DOI: 10.1002/ange.201207744

Gas-Phase Tautomers

All Five Forms of Cytosine Revealed in the Gas Phase**

José L. Alonso, Vanesa Vaquero, Isabel Peña, Juan C. López, Santiago Mata, and Walther Caminati*

The determination of preferred tautomers of nucleobases has been of interest since the structure of nucleic acid and its base pairs was first reported.[1] Molecular-level understanding of their structure can provide important insight into the relationship that exists between the presence of tautomeric forms and spontaneous mutation in DNA.[2] The best experimental approach to address the structural preferences of nucleobases is to place them under isolation conditions in the gas phase, cooled in a supersonic expansion. Under these conditions, the various tautomers/conformers can coexist and are not affected by the bulk effects of their native environments, which normally mask their intrinsic molecular properties.[3] The main restriction to the gas-phase study of these building blocks is the difficulty in their vaporization owing to their high melting points (ranging from 316 °C for guanine to 365 °C for adenine) and associated low vapor pressures. We have shown previously that the use of molecular beam Fourier transform microwave (MB-FTMW) spectroscopy in conjunction with laser ablation (LA) enables these vaporization problems to be overcome and renders the study of the rotational spectra of coded amino acids accessible.[4] The success of these LA-MB-FTMW experiments prompted their application to nucleic acids, and our initial studies on uracil[5] and thymine[6] enabled the determination of the structures of their diketo forms present in the gas phase. A subsequent study of guanine led us to unequivocally identify the four most stable tautomers in the gas phase.[7] The molecular system of cytosine (CY) is even more complex than that of guanine. Figure 1 a shows the five most stable species, in order of stability according to theoretical calculations:[8] enol–amino *trans* (**EA*t***), enol–amino *cis* (**EA*c***), keto–amino (**KA**), keto–imino *trans* (**KI*t***), keto–imino *cis* (**KI*c***).

In 1988, Szczesniak et al.[9] observed the infrared spectra of CY isolated in inert Ar and N_2 matrices and showed that isolated cytosine exists under these conditions as a mixture of the **KA** and **EA** forms (they did not distinguish between **EA*t*** and **EA*c***). Brown et al.[10] reported the free-jet millimeter-wave absorption spectra of three species, which were tentatively assigned as the **KA**, **EA*t***, and **KI** forms. The identification was based on the values of the rotational constants alone. In contrast, Dong and Miller[11] used infrared laser spectroscopy in helium nanodroplets to characterize **EA*t***, **EA*c***, and **KA** species. Nir et al.[12] attributed two features observed in the vibronic spectra to the **KA** and **EA** forms. The electron diffraction pattern[13] was interpreted in terms of a conformation mixture dominated by the **EA** forms. X-ray photoemission spectra provide spectral signatures of oxo and hydroxy populations.[14] In recent experiments in an Ar matrix,[15] photoisomerization processes induced by narrow-band tunable near-infrared[15a] and UV[15b] light were interpreted in terms of the existence of various tautomers of CY. No conclusive experimental evidence for the coexistence of the five predicted forms has yet been reported.

We took advantage of the capabilities of LA-MB-FTMW spectroscopy to investigate the rotational spectra of cytosine in the solvent-free environment of a supersonic expansion. In this technique, the solid samples are vaporized by laser ablation, and the molecules are seeded in a supersonic expansion, in which CY is ideally frozen and the most stable forms trapped in their energy minima. The rotational spectrum of each of these molecular forms can be analyzed separately by Fourier transform microwave spectroscopy. Figure 1 b shows details of the five $1_{1,1}$–$0_{0,0}$ transitions corresponding to five different rotamers of CY observed in the 5100–5300 MHz frequency range. Each rotational transition shows a very complex hyperfine structure composed of tens of quadrupole component lines owing to the presence of three ^{14}N nuclei. This hyperfine structure arises from the coupling of the ^{14}N nuclear-spin angular momenta ($I=1$) to the overall rotational angular momentum through the interaction of the quadrupole moment of each ^{14}N nucleus with the electric-field gradient created at the site of this nucleus by the rest of the molecular charges. Analysis of this hyperfine structure yields the nuclear quadrupole coupling constants $\chi_{\alpha\beta}$ ($\alpha,\beta = a,\ b,\ c$),[16] which are extremely sensitive to the electronic distribution around the quadrupolar nuclei $^{14}N_1$, $^{14}N_3$, and $^{14}N_8$ (see Figure 1 a for nitrogen-atom labeling) and can be used as a valuable tool for the unambiguous identification of tautomers of CY.

[*] Prof. J. L. Alonso, Dr. V. Vaquero, Dr. I. Peña, Prof. J. C. López, S. Mata, Prof. W. Caminati[+]
Grupo de Espectroscopía Molecular (GEM), Edificio Quifima
Laboratorios de Espectroscopia y Bioespectroscopia
Parque Científico UVa, Universidad de Valladolid
47005 Valladolid (Spain)
E-mail: jlalonso@qf.uva.es
Homepage: http://www.gem.uva.es

[+] Present address:
Dipartimento di Chimica "G. Ciamician" dell'Università
via Selmi 2, 40126 Bologna (Italy)

[**] This research was supported by the Ministerio de Ciencia e Innovación (grants CTQ 2006-05981/BQU, CTQ 2010-19008, and Consolider Ingenio 2010 CSD 2009-00038) and the Junta de Castilla y León (grant VA070A08). W.C. thanks the molecular spectroscopy group of the University of Valladolid for their kind hospitality, and the Ministry of Education of the Spanish Government for a grant for a position as "Invited Professor".

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201207744.

© 2013 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

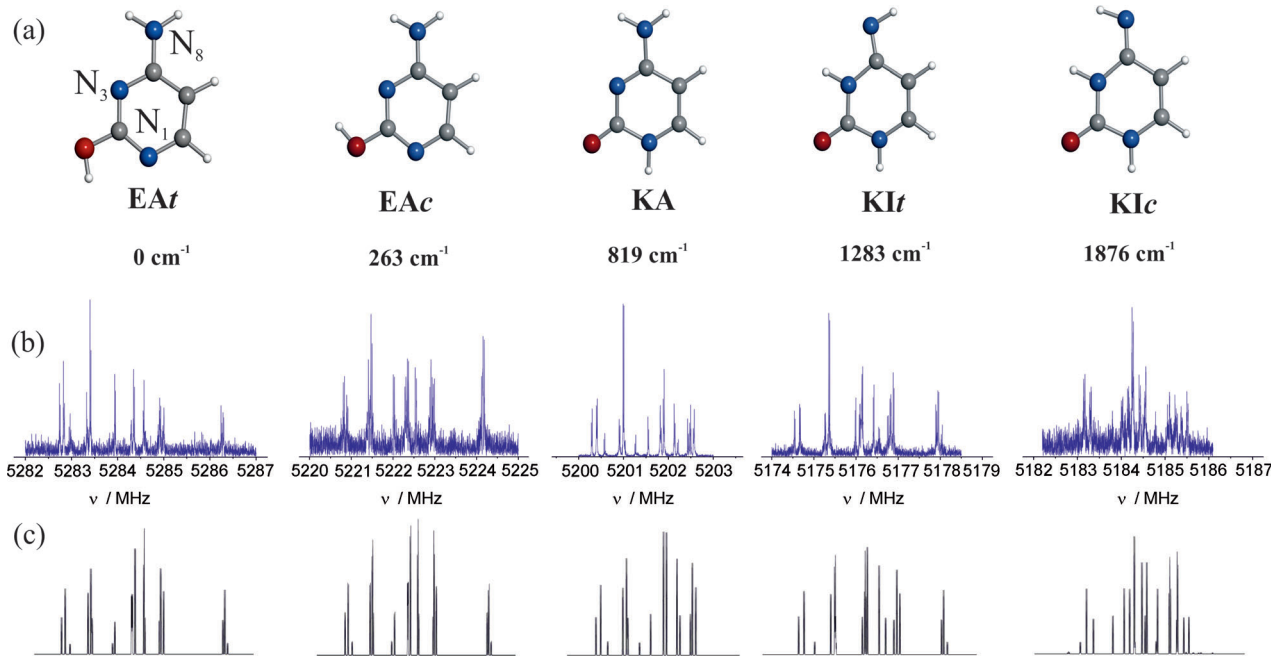

Figure 1. a) The five most stable species of cytosine: enol–amino *trans* (**EA*t***), enol–amino *cis* (**EA*c***), keto–amino (**KA**), keto–imino *trans* (**KI*t***), keto–imino *cis* (**KI*c***). These molecular forms are give in order of stability according to MP2/6-311++G(d,p) ab initio calculations. b) LA-MB-FTMW spectra for the $1_{1,1}$–$0_{0,0}$ rotational transition of the five species. c) Theoretical simulation of the nuclear quadrupole hyperfine structure for the $1_{1,1}$–$0_{0,0}$ rotational transition. The considerable differences among the various patterns act as fingerprints for tautomeric/conformational assignment.

We faced the challenge of analyzing the spectra and their complex hyperfine structures by carrying out geometry optimization[17] of the five predicted species to estimate the rotational and quadrupole coupling constants (see Table S1 of the Supporting Information). On this basis, we first identified structures for the $1_{1,1}$–$0_{0,0}$ transition of each species and later extended this assignment to other μ_a and μ_b-type transitions. All measured hyperfine components (listed in Tables S2–S6 of the Supporting Information) were fitted[18] by the use of a rigid-rotor Hamiltonian supplemented with a term to account for the nuclear-quadrupole-coupling contribution. To our knowledge, the quadrupole hyperfine structure of a spectrum has not been interpreted previously for three ^{14}N nuclei located in non-equivalent positions.

The comparison of experimental and predicted spectroscopic constants (Table 1 and Table S1 of the Supporting Information) led to conclusive identification of the detected rotamers (see also the good match between the experimental and theoretical nuclear-quadrupole-coupling patterns in Figure 1 b,c). Species **EA*t*** and **EA*c*** have very similar values for the three ^{14}N nuclei, since the two conformers differ only in the orientation of the hydroxy group. In these conformers, the positive values of χ_{cc} for $^{14}N_1$ and $^{14}N_3$, in the range from 1.0 to 1.8 MHz, indicate that these atoms are pyridinic nitrogen atoms, whereas the negative values for $^{14}N_8$, close to those observed for aniline[19a] or *p*-toluidine,[19b] indicate that this atom is an amino nitrogen atom. It is possible to discriminate conclusively between the two species on the basis of the trend in the change in the rotational constants. When passing from the **EA*t*** conformer to **EA*c***, the predicted/experimental changes $\Delta A = -58.0/-62.39$ MHz, $\Delta B = 15.6/17.36$ MHz, and $\Delta C = 0.1/0.39$ MHz are in excellent agreement. In going to tautomer **KA**, the χ_{gg} ($g = a, b, c$) values associated with atoms $^{14}N_3$ and $^{14}N_8$ do not change very much; however, those for atom $^{14}N_1$ change radically, in particular the χ_{cc} constant, which has a negative value in the range from −3.1 to −3.9 MHz, similar to those observed for imino nitrogen atoms,[20] including those of uracil[5] and thymine.[6] In passing to conformers **KI**, the quadrupole coupling constants associated with atom $^{14}N_3$ change, and the corresponding value of χ_{cc} reveals that it is now an imino N atom. Finally, the χ_{cc} values of $^{14}N_8$ in **KI*t*** and **KI*c*** are "chemically" different from all other χ_{cc} values. We do not have reference experimental values in this case. However, $^{14}N_8$, with values of 1.35 and 1.15 MHz, respectively, can be considered to have an electronic structure similar to that of a pyrimidine nitrogen atom (C=N–C group), as in the case of $^{14}N_1$ and $^{14}N_3$ of **EA*t*** and **EA*c*** and $^{14}N_3$ of **KA**, but as part of a C=N–H group. It can be seen from Table 1 that although the rotational constants are very similar to each other for the various species, the full sets of quadrupole coupling constants act as fingerprints for the identification of tautomers.

We measured the relative intensity[21] of the corresponding quadrupole component lines of the observed species with the exception of **KI*c***, for which the intensity of the spectrum was too weak. By using the theoretical values in Table S1 (see the Supporting Information) for the dipole-moment components, we obtained the following relative Gibbs energies: 0, 165, 40, and 290 cm^{-1} for the **EA*t***, **EA*c***, **KA**, and **KI*t***, respectively. The **EA** forms were found to be more abundant in the gas phase than the canonical **KA** form. Interestingly,

© 2013 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

Table 1: Spectroscopic constants for the observed tautomers and conformers of cytosine.[a]

		EA*t*	EA*c*	KA	KI*t*	KI*c*
A [MHz]		3951.85325(32)	3889.46510(38)	3871.54618(31)	3848.18174(41)	3861.2966(12)
B [MHz]		2008.95802(12)	2026.31804(12)	2024.97804(11)	2026.31068(31)	2011.41032(62)
C [MHz]		1332.47228(08)	1332.86951(10)	1330.33627(08)	1327.99167(10)	1323.19999(22)
N_1	χ_{aa} [MHz]	−2.6373(13)	−2.8007(18)	1.6211(19)	1.8518(69)	1.898(23)
	χ_{bb} [MHz]	1.1672(28)	1.0340(27)	1.4772(34)	2.0545(40)	2.104(28)
	χ_{cc} [MHz]	1.4701(28)	1.7667(27)	−3.0983(34)	−3.9063(40)	−4.002(28)
N_3	χ_{aa} [MHz]	2.2619(20)	2.2371(23)	2.5217(12)	2.1383(33)	2.105(20)
	χ_{bb} [MHz]	−3.6570(22)	−3.3890(25)	−3.5140(16)	1.6064(42)	1.764(28)
	χ_{cc} [MHz]	1.3951(22)	1.1519(25)	0.9923(16)	−3.7448(42)	−3.870(28)
N_8	χ_{aa} [MHz]	2.2167(17)	2.2237(17)	2.1802(17)	1.8033(87)	−2.091(15)
	χ_{bb} [MHz]	1.9511(20)	1.9832(20)	1.8429(26)	−3.1572(58)	0.940(14)
	χ_{cc} [MHz]	−4.1678(20)	−4.2069(20)	−4.0231(26)	1.3539(58)	1.151(14)
Δ_c [uÅ^2][b]		−0.1676(3)	−0.1767(4)	−0.2212(3)	−0.1789(5)	−0.2023(13)
σ [kHz][c]		1.1	1.4	1.1	1.3	3.9
N[d]		71	74	84	54	45

[a] Errors in parenthesis are expressed in units of the last digit. [b] $\Delta_c = I_c - I_a - I_b$ is the inertial defect. Conversion factor: 505379.1 MHz uÅ^2. [c] Standard deviation of the fit. [d] Number of fitted lines.

the theoretical calculations (see Table S1) did not reproduce the experimental observations.

The slightly negative values of the inertial defects, Δ_c, between −0.167 and −0.221 uÅ^2 (Table 1), show that all five forms of CY are effectively planar. This value would be zero for a hypothetical vibrationless planar molecule, but the out-of-plane torsion (OH group) or bending (C=O bending) or NH_2 inversion can give an additional negative contribution. A comparable value of −0.129 uÅ^2 was obtained for the planar diketo form of uracil.[5]

Herein we have reported the observation of five tautomeric species of cytosine under isolation conditions in the gas phase and their unequivocal identification by LA-MB-FTMW spectroscopy. The values of the inertial defects show that all species are effectively planar. The superior quality of this technique for the determination of different conformational/tautomeric forms of nucleic acid bases has been discussed previously.[22] In the present study of cytosine, we were able to detect keto–imine forms predicted to be more than 1200 cm^{-1} (in Gibbs values) above the global minimum. ^{14}N nuclear quadrupole patterns make it possible to obtain the spectral signatures for each individual tautomer in the complex sample and thus act as a sort of fingerprint. As an extension of the present study, it would be interesting to investigate how the observed tautomeric behavior is altered by the docking of water molecules to cytosine to create a model environment that would more closely resemble the biological medium.

Experimental Section

Details of the LA-MB-FTMW technique have been provided elsewhere;[4,23] we therefore give only a brief description of the spectrometer in this section. The CY molecules were vaporized by laser ablation by using the second harmonic (512 nm) of a Q-switched pulsed Nd:YAG laser. The ablation products were diluted in a light inert carrier gas expanding supersonically to form a molecular beam between the mirrors of a Fabry–Pérot resonator and probed there by Fourier transform microwave spectroscopy. The collinear arrangement of the supersonic jet and the resonator axis causes all lines to be split in two as a result of the Doppler effect. The molecular frequency is the arithmetic mean of the Doppler components.

Received: September 25, 2012
Revised: December 15, 2012
Published online: January 22, 2013

Keywords: cytosine · laser ablation · microwave spectroscopy · supersonic jets · tautomerism

[1] J. D. Watson, F. H. Crick, *Nature* **1953**, *171*, 737–738.
[2] J. D. Watson, F. H. Crick, *Nature* **1953**, *171*, 964–967.
[3] a) J. P. Schermann, *Spectroscopy and Modeling of Biomolecular Building Blocks*, Elsevier, Amsterdam, **2008**; b) B. C. Dian, A. Longarte, T. S. Zwier, *Science* **2002**, *296*, 2369; c) D. W. Pratt, *Science* **2002**, *296*, 2347.
[4] a) J. L. Alonso, C. Pérez, M. E. Sanz, J. C. López, S. Blanco, *Phys. Chem. Chem. Phys.* **2009**, *11*, 617–627, and references therein; b) S. Blanco, M. E. Sanz, J. C. López, J. L. Alonso, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 20183–20188; c) M. E. Sanz, S. Blanco, J. C. López, J. L. Alonso, *Angew. Chem.* **2008**, *120*, 6312–6316; *Angew. Chem. Int. Ed.* **2008**, *47*, 6216–6220.
[5] V. Vaquero, M. E. Sanz, J. C. López, J. L. Alonso, *J. Phys. Chem. A* **2007**, *111*, 3443.
[6] J. C. López, I. Peña, V. Vaquero, M. E. Sanz, J. L. Alonso, *J. Chem. Phys.* **2007**, *126*, 191103.
[7] J. L. Alonso, I. Peña, J. C. López, V. Vaquero, *Angew. Chem.* **2009**, *121*, 6257; *Angew. Chem. Int. Ed.* **2009**, *48*, 6141.
[8] See, for example: A. K. Chandra, D. Michalska, R. Wysokinky, T. Zeegers-Huyskens, *J. Phys. Chem. A* **2004**, *108*, 9593, and references therein.
[9] M. Szczesniak, K. Szczepaniak, J. *S.* Kwiatkowski, K. KuBulat, W. B. Person, *J. Am. Chem. Soc.* **1988**, 110, 8319.
[10] R. D. Brown, P. D. Godfrey, D. McNaughton, A. P. Pierlot, *J. Am. Chem. Soc.* **1989**, *111*, 2308.

 © 2013 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

[11] F. Dong, R. E. Miller, *Science* **2002**, *298*, 1227.
[12] E. Nir, M. Müller, L. I. Grace, M. S. de Vries, *Chem. Phys. Lett.* **2002**, *355*, 59.
[13] D. Shorokhov, PhD dissertation, University of Oslo, **2000**.
[14] V. Feyer, O. Plekan, R. Richter, M. Coreno, M. de Simone, K. C. Prince, A. B. Trofimov, I. L. Zaytseva, J. Schirmer, *J. Phys. Chem. A* **2010**, *114*, 10270.
[15] a) L. Lapinski, M. J. Nowak, I. Reva, H. Rostkowska, R. Fausto, *Phys. Chem. Chem. Phys.* **2010**, *12*, 9615 – 9618; b) L. Lapinski, I. Reva, M. J. Nowak, R. Fausto, *Phys. Chem. Chem. Phys.* **2011**, *13*, 9676 – 9684.
[16] J. K. Bragg, *Phys. Rev.* **1948**, *74*, 533.
[17] Gaussian 03, Revision B.04, M. J. Frisch et al., Gaussian, Inc., Pittsburgh PA, **2003**; see the Supporting Information.
[18] H. M. Pickett, *J. Mol. Spectrosc.* **1991**, *148*, 371.
[19] a) B. Kleibömer, D. Sutter, *Z. Naturforsch. A* **1988**, *43*, 56; b) A. Hellweg, *Chem. Phys.* **2008**, *344*, 281.
[20] a) R. Bohn, K. Hillig, R. Kuczkowski, *J. Phys. Chem.* **1989**, *93*, 3456; b) C. Tanjaroon, R. Subramanian, C. Karunatilaka, S. G. Kukolich, *J. Phys. Chem.* A **2004**, *108*, 9531; c) M. Stolze, D. H. Sutter, *Z. Naturforsch. A* **1987**, *42*, 49; d) R. D. Suenram, F. J. Lovas, G. T. Fraser, *J. Mol. Spectrosc.* **1988**, *127*, 472.
[21] S. Blanco, A. Lesarri, J. C. López, J. L. Alonso, *J. Am. Chem. Soc.* **2004**, *126*, 11675.
[22] W. Caminati, *Angew. Chem.* **2009**, *121*, 9192; *Angew. Chem. Int. Ed.* **2009**, *48*, 9030.
[23] A. Lesarri, S. Mata, J. C. López, J. L. Alonso, *Rev. Sci. Instrum.* **2003**, *74*, 4799.

 © 2013 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim