Relative Stabilities of Three Low-Energy Tautomers of Cytosine: A Coupled Cluster Electron Correlation Study

Géza Fogarasi†

Department of Theoretical Chemistry, Eötvös L. University, H-1518, Pf. 32, Budapest/Hungary

Received: August 7, 2001; In Final Form: November 13, 2001

High-level quantum chemical calculations have been carried out in an effort to reinvestigate the relative stabilities of the three lowest-lying tautomers of cytosine. Geometries were optimized at the CCSD/TZP level and electronic energies calculated at CCSD(T)/cc-pVTZ and vibrational frequencies at MP2/TZP. Comparative DFT calculations were also performed. From these data Gibbs free energies and equilibrium mole ratios were calculated. In agreement with most previous theoretical and experimental results, the amino−hydroxy tautomer **2b** was found to be the most stable structure. As a new result, the amino−oxo form **1** and the imino−oxo form **3a** have very nearly the same electronic energy, about 1.5−1.7 kcal/mol above **2b**. The calculated ΔG values at standard temperature are ∼0.8 kcal/mol relative to **2b**, again for both **1** and **3a**. These results about the stability of the oxo form **1** are in quantitative agreement with experimental estimates in the literature, both from matrix isolation infrared and from molecular beam microwave spectroscopy. However, the calculated stability of the imino form is much higher than suggested by experiment. Our tentative reanalysis of the IR spectrum did not resolve this discrepancy. The widely used MP2 method gives significant deviations from the coupled cluster results and may be not accurate enough for determining close-lying energies of tautomers. Also, density functional theory gives qualitatively different results from traditional wave function methods.

I. Introduction

The primary interest in nucleotide bases is due to their role as constituents of nucleic acids. Apart from this biological allure, these heterocycles are extremely intriguing molecules for the structural chemist as they may exist in a variety of tautomeric forms.[1,2] The treatment of tautomers is a great challenge for theory: as we will see, some isomeric structures lie very close to each other energetically so that an accuracy better than 1 kcal/mol would be desirable. This is the more difficult because—unlike, e.g., conformers—the tautomers to be compared have totally rearranged electronic structures.

The subject of this study is one of the nucleotide bases, cytosine. It has six possible tautomers (plus some rotational isomers for several of them). From numerous earlier studies[3−22] there is general agreement that there are three low-energy tautomers, the other three lying at least 8−10 kcal/mol higher in energy.[3] Only the low-energy forms, shown in Figure 1, will be investigated in this study. We have published earlier computational results on these structures.[23] The notation introduced there will be used also in the present study. Structures **2b** and **3a** have rotamer pairs (**2a** and **3b**), but the previous study showed clearly that the energy difference within a rotamer pair is very stable: **2a** lies 0.7−0.8 kcal/mol above **2b**, and **3b** 1.7−1.9 kcal/mol above **3a**, practically independent of the level of calculation. Therefore the present, very expensive calculations were restricted to the more stable member of each pair. (Compound **2a** cannot be left out of consideration when considering equilibrium; we can use here, however, mainly our earlier results[23] which will be augmented with a few new calculations.[24]) The amino−oxo form (**1**) is the "canonical" structure found in DNA. However, considering *isolated* molecules, the amino−hydroxy form (**2b**) has the lowest energy according to most quantum chemical calculations. (Except for density functional theory (DFT) calculations, see discussion below.) Experimental spectroscopic studies[25−28] also indicate the predominance of tautomer **2b** in the gas phase.

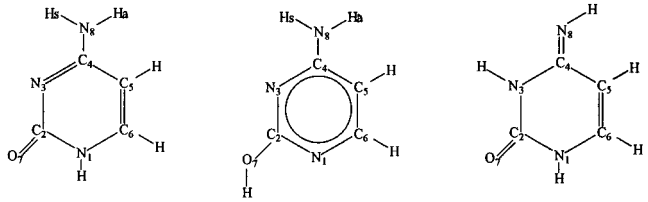

Figure 1. Three tautomers of cytosine investigated, indicating the numbering of atoms and formal bond structures.

The most intriguing result of our previous study[23] was the finding that the *imino*−oxo form (**3a**) may be significantly more stable than thought previously. This effect is not yet present in the perturbation theory, MBPT(2), calculation but appears in the coupled cluster result: the CCSD/DZP//MBPT(2)/DZP calculation (for notations see next section) gave an energy difference of only ∼0.2 kcal/mol relative to the most stable amino−hydroxy form. We thus suggested that the relative energies may show a qualitatively new picture, with stabilities in the order **2b** < **3a** < **1**.

However, our first study—although representing the highest level calculations on cytosine up to that time—had several limitations. Apart from the relatively small basis set, the coupled cluster method was restricted to single−double excitations only; also, geometry optimization was restricted to the MBPT(2) level.

† FAX: +36 1 209 0602. E-mail: fg@para.chem.elte.hu.

10.1021/jp013067x CCC: $22.00
Published on Web 01/22/2002

TABLE 1: Rotational Constants (MHz) of Three Tautomers of Cytosine

	1				**2b**				**3a**			
method[a]	A	B	C	θ^b	A	B	C	θ^b	A	B	C	θ^b
MP2/DZP; plan.[c]	3834	2002	1315		3911	1990	1319		3809	2001	1312	
MP2/TZP; plan.	3870	2019	1327		3951	2006	1330		3846	2015	1322	
MP2/TZP; compl	3866	2016	1326	−0.32	3948	2003	1330	−0.36	3842	2017	1324	−0.48
MP2/cc-pVTZ(−*f*); compl[d]	3897	2027	1335	*e*	3972	2012	1337	*e*	3877	2028	1332	*e*
MP2/cc-pVTZ; plan.	3927	2042	1343		4003	2028	1346		3901	2041	1340	
MP2/cc-pVTZ; compl	3925	2041	1343	−0.17	3999	2026	1346	−0.25	3901	2041	1340	0.00
CCSD/TZP; plan.	3871	2024	1329		3958	2009	1333		3845	2022	1325	
B3LYP/6-311++G(2 d,2p); compl	3882	2026	1331	−0.12	3968	2010	1334	−0.17	3862	2025	1328	0.00
B3PW91/6-311++G (2d,2p); compl	3901	2034	1337	−0.12	3985	2018	1340	−0.18	3881	2034	1335	0.00
exp[f]	3872	2025	1330	−0.22	3952	2009	1332	−0.18	3848	2026	1328	−0.18

[a] Notation: plan., optimization in planarity constraint; compl, complete optimization, without constraint. [b] Inertia defect, $\theta = I_C - I_A - I_B$, in amu Å^2; it is, of course, exact zero for structures with planarity constraint. Note that the four digits quoted here for the rotational constants are not sufficient for θ, which was calculated independently from the Cartesian coordinates. [c] Our earlier results.[23] [d] Results by Kobayashi,[30] the cc-pVTZ basis set was truncated by omitting the *f* functions. [e] The optimized structure was nonplanar, but θ cannot be reproduced from the four-digit rotational constants. [f] Microwave spectroscopic results on a supersonic beam by Brown et al.[28]

We therefore continued that work, trying to go to higher levels of theory. Our preliminary calculations presented at two conferences[29] already indicated that the inclusion of triple excitations, i.e. going from CCSD to CCSD(T), may be important, as well as basis set effects. For example, in CCSD-(T)/TZP calculations both **1** and **3a** are shifted to significantly higher energies than in the CCSD/DZP results.[29] Indeed, in the meantime Kobayashi has published an impressive CCSD(T) study[30] and concluded that our CCSD result quoted above is an artifact because triple excitations cannot be neglected; her CCSD(T) calculations gave the order **2b** < **1** < **3a**.

The aim of the present study is to do further calculations on this intriguing system, trying to determine the effects of basis set and electron correlation in a systematic way and to go as high as possible with the level of theory. Of course, cytosine is too large a molecule for really sophisticated calculations; still several improvements over previous work have been achieved: (a) we use larger basis sets, varying them in a systematic way; (b) geometry optimization is also done at the coupled cluster level; (c) triple excitations are also included, in the approximate CCSD(T) form; (d) zero point energies (ZPE) are calculated at the MP2/TZP level, and beyond ZPE, Gibbs free energies are used for calculating equilibrium; (e) a partial vibrational analysis will also be performed.

Beyond the conventional quantum chemistry techniques, density functional theory calculations will be included for comparison.

II. Computational Details and Notations

In the ab initio quantum chemical calculations, electron correlation was treated at the level of second-order many-body perturbation theory, MBPT(2) (this and the more familiar notation MP2 will be used as synonyms), and coupled cluster (CC) theory with singles−doubles and approximate triples, CCSD and CCSD(T). All CC calculations were carried out by the ACES II program system.[31] The largest MP2 geometry optimizations and vibrational frequency calculations were run by TURBOMOLE.[32] For the DFT calculations the Gaussian 94[33] package was used.

In the ab initio calculations, basis sets were of the Huzinaga−Dunning type, their size increased systematically from double- and triple-ζ polarization (DZP,TZP)[34−36] through triple-ζ two-polarization (TZ2P)[37−39] up to the "correlation-consistent" polarized valence triple-ζ (cc-pVTZ)[40] basis set. The latter includes *f*-functions and means a total of 310 basis functions for cytosine. The *d*- and *f*-functions were of the true spherical, five- and seven-component type, respectively. In the DFT calculations two types of functionals were used, Becke's three-parameter exchange functional (B3)[41] combined with the Lee, Yang, Parr (LYP) correlation functional[42] and with the Perdew, Wang functional PW91,[43] respectively, both as implemented in Gaussian 94. These calculations used the small 6-31G(d,p) basis,[44] as well as the much larger 6-311++G(2d,2p) basis set.[45]

The ab initio geometry optimizations were carried out at several levels. To study the significance of nonplanarity, MBPT-(2) optimizations were done both under planarity assumption and without any constraint. The latter included a full nonplanar MBPT(2)/cc-pVTZ optimization. The most demanding geometry optimizations were done at the coupled cluster CCSD/TZP level. To reduce the extremely high computer cost, in these and all higher level single-point energy calculations, we assumed planarity. Although this means a constraint for some structures, it has no significant influence on the conclusions, as will be explained in the discussion. In the coupled cluster geometry optimizations by ACES II, the GDIIS method[46] in conjunction with "natural coordinates"[47] was used. Natural internal coordinates are generated automatically by the INTC program[48] and are useful in reducing the number of optimization steps, especially in ring systems. Vibrational frequency and intensity calculations by TURBOMOLE[32] were done at the MP2/TZP level, using numerical differentiation of analytic first derivatives.

III. Discussion

Geometry. Some representative results on the rotational constants are compiled in Table 1, with selected geometries listed in Table 2. In the latter, to conserve space, we restricted the list to skeletel parameters and data related to nonplanarity. While all previous literature results on the geometry were restricted at most to the MP2 level, we report here the first coupled cluster results. Before that, however, we show a series of MP2 geometry optimizations with the purpose of (a) checking basis set effects by going with the basis as far as possible and (b) investigating the importance of nonplanarity.

In the first line of Table 1 we reproduced the MP2/DZP results from our earlier study.[23] These relatively simple calculations give already rotational constants which agree within 1% with experiment. When the basis set is increased from DZP to TZP, whether we take the planar or the nonplanar optimization, there is a further, clear improvement: the maximum deviation between theory and experiment in the rotational constants is now 0.5%. Our largest MP2 calculation used the cc-pVTZ basis set. This further increase of the basis set leads, however, to a

TABLE 2: Selected Structural Parameters of Three Cytosine Tautomers[a]

	1			**2b**			**3a**		
parameter	CCSD/ TZP	MP2/ pVTZ[b]	DFT[c]	CCSD/ TZP	MP2/ pVTZ[b]	DFT[c]	CCSD/ TZP	MP2/ pVTZ[b]	DFT[c]
				Skeletal Bond Lengths					
C2−N1	1.4163	1.4083	1.4233	1.3316	1.3259	1.3318	1.3887	1.3820	1.3917
N3−C2	1.3791	1.3684	1.3686	1.3300	1.3232	1.3252	1.3771	1.3684	1.3743
C4−N3	1.3133	1.3079	1.3151	1.3380	1.3298	1.3365	1.4044	1.3920	1.4046
C5−C4	1.4462	1.4259	1.4364	1.4125	1.3971	1.4081	1.4665	1.4444	1.4561
C6−N1	1.3595	1.3461	1.3510	1.3462	1.3370	1.3401	1.3830	1.3675	1.3752
C6−C5	1.3531	1.3503	1.3537	1.3778	1.3724	1.3769	1.3441	1.3415	1.3426
O7−C2	1.2142	1.2147	1.2158	1.3462	1.3417	1.3456	1.2124	1.2129	1.2138
N8−C4	1.3567	1.3556	1.3608	1.3583	1.3613	1.3645	1.2793	1.2790	1.2780
				Skeletal Bond Angles					
N3−C2−N1	116.53	115.98	116.14	128.72	128.27	128.00	113.64	113.24	113.61
C4−N3−C2	119.69	120.21	120.48	115.57	115.93	116.19	127.96	128.25	127.73
C5−C4−N3	124.52	124.25	123.86	121.84	121.54	121.43	113.64	113.68	113.75
C6−N1−C2	123.46	123.75	123.27	114.29	114.59	114.83	123.18	123.25	123.14
C6−C5−C4	115.84	116.04	116.18	115.98	116.53	116.25	119.78	119.89	120.01
C5−C6−N1	119.95	119.76	120.06	123.60	123.13	123.29	121.80	121.70	121.75
O7−C2−N1	118.66	118.84	118.34	116.62	116.82	116.78	122.58	122.65	122.34
O7−C2−N3	124.81	125.18	125.52	114.66	114.91	115.22	123.78	124.11	124.05
N8−C4−N3	117.20	117.10	117.08	116.32	116.38	116.45	117.59	117.27	117.29
N8−C4−C5	118.28	118.64	119.05	121.83	122.05	122.09	128.76	129.05	128.96
				Selected Dihedral Angles					
H_s−N8−C4−N 3	planar constraint	11.7	10.0	planar constraint	15.6	13.3	planar constraint	planar optimum	planar optimum
H_a−N8−C4−C 5	planar constraint	−18.9	−15.8	planar constraint	−22.4	−18.6	planar constraint	planar optimum	planar optimum
N8−C4−N3−C2	planar constraint	176.6	178.1	planar constraint	177.6	178.0	planar constraint	planar optimum	planar optimum
C5−C4−N3−C2	planar constraint	−0.6	−0.5	planar constraint	−0.4	−0.3	planar constraint	planar optimum	planar optimum

[a] Bond lengths in angstroms; angles and dihedral angles in degrees. The coupled cluster calculation (CCSD) was done under planarity constraint; the other two (MP2 and DFT) were complete geometry optimizations. [b] pVTZ is short for cc-pVTZ. [c] DFT = B3LYP/6-311++G(2d,2p).

quite drastic *deterioration* of the rotational constants: all values are systematically too high, by up to 1.4%. Obviously, the calculated bond lengths are too short now, as is characteristic for large basis sets—unless higher-level correlation treatment can be applied. We included in Table 1 Kobayashi's result,[30] obtained with a reduced cc-pVTZ basis set in which the *f*-functions were omitted. Ironically but not unexpectedly, this—somewhat arbitrarily truncated—basis set performs better than the full one, giving results that are between the full cc-pVTZ and the TZP results. We conclude that from a purely empirical point of view the TZP basis set seems to be the best choice: it gives the best results, at a moderate cost.

About the structure of cytosine the question of planarity may be of some interest. In a thorough study Sponer and Hobza emphasized years ago[4,5] that nucleotide bases may be nonplanar. For the amino group in the amino−oxo form of cytosine, their MP2 calculation with a basis set of DZ(2d) type gave values of up to 25−27° for the H_a−N8−C4−C5 dihedral angle (Figure 1). Kobayashi obtained similar results.[30] Nowak et al.[49] have reviewed the question from both spectroscopic and theoretical aspects, suggesting that the IR absorption around 200 cm^{-1} may be related to the amino group inversion. For the present study, possible nonplanarity is of interest only to the extent that it may influence relative energies: in the largest coupled cluster calculations we were forced to assume planarity, and its justification should be checked. We have thus carried out several complete (i.e., nonplanar) optimizations, at MP2 level. In Table 1, the inertia defects θ are listed as a measure of nonplanarity. The TZP basis set—already a fairly big one—gives θ values indicating considerable deviation from planarity. About the geometry parameters (not listed) we note that the H_a−N8−C4−C5 dihedral angle in the two amino forms is 26−28°, in agreement with previous results. Interestingly, in the MP2/TZP results even the imino form (**3a**) is nonplanar and, what is more, θ is largest (−0.48) for this tautomer. This result was so unexpected that we have undertaken a complete vibrational frequency calculation for the planar form: imaginary frequencies have indeed been found—what is more, there are two of them—confirming nonplanarity. Details of the fully optimized structure show that—while the imino group is practically planar—the ring itself gets nonplanar; the largest dihedral angle is $\tau_{2,3,4,5}$, with a magnitude of 10°. In retrospect the result makes sense because tautomer **3a** is the least aromatic of the three isomers, with single bonds all around the ring except for the isolated C5═C6 double bond (see bond lengths in Table 2). For comparison, in the amino tautomers all dihedral angles within the ring are below 1°, suggesting a conjugated structure with aromatic character. In light of the experimental data, however, the TZP results overestimate nonplanarity: the calculated inertia defects are 50−150% larger than the experimental ones and the individual rotational constants have moved mostly in the wrong direction.

When going from TZP to the cc-pVTZ basis set, nonplanarity gets significantly smaller in the amino structures and the imino structure is now exactly planar, as indicated in Table 1 by the inertia defects. For this set of calculations we have listed detailed geometry parameters in Table 2. As to the dihedral angles, H_a−N8−C4−C5 is now 19−22°. Its pair, the H_s−N8−C4−N3 angle, is about 12−16°, with opposite sign. (Only the relative signs are relevant; the two are always opposite, which means that the hydrogens of the amino group are moved in the *same* direction—a "wagging" type distortion using the terminology of molecular vibrations.) The fact that large basis sets may move a structure closer to planarity is well-known. For nucleotide bases, already Sponer and Hobza[4] noted this in their discussion on nonplanarity of the amino group. We have analyzed this question in great detail on formamide.[50] Whatever the case in cytosine, the energetic consequences of (possible) nonplanarity will be discussed in the next section.

Our most demanding geometry optimizations were performed at the CCSD/TZP level. In these huge calculations the use of

TABLE 3: Ab Initio Energies of Three Cytosine Tautomers

method[a]	notation in Figure 2	**1** energy[b]	**1** ΔE[c]	**2b** energy[b]	**3a** energy[b]	**3a** ΔE[c]
MP2/TZP[full]// ∼	S4	−0.115 452	2.13	−0.118 852	−0.114 176	2.93
CCSD/TZP[full]// ∼	S5	−0.144 623	1.57	−0.147 123	−0.145 477	1.03
CCSD(T)/TZP[full]// rfg	S6	−0.203 519	1.82	−0.206 416	−0.203 613	1.76
MP2[f.c.]/TZ2P// rfg		−0.089 658	1.98	−0.092 815	−0.087 729	3.19
CCSD[f.c.]/TZ2P// rfg		−0.105 466	1.34	−0.107 594	−0.105 770	1.14
CCSD(T)[f.c.]/TZ2P// rfg		−0.172 354	1.56	−0.174 849	−0.171 952	1.82
MP2[full]/TZ2P// rfg	S7	−0.233 122	1.88	−0.236 115	−0.231 223	3.07
CCSD[full]/TZ2P// rfg	S8	−0.254 364	1.21	−0.256 296	−0.254 705	1.00
CCSD(T)[full]/TZ2P// rfg	S9	−0.322 487	1.44	−0.324 782	−0.322 134	1.66
MP2[f.c.]/cc-pVTZ// rfg	S10	−0.203 189	2.06	−0.206 469	−0.201 758	2.96
MP2[full]/cc-pVTZ// ∼		−0.331 294	1.86	−0.334 255	−0.329 526	2.97
MP2[full]/cc-pVTZ//∼; nonplanar		−0.331 437	2.01	−0.334 640	−0.329 526	3.21
CCSD[f.c.]/cc-pVTZ// rfg	S11	−0.218 514	1.21	−0.220 439	−0.219 237	0.75
CCSD(T)[f.c.]/cc-pVTZ//rfg	S12	−0.291 353	1.51	−0.293 754	−0.291 378	1.49

[a] Following general practice, the first part of the notation specifies the level of energy calculation; after the symbol // the level of geometry determination is given: fc, frozen core, frozen virtuals; full, no restriction; ∼, geometry optimization done at the same level as the energy calculation; rfg, the CCSD/TZP geometry used in the majority of high-level energy calculations. All geometry optimizations relevant to this table were run under planarity constraint, except, as indicated, for the nonplanar MP2/cc-pVTZ calculation. [b] Absolute energy + 394, in atomic units. [c] Relative energies with respect to tautomer **2b**, in kcal/mol.

planarity constraint was inevitable. As seen in Table 1, electron correlation at the CCSD level brings about small but clear further improvement as compared to the MP2 geometry: the rotational constants agree now within 0.2% with experiment. As the experimental values contain vibrational effects, this agreement is almost surprising and is only possible for quite rigid structures. Anyway, better agreement can certainly not be expected, indicating that deviation from planarity, if any, is physically insignificant.

For comparison we have also carried out several DFT geometry optimizations. The performance of the B3LYP exchange-correlation potential is very impressive: as judged by the rotational constants, the accuracy is close to that of the CCSD results. However, a closer inspection of the individual geometry parameters in Table 2 reveals also discrepancies: individual bond lengths differ by up to 0.01 Å from the CCSD values. Such a difference may be negligible for many purposes, but *is* significant in the present case. Namely, it should be realized that with a force constant of 10−12 aJ/Å^2—a realistic estimate for a stretching coordinate in the ring, this translates into an energy diference of about 0.1 kcal/mol. This effect—especially if combined for several bond lengths—is not negligible when comparing the close-lying energies of tautomers.

About the question of planarity−nonplanarity, Table 2 shows that DFT gives nonplanar structures for tautomers **1** and **2b**, but the relevant dihedral angles are smaller than in the MP2 results. Overall, the B3LYP geometry is excellent. Of course, the DFT method is generally credited with giving good geometries. The only disturbing aspect is the dependence of the results on the form of the exchange-correlation functional. As seen in Table 1, if LYP is replaced by the PW91 potential, the rotational constants are systematically overestimated, indicating that all bond lengths got too short.

Relative Energies. Earlier studies have already shown that relative energies of the three lowest lying tautomers of cytosine are within a narrow range of a few kcal/mol.[3−22] Stabilities present thus a delicate question, and improvement over previous work can only be hoped for if the parameters of calculations are varied in a systematic way until some sort of convergence is found. Therefore, we have investigated the effects of electron correlation as well as basis set size in a series of calculations. The electron correlation treatment extended from MP2 to CCSD-(T) and basis sets were varied from DZP to cc-pVTZ. Since it would be virtually impossible to perform geometry optimizations for each high-level calculation, a uniform reference geometry was used in most of the energy calculations. On the basis of the discussion of the geometry above, we think that the CCSD/TZP geometry, determined under planarity constraint, is a justifiable and good choice of reference geometry. From the pragmatic point of view, the presence of C_s symmetry reduces the requirement for computer recources both in terms of memory and disk capacity by at least an order of magnitude; the largest coupled cluster calculations would be beyond reach otherwise. At the same time, it is this planar geometry that—as discussed above—gives the best, excellent reproduction of the experimental rotational constants.

The results of various calculations on energies are compiled in Table 3. We have listed the absolute energies as well as the energies relative to **2b**, the latter being the lowest-energy tautomer in all calculations. From these data one can judge the effects of various assumptions and approximations that we were forced to use in the largest calculations. The second and third triads of Table 3 differ only in the treatment of core correlation, with the purpose of checking the frozen core approximation. (Note that by the term "frozen core" we actually mean "frozen core−frozen virtuals", i.e. the procedure in which both the core *and* the corresponding antibonding virtual orbitals are frozen.) As can be seen, this approximation affects the relative energies by less than 0.2 kcal/mol. In the last group of data (results obtained with the cc-pVTZ basis set), the first three rows compare MP2 calculations that differ in the treatment of core correlation and/or the assumption on planarity. The range of variation is here 0.20−0.25 kcal/mol. The question of planarity will be discussed in more detail further below.

Figure 2 gives an overview of relative energies, with the amino−hydroxy form **2b** as baseline and the other two tautomers measured from this. Corresponding to the various basis sets as indicated in the picture, four sets of data are shown. Within each triad the level of electron correlation is varied, the three data points referring to MBPT(2), CCSD, and CCSD(T) calculations, respectively. As shown clearly by the graph, there are still quite significant fluctuations—even though we are comparing *relative* energies. Also, the imino−oxo tautomer **3a** is much more sensitive to the level of calculation than is tautomer **1**, the amino−oxo form. Comparing the extreme values for a given tautomer, the range of variation is 3 kcal/mol in the

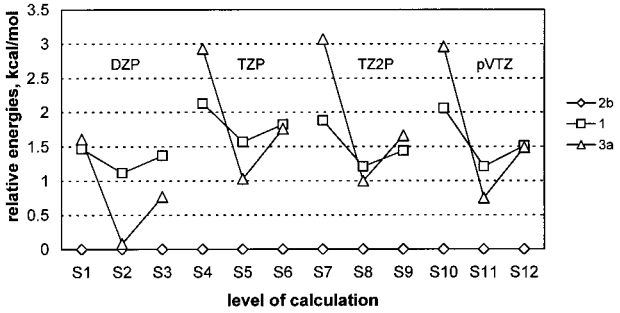

Figure 2. Relative energies of cytosine tautomers from ab initio calculations (Table 3), with isomer **2b** taken as reference. Each triad is a set of MP2, CCSD, and CCSD(T) calculations with a given basis set indicated in the picture, and with the following details: S1, MP2//∼; S2, CCSD//∼; S3, CCSD(T)//CCSD; S4, MP2//∼; S5, CCSD//∼; S6, CCSD(T)//CCSD[=rfg]; S7, MP2//rfg; S8, CCSD//rfg; S9, CCSD(T)//rfg; S10, MP2[f.c.]//rfg; S11, CCSD[f.c.]//rfg; S12, CCSD(T)[f.c.]//rfg. The symbol "∼" indicates that the level of geometry optimization was the same as that of the energy calculation; rfg stands for the CCSD/TZP optimum as reference geometry; [fc] represents frozen core and corresponding frozen virtuals. All calculations in this diagram were performed under planarity constraint.

case of **3a**, while it is only 1 kcal/mol for **1**. Concerning the details, we discuss first basis set effects. It is reassuring to see that these are quite uniform, almost the same at all three levels of electron correlation. The DZP results are obviously out of line, all values being too low: for tautomer **3a** the DZP energies are 1.0−1.3 kcal lower than the counterpart TZP energies, taking either the MP2 or the CCSD or CCSD(T) data points. For tautomer **1** the change is 0.5−0.7 kcal/mol. Use of the DZP basis set was a compromise in our earlier study of some years ago,[23] and it is not surprising that this basis set is too small. From the TZP basis upward, the results are already fairly stable: taking either of the two tautomers, the relative energies change only 0.3−0.4 kcal between the TZP, TZ2P, and cc-pVTZ sets. These residual errors are largely random, perhaps with some tendency toward lower relative energies with increasing basis set size. If one may extrapolate from just three sets of data, basis set effects seem to have converged to about 0.3 kcal/mol and the present values probably represent upper limits.

Considering electron correlation effects, Figure 2 shows a sharp decrease of relative energies between the MP2 and CCSD data points. This is then followed by a bouncing back in the CCSD(T) results. Both steps are characteristic for both tautomers, but their magnitudes are 2−3 times larger in the imino−oxo form (**3a**) than in the amino−oxo form (**1**). These basic effects are very clear with all basis sets, even with the smallest, DZP basis. A more careful scrutiny reveals the following details. The first step, a decrease between MP2 and CCSD, is continuously increasing (in the absolute sense) with increasing basis set size: for **1** the change goes from −0.35 to −0.85 and for **3a** from −1.53 to −2.21 kcal/mol. The second step, an increase between CCSD and CCSD(T), is almost constant, varying between 0.23 and 0.30 for **1** and between 0.66 and 0.74 for **3a**. As a result, the net change from MP2 to CCSD(T) is then increasing in magnitude with increasing basis set: with the largest, cc-pVTZ basis this difference is −0.55 for **1** and −1.47 for **3a**. The methodological conclusion is that second-order perturbation theory—the most widely used method of electron correlation—is far from being satisfactory for the present purpose, where an accuracy of better than 1 kcal/mol is required. Also, as pointed out correctly by Kobayashi,[30] in the coupled cluster treatment triple excitations *are* significant.

In the final highest level, CCSD(T)/cc-pVTZ, calculation (S12 in Figure 2) the imino−oxo and amino−oxo tautomers have practically the same energy, lying about 1.5 kcal/mol above the amino−hydroxy form. As compared to our earlier CCSD/DZP result (S2 in Figure 2), this means a large shift upward mainly for the imino form. This upward shift of **3a** is in part due to the triple excitations in the correlation treatment, raising the relative energy by about 0.7 kcal/mol. At the same time, the basis set effect between DZP and cc-pVTZ is of similar significance, responsible for another rise of ∼0.7 kcal/mol. The conclusion that **1** and **3a** have about the same energy is a revision of our earlier study in which the imino−oxo form was more stable than the amino−oxo isomer, and just a little bit higher in energy than the amino−hydroxy form. Even after this revision, however, the imino isomer's energy is still surprisingly low, indicating a larger stability than usually assumed.

TABLE 4: Energy Differences between Nonplanar and Planar Optimizations, kcal/mol

method	**1**	**2b**	**3a**
MP2/TZP	−0.64	−0.83	−0.06
MP2/cc-pVTZ	−0.09	−0.24	*a*

[a] The completely optimized structure is planar.

Effect of Possible Nonplanarity. The energy calculations discussed above were done in a planar reference geometry. Should the true geometry be, however, not planar for the amino forms **2b** and **1**, while planar for the imino isomer **3a**—as chemical intuition suggests, the energy lowering of the nonplanar structures would push up the relative energy of the planar structure. In the section on geometry we have seen that significant deviations from planarity were found in several cases. To discuss now their energetic effects, we compiled the relevant energy differences in Table 4. With the TZP basis the effect is quite strong: although even **3a** is nonplanar, the energy gain for this tautomer is only 0.06 kcal, while it is 0.83 kcal for **2b**; the relative energy of **3a** would then be raised by almost 0.8 kcal. However, with the larger, cc-pVTZ basis, the effect dramatically decreases: for **2b** the lowering is only 0.24 kcal and this is the total effect now, **3a** being planar. Even larger basis sets may further flatten out the amino group, making this effect insignificant. Still, for the present study it seems reasonable to correct the computed energies for the effect of nonplanarity. On the basis of Table 4, the estimated corrections are 0.24−0.09 = 0.15 kcal/mol for **1** and 0.24 kcal for **3a**. Adding these to the CCSD(T)/cc-pVTZ results in Table 3, our *final estimates* are 1.51 + 0.15 = 1.66 and 1.49 + 0.24 = 1.73 kcal/mol for the electronic energies of tautomers **1** and **3a**, respectively, relative to **2b**.

DFT Calculations. There have been many DFT calculations on cytosine before. Some years ago already Gould et al.[6] pointed out that density functional theory apparently fails to correctly reproduce the relative energies of cytosine tautomers. Their results were based on BLYP/6-31G(d,p) calculations. Kwiatkowski and Leszczynski[9] used the B3LYP potential with the same basis, obtaining results significantly different from the BLYP results but apparantly in better agreement with experiment. However, they also warned that DFT may be not the best choice for calculating relative energies of nucleotide bases. Kobayashi's B3LYP results,[30] obtained with a larger basis set, show differences of up to ∼0.4 kcal/mol relative to the 6-31G-(d,p) results, but give roughly the same scheme of relative energies. To have a clear overview, we have performed our own calculations, extending the basis set up to 6-311++G(2d,2p) and testing also the PW91 correlation potential. The results are compiled in Table 5 and shown in Figure 3.

Basis set effects are varying for the various tautomers. The energy of tautomer **1** relative to **2b** is fairly sensitive: the

TABLE 5: DFT Energies of Three Cytosine Tautomers

method[a]	notation in Figure 3	1		2b	3a	
		energy[b]	ΔE[c]	energy[b]	energy[b]	ΔE[c]
BLYP/6-31G(d,p) plan.	S1	−0.825 522	−1.17	−0.823 663	−0.823 747	−0.05
BLYP/6-311++G(2d,2p) plan.	S4	−0.958 658	−1.80	−0.955 784	−0.955 950	−0.10
B3LYP/6-31G(d,p) plan.	S2	−0.941 350	0.01	−0.941 360	−0.939 344	1.27
B3LYP/6-311++G(2d,2p) plan	S5	−1.066 060	−0.62	−1.065 071	−1.063 241	1.15
B3LYP/6-311++G(2d,2p) compl.	S6	−1.066 124	−0.54	−1.065 267	planar optimum	1.27
B3PW91/6-31G(d,p) plan.	S3	−0.792 977	0.05	−0.793 061	−0.790 544	1.58
B3PW91/6-311++G(2d,2p) plan.	S7	−0.909 387	−0.37	−0.908 793	−0.906 218	1.62
B3PW91/6-311++G(2d,2p) compl.	S8	−0.909 444	−0.28	−0.908 994	planar optimum	1.74

[a] In the DFT calculations, geometry and energy were always calculated at the same level: plan., planarity constraint; compl, complete optimization (allowing nonplanar structure). [b] Absolute energy + 394 in atomic units. [c] Relative energies with respect to tautomer **2b**, in kcal/mol.

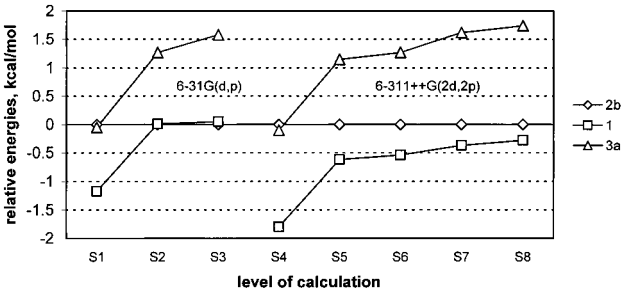

Figure 3. Relative energies of cytosine tautomers from DFT calculations (Table 5), with isomer **2b** taken as reference. Basis sets are indicated in the picture. Type of functional and method of geometry optimization (plan., planar constraint; compl, complete, without constraint): S1, BLYP, plan.; S2, B3LYP, plan.; S3, B3PW91 plan.; S4, BLYP, plan; S5, B3LYP, plan; S6, B3LYP, compl.; S7, B3PW91, plan.; S8, B3PW91, compl.

difference between the small and large basis is quite significant, 0.4−0.6 kcal/mol, the larger basis always lowering the *relative* energy. For tautomer **3a** the basis set effect is much smaller, virtually negligible, only 0.1 kcal. Apparently, the lowering of absolute energies with increasing basis set is closely the same for **2b** and **3a**, while it is slightly larger for **1**. Concerning the exchange-correlation potential, BLYP gives relative energies ∼1.2 kcal lower than B3LYP for both **1** and **3a**. Between B3LYP and B3PW91 the difference is only 0.3−0.5 kcal, the trend being again the same for both tautomers. Looking at the results from another perspective, **3a** is 1.7−2.0 kcal above **1** with all three potentials, and it is the reference **2b** structure whose energy is given too high by BLYP. As to the effect of nonplanarity, it is insignificant, changing relative energies by 0.1 kcal/mol only. With both of the more sophisticated functionals, taking the large basis set results, the final picture is this: the most stable tautomer is the amino−oxo form **1**, with 0.3−0.5 kcal below **2b**, while the imino form **3a** lies 1.3−1.7 kcal above **2b**. As compared to the coupled cluster calculations of conventional quantum chemistry, there is good agreement in the latter result (see, however, the free energies below). However, the stability of the amino−oxo form is far overestimated in the DFT calculations, leading thus to a qualitatively different picture about relative energies.

Equilibrium Composition. *Calculating Thermodynamic Quantities.* To calculate mole fractions of the tautomers, one needs first of all the zero point energies (ZPE). In Table 6 we have compared the results of various calculations. While earlier calculations were either HF or DFT calculations, with present computational facilities it has now become possible to calculate the force field at the MP2/TZP level. The dependence of relative ZPEs is still small for tautomer **1**, but becomes quite dramatic for **3a**. In the latter case, the vibrational energy relative to **2b** changes by 1 kcal/mol from the HF to the MP2 result, and the

TABLE 6: Zero Point Energies, kcal/mol

method	1		2b	3a	
	E	ΔE[e]	*E*	E	ΔE[e]
HF/6-31G(d,p)[a]	66.77	−0.21	66.98	67.60	+0.62
B3LYP/6-31G(d,p)[b]		−0.23			+0.24
B3LYP/truncated pVTZ[c]		−0.21			+0.21
B3LYP/6-311++G(2d,2p)[d]	61.65	−0.21	61.86	62.12	+0.25
MP2/TZP[d]	61.94	−0.37	62.31	61.97	−0.35

[a] Gould et al.[6] [b] Kwiatkowski and Leszczynski.[9] [c] Kobayashi,[30] only the differences listed; *f* functions omitted from the cc-pVTZ basis set. [d] Present work. [e] Relative to **2b**.

TABLE 7: Thermodynamic Quantities at $T = 298$ K, kcal/mol

	$TS_{rot.}$	H_{vib}	TS_{vib}	G_{nucl}[c]	ΔG_{nucl}[d]	$\Delta G_{tot.}$[e]
Conventional Quantum Chemistry[a]						
1	8.291	64.369	4.023	42.496	−0.802	0.86
2b	8.286	64.566	3.423	43.298		
3a	8.295	64.389	4.141	42.393	−0.904	0.83
Density Functional Theory[b]						
1	8.290	64.060	3.808	42.403	−0.432	−0.97
2b	8.285	64.137	3.457	42.835		
3a	8.292	64.363	3.594	42.918	+0.083	1.35

[a] Geometries (moments of inertia) from CCSD/TZP, vibrational frequencies from MP2/TZP, electronic energies from CCSD(T)/cc-pVTZ. [b] B3LYP/6-311++G(2d,2p). [c] Total Gibbs free energy from nuclear motions, including the constant contributions from translation and $H_{rot.}$. [d] Relative to tautomer **2b**. [e] After adding the electronic energies, from Tables 3 and 5, respectively, including the corrections for nonplanarity from Table 4, see also text.

difference is 0.6 kcal/mol even between DFT and MP2. These fluctuations are comparable to the total (relative) electronic energies of ∼1.5 kcal/mol discussed above (!). HF and DFT calculations are widely used with good success to determine vibrational frequencies, either for spectroscopic purposes or to obtain ZPE. In the present case, however, small *differences* in ZPE are needed and these are—not surprisingly—very sensitive to the method.

Beyond the zero point energies, we present for the first time results for cytosine on the thermal enthalpy and entropy contributions, needed to calculate Gibbs free energies and thus true equilibrium constants. As usual, rotation was considered classically and vibrations were treated in the harmonic approximation. Table 7 shows details of such calculations at standard temperature first. As expected, the rotational contribution is practically constant for the various tautomers. In the vibrational part, especially the entropy shows significant dependence on structure. In the conventional quantum chemistry results, the total nuclear contribution ΔG_{nucl} (relative to **2b**) is −0.8 − (−0.9) kcal/mol, in magnitude about half of the (relative) electronic energies! The DFT results are, as was the

TABLE 8: Calculated Mole Ratios[a]

tautomer	$T = 200$ K	$T = 298$ K	$T = 470$ K	$T = 570$ K
1	0.07	0.23	0.59	0.78
2b	1.00	1.00	1.00	1.00
3a	0.07	0.24	0.64	0.86
2a[b]	0.15	0.29	0.45	0.52

[a] Relative to tautomer **2b**, as obtained from the conventional quantum chemistry calculations of Gibbs free energies; for details see the footnotes to Table 7. [b] See text.

case with the electronic energies, again totally different: the negative (in all probability wrong) relative electronic energy of **1** is further increased here in magnitude, with a final $\Delta G_{\text{tot.}} \sim -1$ kcal/mol relative to **2b**. The nuclear contributions in **3a** are +0.1 kcal/mol, compared to −0.9 kcal/mol in the MP2 calculations.

When calculating the mole ratios of various isomers, we have to include the rotamer pair **2a** of the hydroxy tautomer because it lies also in the low-energy range.[23] In **2a** the OH bond is in trans position relative to the N1C2 bond. We mentioned in the Introduction that the electronic energy difference between rotamers is very stable, independent of the level of calculation. Thus, saving the expensive additional calculations, we used first the value of $\Delta E_{\text{el}}(\mathbf{2a}-\mathbf{2b}) = 0.75$ kcal/mol, based on our earlier study[23] and assumed that the rotational−vibrational contributions are the same in the rotamers. Later, we still did some explicit calculations.[24] The CCSD(T)/TZP energy is $\Delta E_{\text{el}} = 0.74$ kcal/mol, and the MP2/TZP zero-point energy, ZPE = 62.29 kcal/mol (as compared to 62.31 for **2b**; see Table 6), perfectly confirming the original assumption. In Table 8 we have compiled the calculated mole ratios of isomers at four selected temperatures as obtained from the ΔG values—the latter recalculated, of course, independently for each temperature. Beside the standard temperature, the low temperature was arbitrarily selected to indicate the sensitivity of results; the two higher temperatures correspond to the temperatures of the spectroscopic studies (see below). Once we have seen above that DFT gives totally different free energies, only the results from conventional quantum chemistry are listed here.

The basic theoretical result about stabilities is then the following. The most stable tautomer in the gas phase is the amino−hydroxy form **2b**. *Both* the amino−oxo form **1** and the imino−oxo form **3a** should be present in significant amounts, and the two have about the same stability. Rotamer **2a** of the hydroxy form can have a concentration comparable to that of the latter two tautomers and should also be taken into account. For **2b** we agree with most previous results, but our calculated mole ratio for **1** is more, and that for **3a** is *much* more than in previous studies, either theoretical or experimental. Because the present theoretical calculations represent the highest level up to now, one could easily say that they are the best. However, there are discrepancies in the light of experimental results, which will now be discussed.

Structural studies on cytosine in the gas phase are difficult because it decomposes at temperatures above 300 °C, needed to give sufficient vapor pressure. Still, by infrared (IR) spectroscopy there are three matrix isolation studies.[25−27] Also, in an elegant molecular beam study the microwave (MW) spectrum was successfully recorded and analyzed.[28] Concerning the infrared measurements, Radchenko et al.[25] already concluded that—contrary to solution where the amino−oxo form dominates—cytosine in the inert gas matrix exists mainly in the hydroxy form with significant amounts of the oxo form also present. More quantitative results were given by two (practically simultaneous) studies by Nowak et al.[26] and Szczesniak et al.[27] Using infrared intensities and helped by HF calculation of the spectra,[27] they estimated the ratio oxo:hydroxy as about (0.4−0.5):1. They have also identified small amounts of the imino form, with a ratio to the oxo form about one-fourth[26] and one-tenth,[27] respectively. Nowak et al.[26] also mention that some splittings in the spectra may be due to the presence of the rotamer **2a**, but no further investigation was done. The temperature of evaporation was 170−220 °C in these studies. (Of course, the matrix itself is only at about 10 K.) The MW investigation—also helped by HF calculations—identified all three tautomers, and estimated the ratio amino−hydroxy:amino−oxo:imino−oxo as about 1:1:(1/4). The nozzle temperature was higher here, close to 300 °C.

As to the oxo form, our results in Table 8 conform quite agreeably to the experiments. At 470 K, the temperature of IR spectroscopic measurements, the ratio of **1** to (**2b** + **2a**) is ∼0.4, in accord with the above spectroscopic estimates. (As noted above, the rotamers cannot really be distinguished in the IR spectrum; thus we considered their sum here.) At 570 K the ratio **1**:**2b** is ∼0.8:1, again in accord with the microwave result that the two "have similar abundancies". (As to rotamer **2a**, in the MW spectrum individual rotamers should, of course, easily be seen separately. In our calculations (CCSD/TZP), **2a** has a dipole moment even somewhat larger than **2b**, 4.9 vs 3.6 D. We find no explanation for why **2a** was not reported on, except that perhaps it was not even looked for.)

Concerning the imino form **3a**, however, we are seriously at variance with the experiments. For its mole ratio we obtain ∼0.4−0.6 at 470 K (depending on whether we consider the ratio to **2b** only or to the sum of **2b** and **2a**). At 570 K the result is ∼0.8 (taking the ratio to **2b** now, this being the distinct rotamer reported in the MW spectrum). One may note that the experimental temperatures are somewhat uncertain, being evaporation temperatures, not sample temperatures. Still, we predict equal abundance for **1** and **3a**, largely independent of temperature. The discrepancy thus persists.

Using Infrared Spectral Information. Given the above uncertainties, we have also tried an independent way to determine relative stabilities. Obviously, infrared spectra represent the best, direct information on the composition of a mixture, *if* the assignment is available *and* intensities are accurate enough, *both* experimentally and theoretically. Of course, the difficulties in this respect are well-known.

A complete vibrational study is beyond the scope of this paper, but we have analyzed the experimental and theoretical spectra to the extent needed for the present purpose. Of the three matrix isolation infrared studies[25−27]—which agree basically—we have taken the data from Nowak et al.[26] The theoretical spectra are based on the results of the MP2/TZP calculations discussed above in connection with the thermodynamic quantities. For spectroscopic use, we made a minor empirical adjustement of the force field, using the SQM procedure.[51] The Cartesian force constants were transformed to internal coordinates (the latter generated automatically by our INTC program[47,48]) and the internal force field was "scaled". Only three scale factors were introduced, 0.91 for X−H stretchings, 0.93 for X−Y stretchings, and 0.98 for the rest. (Note again: these refer to the force constants, not the frequencies; for the latter these correspond to about 0.95−0.99, closer to 1.) The three scale factors were determined by a rough adjustment to the experimental frequencies, and no effort was made to do a detailed refinement.

In Table 9 we have compiled two regions of the spectra, those that seem best suitable for identification: the high-frequency

TABLE 9: Two Selected Regions of the Infrared Spectrum of Cytosine[a]

calculated				experimental			
ν[b]	I[c]	isomer	mode description	ν[b,d]	I[e]	isomer[f]	mode description
			OH, NH Stretchings (3300−3800 cm^{-1})				
3628	99	**2a**	OH str	3618			
3620	126	**2b**	OH str	3609	248	H	OH str
3571	48	**2b**	NH_2 as str	3575	176	H,O	NH_2 as str
3555	51	**1**	NH_2 as str				
3548	51	**2a**	NH_2 as str	3527	−	−	−
3490	139	**3a**	N1−H str	3501	26	I	N1−H str
3463	108	**1**	N1−H str	3474	86	O	N1−H str
3442[g]	76	**3a**	N3−H str				
3445[g]	78	**2b**	NH_2 sym str	3461	457	H	NH_2 sym str
3427	89	**1**	NH_2 sym str	3457		O	NH_2 sym str
3425	74	**2a**	NH_2 sym str				
3382	20	**3a**	N8−H str				
			C═O Stretchings, etc. (1500−1800 cm^{-1})				
1761	817	**3a**	C═O str	1770	62	O	−
				1760	69	O	−
1730	777	**1**	C═O str	1730, 1725	219, 150	O	C═O str
				1686	74	H	−
1675	296	**3a**	C═N, C═C	1678	129	I	N═C?
1665	353	**1**	C═C	1668,1659	109, 146	O	NC, CC str, CH bend
				1642	64	−	−
1631	379	**2a**					
1629	433	**2b**	NH_2 scis, ring str	1625	493	H	CN, CC str, CH bend
1609	64	**1**	NH2 scis				
1606	18	**2b**					
1605	52	**2a**	NH2 scis, ring str				
1601	16	**3a**	C═N,C═C				
1590	227	**2a**		1602, 1597	198, 26	O	NH_2 scis
1585	213	**2b**	ring str	1592	81	H	NH_2 scis
1551	142	**1**	ring str	1576, 1569, 1563	93, 88, 98		ring str
				1554	33	−	−
				1540	48	O	NC str, NH bend

[a] Calculated at the MP2/tzp level; force constants scaled by the SQM procedure[51] with scale factors: 0.91 for X−H str, 0.93 for X−Y str, 0.98 for all the rest. Experimental results from Nowak et al.[26] [b] Frequencies in cm^{-1}. [c] Intensities in km/mol. [d] Pairs or triples of bands that were assigned in the experimental work as a group to one mode are listed in one row. [e] Relative values, arbitrary units. Note that different scaling (normalization) was used by Nowak et al.[26] for the intensities of the hydroxy and oxo forms, respectively. To make comparison of tautomers possible, we scaled "back" all values for the oxo form by multiplying with 1.89, as quoted in the footnote to Table 1 of that reference. [f] We have kept the original notation:[26] H, hydroxy; O, oxo; I, imino form, corresponding to our **2b** (**2a** not distinguished), **1**, and **3a**. [g] Sequence of two rows interchanged for better overview.

region of the OH and NH stretching frequencies (the CH stretching bands are too weak) and the medium region 1500−1800 cm^{-1} which includes C═O stretchings and various ring vibrations.

Concerning the assignment, in the OH, NH region we agree with Nowak et al.,[26] except that not all the individual, close-lying calculated frequencies are observed. Notable is the band observed at 3501 cm^{-1} which was tentatively assigned to the imino form. The calculated value of 3490 cm^{-1} supports this interpretation. In the region of C═O frequencies our calculations have delivered important new information. In the experimental studies, the relatively strong pair of bands at 1760−1770 cm^{-1} was assigned to the amino−oxo form (**1** in our notation), but no mode description was found. We find no reason to doubt the computational result that this is in fact the C═O mode of the *imino*−oxo form, **3a**. Its frequency lying higher than in **1** is also in line with the slightly shorter bond length (Table 2). For the rest of vibrations in this region we basically agree with the experimental assignment. Notably, we confirm the assignment of the 1678 cm^{-1} band to the imino tautomer.

We can now try to use the observed and calculated intensities for determining the mole ratios of isomers. From Table 9 we selected several pairs of bands, and normalized the measured intensities by the calculated ones. Skipping details, in the region of OH and NH stretchings we obtained the ratios hydroxy:oxo:imino roughly as 1:0.4:0.1, in agreement with the results by Szczesznak et al.[27] and Nowak et al..[26] The agreement is, however, not surprising since they used a similar approach, except that they did not have theoretical frequencies for the imino form and they used Hartree−Fock, while we used MP2 calculations. Unfortunately, however, our further checks showed that the results differ considerably if other regions of the spectrum are used. To give just one example, we calculate the imino:oxo ratio from two pairs of bands in the C═O stretching region:

$$I(\nu{=}1678)/I(\nu{=}1725{-}1730) = (129/296)/(369/777) = 0.92;$$
$$I(\nu{=}1678)/I(\nu{=}1659{-}1668) = (129/296)/((109 + 146)/353) = 0.60$$

These would indicate much larger concentration of the imino form. (At the same time, it is noticeable that these values are closer to our result above, obtained from free energies.)

As a further check, we examined the relative intensities of selected bands within the *same* tautomer. Comparison of theoretical and experimental values showed deviations by up to a factor of 2−3. The discrepancies may be due to both theoretical and experimental uncertainties. It is well-known that ab initio calculations are much less reliable for the intensities than for the frequencies. In the present case, our MP2 intensities and the HF intensities from Szczesniak[27] agree semiquantitatively, within about 50%. At the same time, experimental

intensities may be distorted due to matrix effects, and, in the high-frequency region, Fermi resonance is also always an uncertainty factor. We conclude that the calculated and experimental intensities are not consistent.

IV. Conclusion

We have performed extensive, high-level quantum chemical calculations trying to determine the relative stabilities of three cytosine tautomers. Geometries have been optimized at the CCSD/TZP level and electronic energies calculated at CCSD(T)/cc-pVTZ and vibrational frequencies at MBPT(2)/TZP. Comparative DFT calculations were also performed. From these data we have obtained Gibbs free energies and equilibrium mole ratios. Comparing different levels of theory, the following conclusions can be drawn. (When quoting energies below, they always refer to *relative* values between tautomers.)

(a) Concerning equilibrium geometry, CCSD optimization brings about a small but clear improvement over MBPT(2). The final CCSD/TZP rotational constants agree with experiment within 0.2%.

(b) When calculating electronic energies, the basis set must be at least TZP. Upon further increase of the basis, *relative* energies stay already stable within about 0.3 kcal/mol.

(c) In the treatment of electron correlation, the fluctuation of relative electronic energies between MBPT(2), CCSD, and CCSD(T) is still big, up to 1 kcal/mol. In the highest-level, CCSD(T)/cc-pVTZ results the amino−oxo (**1**) and the imino−oxo (**3a**) tautomers are about equal, 1.5−1.7 kcal/mol above the most stable, amino−hydroxy form (**2b**).

(d) Zero point energies are again sensitive to the method applied. Our MP2/TZP relative energies differ from Hartree−Fock results by up to 1 kcal/mol.

(e) The determination of thermal equilibrium to be reasonable, it is necessary to calculate Gibbs free energies: ΔG_{nucl} contributions are 2−3 times larger than ZPE and about half of the electronic values (considering, as always, the *differences* between tautomers).

(f) DFT calculations perform very well for the geometries. However, concerning electronic energies, they seem to heavily overestimate the stability of the oxo form. Also, ZPE and nuclear contributions ΔG_{nucl} differ by up to 1 kcal/mol from the MP2 results. As a whole, DFT gives a picture qualitatively different from that obtained by conventional quantum chemistry.

As compared to the available experimental (matrix isolation infrared and molecular beam microwave spectroscopy) information, the agreement between theory and experiment is still not satisfactory. In accord with previous computations, we agree that in the gas phase the most stable tautomer is **2b**, the hydroxy form, and also that the oxo form **1** must be present in large amounts; however, we predict about the same stability also for the imino form **3a**, with a free energy of only ∼0.8 kcal/mol above **2b**, for both. The large stability of the imino form cannot be supported by experiment at present.

Given the small energy differences, we cannot state definitely that the theoretical results—however much computational effort was made—are accurate enough. We tried to check these results also by analysis of the vibrational spectra, without satisfactory agreement. Beyond the matrix isolation infrared spectra, gas-phase spectra would be of great value, but, of course, these would be extremely difficult to obtain. In an effort to collect new, independent experimental information, we have recently carried out a gas electron diffraction study—the first of this kind on cytosine—which is now being evaluated.[52] At present, however, the final answer to the question of stabilities of cytosine tautomers remains elusive.

Acknowledgment. This work has been supported by grants from the National Scientific Research Foundation of Hungary (OTKA), Grant No. T-030815, and the Ministry of Education, Grant No. FKFP 0511/1999.

References and Notes

(1) Elguero, J.; Marzin, C.; Katritzky, A. R.; Linda, P. In *The Tautomerism of Heterocycles; Advances in Heterocyclic Chemistry*; Katritzky, A. R., Boulton, A. J., Eds.; Academic Press: New York, 1976; Supplement 1.

(2) Kwiatkowski, J. S.; Pullman, B. *Adv. Heterocycl. Chem.* **1975**, *18*, 200.

(3) Ha, T.-K.; Gunthard, H. H. *J. Mol. Struct.* **1993**, *300*, 619.

(4) Sponer, J.; Hobza, P. *J. Phys. Chem.* **1994**, *98*, 3161.

(5) Sponer, J.; Hobza, P. *J. Am. Chem. Soc.* **1994**, *116*, 709.

(6) Gould, I. R.; Burton, N. A.; Hall, R. J.; Hillier, I. H. *J. Mol. Struct. (THEOCHEM)* **1995**, *331*, 147.

(7) Colominas, C.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc.* **1996**, *118*, 6811.

(8) Florian, J.; Baumruk, V.; Leszczynski, J. *J. Phys. Chem.* **1996**, *100*, 5578.

(9) Kwiatkowski, J. S.; Leszczynski, J. *J. Phys. Chem.* **1996**, *100*, 941.

(10) Hobza, P.; Sponer, J.; Polasek, M. *J. Am. Chem. Soc.* **1995**, *117*, 792.

(11) Sponer, J.; Leszczynski, J.; Hobza, P *J. Comput. Chem.* **1996**, *17*, 841.

(12) Florian, J.; Leszczynski, J.; Johnson, B. G. *J. Mol. Struct.* **1995**, *349*, 421.

(13) Florian, J.; Leszczynski, J.; Scheiner, S. *Mol. Phys.* **1995**, *84*, 469.

(14) Bakalarski, G.; Grochowski, P.; Kwiatkowski, J. S.; Lesyng, B.; Leszczynski, J. *Chem. Phys.* **1996**, *204*, 301.

(15) Florian, J.; Leszczynski, J. *J. Am. Chem. Soc.* **1996**, *118*, 3010.

(16) Florian, J.; Leszczynski, J. *Int. J. Quantum Chem. Suppl.* **1995**, *22*, 207.

(17) Paglieri, G.; Corongiu, G.; Estrin, D. A. *Int. J. Quantum Chem.* **1995**, *56*, 615.

(18) Sobolewski, A. L.; Adamowicz, L. *J. Chem. Phys.* **1995**, *102*, 5708.

(19) Fulscher, M. P.; Roos, B. O. *J. Am. Chem. Soc.* **1995**, *117*, 2089.

(20) Smets, J.; Adamowicz, L.; Maes, G. *J. Phys. Chem.* **1996**, *100*, 6434.

(21) Ha, T.-K.; Keller, H. J.; Gunde, R.; Gunthard, H. H. *J. Mol. Struct.* **1996**, *376*, 375.

(22) Santamaria, R.; Quirozgutierrez, A.; Juarez, C. *J. Mol. Struct. (THEOCHEM)* **1995**, *357*, 161.

(23) Fogarasi, G. *J. Mol. Struct.* **1997**, *413*−*414*, 271.

(24) Upon the suggestion of one of our reviewers, we *have* done some explicit calculations on rotamer **2a**; see the section on equilibrium.

(25) Radchenko, E. D.; Sheina, G. G.; Smorygo, N. A.; Blagoi, Yu. P. *J. Mol. Struct.* **1984**, *116*, 387.

(26) Nowak, M. J.; Lapinski, L.; Fulara, J. *Spectrochim. Acta, Part A* **1989**, *45A*, 229.

(27) Szczesniak, M.; Szczepaniak, K.; Kwiatkowski, J. S.; Kubulat, K.; Person, W. B. *J. Am. Chem. Soc.* **1988**, *110*, 8319.

(28) Brown, R. D.; Godfrey, P. D.; McNaughton, D.; Pierlot, A. P. *J. Am. Chem. Soc.* **1989**, *111*, 2308.

(29) (a) International Conference on Electron Correlation Effects in Spectroscopy and Dynamics, Kaiserslautern/Germany, Dec 9−12, 1998. (b) 8th European Symposium on Gas Electron Diffraction, Blaubeuren/Germany, June 20−23, 1999.

(30) Kobayashi, R. *J. Phys. Chem.* **1998**, *102*, 10813.

(31) Stanton, J. F.; Gauss, J.; Watts, J. D.; Lauderdale, W. J.; Bartlett, R. J. *ACES II*, Quantum Theory Project; University of Florida: Gainesville, FL, 1991.

(32) Ahlrichs, R.; Bär, M.; Baron, H.-P.; Bauernschmitt, R.; Böcker, S.; Ehrig, M.; Eichkorn, K.; Elliott, S.; Furche, F.; Haase, F.; Häser, M.; Horn, H.; Huber, Ch.; Huniar, U.; Kattannek, M.; Kölmel, Ch.; Kollwitz, M.; May, K.; Ochsenfeld, Ch.; Öhm, H.; Schäfer, A.; Schneider, U.; Treutler, O.; Arnim, v. M.; Weigend, F.; Weis, P.; Weiss, H. *TURBOMOLE*, Version 5; University of Karlsruhe: Karlsruhe, Germany, 2000.

(33) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T. A.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.;

Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzalez, C.; and Pople, J. A. *Gaussian 94* (Revision B.2); Gaussian, Inc.: Pittsburgh, PA, 1995.

(34) Dunning, T. H. *J. Chem. Phys.* **1970**, 53, 2823.

(35) Huzinaga, S. *J. Chem. Phys*. **1965**, *42*, 1293.

(36) Rednon, L. T.; Purvis, G. D.; Bartlett, R. J. *J. Am. Chem. Soc*. **1979**, *101*, 2856.

(37) Dunning, T. H. *J. Chem. Phys*. **1971**, *55*, 716.

(38) Dunning, T. H. *J. Chem. Phys.* **1971**, *55*, 3958.

(39) Gauss, J.; Bartlett, R. J. *ACES II*[31]program basis library.

(40) Dunning, T. H. *J. Chem. Phys*. **1989**, *90*, 1007.

(41) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.

(42) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev.* **1988**, *B41*, 785.

(43) (a) Perdew, J. P.; Wang, Y. *Phys. Rev.* **1992**, *B45*, 13244. (b) Perdew, J. P.; Chevary, J. A.; Vosko, S. H.; Jackson, K. A.; Pederson, M. R.; Singh, D. J.; Fiolhais, C. *Phys. Rev.* **1992**, *B46*, 6671.

(44) (a) Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257. (b) Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta* **1973**, *28*, 213.

(45) (a) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. *J. Chem. Phys*. **1980**, *72*, 650. (b) Frisch, M. J.; Pople, J. A.; Binkley, J. S. *J. Chem. Phys*. **1984**, *80*, 3265.

(46) (a) Pulay, P. *Chem. Phys. Lett.* **1980**, *73*, 393. (b) Császár, P.; Pulay, P. *J. Mol. Struct.* **1984**, *114*, 31.

(47) Fogarasi, G.; Zhou, X.; Taylor, P. W.; Pulay, P. J. *Am. Chem. Soc.* **1992**, *114*, 8191.

(48) Pulay, P.; Fogarasi, G. The *INTC program* for generating natural internal coordinates, Fayetteville, AR, and Budapest, Hungary, 1991−1992.

(49) Nowak, M. J., Lapinski, L., Kwiatkowski, J. S., Leszczynski, J. *Molecular Structure and Infrared Spectra of DNA Bases and Their Derivatives: Theory and Experiment;* in Leszczynski, J., Ed.; Computational Chemistry: Reviews of Current Trends; World Scientific: Singapore, 1997; pp 140−216.

(50) Fogarasi, G.; Szalay, P. G. *J. Phys. Chem. A* **1997**, *101*, 1400.

(51) Pulay, P.; Fogarasi, G.; Pongor, G.; Boggs, J. E.; Vargha, A. *J. Am. Chem. Soc.* **1983**, *105*, 7037.

(52) Samdal, S.; Shorokhov, D. J.; Strand, T. G.; Volden, H. V.; Fogarasi, G. Manuscript in preparation.