J|A|C|S
ARTICLES
Published on Web 12/10/2003

Unique Tautomeric Properties of Isoguanine

José Ramón Blas,[†] F. Javier Luque,*[,‡] and Modesto Orozco*[,†,§]

Contribution from the Unitat de Modelització Molecular i Bioinformática, Institut de Recerca Biomédica, Parc Científic de Barcelona, Josep Samitier 1-5 Barcelona 08028, Spain, Departament de Fisicoquímica, Facultat de Farmacia, Universitat de Barcelona, Avgda Diagonal 643, Barcelona 08028, Spain, and Departament de Bioquímica i Biologia Molecular, Facultat de Química, Universitat de Barcelona, Martí i Franquès 1, Barcelona 08028, Spain

Received June 20, 2003; E-mail: modesto@mmb.pcb.ub.es; javier@far1.far.ub.es

Abstract: The tautomeric properties of isoguanine (also named 2-oxoadenine or 2-hydroxyadenine) have been studied in the gas phase, in different pure solvents, and in the DNA environment using state of the art theoretical methods. Our results show that isoguanine constitutes an unique example of how tautomerism can be modulated by the environment. Compared to the tautomeric preference in the gas phase, both polar solvents and the DNA microenvironment dramatically change the intrinsic tautomeric properties of isoguanine. Tautomers which are important in physiological conditions are less than $1/10^5$ of the total population of isoguanine in the gas phase. The impact of the present findings in the understanding of spontaneous mutations and in the design of new nucleobases with multiple recognition properties is discussed.

Introduction

Watson and Crick were the first to notice that the recognition pattern of nucleobases is determined by their tautomeric state and that accordingly the DNA architecture stems from the predominance of specific tautomeric forms of nucleobases.[1] Further studies have shown that the change in the tautomeric equilibrium can promote mismatchings in the DNA, leading to spontaneous mutations in the genome.[2−4] Due to its biological impact, the tautomerism of nucleobases has been largely studied both in the gas phase and in solution using a variety of experimental and theoretical methods (for reviews, see ref 5). In physiological environments, most nucleobases exist in the keto and amino tautomeric forms, though in some cases other species are populated in the gas phase. Exceptions to this behavior are modified nucleobases such as *N*-methylated cytosine, which has a preference for the imino form,[6] nucleobases in the presence of metal cations,[7] or in very special sequence environments, like Hoogsteen cytosines in triplexes containing polyd(C-G·C) tracks[8] or thymines paired to other thymines in mutated DNAs.[9] With some exceptions,[6] the tautomeric pattern of modified nucleobases has not been thoroughly studied, though they are assumed to be in the keto−amino tautomeric form.

Figure 1. Structure of isoguanine and guanine.

One of the most interesting noncanonical purine nucleobases is isoguanine (isoG; see Figure 1), a mutagenic molecule which can be spontaneously formed by oxidative stress of adenine,[10] either at the DNA structure or more commonly at the nucleotide level (AMP; 10). In the 1960s, Rich and co-workers[11] realized that this nucleobase was able to form a very stable complex with isocytosine, which closely resembles the canonical G·C pair (see refs 10 and 12). At that time, this finding opened the

[†] Institut de Recerca Biomédica.
[‡] Facultat de Farmacia, Universitat de Barcelona.
[§] Facultat de Química, Universitat de Barcelona.

(1) Watson, J. D.; Crick, F. H. C. *Nature* **1953**, *171*, 964.
(2) Goodman, M. F. *Nature* **1995**, *378*, 237.
(3) Topal, M. D.; Fresco, J. R. *Nature* **1976**, *263*, 285.
(4) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984.
(5) (a) Kwiatkowski, J. S.; Pullman, B. *Adv. Heterocycl. Chem.* **1975**, *18*, 199. (b) Kwiatkowski, J. S.; Person, W. B. In *Theoretical Biochemistry and Molecular Biology*; Beveridge, D. L., Lavery, R., Eds.; Adenine Press: New York, 1990; pp 153−171.
(6) (a) Anand, N. N.; Brown, D. M.; Salisbury, S. A. *Nucleic Acids Res.* **1987**, *15*, 8167. (b) Fazakerley, G. V.; Gdaniec, Z.; Sowers, L. C. *J. Mol. Biol.* **1993**, *230*, 6. (c) Schuerman, G. S.; van Meervelt, L.; Loakes, D.; Brown, D. M.; Lin, P. K. T.; Moore, M. H.; Salisbury, S. A. *J. Mol. Biol.* **1998**, *282*, 1005.
(7) (a) Lippert, B.; Schöllhorn, H.; Thewalt, U. *J. Am. Chem. Soc.* **1986**, *108*, 6616. (b) Sponer, J.; Sponer, J. E.; Gorb, L.; Leszczynski, J.; Lippert, B. *J. Phys. Chem. A* **1999**, *103*, 11406.
(8) (a) Colominas, C.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc.* **1996**, *118*, 6811. (b) Soliva, R.; Laughton, C. A.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc.* **1998**, *120*, 11226. (c) Soliva, R.; Luque, F. J.; Orozco, M. *Nucleic Acids Res.* **1999**, *27*, 2248.
(9) Chattopadhyaya, R.; Ikuta, S.; Grzeskowiak, K.; Dickerson, R. E. *Nature* **1988**, *334*, 175.
(10) (a) Maki, H.; Kornberg, A. *J. Biol. Chem.* **1985**, *260*, 12987. (b) Kamiya, H. *Nucl. Acids Res.* **2003**, *31*, 517.
(11) Rich, A. In *Horizons in Biochemistry*; Kasah, M., Pullman, B., Eds.; Academic Press: New York, 1962; pp 103−126.
(12) (a) Roberts, C.; Banduru, R.; Switzer, C. *J. Am. Chem. Soc.* **1997**, *119*, 4640. (b) Chen, X.; Kierzek, R.; Turner, D. H. *J. Am. Chem. Soc.* **2001**, *123*, 1267.

 10.1021/ja036806r CCC: $27.50

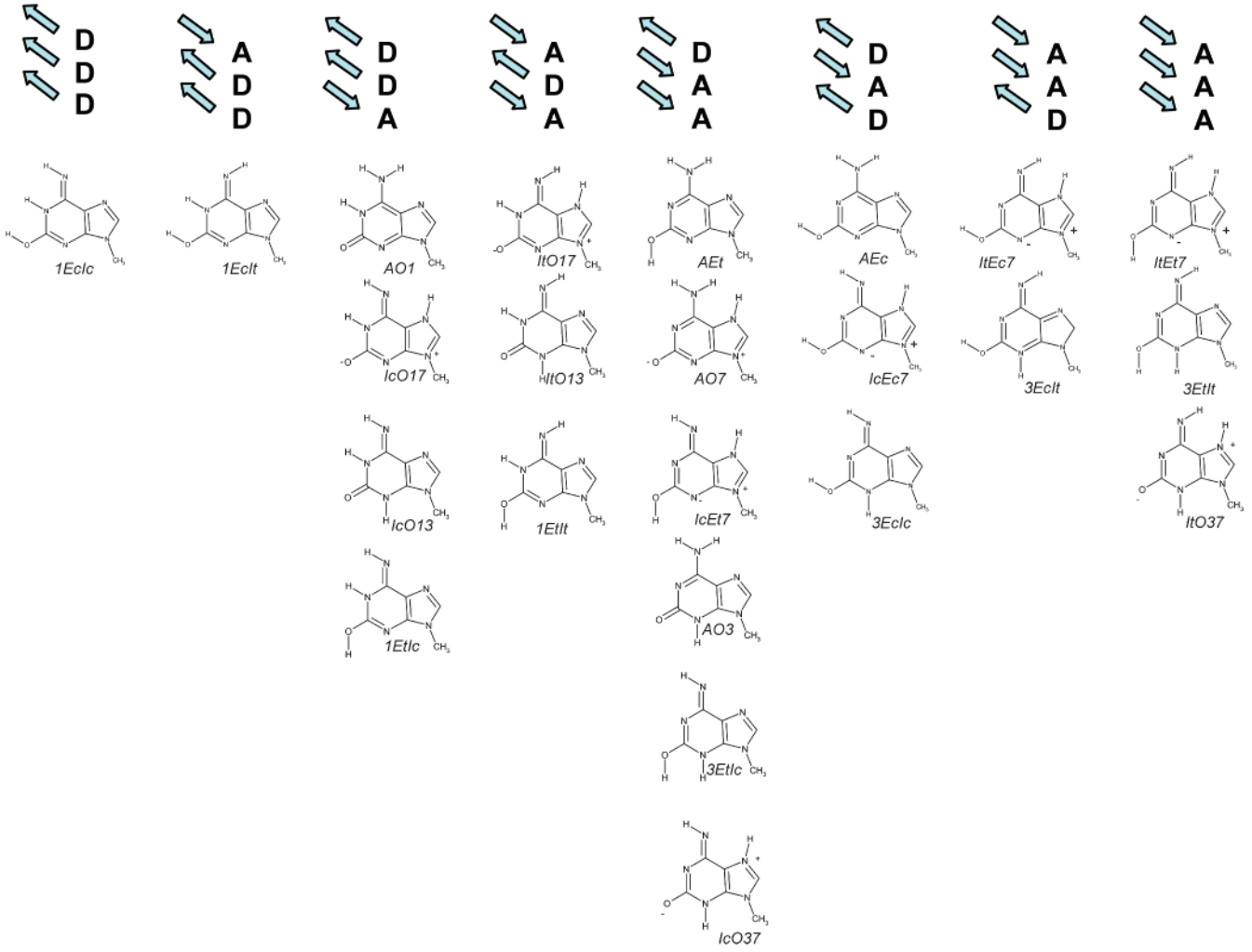

Figure 2. Representation of the 23 possible tautomers of isoguanine and the H-bond recognition pattern defined by them along the Watson−Crick side.

possibility to expand the genetic alphabet using nonstandard nucleobases, an issue that is still under intense research.[13] The insertion of isoG in the DNA increases the rate of mutagenesis,[12−14] which has been related to the similar ability of isoG to recognize cytosine and thymine, leading to pyrimidine transitions. The ability of isoG to recognize guanine (see ref 14d,f,g) and even adenine[10b] also explains the cytosine→guanine and cytosine→adenine transversions. Furthermore, isoG also modulates the stability of anomalous forms of DNA, like tetraplexes[15] or parallel-stranded DNA duplexes.[16] In summary, isoG is one of the most promiscuous and versatile nucleobases in nature.

The promiscuity of isoG has been related to its richness in tautomeric possibilities. Thus, the 23 possible tautomers of isoG (methylated at N9; see Figure 2) can lead to eight different Watson−Crick-like H-bonding patterns, that is, all the possible combinations of three donor/acceptor sites in the Watson−Crick side of a purine. Experimental studies[17] suggested that the keto N1-H form (AO1 in Figure 2) is the preferred form in water, while the enol O2-H forms (AEc and AEt) are the most stable tautomers in apolar solvents. Other experimental studies[18] pointed out that the keto N3-H form (AO3) can be also very stable in polar solvents, suggesting the possibility that isoG might populate at least four major tautomers depending on the environment. Medium-level ab initio calculations[12a] suggested that the two enol tautomers are the only detectable species in the gas phase and that the relative stability of the amino−oxo and imino−oxo tautomers increased in polar environments. On the basis of simplified models of physiological environments, it was also suggested that the molecule might exist as a mixture

(13) (a) Strazewski, P.; Tamm, C. *Angew. Chem., Int. Ed. Engl.* **1990**, *29*, 36. (b) Thuong, N. T.; Helene, C. *Angew. Chem., Int. Ed. Engl.* **1993**, *32*, 666. (c) Seela, F.; Debelak, H. *Nucleic Acids Res.* **2000**, *28*, 3224. (d) Amosova, O. A.; Fresco, J. R. *Nucleic Acids Res.* **1999**, *27*, 4632. (e) Robles, J.; Grandas, A.; Pedroso, E.; Luque, F. J.; Eritja, R.; Orozco, M. *Curr. Org. Chem.* **2002**, *6*, 1333.

(14) (a) Switzer, C.; Moroney, S. E.; Benner, S. A. *J. Am. Chem. Soc.* **1989**, *111*, 8322. (b) Kamiya, H.; Kasai, H. *J. Biol. Chem.* **1995**, *270*, 19446. (c) Kamiya, H.; Kasai, H. *FEBS Lett.* **1996**, *391*, 113. (d) Krishnamurthy, R.; Pitsch, S.; Minton, M.; Miculka, C.; Windhab, N.; Eschenmoser, A. *Angew. Chem., Int. Ed. Eng.* **1996**, *35*, 1537. (e) Kamiya, H.; Kasai, H. *Biochemistry* **1997**, *36*, 11125. (f) Kamiya, H.; Maki, H.; Kasai, H. *Biochemistry* **2000**, *39*, 9508. (g) Kamiya, H.; Kasai, H. *Nucleic Acid Res.* **2000**, *28*, 1640. (h) Kawakami, J.; Kamiya, H.; Yasuda, K.; Fujiki, H.; Kasai, H.; Sugimoto, N. *Nucleic Acid Res.* **2001**, *29*, 3289.

(15) Roberts, C.; Chaput, J. C.; Switzer, C. *Chem. Biol.* **1997**, *4*, 899.

(16) (a) Sugiyama, H.; Ikeda, S.; Saito, I. *J. Am. Chem. Soc.* **1996**, *118*, 9994. (b) Yang, X.-L.; Sugiyama, H.; Ikeda, S.; Saito, I.; Wang, A. H. *Biophys. J.* **1998**, *75*, 1163. (c) Cubero, E.; Luque, F.J.; Orozco, M. *J. Am. Chem. Soc.* **2001**, *123*, 12018.

(17) (a) Sepiol, J.; Kazimierczuk, Z.; Shugar, D. *Z. Naturforsch.* **1976**, *31C*, 361. (b) Seela, F.; Wei, C.; Kazimierczuk, Z. *Helv. Chim. Acta* **1995**, *78*, 1843.

(18) (a) Pitsch, S.; Krishnamurthy, R.; Bolli, M.; Wendeborn, S.; Holzener, A.; Minton, M.; Lesueur, C.; Schlonvogt, I.; Jaun, B.; Eschenmoser, A. *Helv. Chim. Acta* **1995**, *78*, 1621. (b) Eschenmoser, A. *Pure Appl. Chem.* **1993**, *65*, 1179.

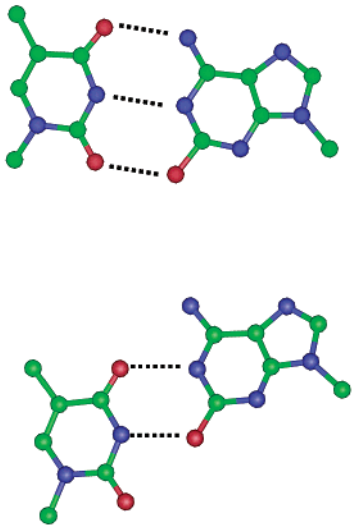

Figure 3. Schematic representation of the type of d(IsoG·T) pairings detected experimentally by X-ray crystallography in a piece of DNA (PDB entry 1BHR). In accordance with X-ray data only the position of the heavy atoms is displayed.

of amino−enol (AE), oxo−imino (ItO), and amino−oxo (mainly AO3) tautomers.[12a]

The complexity of the tautomeric scenario of isoG was reinforced by the identification by high-resolution X-ray data[19] of two different recognition patterns for the d(IsoG·T) pair in the same duplex (see Figure 3). The first recognition mode found in the crystal (PDB entry 1BHR) is consistent with a triple H-bond scheme, which should be related to an amino−enol or imino−enol tautomer (AEc or IcEc in Figure 2). The second recognition mode shows a wobble-like double H-bond scheme, where N1 and O2 groups of isoG are at H-bond distance of the O4 and N3 groups of thymine. This later scheme is consistent with the presence of isoG in the AO1 form or in any of the imino−oxo tautomers (see Figure 2). Furthermore, considering the accuracy of the electron density map, a H-bond scheme with interactions between N6 and N1 of isoG and O4 and N3 of thymine cannot be completely ruled out, opening the possibility that other tautomeric forms such as AO3 can play a stabilizing effect in the duplex. In summary, high-resolution X-ray data demonstrate that isoG exists in different tautomeric forms in physiological DNA, something unique, to the best of our knowledge. Unfortunately, X-ray data do not clarify which are the major tautomers of isoG in the DNA or the reasons for their predominance in front of other species.

In this paper, we present a systematic study of the tautomeric preferences of isoG in the gas phase and in different solvents using high-level ab initio calculations in conjunction with state-of-the-art self-consistent reaction field (SCRF) and molecular dynamics−thermodynamic integration calculations (MD/TI). The intrinsic pairing properties of selected tautomers of isoG were also investigated from ab initio, density functional, and classical force-field calculations. Finally, MD/TI simulations combined with ab initio calculations allowed us to determine the population of different tautomers of isoG in the DNA environment. Overall, we report here for the first time a complete theoretically derived picture of the unique tautomeric properties of isoG.

(19) Robinson, H.; Gao, Y.-G.; Bauer, C.; Roberts, C.; Switzer, C.; Wang, A. H.-J. *Biochemistry* **1998**, *37*, 10897.

Methods

Gas-Phase Calculations. Structures of the 23 possible tautomers of isoG (methylated at N9) were optimized at the HF/6-31G(d) level. The minimum nature of the optimized geometries was verified by the lack of imaginary vibrational frequencies. Zero-point, thermal, and entropic corrections to the electronic energies were introduced using the standard classical oscillator model and HF/6-31G(d) frequencies. Such classical corrections to the free energy of tautomerization were always small. To obtain more accurate electronic energies, single-point energies were perfomed using larger basis sets: 6-311++G-(d,p), aug-cc-pVDZ, and aug-cc-pVTZ[20] and introducing correlation effects. The later were first introduced at the MP2 level using the above-mentioned basis sets. To account for higher order effects, MP4 and CCSD(T) corrections to the MP2 values (for example: $E(\mathrm{MP4}^{\mathrm{corr}})_{\mathrm{large\ basis}} = E(\mathrm{MP2})_{\mathrm{large\ basis}} + \Delta\mathrm{MP4}_{\mathrm{small\ basis}}$) were computed using the 6-31G(d) basis set. On the basis of our previous experience, MP4- and CCSD(T)-corrected MP2/aug-cc-pVTZ values are very close to what should be the limit of accuracy for ab initio calculation of tautomerization energies in the gas phase.[8a,21] Higher order correlation effects were introduced only in those cases where the tautomer under study might have an physiological role.

Dimerization energies of selected tautomers of isoG with thymine or cytosine were computed at both classical and quantum mechanical levels. Starting structures for the dimers were modeled to mimic Watson−Crick pairs and to maximize the H-bond interactions. The structures were optimized using the B3LYP/6-31G(d) functional[22] and subsequently subjected to single-point calculations at the MP2/aug-cc-pVDZ level. For comparison purposes, additional calculations were performed using the AMBER-99 force-field.[23,24] Distortion effects were computed at the B3LYP/6-31G(d) level. The basis set superposition error in quantum mechanical calculations was corrected using Boys and Bernardi couterpoise method.[25]

SCRF Solvation Calculations. The impact of solvent in the tautomerism of the *N*-methylated isoG was examined from SCRF calculations for water, octanol, chloroform, and carbon tetrachloride[26] by using the HF/6-31G(d)-optimized version of

(20) (a) Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta* **1973**, *28*, 213. (b) Dunning, T. H. *J. Phys. Chem. A* **2000**, *104*, 9062.
(21) Trygubenko, S. A.; Bogdan, T. V.; Rueda, M.; Orozco, M.; Luque, F. J.; Sponer, J.; Slavicek, P.; Hobza, P. *Phys. Chem. Phys.* **2002**, *4*, 4192.
(22) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.
(23) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Fergurson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.
(24) Cheatham, T. E.; Cieplak, P.; Kollman, P. A. *J. Biomol. Struct. Dyn.* **1999**, *16*, 845.
(25) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553.
(26) (a) Bachs, M.; Luque, F. J.; Orozco, M. *J. Comput. Chem.* **1994**, *15*, 446. (b) Orozco, M.; Bachs, M.; Luque, F. J. *J. Comput. Chem.* **1995**, *16*, 563. (c) Luque, F. J.; Zhang, Y.; Aleman, C.; Bachs, M.; Gao, J.; Orozco, M. *J. Phys. Chem.* **1996**, *100*, 4269 (d) Luque, F. J.; Alemán, C.; Bachs, M.; Orozco, M. *J. Comput. Chem.* **1996**, *17*, 806. (e) Curutchet, C.; Orozco, M.; Luque, F. J. *J. Comput. Chem.* **2001**, *22*, 1180.
(27) (a) Miertus, S.; Tomasi, J. *Chem. Phys.* **1982**, *65*, 239. (b) Miertus, S.; Scrocco, E.; Tomasi, J. *Chem. Phys.* **1981**, *55*, 117.
(28) (a) Orozco, M.; Luque, F. J. *J. Am. Chem. Soc.* **1995**, *117*, 1378. (b) Luque, F. J.; Alambra, C.; Estelrich, J.; Orozco, M. *J. Org. Chem.* 1995, *60*, 969. (c) Hernández, B.; Luque, F. J.; Orozco, M. *J. Org. Chem.* **1996**, *61*, 5964. (d) Hernández, B.; Orozco, M.; Luque, F. J *J. Comput.-Aided Mol. Des.* **1996**, *10*, 535. (e) Hernández, B.; Orozco, M.; Luque, F. J. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 153. (f) Luque, F. J.; López-Bes, J. M.; Cemelí, J.; Aroztegui, M.; Orozco, M. *Theor. Chem. Acc.* **1997** *96*, 105. (g) Orozco, M.; Hernández, B.; Luque, F. J. *J. Phys. Chem. B* **1998**, *102*, 5228. (h) Forés, M.; Durán, M.; Solà, M.; Orozco, M.; Luque, F. J. *J. Phys. Chem. A* **1999**, *103*, 4525. (i) de Rosa, M.; Issac, R. P.; Márquez, M.; Orozco, M.; Luque, F. J.; Timken, M. D. *J. Chem. Soc., Perkin Trans. 2* **1999**, 1433.

the MST method.[27,28] This method yields solvation free energies for a large variety of neutral organic solutes with average root-mean-square errors below 0.7 kcal/mol[26e] and has proved to be very powerful to describe the influence of solvation on the tautomeric equilibrium of different nucleobases.[21,28] As described elsewhere, MST is a derivation of the PCM method[27] where cavities shapes and steric contributions to solvation (van der Waals and cavitation) have been parametrized for a series of solvents. In our experience for water, MST and PCM provide similar results. PCM has not been extensively parametrized for non aqueous solvents.

Following our standard procedure,[8,21,26,28] the free energy of solvation was computed by using the gas-phase optimized geometries. The relative stability of the tautomers in solution (eq 1) was estimated by combining gas-phase tautomerization free energies (ΔG^{gas}_{taut}) and differential solvation free energies ($\Delta\Delta G_{sol}$). In general, the tautomerization free energies determined at the MP2/6-31++G(d,p) in the gas phase were used, but for the most stable tautomers in the gas phase the averaged MP4/aug-cc-pVTZ and CCSD(T)/aug-cc-pVTZ values, which were very close in all cases, were used.

$$\Delta G^{sol}_{taut} = \Delta G^{gas}_{taut} + \Delta\Delta G_{sol} \quad (1)$$

As discussed elsewhere,[26,28] MST has been parametrized to obtain good solvation free energies from gas-phase geometries. For rigid molecules such as those considered here, the effect of geometry reoptimization in solution is expected to be small, leading to a constant increase in solvation free energy for all tautomers. In summary, geometry re-optimization in solution is expected to increase the complexity of the calculation making no significant contribution to the differential solvation free energy between tautomers.[28f] To verify this assumption, we computed the changes in the differential free energy of hydration ($\Delta\Delta G_{sol}$ for sol = water) for gas phase and solvent-optimized geometries in a few cases. For AO1→AEC equilibrium the effect of optimizing geometry in water was 0.3 kcal/mol; for AO1→AO3, 0.1 kcal/mol; for AO1→IcO13, 0.0 kcal/mol; and for AO1→AEt, 0.3 kcal/mol. In summary, these are negligible effects considering the magnitude of other errors in the calculation.

In addition, to gain extra confidence in key cases (those of potential biological impact) differential hydration free energy between corresponding tautomers was determined (mutations between AO1 and AEc, AEt, AO3, ItO13, and 1EcIt) using molecular dynamics calculations coupled to thermodynamic integration simulations (MD/TI; see below and ref 29).

Gas-phase calculations were carried out using the Gaussian-98 package.[30] MST calculations were performed using a locally modified version of the Monster-Gauss computer program.[31]

Molecular Dynamics and Thermodynamic Integration Calculations. Molecular dynamics (MD) simulations were first used to further explore the solvent effect on the tautomerization of isoG. For this purpose, a simulation system containing one N9-methylated isoG molecule (in the AO1 tautomeric form) surrounded by 477 water molecules in a cubic box was defined. The system was optimized (2000 cycles of steepest descent and 2000 cycles of conjugated gradient), heated (at 298 K), and equilibrated for 100 ps. The differences in hydration free energy between the AO1 tautomer and the AEc, AEt, AO3, IcO13, and ItO13 forms[29] were determined using the thermodynamic integration (TI) technique[32] and considering 21 and 41 windows of 20 ps each (divided in two sub-windows of 10 ps) for a total simulation time of 420 and 820 ps. The structures at the end of the trajectories (corresponding to the AEc, AEt, AO3, IcO13, ItO13, and 1EcIt tautomers) were then further equilibrated for 100 ps, and mutations were repeated starting from each tautomer and ending in the AO1. This lengthy procedure leads up to eight independent estimates of the hydration free energy change associated with each tautomeric change. Once the $\Delta\Delta G_{solv}$ is known from TI calculations, the tautomerization free energy in solution is easily determined using eq 1.

MD simulations were also used to analyze the characteristics of DNAs containing different tautomers of isoG in front of either cytosine or thymine. To this end, two 11-mer sequences, d(GATGC**isoG**AACAC)·d(GTGTT**T**CGCTC) (denoted d(isoG·T) DNA) and d(GATGC**isoG**AACAC)·d(GTGTT**C**CGCTC) (denoted d(isoG·C) DNA), were built taken the central 5-mer structure of Robinson's crystal structure[19] and extended three base pairs in each direction using standard fiber data of B-DNA.[33] These starting configurations were manipulated, when necessary, to generate the different tautomeric pairs. The structures were surrounded by 2190 water molecules and 20 Na^+ counterions to maintain neutrality. The final system was then optimized, heated, and equilibrated using our standard multistage protocol[34] and then subjected to 500 ps of unrestrained MD simulation (T = 298 K).

The structures obtained after the MD simulations were then used in 420 and 820 ps MD/TI simulations (21 or 41 (10 + 10 ps) windows; see above), where different tautomers of isoG were interconverted in the presence of either cytosine or thymine. Changes in both intermolecular and nonbonded intramolecular interactions were considered, and the intramolecular terms were corrected using simulations of the isolated nucleotide in the gas phase to determine the effect of DNA environment in the tautomerism of isoG. In a few cases, the mutation from one tautomer to the other implies a pathway where one group of

(29) Following the advice of one of the referees we determine the solvent effect on the AO1→ItO37 and AO1→ItEt7 also by MD/TI calculations. The idea was to determine the convergence of MST and MD/TI methods in cases of dramatic differences in solvation. The results of $\Delta\Delta G_{sol}$ obtained were for the AO1→ItO37 −28.5 (MD/TI) and −31.0 kcal/mol (MST); for the AO1→ItEt7 the values were −13.5 (MD/TI) and −14.1 kcal/mol (MST). Considering these extreme cases, in the analysis the RMSd between MST and MD/TI values is 0.7 kcal/mol, the scaling coefficient is 0.94 and the determination coefficient r^2 is 0.999 (compared with 0.5 kcal/mol, 1.00, and 0.98 for chemically revelevant tautomers). In summary, even in cases of large differences in $\Delta\Delta G_{sol}$ MST and MD/TI calculations agree well, at least for the compounds considered here.

(30) Gaussian 98, Rev. A.7: Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.. Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. Gaussian, Inc., Pittsburgh, PA, 1998.
(31) Peterson, M.; Poirier, R. MonsterGauss, Department of Biochemistry. University Toronto, Canada. Cammi, R.; Tomasi, J. Modified version, 1987, Luque, F. J.; Orozco, M. Modified version, 2002.
(32) Kirkwood, J. *J. Chem. Phys.* **1935**, *3*, 300.
(33) Arnott, S.; Bond, P. J.; Selsing, E.; Smith, P. J. C. *Nucleic Acids Res.* **1976**, *133*, 1405.
(34) (a) Shields, G. C.; Laughton, C. A.; Orozco, M. *J. Am. Chem. Soc.* **1997**, *119*, 7463. (b) Soliva, R.; Laughton, C. A.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc.* **1998**, *120*, 11126.

Table 1. Free Energy Difference (kcal/mol) between Tautomers of Isoguanine[a]

tautomer	ΔG(MP2) 6-311++G(d,p)	ΔG(MP2) aug-cc-pVDZ	ΔG(MP2) aug-cc-pVTZ	ΔG(MP4)[b] aug-cc-pVTZ	ΔG(CCSD(T))[b]aug-cc-pVTZ	best estimate[c]
IcO17	19.0					
ItO17	23.7					
AEc	−9.0	−8.7	−8.6	−7.1	−7.7	−7.4
AEt	−8.8	−8.7	−8.5	−6.9	−7.5	−7.2
AO7	29.2					
IcEc7	22.7					
IcEt7	24.8					
AO3	−0.2	−0.8	−0.7	−0.1	0.1	0.0
IcO13	6.6	6.3	6.3	6.5	6.6	6.6
ItO13	0.9	0.9	0.9	1.3	1.3	1.3
ItEc7	32.4					
ItEt7	35.4					
1EcIc	18.1					
1EcIt	10.6	10.3	10.0	11.0	10.3	10.7
1EtIc	9.8	10.0	9.9	10.8	10.1	10.5
1EtIt	3.1	3.6	3.6	4.7	4.1	4.5
3EcIc	21.3					
3EcIt	20.5					
3EtIc	29.6					
3EtIt	29.7					
IcO37	38.3					
ItO37	51.7					

[a] Values Relative to the AO1 Tautomer. A positive value means the AO1 tautomer is preferred. [b] Values obtained by adding the MP4-MP2/6-31G(d) or CCSD(T)-MP2/6-31G(d) corrections to the MP2-aug-cc-pVTZ estimates. [c] Value obtained by averaging MP4 and CCSD(T) estimates.

isoG is transitorily involved in unfavorable electrostatic interactions (amino−amino contacts between isoG and cytosine), leading to irreversible pathways and large hysteresis. To correct this problem, difficult mutations A→B (A and B being two tautomers of isoG) were carried out in three steps: (i) the charge at the 4-amino group of the paired cytosine was annihilated by changing also the charge at C4 to obtain a neutral molecule, (ii) the mutation A→B (B→A) was performed keeping the collapsing group uncharged, and (iii) the change made in step (i) was reverted. Using this procedure, which implies 200 ps more of MD simulation (100 for step (i) and 100 for step (iii)), reversible pathways were obtained. Once the "environmental" effect of DNA in the tautomerism of isoG ($\Delta\Delta G_{solv}$) was determined, eq 1 (where DNA = solvent) provides the free energy of tautomerization in the DNA.

All MD simulations were carried out in the isothermic/isobaric ensemble (1 atm, 298 K) using periodic boundary conditions and the PME technique to account for long-range electrostatic effects.[35] SHAKE[36] was used to maintain all the bonds at their equilibrium distances, which allowed us to use 2 fs time step for integration of Newton laws of motion. All simulations were carried out using the AMBER-6.0 computer program.[37] Molecular interactions were computed using TIP3P and AMBER-99[23,24,38] force fields supplemented with specific parameters for isoG derived using the RESP methodology and HF/6-31G(d) wave functions.[39]

Results and Discussion

Intrinsic Tautomeric Preferences. State-of-the-art ab initio calculations were performed to determine with high accuracy the intrinsic tautomeric preferences of isoG. Results in Table 1 show that at the MP2/6311++G(d,p) level of theory only eight tautomers are within 11 kcal/mol of the free energy of the reference form AO1 of N9-methyl isoG (see Figure 2). These tautomeric species correspond to two amino−oxo tautomers (AO1 and AO3), two amino−enol (AEc and AEt), two imino−oxo (IcO13 and ItO13), and three imino−enol tautomers (1EtIt, 1EtIc, and 1EcIt). Higher level calculations were performed only for these nine tautomers.

Comparison of MP2 results with 6-311++G(d,p), aug-cc-pVDZ, and aug-cc-pVTZ basis sets suggest that the tautomerization free energies are well converged with respect to the extension of the basis set (Table 1). Inclusion of higher order correlation terms (MP4 and CCSD(T)) seems, however, necessary to capture with good accuracy the relative stabilities between tautomers (Table 1). Thus, lower level calculations might overestimate the stability of enol tautomers in almost 2 kcal/mol. The similarity between MP4 and CCSD(T) results suggests, nevertheless, that further inclusion of higher order correlation terms should not lead to changes beyond a few tenths of kcal/mol in our better estimates of the tautomerization free energy in the gas phase.

The most populated tautomers in the gas phase are clearly the amino−enol forms: AEc and AEt. All calculations suggests that the cis conformer (AEc) is slightly more stable (by 0.3 kcal/mol) than the trans one. Very interestingly, the amino−oxo tautomers (AO1 and AO3) are very disfavored in the gas phase (they represent a negligible 0.001% of the population of isoG in the gas phase). The imino−oxo tautomer ItO13 shows a stability not very different to that of the amino−oxo tautomers (1.3 kcal/mol less stable than AO1). The enol−imino forms are largely disfavored, the preferred species being 5 kcal/mol less stable than the AO1 form. In summary, the only significant species of isoG in the gas phase are predicted to be the amino−

(35) Darden, T. A.; D. M.York, D. M.; Pedersen, L. G. *J. Chem. Phys.* **1993**, *98*, 10089.
(36) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.
(37) Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham, T. E., III; Ross, W. S.; Simmerling, C. L.; Darden, T. L.; Marz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Tsui, V.; Radmer, R. J.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. AMBER6, University of California, San Francisco, 1999.
(38) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
(39) Bayly, C. E.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269.

Table 2. Relative Free Energy of Solvation ($\Delta\Delta G_{sol}$; kcal/mol) in Water, *n*-Octanol, Chloroform, and Carbon Tetrachloride of the Different Tautomers of Isoguanine with Respect to the AO1 Tautomer[a]

tautomer	water	*n*-octanol	$CHCl_3$	CCl_4
IcO17	−8.0	−5.4	−3.0	−1.3
ItO17	−12.0	−8.5	−5.5	−2.7
AEc	8.5	5.0	3.4	1.0
	8.8 ± 0.1			
AEt	8.1	4.4	3.4	0.8
	7.8 ± 0.4			
AO7	−20.0	−12.7	−9.6	−5.1
IcEc7	−3.9	−3.7	−1.9	−1.4
IcEt7	−5.0	−4.8	−2.9	−1.9
AO3	0.2	0.4	0.1	−0.1
	0.6 ± 0.2			
IcO13	−0.1	0.0	0.9	0.7
	0.4 ± 0.3			
ItO13	3.9	2.6	2.6	1.5
	3.0 ± 0.1			
ItEc7	−11.5	−9.1	−5.8	−3.4
ItEt7	−14.1	−11.2	−7.5	−4.3
1EcIc	−3.6	−2.9	−1.3	−0.9
1EcIt	2.4	1.2	1.8	0.5
	3.2 ± 0.2			
1EtIc	0.9	1.0	1.0	0.2
1EtIt	5.9	3.9	3.7	1.4
3EcIc	−2.2	−1.5	−0.3	−0.7
3EcIt	−2.0	−1.2	−0.2	−0.6
3EtIc	−5.5	−4.1	−2.1	−1.5
3EtIt	−6.1	−4.2	−2.3	−1.5
IcO37	−18.4	−12.7	−8.8	−4.5
ItO37	−31.0	−20.9	−14.9	−7.7

[a] Values in roman correspond to MST calculations. Values in italics (standard errors indicated) correspond to MD/TI calculations. Positive values mean better solvation of the AO1 tautomer.

Table 3. Free Energy of Tautomerization (kcal/mol) in Solution Obtained by Combining the Best ab Initio Estimate of the Gas Phase Tautomerization Free Energy and MST (Roman) or MD/TI (Italics) Values of Differential Solvation Free Energies (See eq 1)[a]

tautomer	water	*n*-octanol	$CHCl_3$	CCl_4
IcO17	11.0	13.6	16.0	17.7
ItO17	11.7	15.2	18.2	21.0
AEc	1.1	−2.4	−4.0	−6.4
	1.4			
AEt	0.9	−2.8	−3.8	−6.4
	0.6			
AO7	9.2	16.5	19.6	24.1
IcEc7	18.9	19.0	20.9	21.3
IcEt7	19.8	20.0	21.9	22.9
AO3	0.2	0.4	0.1	−0.1
	0.6			
IcO13	6.5	6.7	7.5	7.3
	7.0			
ItO13	5.2	3.9	4.0	2.8
	4.3			
ItEc7	20.9	23.2	26.6	29.0
ItEt7	21.4	24.2	28.0	31.1
1EcIc	14.5	15.2	16.8	17.2
1EcIt	13.1	11.8	12.4	11.1
	13.9			
1EtIc	11.3	11.4	11.7	10.7
1EtIt	10.3	8.3	8.1	5.8
3EcIc	19.1	19.8	20.9	20.6
3EcIt	18.5	19.3	20.3	19.9
3EtIc	24.1	25.5	27.5	28.1
3EtIt	23.6	25.4	27.4	28.0
IcO37	19.9	25.6	29.5	33.8
ItO37	20.7	30.7	36.8	44.0

[a] All values are referred to the AO1 tautomer. Positive values mean that the AO1 tautomer is favored.

enol tautomers, in good agreement with experimental findings by Sepiol et al. and Seela et al.[17]

Solvent Effect. The tautomers of isoG have very different dipole moments in the gas phase (from 2 D in AEc to 14.2 D in ItEt7) and display very different patterns of H-bond interactions, suggesting that the solvent can largely modulate the tautomeric population.[12,17,18] To investigate this point, SCRF-MST calculations were performed considering four representative solvents of different permittivity: water, octanol, chloroform, and carbon tetrachloride. Results are summarized in Table 2, where the total solvation free energy (relative to the canonical AO1 tautomer) is given. As expected, the magnitude of the solvent effect changes with the polarity of the solvent. For CCl_4 the solvent effect is small, but not negligible (up to 7 kcal/mol) for some minor tautomers having large charge separation. The increase in the solvent's polarity enlarges the magnitude of the solvent effect, which for water can be as large as 30 kcal/mol (AO1 taken as reference, see Table 2). The amino−enol tautomers are poorly solvated compared with the rest of tautomers, especially in water, as expected from their small dipole moments (2−3 D). The amino−oxo tautomers (AO1 and AO3) are better solvated than the amino−enol forms, but still worse solvated than most of the imino (oxo and enol) tautomers. Not surprisingly, tautomers such as AO7, ItO17, and ItEt7, which display a clear charge separation between rings (see Figure 2) and are then very unstable in the gas phase (up to 30 kcal/mol less stable than AO1; seeTable 1), are very well solvated in aqueous solution.

Combination of results in Tables 1 and 2 provides tautomerization free energies of isoG in different pure solvents (see Table 3). Clearly, solvation reduces the large differences in stability found between tautomers in the gas phase (the range of 60 kcal/mol difference in stability in the gas phase is reduced to 24 kcal/mol in aqueous solution). The amino−enol tautomers AEc and AEt are expected to be the major species of isoG in CCl_4 and in chloroform (see Table 3). In *n*-octanol, the two amino−oxo tautomers (AO1 and AO3) are within 2 kcal/mol of the most stable amino−enol species. Finally, in water the tautomeric preference is shifted toward the amino−oxo forms, though sizable amounts of amino−enol tautomers (AEc, AEt) are also expected. Imino species (like ItO13), which have a sizable population in very apolar solvents, are not expected to be significantly populated in aqueous solution. In summary, the results indicate that solvation in water completely alters the tautomeric scenario of isoG.

As noted above, the HF/6-31G(d)-optimized version of the MST algorithm used here has been successfully applied to study a variety of tautomeric process in solution,[28] which gives confidence on the results displayed in Tables 2 and 3. However, to verify the quality of the results,[40] the hydration effect on the equilibrium between selected tautomers (AO1, AO3, AEc, AEt, IcO13, ItO13, and 1EcIt) was also computed using MD/TI calculations with explicit solvent representation. The relative free energies of hydration and the corresponding tautomerization free energies in aqueous solution are shown in Tables 2 and 3. There is a very good agreement between MD/TI and MST results. Thus, the RMSd between both the corresponding values is only 0.5 kcal/mol for the compounds shown in Table 2 (see ref 29), the optimum scaling factor (MST vs MD/TI) is 1.005,

(40) Orozco, M.; Luque, F. J. *Chem. Rev.* **2000**, *100*, 4187.

and the determination coefficient (r^2) is 0.98 (see ref 29) if the extended series is considered). This almost perfect identity between the results provided from two very different solvation approaches suggests that the conclusions drawn here on the effect of solvation on the tautomerism of isoG are solid and clearly out of the range of error expected in our simulations.

Theoretical calculations presented here agree very well with previous experimental data,[17,18] which suggested that while the enol AEc/AEt forms were the only detectable tautomers in apolar solvents, the enol forms AO1 and AO3 became the most popular ones in water. Our calculations can be also qualitatively compared with previous theoretical calculations by Switzer and co-workers,[12a] who used a low-level SCRF method to mimic the solvent effect on isoG tautomerism. These authors found also a strong stabilization of the keto form in polar solvent (a generic polar solvent with $\epsilon = 40$ was used in Switzer's work), but without the shift in tautomeric preferences found here. In summary, our different calculations are self-consistent and agree very well with all the available experimental data, and also qualitatively well with the general trends reported in previous lower level theoretical calculations. Overall, this general agreement gives strong confidence on the quality of our calculations.

Pairing Properties of Isoguanine. As shown in Figure 2, isoG can in principle display all possible H-bond interaction patterns in the Watson−Crick side. Six H-bond patterns fulfill the requirement for "Watson−Crick-like" recognition of thymine (a donor (D)−acceptor (A) pattern contiguous in the space). The analysis of all the pairings arising from these six H-bond patterns would imply the study of 18 different dimers, but most of the tautomers involved are very unstable, and the corresponding dimers are then meaningless. Accordingly, only six dimers (see Figure 4) representative of four isoG·T interaction patterns were built and optimized at the B3LYP/6-31G(d) level: AO1·T and IcO13·T (DDA pattern); ItO13·T (ADA pattern); AEt·T and AO3·T (DAA pattern), and finally the AEc·T pair (DAD pattern). The optimized geometries are shown in Figure 4, which also shows the interaction energy determined from DFT, MP2 and AMBER calculations (see below). Inspection of the optimized geometries reveals good H-bond geometries for the six dimers, with standard H-bond distances and angles. With the exception of the AO3·T pair, which shows large propeller twist (C6−N1−N3−C2 dihedral angle of 3°), all the dimers are planar, as desired for interactions in the DNA environment. Three of the pairs show a standard Watson−Crick geometry, while the other three correspond to Watson−Crick wobble pairings.

The interaction energies computed at different levels of theory are very similar (see Figure 4). The agreement between methods is even larger if relative values are considered (comparison of MP2, DFT, and AMBER values yields correlation coefficients between 0.98 and 0.99). There are not large differences between the interaction energies of wobble and standard Watson−Crick pairs (see below), though the later are expected to be better incorporated into a regular helical structure. Interestingly, the interaction energies isoG·T vary from −9 to −18 kcal/mol, the formation of some isoG·T dimers in the gas phase is competitive with that of the canonical A·T pair, whose interaction energy ranges from −12 to −14 kcal/mol.

The most stable pairings are AEc·T and AO1·T, the energy difference between them being within the expected level of accuracy of the methods. However, if the relative stability of the AE_C and AO1 forms is accounted for, we can conclude that the AEc·T pair is the major form of the isoG·T dimer in the gas phase (see Table 1). Other possible isoG·T pairings are clearly less stable, since they often imply the existence of imino forms. Based on our best estimates of interaction and tautomerization energies, the following ordering of stability for the isoG·T dimer in the gas-phase arises: AEc·T (−16.8 kcal/mol) ≫ AEt·T (−11.5 kcal/mol) > AO1·T (−9.7 kcal/mol) ≫ AO3·T (−4.8 kcal/mol) ≫ ItO13·T (−1.7 kcal/mol) > IcO13·T (1 kcal/mol). However, considering the tremendous impact of the environment in the tautomerism of isoG, important changes in the relative stability of the dimmers might occur in the DNA environment (see below).

Analysis of Figure 2 shows that up to five H-bond patterns are consistent with a Watson−Crick-like double H-bond interaction between isoG and cytosine (acceptor−donor or donor−donor contiguous in the space), and one H-bond pattern is compatible with a triple H-bond (1EcIt tautomer). Therefore, 14 tautomers of isoguanine could a priori be involved in Watson−Crick-like pairing with cytosine. As noted above, by excluding pairings involving very unstable tautomers of isoG, only five pairs (see Figure 5) can be defined. The corresponding dimmers were generated and optimized at the B3LYP/6-31G-(d) level: AO1·C and IcO13·C (DDA pattern); ItO13·C (ADA pattern); AEc·T pair (DAD pattern); and finally the 1EcIt·C pair (ADD pattern). Geometry optimization in the gas-phase yields structures which preserve the expected H-bond pattern, but where large buckle and propeller twist movements signal the existence of amino−amino or keto−keto repulsions between isoG and cytosine dihedral angles between nucleobases of 45−45° are found for three dimers in Figure 5). In fact, only two pairing schemes lead to planar dimers (IcO13·C and 1EcIt·C), and both imply the existence of minor tautomeric forms of isoG.

The DFT, MP2, and AMBER interaction energies for the isoG·C pairs in the gas phase agree well, as found for the isoG·T pairings (see above), with AMBER results slightly overestimating the MP2 and DFT values. As noted above, the agreement is even better if relative values are considered (determination coefficients (r^2) larger than 0.98). Such an agreement reinforces the confidence in the estimated interaction energies for isoG·T and isoG·C pairing (Figures 4 and 5) and gives particularly confidence in the quality of the AMBER force-field to represent interactions involving isoG.

The four pairing schemes having a double H-bond contact have averaged interaction energies in the range from −11 to −15 kcal/mol (i.e., similar to an A·T pair), which are slightly less stable than the best isoG·T interactions (see Figure 4) and clearly less stable than the canonical G·C pair (from −25 to −29 kcal/mol; Figure 5). On the contrary, the 1EcIt·C dimerization energy is very large and negative (from −30 to −35 kcal/mol; Figure 5) due to the existence of three hydrogen bonds and two favorable secondary interactions. If the intrinsic stability of the tautomers is considered (Table 1), the following order of stability is found for the dimers: AEc·C (−12.8 kcal/mol) ≈ 1EcIt·C (−12.2 kcal/mol) ≫ AO1·C (−6.7 kcal/mol) > ItO13·C (−4.1 kcal/mol) ≫ IcO13·C (2.1 kcal/mol). Therefore, the isoG·C pair in the gas phase should coexist as a mixture of 1EcIt·C and AEc·C dimers. The occurrence of the amino−enol form is not surprising considering the large population of this

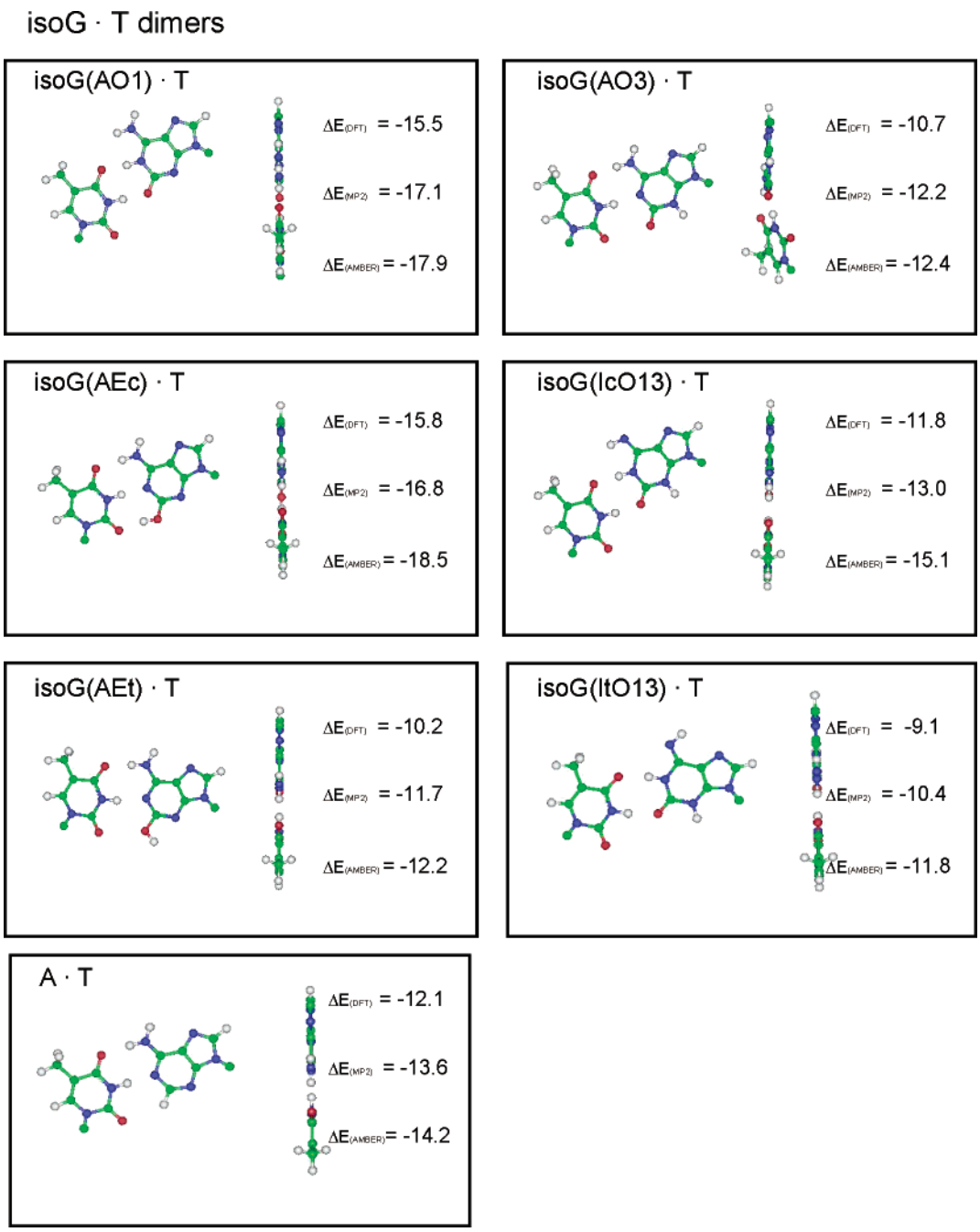

Figure 4. Front and side views of the different B3LYP/6-31G(d)-optimized dimers of isoguanine and thymine. Interaction energies (in kcal/mol) computed at different levels of theory (DFT: B3LYP/6-31G(d), MP2: MP2/aug-cc-pVDZ, and AMBER) are also displayed. The canonical Watson−Crick adenine-thymine dimer is shown as reference.

tautomer in isolated isoG. The involvement of the imino−enol tautomer 1EcIt is, however, unexpected because of its reduced stability for isolated isoG. Clearly, the tautomeric plasticity of isoG leads to a very complex scenario of interactions, where the complementary nucleobase as well as the surrounding environment can change the tautomeric state of the molecule. Interestingly, the stabilization energy of the dimer (interaction energy corrected by tautomerization energy) is around −20 kcal/mol, i.e., around 5 kcal/mol less stable than the canonical G·C pair, but still 8 and 3 kcal/mol more stable than the Watson−Crick A·T dimmer and the best isoG·T pair.

DNA Simulations. The preceding results show that the change from the gas phase to a polar solvent like water or the presence of a complementary base can alter the tautomeric preferences of isoG in the gas phase. What is then the impact of the DNA environment in the tautomerism of isoguanine?. To investigate this point, MD/TI calculations for the mutations between selected tautomers of isoG when paired to either thymine or cytosine in the middle of a 11-mer DNA duplex (see Methods) were performed.

For the isoG·T pairings, the AO1, AEc, IcO13, and ItO13 tautomers were considered, while for the isoG·C pairings the tautomers considered were 1EcIt, AO1, AEc, IcO13, and ItO13. Analysis of the different DNA duplexes suggests that the presence of a given tautomer of isoG does not induce major changes in the helical structure, which demonstrates the large plasticity of DNA. Interestingly, the expected H-bond pattern persists along the entire trajectory, suggesting that all the pairings

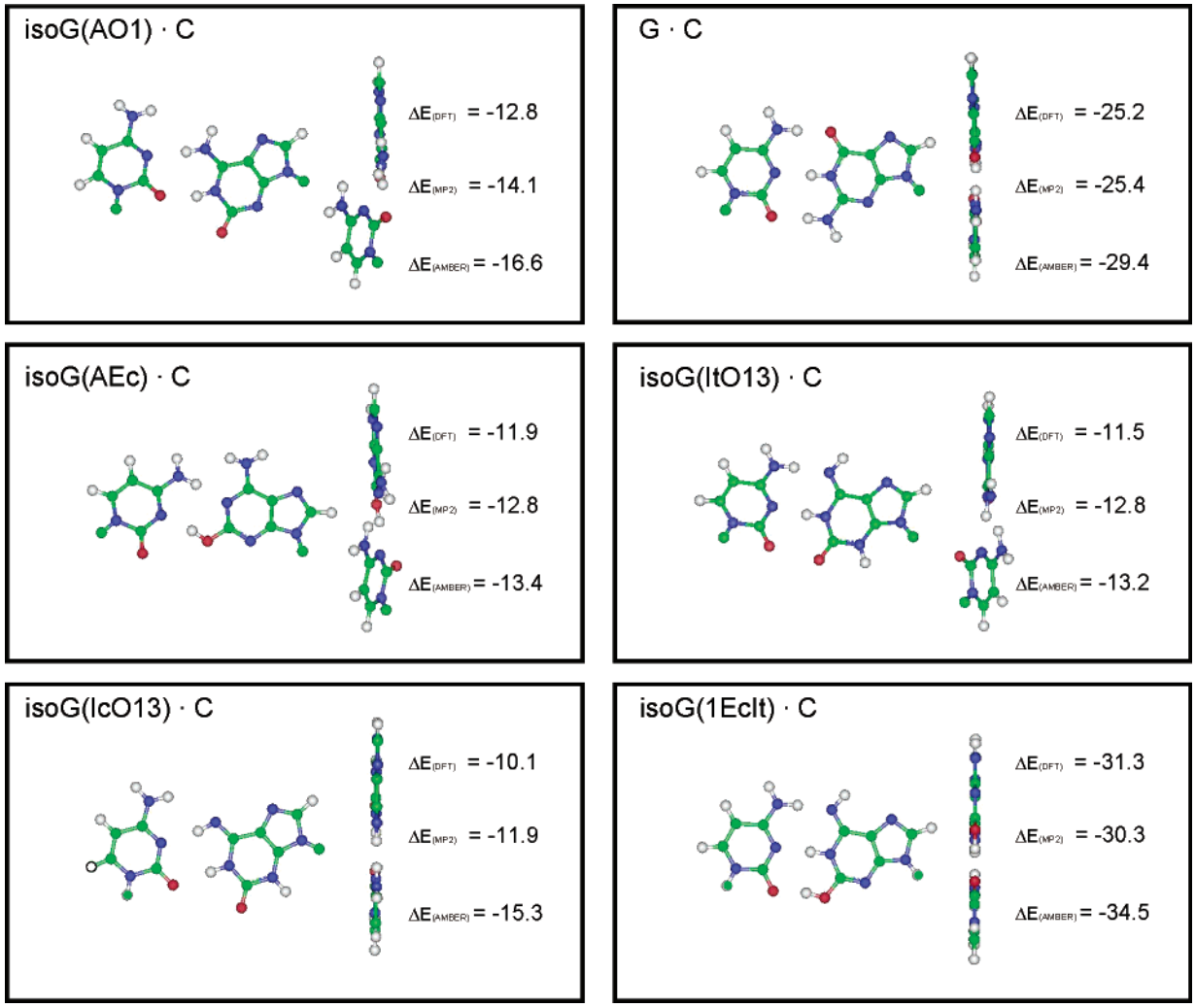

Figure 5. Front and side views of the different B3LYP/6-31G(d)-optimized dimers of isoguanine and cytosine. Interaction energies (in kcal/mol) computed at different levels of theory (DFT: B3LYP/6-31G(d), MP2: MP2/aug-cc-pVDZ, and AMBER) are also displayed. The canonical Watson−Crick cytosine-guanine dimer is shown as reference.

Table 4. Contribution of the DNA (kcal/mol) to the Differential Stabilization of a Given Tautomer (Relative to AO1) of Isoguanine in DNA in Front of Thymine or Cytosine ($\Delta\Delta G_{solv}$) and Total Free Energy of Tautomerization in the DNA ($\Delta\Delta G_{taut}$)[a]

mutation	pairing	$\Delta\Delta G_{solv}$	$\Delta\Delta G_{taut}$
AO1→AEc	d(isoG·T)	7.3 ± 0.3	−0.1 ± 0.4
AO1→IcO13	d(isoG·T)	5.1 ± 0.4	11.7 ± 0.4
AO1→ItO13	d(isoG·T)	6.8 ± 0.8	8.1 ± 0.8
AO1→AEc	d(isoG·C)	4.6 ± 0.3	−2.1 ± 0.4
AO1→IcO13	d(isoG·C)	6.9 ± 0.4	12.8 ± 0.4
AO1→ItO13	d(isoG·C)	11.2 ± 0.6	12.8 ± 0.6
AO1→1EcIt	d(isoG·C)	−2.5 ± 0.3	8.2 ± 0.5

[a] A positive sign means that the AO1 tautomer is favored. Errors in $\Delta\Delta G_{taut}$ are computed by propagating those in solvation simulations and those in the ab initio gas-phase calculation.

are sterically permitted in the duplex structures. The equilibrated structures were used to perform MD/TI simulations. In all the cases, smooth perturbation profiles, without apparent discontinuities, were found. Mutations in the AO1→X and X→AO1 directions yielded very similar values, and short and long simulations provided also similar $\Delta\Delta G_{sol}$ values (see Table 4). In summary, despite all the uncertainties intrinsic to these calculations, we are confident on the statistical quality of our estimates of the "DNA" effect ($\Delta\Delta G_{sol}$) on the tautomeric equilibrium of isoG.

The DNA stabilizes the AO1 form with respect to other tautomers of isoG (see $\Delta\Delta G_{sol}$ values in Table 4) when paired in front of thymine. The "solvation" effect of DNA is very strong, and in some cases quite different to that of water. For example, the solvent-induced stabilization of AO1 relative to IcO13 and ItO13 forms is around 5 and 7 kcal/mol in the DNA, but around 0 and 3 kcal/mol, respectively, in bulk water. The effect of the DNA environment on the tautomerism of isoG is also different to that expected from simple H-bond interaction energies in the gas phase (see Figure 4). For example, the AEc·T and AO1·T dimers are equally stable in the gas phase (Figure 4), but the DNA mainly stabilizes (by around 7 kcal/mol) the latter pairing (Table 4).

To obtain a more detailed picture of the influence of nucleobase interactions in the stability of a given tautomeric form of isoG in the DNA, we computed the MD-averaged nucleobase−nucleobase interaction energy (H-bond and stacking (both intra- and inter-) for the central trimer of base-pairs containing isoG paired to thymine (Table 5). The strongest H-bonds occur for the trimer containing the AEc·T pair, as expected from gas phase results (Figure 4), and the remaining trimers show similar H-bond energies, smoothing the differences expected from gas phase calculations. The best stacking

Table 5. MD-Averaged Total Nucleobase−Nucleobase Interaction Energy (⟨*E*(tot)⟩; kcal/mol) for the Central Trimer of the Different Duplexes Containing Isoguanine (in Different Tautomeric Forms) Paired to Cytosine and Thymine[a]

pairing	⟨*E*(H-bond)⟩	⟨*E*(stack)⟩	⟨*E*(tot)⟩
AO1·T	−50.8	−28.9	−79.7
AEc·T	−53.6	−29.8	−83.4
IcO13·T	−49.6	−32.1	−81.7
ItO13·T	−48.9	−30.0	−78.9
AO1·C	−53.3	−31.7	−85.0
AEc·C	−46.5	−29.2	−75.7
IcO13·C	−48.6	−25.5	−74.1
ItO13·C	−45.1	−25.1	−70.2
1EcIt·C	−68.3	−27.8	−96.1

[a] The total interaction energy is divided into hydrogen bonding and stacking (both intra- and interstrand) contributions for the three central base pairs. Values are obtained by averaging the last 0.5 ns of 1.5 ns trajectories. Calculations are done with the nucleobases capped and neutralized by an UA Me Group and the usual AMBER parameters.

interactions occur in the trimer containing the IcO13·T pair, the remaining trimers showing similar values. Overall, the total nucleobase-nucleobase interaction energies (see values in Table 5) suggest the following order of stability: AEc·T > IcO13·T > AO1·T > ItO13·T. This order does not reflect the DNA-induced stabilization of the different tautomers suggested by free energy calculations (AO1 > IcO13 > ItO13 > AEc; Table 4). Therefore, not only direct contacts, but also the entropic term, bulk solvent, and other remote interactions modulate the preferential stabilization of tautomers of isoG paired to thymine in DNA duplexes.

Tautomerization free energies ($\Delta\Delta G_{taut}$) in the DNA can be obtained by combining the $\Delta\Delta G_{sol}$ values in Table 4 with the intrinsic tautomeric preferences in Table 1. Results (see Table 4) indicate that the AO1 and AEc tautomers coexist in the DNA paired with thymine in almost equal percentages ($\Delta\Delta G_{taut}$ close to zero), while the imino tautomers (IcO13 and ItO13) have a negligible population. Accordingly, with the obvious cautions due to the expected errors in the simulations, we can assume that both AO1·T and AEc·T pairings might occur in the DNA as a consequence of a unique cancellation of large DNA-specific solvation and intrinsic tautomeric terms. This finding allows us to realize the surprising finding of the existence of both wobble (that we assigned here to the AO1 tautomer) and Watson−Crick (that we assigned to the AEc form) pairings in the crystal structure of a duplex containing isoG paired to thymine.[19] With such a small stability difference between the AO1 and AEc tautomers in the DNA, small changes in the duplex sequence, structure, or solvent composition can alter the population of these two species. We can also speculate that the enzymatic incorporation of isoG in front of T (or T in front of isoG) the AO1 ↔ AEc equilibrium should occur based on the AEc form, since this species would yield Watson−Crick-like pairing easy to be recognized by DNA(RNA) polymerases, in contrast to the wobble pairing induced by the AO1 tautomeric form.

Once the pairing of isoguanine with thymine in DNA was characterized, we should considered the alternative pairing between isoguanine and cytosine. Also in this case the duplexes containing isoguanine in any of its tautomeric states (AO1, AEc, IcO13, ItO13, and 1EcIt) were stable, the sampled structures were similar in all the cases, and the isoG·C pattern of interaction is that expected from Figure 5. In summary, as previously noted for the isoG·T duplexes the large plasticity of DNA allows it to easily accommodate local changes in the structure induced by the presence of isoG·C pairs. Mutations between tautomers of isoguanine (AO1, AEc, IcO13, ItO13, and 1EcIt) were smooth and reversible and preserved well the expected hydrogen-bond pattern. In summary, as found previously for the isoG·T duplexes, free energy estimates in Table 4 seem statistically reliable and well converged.

The DNA stabilizes the tautomers of isoG paired to cytosine in the order 1EcIt > AO1 > IcO13 > ItO13 (see Table 4). The "environment" effect of DNA is strong and different from that of bulk water or from that expected from gas-phase H-bond interaction energies (see Tables 2 and 4 and Figure 5). MD-averaged nucleobase−nucleobase interaction energies (Table 5) shows that the strongest H-bonds occur for the trimer containing the 1EcIt·C pair, followed by that containing the AO1·C pair. This finding was in fact expected from the gas-phase values in Figure 5 but it is quite remarkable to notice how that the DNA environment reduces the H-bonding preference of the 1EcIt·C pair with respect to the AO1·C one from nearly 30 kcal/mol in the gas phase to 15 kcal/mol in the DNA environment. The stacking energy clearly favors the trimers involving amino tautomers of isoG, particularly the AO1 one. Overall, the stability of the dimmers in terms of nucleobase−nucleobase interaction energy is 1EcIt·C ≫ AO1·C > AEc·C > IcO13·C > ItO13·C (see values in Table 5). This ordering qualitatively agrees with the expected role of DNA in stabilizing specific tautomers of isoG predicted by free energy calculations (see Table 4), but with some relevant quantitative differences. Thus, Table 4 suggests that the DNA stabilizes around 2 kcal/mol the 1EcIt form with respect to the AO1 tautomer, while Table 5 suggests that this difference is around 15 kcal/mol. Again, this demonstrates that not only direct contacts, but the whole DNA atmosphere, including remote effects, bulk solvent, and entropic terms, influence the preferential stabilization of tautomers of isoG in the DNA. These findings combined with those noted above for the isoG·T pairing demonstrate that neither isolated dimers in the gas phase nor pure polar solvents (like water) are good models to represent the effect of DNA on tautomeric equilibrium.

Combination of "solvation" free energies (Table 4) with intrinsic tautomeric preferences (Table 1) allowed us to determine the tautomerization free energies of isoG in the interior of a DNA duplex paired to cytosine. Results (Table 4) suggest that the AEc tautomer is the major species of isoG when paired with cytosine in a piece of DNA, due mostly to its greater intrinsic stability. Some population of the AO1 tautomer might be expected in certain sequences, or in specific environmental conditions, but the rest of tautomers (including the 1EcIt one) are not expected to play any major role in stabilizing isoG·C pairs in physiological DNA.

No high-resolution data exist for isoG·C dimers in DNA, and recognition models were built based mostly on chemical intuition. Thus, Switzer and co-workers[12a] reported melting experiments showing the stability of isoG·C pairs and suggested, based on simple modeling considerations, that the ItO13 tautomer should be the major species, with a minor role for 1EcIt. On the contrary, Kamiya et al.,[14f] after analysis of misinsertion experiments of isoG by two DNA polymerases, postulated two possible pairing modes involving wobble interactions between cytosine and the AO1 or AEc tautomers. Our

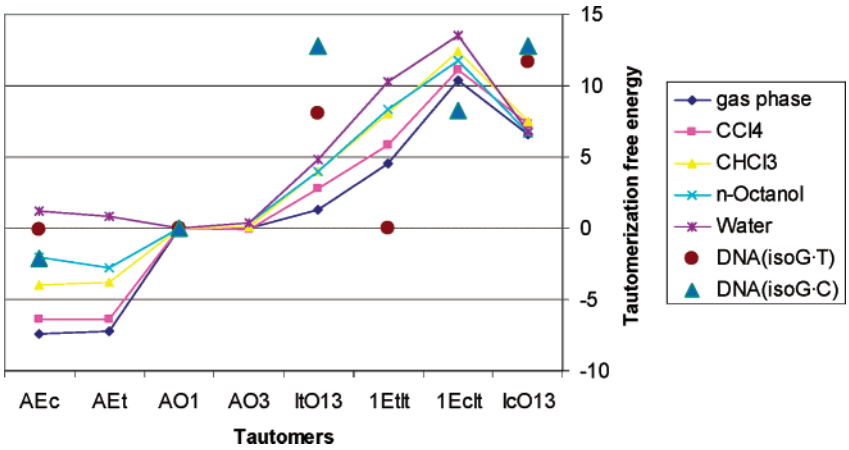

Figure 6. Graphical representation of the relative stability (tautomerization free energy in kcal/mol) of the 8 most important tautomers of isoguanine in the gas phase, different pure solvent, and in the DNA paired with thymine and cytosine.

calculations clearly support the hypothesis raised by Kamiya et al.,[14f] suggesting the Aec tautomer as the most stable species. We also support the hypothesis that only wobble pairings are involved in the isoG·C pairing in DNA. We should consider that wobble pairs might be stable (see Table 5 and Figure 5) in already formed DNA, but can represent a problem for the correct recognition of the isoG·C pair by DNA-polymerases. We believe that this can explain the problems of polymerases to incorporate isoG in front of C (or C in front of isoG) despite the stability of the resulting duplex.[12a,14a,f,41]

Conclusions

Isoguanine is an unique example of the importance of the environment in modulating the tautomeric properties of molecules and of the impact of tautomerism in the recognition properties of nucleobases. Our results clearly demonstrate that tautomers which are completely negligible in the gas phase can be the dominant species in polar solvents such as water and even be crucial in the stabilization of isoG·C and isoG·T pairs in duplex DNA (see Figure 6). Our calculations also suggest that wobble (related to the AO1 tautomer) and Watson−Crick-like (due to the AEc) tautomers exist when isoG is paired to T. When isoG is inserted in front of cytosine the Watson−Crick pair is disfavored and a wobble pairing involving mostly the AEc tautomer occurs. The different pattern of interactions might explain why polymerases incorporate isoG in front of T (or T in front of isoG) better than isoG in front of C (or C in front of isoG),[14a,14f,38] despite the fact that melting temperatures indicate that oligonucleotides containing isoG·T pairs are slightly less stable than those containing isoG·C pairs.[12a] Clearly, the tautomeric capabilities of isoG make this molecule to be a unique case to modulate recognition, stability and enzymatic susceptibility.

Acknowledgment. J.R.B. is a fellowship of the Catalan Government. This work was supported by the Spanish Ministry of Science and Technology (PB99-0046, SAF2002-04282 and GEN2001-4758) and the Centre de Supercomputació de Catalunya (CESCA; Mol. Recog. Project).

JA036806R

(41) Switzer, C. Y.; Moroney, S. E.; Benner, S. A. *Biochemistry* **1993**, *32*, 10489.