	1.7
Cytosine	1.3	2.0	1.1
Guanine	9.1	8.3	8.2
Thymine	1.8	1.8	2.5
Uracil	1.1	1.8	1.5
Isoguanine	2.1	2.5	2.5
Average	2.9	3.0	2.9

[a] The hydration of a base is always considered as specific, [b] The hydration of a base is considered to be specific only if the water dimer criterion is fulfilled.

Table 8. Relative Gibbs energies in the gas phase (ΔG), relative Gibbs energies of the complex formation ($\Delta\Delta G_{FORM}$) and relative hydration-Gibbs energies ($\Delta\Delta G_{HYD}$) of the tautomers studied.

	1	2	3	4	5	6	7
Base	**Structure** [a]	**ΔG**	$\Delta\Delta G_{FORM}$ [b]	$\Delta\Delta G_{HYD}$ COSMO [c]	$\Delta\Delta G_{HYD}$ COSMO-HYBRID [d]		$\Delta\Delta G_{HYD}$ TI [e]
					a	b	
Adenine	a9	0.0	0.0	0.0	0.0	0.0	0.0
	a1	18.4	-6.1	-13.4	-14.9	-14.9	-11.5
	a17r	16.8	0.4	-7.3	-8.0	-8.0	-7.8
	a19r	11.8	-0.3	-4.1	-4.2	-4.2	-4.0
	a3	8.5	-1.6	-4.6	-4.9	-4.9	-5.0
	a37l	17.8	-3.0	-4.4	-7.4	-7.4	-5.1
	a39r	30.5	-1.2	-14.6	-15.6	-15.6	-12.2
	a7	7.3	0.6	-6.8	-3.4	-3.4	-4.7
	a79l	37.0	-7.6	-14.7	-18.6	-18.6	-21.3
Cytosine	c1	0.0	0.0	0.0	0.0	0.0	0.0
	c2a	-2.2	0.8	6.5	7.5	6.1	3.7
	c2b	-1.5	1.4	6.0	6.4	5.5	7.3
	c3a	2.0	0.4	4.3	0.9	3.9	3.8
	c3b	0.5	0.5	5.2	5.2	4.7	4.7
Guanine	g19	0.0	0.0	0.0	0.0	0.0	0.0
	g17	0.3	-0.7	1.5	-0.1	-0.1	-0.9
	g39	19.0	-0.7	-12.7	-11.4	-11.4	-24.8
	g37	7.2	-1.7	-1.5	-3.9	-3.9	-18.5
	g79	23.2	-2.1	-10.6	-11.1	-11.1	-30.7
	g7o2	4.5	1.0	5.1	3.8	3.7	3.1
	g9o2	1.0	1.8	6.8	6.0	5.4	5.6
Thymine	t1	0.0	0.0	0.0	0.0	0.0	0.0
	t1o2r	9.8	0.3	0.1	2.5	0.1	1.0
	t1o4l	22.3	-1.5	-4.6	-5.3	-6.5	-5.5
	t3o2l	18.0	-2.6	-5.7	-6.5	-7.7	-2.5
	t3o4r	12.9	-1.4	-4.0	-4.1	-5.2	-3.5
	to2lo4r	12.0	2.2	0.1	0.6	0.1	3.0
	to2ro4r	10.9	-0.6	1.3	1.5	0.4	3.9
Uracil	u1	0.0	0.0	0.0	0.0	0.0	0.0
	u1o2r	10.3	-1.3	0.4	-0.9	-0.3	0.8
	u1o4l	20.8	-2.5	-7.3	-9.2	-8.7	-6.5
	u3o2l	18.7	-2.5	-5.7	-6.6	-6.0	-3.2
	u3o4r	12.1	-1.1	-3.1	-3.5	-2.9	-3.7
	uo2lo4r	11.7	1.5	1.2	0.8	1.2	2.3
	uo2r04r	10.6	-0.4	2.1	1.8	2.3	3.1
Isoguanine	ao1	0.0	0.0	0.0	0.0	0.0	0.0
	ao7	28.6	0.1	-21.7	-22.0	-22.0	-20.3
	c37	38.8	-6.3	-20.4	-20.7	-20.7	-24.7

[a] Figures 1 and 2; [b] ΔG_{FORM} of water dimer, a9, c1, g19, t1, u1 and ao1 amounts to 4.85; 2.97; 3.62; 2.88; 4.02; 3.93 and 1.61 kcal/mol; [c] COSMO method used for bare tautomers; [d] hybrid model, 1- the hydration of a base is always considered as specific, 2- the hydration of a base is considered being specific if the water dimer criterion is fulfilled; [e] the thermodynamic integration method. All energies are in kcal/mol.

4.2 Total Interaction Energies of Tautomeric Nucleic Acid Base Pairs

Interaction energies of adenine...thymine and guanine...cytosine base pairs are presented in Table 9. First line in both parts contains results for canonical bases and Hoogsteen (H) and Watson-Crick (WC) arrangement of bases.

GC base pair will be discussed first. First column shows MP2 interaction energies evaluated at the complete basis set limit. Evidently, the WC structure is by far most stable and the stabilization energy difference with the second structure (c3a_g39) is about 9 kcal/mol. Remaining structures in Table 9 are even less stable. H-bonding pattern of structures 2-4 is the same as that of the GC WC one but the respective stabilization energies diiffer considerably. Evidently, the combination of canonical amino-keto/amino-keto tautomers is energetically most favorable. Passing to imino-keto/amino-keto tautomers pairs reduces the stabilization and this reduction is the largest when imino-keto/amino-enol tautomers are considered. Structure 1 as well structures 2-4 contain three H-bonds while in case of structures 5-7 one (attractive) H-bond is replaced by repulsive electrostatic interaction between N and O atoms. Consequently, stabilization energies of these three base pairs is the smallest one. All structures considered are planar H-bonded ones and these structures are characteristic by small difference between MP2 and CCSD(T) interaction energies.

The third column of the table confirms this result; all values are smaller than 0.7kcal/mol what is less than 3% of MP2 stabilization energy (The same is true for the AT base pairs). Evidently, the very expensive CCSD(T) calculations can be omitted and much less expensive MP2/CBS calculations yield sufficiently accurate stabilization energies. Let us add that the same conclusion was found for about 100 H-bonded DNA base pairs when canonical tautomers were considered.[24-26] The present data extend this important finding even for H-bonded pairs containing any tautomeric form.

The second column of the Table 9 shows the relative total energies of all GC pairs reflecting the energy penalization when passing from canonical to any tautomeric form. Evidently, all pairs are strongly penalized relative to canonical WC pair. Giving together values of the second and the fourth column we obtain the total (column 5) and relative (column 6) stabilization energies covering the tautomeric penalization. Energy difference between the most stable GC WC pair and other pairs containing tautomers is huge, more than 30 kcal/mol. This large difference clearly exclude any possibility of finding any other pair in

Table 9. Interaction Energies of Base Pairs

E_{int}^{CBS}: MP2 interaction energies evaluated at the complete bases set limit

$E^{RI\text{-}MP2}$: relative total energies

$\Delta E^{CCSD(T)} - \Delta E^{MP2}$: difference between MP2 and CCSD(T) interaction energies

$\Delta E^{CCSD(T)}_{CBS}$: E_{int}^{CBS} corrected for higher order correlation contributions

$\Delta E_{CCSD(T)}{}^{CBS} + E^{RI\text{-}MP2}$: total stabilization energies covering the tautomeric penalization

$\Delta E_{CCSD(T)}{}^{CBS} + E^{RI\text{-}MP2}$ rel.: relative stabilization energies covering the tautomeric penalization

[a] cf. Figure 5

All energies are in kcal/mol.

A ... T [a]		E_{int}^{CBS} (kcal/mol)	$E^{RI\text{-}MP2}$ (kcal/mol)	$\Delta E^{CCSD(T)} - \Delta E^{MP2}$ (kcal/mol)	$\Delta E^{CCSD(T)}_{CBS}$	$\Delta E_{CCSD(T)}{}^{CBS} + E^{RI\text{-}MP2}$	$\Delta E_{CCSD(T)}{}^{CBS} + E^{RI\text{-}MP2}_{rel.}$
1	***t1_a9H***	-16.8	0.0	0.02	0.02	-16.8	0.0
2	***t1_ai2_3H9H***	-13.9	34.8	0.12	34.96	21.0	37.8
3	***t1_ai2_7H9H***	-13.7	37.0			-13.7	3.1
4	***t3o4l_ai2_1H9H***	-13.8	38.2	-0.21	37.99	24.2	41.0
5	***t3o4r_ai1_1H9H***	-32.9	11.8	0.66	12.45	-20.5	-3.7
6	***t3o2r_ai2_1H9H***	-14.3	47.9	0.01	47.91	33.6	50.4
7	***t3o2l_ai2_1H9H***	-13.4	38.8	0.06	38.87	25.5	42.3
G ... C [a]		E_{int}^{CBS} (kcal/mol)	$E^{RI\text{-}MP2}$ (kcal/mol)	$\Delta E^{CCSD(T)} - \Delta E^{MP2}$ (kcal/mol)	$\Delta E^{CCSD(T)}_{CBS}$	$\Delta E_{CCSD(T)}{}^{CBS} + E^{RI\text{-}MP2}$	$\Delta E_{CCSD(T)}{}^{CBS} + E^{RI\text{-}MP2}_{rel.}$
1	***c1_g19***	-30.8	0.0	-0.47	-0.5	-31.2	0.0
2	***c3a_g39***	-21.6	30.9	-0.18	30.7	9.1	40.3
3	***c3a_g79***	-21.1	30.8	-0.17	30.6	9.5	40.8
4	***c3a_g9o1***	-15.9	16.7	-0.09	16.6	0.7	32.0
5	***c3b_g39***	-13.9	37.2		37.2	23.3	54.5
6	***c3b_g79***	-13.3	37.1	0.58	37.7	24.4	55.7
7	***c3b_g9o1***	-11.7	19.4		19.4	7.7	39.0
8	***c3b_g9o2***				0.0	0.0	31.2

the gas phase but also in any environment where solovation/desolvation energies can modify the gas-phase results.

Situation with AT pairs is completely different. Investigating the first column of the Table 9 for AT pairs we find that structure 1 (Hoogsteen pair) does not correspond to the global minimum at the MP2/CBS interaction energy surface. Structure 5 (t3o4r_ai1_1H9H) possess very large stabilization energy of almost 33 kcal/mol. From the first column of Table 9 follows that all tautomer pairs are penalized with comparison with the canonical pair. This penalization is smallest for the structure 5. Putting together both values and also the CCSD(T) correction terms we find (columns 5 and 6) that structure 5 is even more stable than the canonical AT H structure. All other structures are energetically highly above structures 1 and 5. We can conclude that structure 5 will be dominantly populated in the gas phase. Energy difference of about 4 kcal/mol is substantial what indicate that even in an environment (which might destabilize structure 5 over structure 1) both structures will be populated.

It must be added that investigating the whole PES of GC base pairs formed by canonical forms we found that WC structure is by far the most stable. Different situation exist for the AT pair where the global minimum correspond to the structure possessing H-bonds between N(9)H of A and CO(10) of T.[27] This structure corresponding to the global minimum in the gas phase cannot, however, exist in DNA since the position N(9) in T is blocked by the sugar group.

It is inevitable to mention here that all our structures were designed in order to suit for incorporation to the DNA molecule. All positions essential for incorporation are free, just CH_3 group was introduced. Our goal was to study only those systems really biologically relevant.

From this point of view our results are of a great importance since they reveal a serious possibility of incorporation tautomeric base pairs into DNA and again turn the case of tautomerism inside DNA (proposed already by Crick) or the origin of point mutations open.

5 Conclusions

Present work was aimed to explore equilibrium properties of tautomeric forms of nucleic acid bases in the gas-phase, in a microhydrated environment and in aqueous solution. The study also presents thorough comparison of different methods used for calculations in water environment. Such calculations were supposed to reveal the accuracy of these methods and the possibility of their usage for more biologically relevant calculations calculations. This was true for the last presented paper where the knowledge of equilibrium properties together with geometry properties of different tautomers was used for base pairs calculations.

We explored the tautomeric equilibrium properties of adenine, uracil and thymine. Adenine case ended in the theoretical prediction of coexistence of several tautomers in water phase. Uracil/thymine case confirmed that population of rare *enol* forms in bulk water is very low in this case and canonical structure is also clearly dominant in this phase.

The hydration free energies of nucleic acid bases tautomers were calculated using two different methods bare COSMO and MD-TI. Although results are in a good agreement in our calculations, some minor differences still remain. This difference becomes smaller when optimization of the geometry in the water phase was performed.

We believe that correctly described specific hydration can also minimize this difference. This belief lead us to introduce the hybrid model which should more or less cover the specific hydration. In the most cases the inclusion of explicit water molecules neither improved nor deteriorated results obtained for rare tautomers. It can thus be concluded that introduction of specific hydration, which is physically fully adequate, does not bring any improvement over bare COSMO. The use of our hybrid model is thus recommended only if the solute dipole moment becomes very large (approximately >10D).

The consequential part of this work was focused on broadening our knowledge about nucleic acid base pairs constructed from unusual nucleic acid base tautomers. Their common presence in the native DNA was not discovered yet althought some structures containing a single nucleic acid base tautomer are known. This was the first thorough study of these tautomeric base pairs which ment to be a solid background for forthcoming studies (e.g. hydration, nucleotide studies). It was mainly concerned about geometries and accurate interaction energies.

We optimized all sterically possible tautomeric base pairs which according to their spatial properties could be incorporated into DNA molecule . For all structures we calculated

total interaction energies including complete basis set extrapolation and including the higher-order correlation effects energy contributions.

Some optimized structers of studied base pairs posses interesting motifs which we believe become more important after inclusion of a water molecule. Stabilization energies of some tautomeric base pairs were very high (even higher than in the case of Watson-Crick base pair) and those structures deserves further investigation. Even for optimized H-bonded tautomeric base pairs it is evident that the CCSD(T) correction term is negligible.

References

1. J. D. Watson, *The Double Helix*, A Mentor Book, New York, **1969**.
2. W. Saenger, *Principles of Nucleic acid Structure*, Springer Verlag, New York, **1984**.
3. B. H. Geirstanger and D. E. Wemmer, *Annu. Rev. Biophys. Biomol. Struct.*, **1995**, 24, 463.
4. J. D. Watson and F. H. C. Crick, *Nature*, **1953**, 171, 737.
5. G. J. Fogarasi, J. Mol. Struct., 1997, 413, 271.
6. C. Aleman, *Chem. Phys.*, **2000**, 253, 13.
7. M. Mons, I. Dimicoli, F. Piuzzi, B. Tardivel, M. Elhanine, *J. Phys. Chem. A*, **2002**, 106, 5088.
8. V. H. Harris, C. L. Smith, W. J. Cummins, A. L. Hamilton, H. Adams, M. Dickman, D. P. Hornby, D. M. Williams, *J. Mol. Biol.* **2003**, *326*, 1389.
9. J. Florián, J. Leszczynski, *J. Am. Chem. Soc.* **1996**, *118*, 3010.
10. L. Stryer, *Biochemistry*, W.H. Freeman and Company, New York, **1999**.
11. W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman, *J. Am. Chem. Soc.* **1995**, *117*, 5179.
12. H. J. C. Berendsen, D. v. d. Spoel, R. v. Drunnen, GROMACS: A Message-Passing Parallel Molecular Dynamics Implementation., *Comp. Phys. Commun.*, **1995**, 91, 43.
13. L. J. Onsager, *J. Am. Chem. Soc.*, **1936**, 58, 1486.
14. O. Tapia, O. Goscinski, *Mol. Phys.*, **1975**, 29, 1653.
15. A. Klamt, G. Schüürmann, *J. Chem. Soc. Perkin Trans. 2*, **1993**, 799.
16. S. F. Boys, F. Bernardi, *Mol. Phys.*, **1970**, 19, 553.
17. T. H. Dunning Jr., *J. Chem. Phys.*, **1989**, 90, 1007.
18. A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, H. Koch, O. Jeppe, A.K. Wilson, *Chem. Phys. Lett.* **1998**, *286*, 243.
19. A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, J. Olsen, *Chem. Phys. Lett.* **1999**, *302*, 437.
20. J. Šponer, J. Leszczynski, P. Hobza, *J. Phys. Chem.*, **1996**, 100, 1965.
21. P. Hobza, J. Šponer, *J. Chem. Rev.*, **1999**, 99, 3247.
22. S. A. Trygubenko, T. V. Bogdan, M. Rueda, M. F. J. Orozco, Luque, J. Šponer, P. Slavíček, P. Hobza, *Phys. Chem. Chem. Phys.* **2002**, 4, 4192.

23. M. Hanus, F. Ryjáček, M. Kabeláč, T. Kubař, T. V. Bogdan, S. A. Trygubenko, P. Hobza, *J. Am. Chem. Soc.* **2003**, 125, 7678.
24. P. Jurečka, P. Hobza, *J. Am. Chem. Soc.*, **2003**, 125, 15608.
25. I. Dabkowska, H. V. Gonzalez, P. Jurečka, P. Hobza, *J. Phys. Chem. A*, **2005**, 109, 1131.
26. J. Šponer, P. Jurečka, *J. Am. Chem. Soc.*, P. Hobza, **2004**, 126, 10142.
27. M. Kabeláč, P. Hobza, *J. Phys. Chem. B*, **2001**, 105, 5804.