Spontaneous DNA Mutations Induced by Proton Transfer in the Guanine·Cytosine Base Pairs: An Energetic Perspective

Jan Florián*,†,‡,§ and Jerzy Leszczyński*,†

Contribution from the Department of Chemistry, Jackson State University, 1400 Lynch Street, Jackson, Mississippi 39217, and Institute of Physics, Charles University, Ke Karlovu 5, CZ-12116 Prague, Czech Republic

Received June 19, 1995. Revised Manuscript Received December 29, 1995[⊗]

Abstract: The energetic provisions for Löwdin's DNA mutational mechanism (Löwdin, P. O. *Rev. Mod. Phys.* **1963**, *35*, 724) of the formation of substitution DNA mutations were investigated for the guanine·cytosine Watson−Crick base pair. The structures studied involve the canonical base pair (GC1), rare base-pair tautomers that are formed from GC1 by the antiparallel simultaneous transfer of two protons in hydrogen bonds, and ion-pair structures that are formed by the transfer of a single proton. The geometries of these complexes were optimized by *ab initio* Hartree−Fock (HF) calculations using the 6-31G* basis set. At the same level, harmonic vibrational frequencies were determined. Nonplanar geometries featuring considerable propeller-twist angles and a pyramidal guanine amino group were found for base pairs involving the guanine anion and 6-hydroxyguanine. The relative stabilities and dissociation energies of the base pairs were determined at the higher MP2/6-31G**//HF/6-31G* level of theory. These methods were also used to locate transition states on the potential energy surface of the guanine·cytosine base pair. Starting from the geometries of two different transition states lying close to the ion-pair G^-C^+ minimum, the intrinsic reaction coordinate for the proton transfer from the canonical to the 6-hydroxyguanine·4-iminocytosine tautomer (GC2) was evaluated. We concluded that, in contrast to the adenine·thymine base pair (for which Löwdin's mutational mechanism is not supported by the present theoretical data), the GC1 → GC2 tautomeric transition is likely to occur in 1 in 10^6-10^9 guanine·cytosine base pairs. This frequency is significant from the point of view of the fidelity of DNA replication.

Introduction

Two basic molecular mechanisms are recognized as being responsible for the formation of substitution mutations at the DNA-synthesis level. First, base mispairs may occur between the canonical (amino, keto) and the minor tautomeric (imino, enol) forms of nucleobases during catalytic incorporation of the new base on the growing DNA strand.[1,2] For example, guanine in its enol form (G*) could bind with thymine (T), or iminocytosine (C*) with adenine (A). In this case, the frequency of the mutation event is governed by the concentration of free nucleotide triphosphates in their minor tautomeric forms in solution.[2] To date, however, no experimental evidence has been given in support of this mechanism. Alternatively, ionized[3] and/or wobble base pairing[4] have been suggested to play a major role in mispair formation. A number of NMR and X-ray crystallographic studies have shown the presence of ionized and wobble mispairs in duplex DNA (see, for example, ref 5 and references therein). However, most of these structures were prepared by cocrystallizing short complementary oligonucleotides containing a single mismatched base pair. These conditions are far from those required for DNA replication.[6] In addition, conformational DNA dislocations brought about by wobble base pairing could be more easily located and excised by the polymerase proofreading mechanisms than could mispairs involving rare tautomers.[7]

The present paper deals with energetic provisions of the *modified tautomeric mechanism* for the formation of spontaneous substitution mutations, which does not require the presence of the free rare tautomers in solution. The basis of this hypothesis stems from the possibility for rare tautomers to be formed in the *template* via the concerted transfer of two protons in the interbase hydrogen bonds in DNA.[8−11] This mechanism assumes that an evolutionally significant number of imino/enol tautomers will be formed in this way, and that these tautomers will remain stable during DNA unwinding and strand separation, which are the prerequisite steps for the synthesis of the new DNA strand by polymerase. Quantitatively, in order for this mechanism to be capable of introducing appreciable genetic instability, an energy difference smaller than ∼13 kcal/mol between the canonical and rare base-pair tautomers is required. In addition, the barrier height for the proton transfer reaction, determined by the energy of the transition state on the potential energy surface, should be in the appropriate range. Using classical transition-state theory, along with the simple reasoning that (i) the lifetime of the canonical base pair should be shorter than the reproduction period of the given species ($\sim10^8$ s) and (ii) the lifetimes of the rare tautomeric forms should exceed the characteristic time for base-pair opening ($\sim10^{-10}$ s), one can estimate that the barriers for the forward and reverse proton-

* To whom correspondence should be addressed.
† Jackson State University.
‡ Charles University.
§ Present address: Department of Chemistry, University of Southern California, Los Angeles, CA 90089-1062.

⊗ Abstract published in *Advance ACS Abstracts,* March 1, 1996.

(1) Watson, J. D.; Crick, F. H. C. *Nature* **1953**, *171*, 964.
(2) Topal, M. D.; Fresco, J. R. *Nature* **1976**, *263*, 285.
(3) Sowers, L. C.; Shaw, B. R.; Veigl, M. L.; Sedwick, W. D. *Mutation Res.* **1987**, *177*, 201.
(4) Sinha, N. K.; Haimes, M. D. *J. Biol. Chem.* **1981**, *256*, 10671.
(5) Echols, H.; Goodman, M. F. *Annu. Rev. Biochem.* **1991**, *60*, 477.
(6) Kornberg, A.; Baker, T. A. *DNA Replication*; W. H. Freeman: New York, 1992.
(7) Strazewski, P.; Tamm, C. *Angew. Chem., Int. Ed. Engl.* **1990**, *29*, 36.
(8) Löwdin, P. O. *Rev. Mod. Phys.* **1963**, *35*, 724.
(9) Löwdin, P. O. In *Electronic Aspects of Biochemistry*; Pullman, B., Ed.; Academic Press: New York, 1964; pp 167.
(10) Löwdin, P. O. *Mutation Res.* **1965**, *2*, 218.
(11) Löwdin, P. O. *Adv. Quantum Chem.* **1965**, *2*, 213.

0002-7863/96/1518-3010$12.00/0

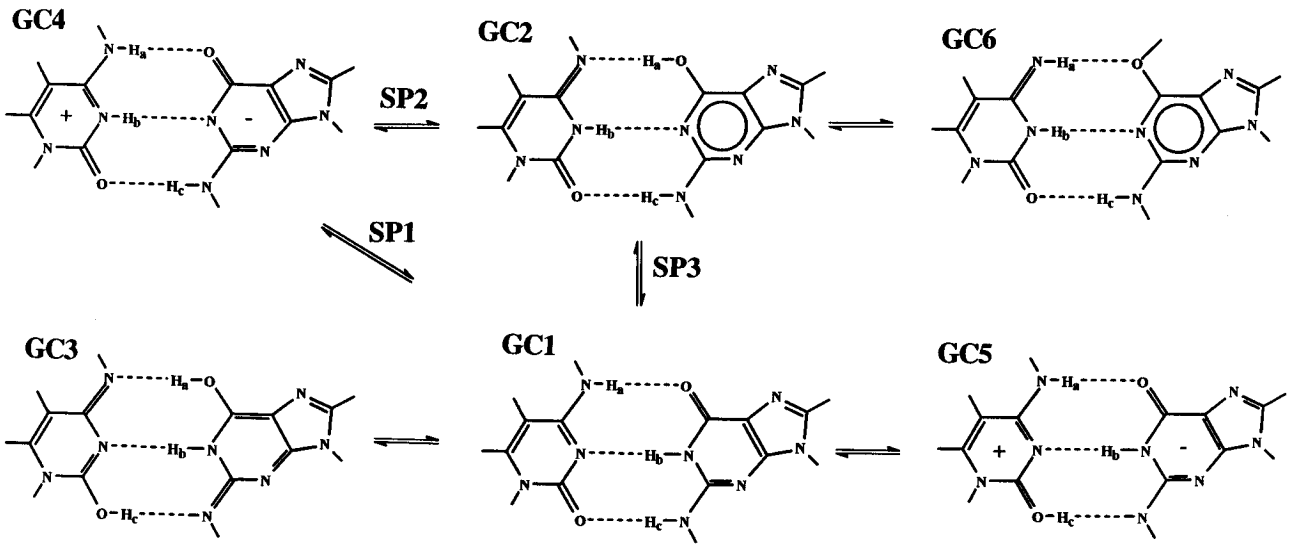

Figure 1. The studied forms of the guanine·cytosine base pair.

transfer reactions should be smaller than ~28 kcal/mol and larger than ~3 kcal/mol, respectively.

Because the mutationally significant concentrations of rare tautomeric base pairs in DNA fall below the detection limits of available experimental techniques, accurate *ab initio* quantum-mechanical predictions of tautomeric equilibria are needed. Their use is facilitated by the fact that proton transfer within DNA base pairs occurs without the direct influence of water molecules. Although *ab initio* methods represent a well-established research tool for many fields of chemistry, their application to biologically relevant problems, the size of which requires the use of lower-end approximations, is still in its infancy. Fortunately, the errors that arise from such approximations, predominantly those attributed to the use of limited basis sets or to the lack of electron correlation, tend to be preserved for chemically related systems and properties. Thus, qualitatively correct results can be obtained for large systems, assuming these results are supported by some sort of method calibration. More specifically, for the assessment of quantum mechanical calculations of nucleic acid base pairs, for which geometry optimizations at a correlated level are still very computationally demanding, the results of previous *ab initio* studies on the energetics of the keto−enol and amino−imino tautomeric transitions in formamidine,[12,13] formamide/formamidic acid,[14,15] the cyclic formamide dimer,[16,17] uracil,[18−20] cytosine,[20−25] 1-methylcytosine,[26] adenine,[27] and guanine,[27−29] obtained using different basis sets and methods ranging from Hartree−Fock (HF) to post-HF (many-body perturbation theory, coupled-cluster, local correlation) and density functional[17,24] methods, can be utilized. It follows from these studies that full geometry optimizations and the use of nonempirical computational methods are necessary for obtaining meaningful relative stabilities of base-pair tautomers. In addition, comparison with experimental relative stabilities of tautomers of cytosine,[21] 1-methylcytosine,[26] guanine,[30] and 9-methylguanine[30,31] is available. For these systems, HF calculations with polarized basis sets provide results accurate to within 1 kcal/mol. The inclusion of electron correlation, which is accomplished in this work using second-order perturbation theory (MP2), generally improves the accuracy of the calculated energies. This contribution is especially important for cyclic hydrogen-bonded systems and for energies of the transition states of proton-transfer reactions.[12,13,16,17]

Due to the large size of the systems in question, previous quantum mechanical studies on this topic were limited to the use of rigid monomer approximations,[32−35] or they involved HF calculations with the insufficiently extended MINI-1 basis set.[36−38] The latter studies provided a relatively large equilibrium constant (10^{-7}) for the formation of the rare tautomer pair, A*T*, from the canonical AT base pair by a double-proton-transfer reaction. However, the energy of the transition state for this reaction was predicted to be only 0.2 kcal/mol larger than the energy of the corresponding A*T* minimum structure.[37] Moreover, the separation of nucleobases from the A*T* complex toward the isolated A* and T* bases was penalized by a larger dissociation energy than that for the unpairing of the canonical AT base pair.[36] At the same HF/MINI-1 level of theory, a more probable path for the formation of rare tautomers in DNA was found to involve the 6-hydroxyguanine·4-iminocytosine base pair.[38] These results provided us with an incentive for a more

(12) Bell, R. L.; Truong, T. N. *J. Chem. Phys.* **1994**, *101*, 10442.
(13) Zhang, Q.; Bell, R.; Truong, T. N. *J. Phys. Chem.* **1995**, *99*, 592.
(14) Kwiatkowski, J. S.; Bartlett, R. J.; Person, W. B. *J. Am. Chem. Soc.* **1988**, *110*, 2353.
(15) Kwiatkowski, J. S.; Leszczynski, J. *J. Mol. Struct.* **1992**, *270*, 67.
(16) Hrouda, V.; Florián, J.; Polásek, M.; Hobza, P. *J. Phys. Chem.* **1994**, *98*, 4742.
(17) Florián, J.; Johnson, B. G. *J. Phys. Chem.* **1995**, *99*, 5899.
(18) Leszczynski, J. *J. Phys. Chem.* **1992**, *96*, 1649.
(19) Broughton, J. W.; Pulay, P. *Int. J. Quantum Chem.* **1993**, *47*, 49.
(20) Estrin, D. A.; Paglieri, L.; Corongiu, G. *J. Phys. Chem.* **1994**, *98*, 5653.
(21) Szczesniak, M.; Szczepaniak, K.; Kwiatkowski, J. S.; KuBulat, K.; Person, W. B. *J. Am. Chem. Soc.* **1988**, *110*, 8319.
(22) Gould, I. R.; Hillier, I. H. *Chem. Phys. Lett.* **1989**, *161*, 185.
(23) Les, A.; Adamowicz, L.; Bartlett, R. J. *J. Phys. Chem.* **1989**, *93*, 4001.
(24) Hall, R. J.; Burton, N. A.; Hillier, I. H.; Young, P. E. *Chem. Phys. Lett.* **1994**, *220*, 129.
(25) Sobolewski, A. L.; Adamowicz, L. *J. Chem. Phys.* **1995**, *102*, 5708.
(26) Szczesniak, M.; Leszczynski, J.; Person, W. B. *J. Am. Chem. Soc.* **1992**, *114*, 2731.
(27) Kwiatkowski, J. S.; Leszczynski, J. *J. Mol. Struct.* **1990**, *208*, 35.
(28) Leszczynski, J. *Chem. Phys. Lett.* **1990**, *174*, 347.
(29) Leszczynski, J. *THEOCHEM* **1994**, *311*, 37.
(30) Szczepaniak, K.; Szczesniak, M. *J. Mol. Struct.* **1987**, *156*, 29.
(31) Szczepaniak, K.; Szczesniak, M.; Szajda, W.; Person, W. B.; Leszczynski, J. *Can. J. Chem.* **1991**, *69*, 1705.
(32) Scheiner, S.; Kern, C. W. *J. Am. Chem. Soc.* **1979**, *101*, 4081.
(33) Clementi, E.; Mehl, J.; Niessen, W. v. *J. Chem. Phys.* **1971**, *54*, 508.
(34) Clementi, E. *Proc. Natl. Acad. Sci. U.S.A.* **1972**, *69*, 2942.
(35) Kong, Y. S.; Jhon, M. S.; Löwdin, P.-O. *Int. J. Quantum Chem., Quantum Biol. Symp.* **1987**, *14*, 189.
(36) Hrouda, V.; Florián, J.; Hobza, P. *J. Phys. Chem.* **1993**, *97*, 1542.
(37) Florián, J.; Hrouda, V.; Hobza, P. *J. Am. Chem. Soc.* **1994**, *116*, 1457.
(38) Florián, J.; Leszczynski, J. *Mol. Phys.* **1995**, *84*, 469.

detailed and accurate study of the role of single and double proton transfer in the initiation of spontaneous point mutations.

Guanine·cytosine base pairs featuring the proton donor/acceptor patterns presented in Figure 1 were optimized in this study with the goal to find energies, geometries, and vibrational frequencies of the corresponding stationary states on the potential energy surface (PES). Except for the GC6 structure, only the tautomers arising from proton transfers in hydrogen bonds were considered. Because there are three parallel hydrogen bonds in the guanine·cytosine base pair, there are two ways in which protons can rearrange by a double proton transfer while keeping each monomer in its neutral form. The minor tautomers of the base pairs formed in this way are denoted as GC2 and GC3 (Figure 1). In addition, the zwitterionic tautomers GC4 and GC5 were considered in this study. These structures can be formed from the canonical base pair (GC1) by the transfer of a single proton from guanine to cytosine. The last tautomer, GC6, is formed from the GC2 base pair by coupled rearrangement of hydrogens in the *N*4-imino and *O*6-hydroxy groups. The inclusion of the GC6 tautomer in our calculations was promoted by the recent claim that quantum mechanical "flip-flop" of hydrogens between the GC2 and GC6 structures is capable of altering the genetic specificity of the DNA template via enhanced stabilization of the GC2 base pair.[39]

Since the proton-transfer barriers are important for unraveling proton-transfer mechanisms and kinetics, we also calculated the structures of saddle points (SP1, SP2, SP3) on the reaction path interconnecting the GC1, GC2, and GC4 minima. Starting from the transition states (as discussed below), we evaluated the intrinsic reaction coordinate (IRC)[40−42] for the stepwise proton transfer between the GC1 and GC2 tautomers. The resulting potential energy profile and internal coordinate variations along the reaction coordinate were used to visualize the actual course of the proton transfer reaction. Finally, we examined the nonplanar character of the structure of the GC2 and GC4 base pairs induced by the guanine pyramidal amino group, and the relationships among the geometry of the guanine amino group, hydrogen bonding, and protonation.

Computational Methods

The search for stationary states (minima and transition states) on the sections of the potential energy surface (PES) corresponding to the single and double proton transfers in the GC base pair was carried out at the HF/6-31G* level, using the Berny gradient full optimization method implemented in the Gaussian92 program.[43] The single-point calculations were carried out at the HF/6-31G* geometry, using second-order many-body perturbation theory and a basis set augmented by d-polarization functions on heavy elements and p-polarization functions on hydrogens (MP2/6-31G**). This method was used to evaluate the electron correlation contributions to the relative tautomeric stabilities and interaction enthalpies.

The first-order saddle point (transition state) of a chemical reaction determines the lowest barrier separating reactants and products on the PES. Calculation of the structures of transition states was done in three steps. First, the positions of transferring protons were fixed near the middle of the hydrogen bonds while the remaining degrees of freedom were relaxed by geometry optimization. Next, the Hessian matrix was calculated. Information from the Hessian matrix and the normal vector corresponding to the largest imaginary frequency were used in the subsequent full optimization to the transition state, as implemented in the Gaussian92 program. Due to the flat PES, this procedure had to be repeated several times, until convergence in the last step was reached. The correspondence of the optimized structures to the minimum or to a saddle point was verified by the calculation of harmonic vibrational frequencies, from which the zero-point vibrational energy (ZPE) and the thermal contributions to the enthalpy were subsequently determined. During the search for saddle points, the C_s symmetry constraint was assumed. The calculated small magnitudes of imaginary frequencies corresponding to the out-of-plane vibrations in saddle-point structures indicate that the assumption of planarity is a reasonable approximation that negligibly influences the calculated energy barriers.

The interaction energies were corrected for the basis set superposition error (BSSE). The standard Boys−Bernardi counterpoise (CP) correction scheme[44,45] was slightly modified to take into account the geometry reorganization when going from the isolated subsystems to the complex; namely, the CP correction for each monomer was determined as the difference between the energy of the monomer in the complex geometry with the basis set of the whole complex and that of the same monomer without ghost orbitals.[36,46]

To estimate the effects of the polar medium upon the relative stabilities of tautomers, we carried out self-consistent reaction field (SCRF) calculations. The Onsager reaction field model,[47] as implemented in the Gaussian92 program,[43] was used for these calculations. In this model, the solvent is viewed as a continuous dielectric medium of uniform dielectric constant ϵ_r. The solute occupies a spherical cavity within the solvent. We have used the relative permittivity $\epsilon_r = 40$ and cavity radii of 5.0, 4.1, and 3.9 Å in the SCRF calculations of GC, guanine, and cytosine tautomers, respectively.

The intrinsic reaction coordinates (IRC) tracing the steepest-descent path from the transition states toward the reactants and products were evaluated for the GC1 ↔ GC4 and GC4 ↔ GC2 single-proton-transfer pathways. The planar (C_s) symmetry was assumed for the GC base pair during the IRC calculations. In determining IRC, the first step was taken from the transition state along the normal mode corresponding to the negative eigenvalue of the Hessian (second energy derivative) matrix. The IRC was computed in mass-weighted internal coordinates by the method of Gonzales and Schlegel,[42] using a step size of 0.1 amu$^{1/2}$ bohr, which is the default step size in the Gaussian92 program.[43] From SP2, the GC4 minimum was reached in 10 steps. In the SP2 → GC2 direction, 34 steps along the IRC path were computed, and the remaining part of the IRC was extrapolated. From SP1, 24 steps in both directions were taken. The rest of the IRC leading to the GC1 and GC4 minima was extrapolated, in part using the calculated energy profile for the SP2 → GC2 path. Such an extrapolation represents the standard procedure in studies of this type.[48] Extrapolation is necessitated by convergence problems near minima on the flat PES, and also by the extreme computational demands connected with the restricted geometry optimizations carried out in each IRC step.

Results and Discussion

In this section, we proceed from the more technical paragraphs that cover the method dependence of the calculated energies, over a description of the calculated proton-transfer reaction coordinate and important features of the structure of the base pairs, to the more biologically-oriented discussion of the accuracy of our results and their implications for mutations that occur in DNA.

Relative Energies. The total and relative energies of the stationary points on the PES of guanine·cytosine that account for different proton arrangements are presented in Table 1 and Figure 2. Apparently, the most stable conformer corresponds to the canonical keto/amino form of the guanine·cytosine base

(39) Cooper, W. G. *Biochem. Genet.* **1994**, *32*, 383.
(40) Fukui, K. *J. Phys. Chem.* **1970**, *74*, 4161.
(41) Fukui, K. *Acc. Chem. Res.* **1981**, *14*, 363.
(42) Gonzales, C. *J. Phys. Chem.* **1990**, *94*, 5523.
(43) Frisch, M. J.; Trucks, G. W.; Head-Gordon, M.; Gill, P. M. W.; M. W. Wong; Foresman, J. B.; Johnson, B. G.; Schlegel, H. B.; Robb, M. A.; Replogle, E. S.; Gomperts, R.; Andres, J. L.; Raghavachari, K.; Binkley, J. S.; Gonzalez, C.; Martin, R. L.; Fox, D. J.; Defrees, D. J.; Baker, J.; Stesart, J. J. P.; Pople, J. A. Gaussian, Inc., Pittsburgh, PA, 1992.
(44) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553.
(45) Meumier, A.; Levy, B.; Berthier, G. *Theor. Chim. Acta* **1973**, *29*, 49.
(46) Mayer, I.; Surjan, P. R. *Chem. Phys. Lett.* **1992**, *191*, 497.
(47) Onsager, L. *J. Am. Chem. Soc.* **1936**, *58*, 1486.
(48) Isaacson, A. D.; Wang, L.; Scheiner, S. *J. Phys. Chem.* **1993**, *97*, 1765.

Table 1. Comparison of the Energies of the Stationary Points on the Potential Energy Surface of the GC Base Pair

structure[a]	method[b]	ΔE[c]	E_0^{vib}[d]	$E_{298}^{vib} - TS$[e]
GC1	MINI-1	0	139.7	
	6-31G*	0	148.5	120.0
	6-31G**	0		
	MP2	0		
GC2 (C_1)	6-31G*	11.11	148.6	120.4
	6-31G**	9.6		
	MP2	9.6		
GC2	MINI-1	0.5	139.8	
	6-31G*	11.14	(1)	
	6-31G**	9.6		
	MP2	9.0		
GC3	6-31G*	41.2	148.1	120.9
	6-31G**	37.7		
	MP2	32.5		
GC4 (C_1)	6-31G*	24.7	147.4	119.7
	6-31G**	23.6		
	MP2	17.9		
GC4	6-31G*	25.2	(1)	
	6-31G**	23.8		
	MP2	18.6		
GC5	6-31G*	not a stationary point		
GC6	6-31G*	17.6	148.2	118.4
	6-31G**	16.1		
	MP2	16.6		
SP1	6-31G*	27.4	(2)	
	6-31G**	25.3		
	MP2	17.0		
SP2	6-31G*	25.3	(2)	
	6-31G**	23.4		
	MP2	17.4		
SP3	MINI-1	3.8	136.48	
	6-31G*	31.0	(3)	
	6-31G**	27.0		
	MP2	14.6		

[a] See Figure 1. If not indicated otherwise (by C_1 in parentheses), energies for planar structures are given. [b] MINI-1 and 6-31G* denote HF/MINI-1 (ref 32) and HF/6-31G* results, respectively. 6-31G** and MP2 denote HF/6-31G**//HF/6-31G* and MP2/6-31G**//HF/6-31G* results (energy//geometry), respectively. [c] Relative energy (kcal/mol). The total energy of the reference structure (i.e., the sum of the electronic energy and nuclei repulsion of the GC1 base pair) amounts to −932.050 755 and −934.887 319 hartrees for HF/6-31G* and MP2/6-31G** computational levels, respectively. E_0^{vib} and $E_{298}^{vib} - TS$ contributions are not included in ΔE. The relative gas phase free energy at 298 K can be obtained as $\Delta G_{298} = \Delta E + \Delta(E_{298}^{vib} - TS)$. [d] Zero-point vibrational energy (ZPE) (kcal/mol). For structures that are not minima on PES, the calculated number of imaginary frequencies is given in parentheses. The calculated magnitudes of imaginary frequencies amount to 32 cm^{-1} (GC2 (C_s)), 210 cm^{-1} (GC4 (C_s)), 1440 and 88 cm^{-1} (SP1), 464 and 192 cm^{-1} (SP2), and 1971, 1091, and 35 cm^{-1} (SP3). [e] Sum of vibrational (E_{298}^{vib}) and entropic (TS) contributions to the gas phase free energy at a temperature of 298 K. These contributions were evaluated using ideal gas, rigid rotor, and harmonic approximations from calculated vibrational frequencies, moments of inertia, and molecular masses.

pair (Figure 1), henceforth denoted as GC1. This complex is planar, although the calculated frequencies of intermolecular vibrational modes that are as low as 20 cm^{-1} indicate that the planar structure can be easily distorted by propeller twist, buckle, and stagger displacements. A detailed account of the vibrational spectra of the GC1 base pair and comparison of its geometry with the geometries of individual guanine and cytosine molecules has been given elsewhere.[38,49,50] The HF/6-31G* geometry and interaction energy of the 9-methylguanine·1-methylcytosine complex were published by Gould and Kollman.[51]

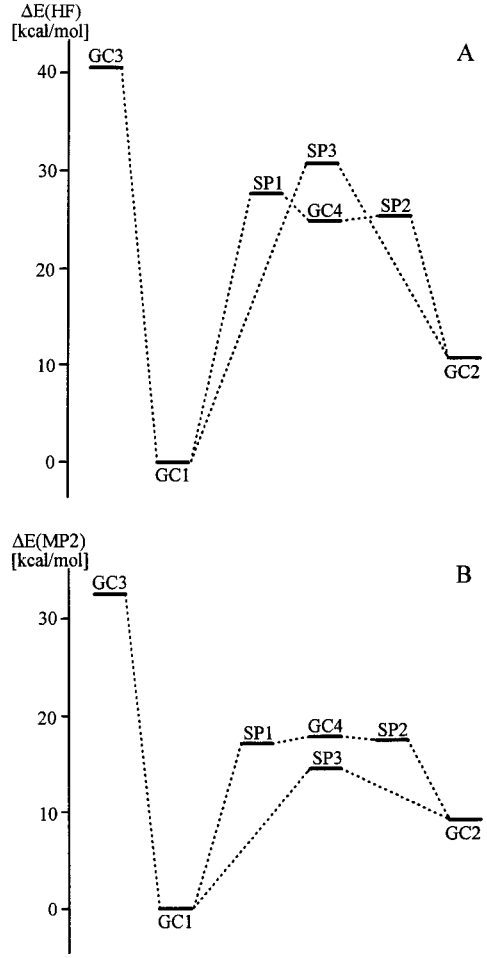

Figure 2. (a) Diagram of relative HF/6-31G* energies of stationary states on the PES of the guanine·cytosine base pair. (b) Diagram of relative MP2/6-31G**//HF/6-31G* energies of stationary states on the PES of the guanine·cytosine base pair.

The second most stable conformer corresponds to the GC2 base pair (Figure 1). This base pair is slightly nonplanar. Its HF/6-31G* energy is 11.1 kcal/mol higher than that of the GC1 base pair. Enlargement of the basis set and inclusion of electron correlation by the MP2 method decrease this energy difference to 9.0 kcal/mol. The effect of vibrational energy and entropy differences on relative tautomer stabilities is negligible (Table 1; see also ref 52). Also, only a small contribution to the relative stabilities of GC1 and GC2 is associated with solvent effects. Use of the polarizable continuum model (SCRF) results in an additional 0.6 kcal/mol stabilization of GC1 relative to GC2 at the HF/6-31G* level (not shown in the table). Because of large oversimplifications inherent to the SCRF model, this value should be considered with caution. Moreover, aside from this polarization stabilization, there could be a wide range of more specific intermolecular interactions leading to the additional stabilization of either the canonical or the GC2 base pair. For example, *N*4-imino-1-methylcytosine was found to be stabilized by coordination of N4 to a platinum(IV) ion,[53] and *cis*-platinum binding to the N7 site of 9-ethylguanine was shown to ease ionization of its N1 proton[54] (for atom numbering see Figure 3). In addition, small perturbations of the GC1−GC2 energy difference can be caused by vertical electrostatic interactions among stacked base pairs in DNA. We expect that these effects

(49) Florián, J.; Leszczynski, J. *Int. J. Quantum Chem., Quantum Biol. Symp.* **1995**, *22*, 207.
(50) Florián, J.; Leszczynski, J.; Johnson, B. G. *J. Mol. Struct.* **1995**, *349*, 421.
(51) Gould, I. R.; Kollman, P. A. *J. Am. Chem. Soc.* **1994**, *116*, 2493.
(52) Kwiatkowski, J. S.; Leszczynski, J. *Chem. Phys. Lett.* **1993**, *204*, 430.
(53) Lippert, B.; Schollhorn, H.; Thewaldt, U. *J. Am. Chem. Soc.* **1986**, *108*, 6616.
(54) Lippert, B. *J. Am. Chem. Soc.* **1981**, *103*, 5691.

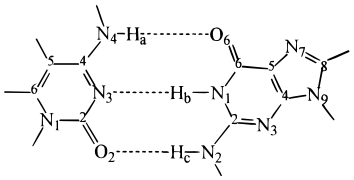

Figure 3. Structure and numbering of the canonical guanine·cytosine base pair (GC1). This numbering scheme is also used for other tautomeric forms.

will result in a sequence dependence of the propensity toward formation of spontaneous point mutations. A rigorous treatment of such stacking-dependent effects is, however, beyond the limits of current computer technology.

The antiparallel double proton ransfer of H_a and H_c protons in the outer hydrogen bonds mutates the canonical base pair into the imino−enol−imino−enol tautomer GC3. At all computational levels, the energy of this tautomer is substantially higher than the GC1 and GC2 energies, so it need not be further considered. This is also true for the GC6 structure, evolved from GC2 by a complex and energetically improbable two-proton transition. Due to the large energy gap between the GC2 and GC6 structures, location of the transition state between these structures was not attempted by us.

The ion-pair GC4 tautomer turned out to be more stable than the GC3 form, even in the gas phase. Because of its zwitterionic character, the GC4 structure should be stabilized by solvent effects. Another means of GC4 stabilization could involve guanine protonation or methylation.[3,55]

As for the other ion-pair structure, GC5, it does not represent a stationary point on the PES. This was somewhat unexpected because the gas-phase proton affinities of O2 and N3 sites of cytosine differ by less than 1 kcal mol,[56] and because O2-protonated cytidine monophosphate has been observed in acidic aqueous solution.[57] At first sight, the reason for the instability of GC5 and GC3 seems to reside in the large energy requirement for deprotonation of the guanine amino group. To determine the validity of this presumption, we calculated at the HF/6-31G* level the energies required to remove H_b and H_c protons from isolated guanine. However, the resulting deprotonation energies (357.8 and 357.5 kcal/mol, respectively) were nearly the same. Thus, the instability of GC5 and GC3 compared to the GC2 and GC4 base pairs most probably originates from repulsive dipole−dipole interactions between the monomers forming the GC3 and GC5 base pairs.

Transition States. In addition to the mentioned tautomeric base pairs representing minima on the PES, we located three saddle-point structures (Figure 1, Table 1). If the small out-of-plane imaginary modes originating from the C_s symmetry constraint are disregarded, these saddle points can be classified as the first- (SP1, SP2) and second-order (SP3) saddle points. (Note that the terms "transition state" and "first-order saddle point" have the same meaning.) The energies of the SP1 and SP2 transition states determine the barriers separating the GC1 and GC4 minima and GC4 and GC2 minima on the PES, respectively. The corresponding chemical reactions can be characterized as single proton transfers. At the HF/6-31G* level, the SP1 and SP2 transition states lie only a few tens of kilocalories per mole above the GC4 tautomer (Figure 2a). The single-point MP2/6-31G* calculations further decrease the relative energies of the SP1 and SP2 structures so that they become lower than the GC4 energy (Figure 2b). This finding might indicate that, at the correlated level, the GC4 structure is not a minimum on the PES, but rather a saddle point.

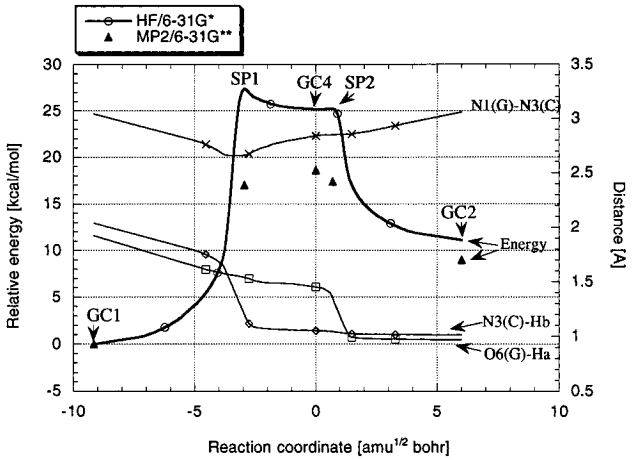

Figure 4. Energy profile and hydrogen bond lengths along the HF/6-31G* minimum-energy path (MEP, IRC) connecting GC1, GC4, and GC2 structures. The MP2/6-31G** energies calculated at HF geometries of stationary states are also indicated.

The HF/6-31G* barrier for the simultaneous transfer of H_a and H_b protons between the GC1 and GC2 structures, i.e., for the reaction GC1 → SP3 → GC2, is rather high (31 kcal/mol, Table 1). The top of this barrier, denoted as SP3, is a second-order saddle point. The calculated imaginary frequencies (1971 and 1091 cm^{-1}) correspond to the pathways leading from SP3 to neutral and zwitterionic minima, respectively. The SP3 barrier was calculated to be 3.6 kcal/mol higher than the highest barrier along the stepwise GC1 ↔ SP1 ↔ GC4 ↔ SP2 ↔ GC2 pathway. However, at the MP2 level of theory, the relative energy of the SP3 structure is significantly decreased (to 14.6 kcal/mol), so the simultaneous double-proton-transfer mechanism for the GC1 → GC2 transition becomes more probable than the stepwise mechanism (Figure 2).

Proton-Transfer Reaction Coordinate. In Figure 4, we have drawn the HF/6-31G* energy profile, along with the variations both in length of the middle hydrogen bond and in the positions of transferring protons along the reaction coordinate corresponding to the GC1 ↔ SP1 ↔ GC4 ↔ SP2 ↔ GC2 stepwise proton transfer. These data were obtained in two separate intrinsic reaction coordinate (IRC) calculations, initiated from the SP1 and SP2 transition states. The origin of the reaction coordinate was placed at the GC4 minimum. Due to the extrapolation procedure undertaken in the vicinity of the GC1 and GC2 minima, the positions of these minima on the reaction coordinate should be considered approximate.

Starting from the canonical base pair, the proton-transfer reaction is commenced by the guanine and cytosine coming closer to each other. During this stage, the N4−H_a and N1−H_b bond lengths (Figure 3) remain practically constant. When the N1−N3 distance approaches 2.66 Å, the value it attains in the SP1 transition state, the H_b proton quickly transfers from guanine to cytosine. In this way, GC4 is formed. Due to low inherent energy barriers, GC4 represents an unstable reaction intermediate that is quickly "decomposed" into the GC2 tautomer by the transfer of the proton H_a toward the guanine oxygen. In the final stage, N3−H_b and O6−H_a distances remain constant and the hydrogen bonds lengthen to their GC2 values.

Geometry of the Complexes. The geometries of the molecular fragments involved in the hydrogen bonding for all stationary states treated in this study are compared in Table 2. Additional information concerning intermolecular parameters

(55) Ford, G. P.; Wang, B. *Int. J. Quantum Chem.* **1992**, *44*, 605.
(56) Florián, J.; Baumruk, V.; Leszczynski, J. *J. Phys. Chem.* **1996**, in press.
(57) Purrello, R.; Molina, M.; Wang, Y.; Smulevich, G.; Fossella, J.; Fresco, J. R.; Spiro, T. G. *J. Am. Chem. Soc.* **1993**, *115*, 760.

Table 2. Calculated Geometry of Hydrogen Bonds (Å, deg)

bond/angle[a]	GC1	SP3(Cs)	GC2(Cs)	GC2(C1)	SP1(Cs)	SP2(Cs)	GC4(Cs)	GC4(C1)	GC3
N4−O6	2.929	2.442	2.871	2.872	2.583	2.486	2.534	2.564	2.814
N4−H_a	1.009	1.282	1.911	1.912	1.045	1.132	1.086	1.071	1.835
O6−H_a	1.922	1.160	0.967	0.967	1.539	1.355	1.449	1.495	0.980
N4−H_a−O6	177.0	177.5	171.4	171.5	177.0	175.4	177.2	176.4	176.7
C4−N4−H_a	120.4	123.5	125.6	125.5	121.3	123.7	122.7	122.0	125.9
N3−N1	3.043	2.608	3.057	3.056	2.661	2.848	2.840	2.820	2.862
N3−H_b	2.036	1.283	1.012	1.012	1.204	1.042	1.049	1.051	1.853
H1−H_b	1.008	1.326	2.046	2.046	1.457	1.809	1.793	1.771	1.011
N3−H_b−N1	176.4	176.8	176.1	176.0	178.3	174.6	176.0	175.4	175.4
C2−N1−H_b	118.8	121.3	119.4	119.3	124.2	123.9	124.6	116.3	114.9
O2−N2	3.018	2.867	3.123	3.134	2.897	3.160	3.125	3.168	2.669
O2−H_c	2.016	1.872	2.127	2.138	1.911	2.171	2.137	2.186	1.013
N2−H_c	1.002	0.999	0.997	0.997	0.994	0.993	0.993	0.996	1.656
O2−H_c−N2	178.2	173.9	179.8	178.1	171.4	173.2	172.7	168.5	177.3
C2−N2−H_c	123.1	123.2	122.4	121.3	122.8	122.0	122.1	117.6	131.7

[a] For atom numbering see Figure 3. Notations atom 1−atom 2, atom 1−atom 2−atom 3, and atom 1−atom 2−atom 3−atom 4 denote bond length and bond angle, respectively. The propeller twist (buckle) amounts to 4.9° (3.2°) and 11.5° (3.3°) in the GC2 (C_1) and GC4 (C_1) structures, respectively. The propeller-twist angle was defined as the angle between the line intersecting C2 and C4 atoms of cytosine and the plane defined by the N1, C6, and N9 atoms of guanine. The buckle angle was defined as the angle between the line intersecting C6 and N3 atoms of cytosine and the plane defined by the N1, C6, and N9 atoms of guanine. The N1−C2−N2−O2 torsional angle amounts to 10.8° and 17.0° in GC2 (C_1) and GC4 (C_1), respectively.

Table 3. Comparison of Characteristic Parameters[a] of the Guanine Amino Group Present in Different Structures

	Gua	GC1	Gua*	GC2 (C_1)	Gua$^{(-)}$ [b]	GC4 (C_1)	Gua(N7$^+$)[c]
ΔE (kcal/mol)	0.5		0.2	0.03	1.9	0.5	
C2−N2 (Å)	1.363	1.336	1.357	1.347	1.395	1.367	1.327
C6−N1−C2−N2 (deg)	177.6	180	178.0	178.1	178.2	177.3	180
N1−C2−N2−H_c (deg)	31.4	0	17.6	12.3	23.6	24.8	0
N1−C2−N2−H (deg)	169.9	180	164.5	171.3	152.5	164.9	180

[a] HF/6-31G* energy difference between the planar (C_s) and nonplanar (C_1) stationary points (ΔE), C−NH_2 bond length of guanine, and amino group torsional angles. Geometry parameters are given only for the structures corresponding to the minima on the PES. For atom numbering see Figure 3. [b] Guanine deprotonated at the N1 nitrogen. [c] Guanine protonated at the N7 nitrogen.

for the nonplanar GC2 and GC4 tautomers can be found in Table 3. For the sake of brevity, we have omitted the presentation of intramolecular geometric parameters. For GC1, these data can be found in our previous papers.[38,49] For other complexes, the optimized geometries can be obtained from us upon request.

One can observe from Table 2 that all hydrogen bonds remain nearly linear regardless of the actual proton positions. In the GC1 base pair, the upper N4···O6 hydrogen bond is notably shorter than the N2···O2 one. This feature is further amplified in the GC2 and GC4 tautomers. In addition, it remains in effect during the whole stepwise proton-transfer reaction, and it also occurs for the SP3 saddle point. The calculated N4···O6 and C4−N4 distances are shorter than the N2···O2 and C2−N2 ones, respectively; this also explains the observed differences in the amino group rotation rates of guanine and cytosine forming the Watson−Crick base pair.[58] For a more detailed discussion on this topic, see refs 38 and 59.

The intermolecular separation and the strength of hydrogen bonding can be characterized best by the length of the middle (N1···N3) hydrogen bond. It is about the same in the GC1 and GC2 structures, but becomes 0.2 Å shorter in the zwitterionic base pair (GC4) and 0.4 Å shorter in the SP1 and SP3 saddle-point structures.

The geometry optimizations of the GC1, GC3, and GC6 base pairs resulted in planar structures, whereas the GC2 and GC4 base pairs were found to be nonplanar. The SP1, SP2, and SP3 saddle points, which were optimized under the planar symmetry constraint, are also slightly nonplanar, as indicated by out-of-plane imaginary frequencies (Table 1). To reveal the geometry differences between the planar and the more stable nonplanar

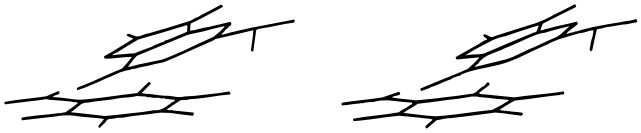

Figure 5. Stereoview of the structure of the GC4 base pair.

arrangements, the geometries calculated with and without the C_s symmetry constraint are compared in Table 2. For GC2 and GC4, there are only insignificant differences in the lengths of hydrogen bonds between their planar (C_s) and nonplanar (C_1) structures, although notable mutual twists are seen in the C_1 base pair structures (see Table 3 and Figure 5). In spite of this relatively large nonplanarity, the HF/6-31G* energy difference between the planar and nonplanar forms of GC2 and GC4 amounts to only 0.03 and 0.5 kcal/mol, respectively. The largest deviation from planarity was calculated for the hydrogens of the guanine amino group, which exhibits pyramidal geometry in the GC2 and GC4 base pairs.

In fact, the tendency of the $-NH_2$ group on guanine to be distorted from planarity, combined with repulsive electrostatic secondary interactions, is the actual driving force for the propeller twist in these base pairs. As a result, the molecular planes of cytosine and guanine in the GC2 and GC4 structures are twisted asymmetrically; i.e., they intersect along the outer N4···O6 hydrogen bond (Figure 5). The symmetric propeller-twist structure has been recently reported for the nonclassical $\pi\kappa$ base pair, the κ base of which possesses two pyramidal amino groups.[60] In accordance with previous studies,[61,62] we found

(58) Williams, L. D.; Williams, N. G.; Shaw, B. R. *J. Am. Chem. Soc.* **1990**, *112*, 829.
(59) MacPhail, R. A.; Williams, L. D.; Jones, D. A.; Shaw, B. R. *J. Biomol. Struct. Dyn.* **1992**, *9*, 881.
(60) (a) Florián, J.; Leszczynski, J. *J. Biomol. Struct. Dyn.* **1995**, *12*, 1055. (b) Leszczynski, J. *Chem. Phys. Lett.* **1990**, *173*, 371. (c) Leszczynski, J. *Chem. Phys. Lett.* **1991**, *181*, 123.
(61) Sponer, J.; Hobza, P. *J. Phys. Chem.* **1994**, *98*, 3161.
(62) Sponer, J.; Hobza, P. *J. Mol. Struct.: THEOCHEM* **1994**, *304*, 35.

the amino group of isolated guanine to be strongly nonplanar, with a 0.5 kcal/mol barrier (HF/6-31G*) for the wagging flip-flop of amino group hydrogens from below to above the molecular plane. As can be seen from Table 3, the extent of this nonplanarity is primarily determined by the C2−N2 bond order (bond length). Hence, guanine ionization stabilizes the pyramidal character of the $-NH_2$ group, whereas protonation at the N7 position results in a planar structure of guanine (Table 3). The ring protonation-induced increase in the $C-NH_2$ bond order is a general effect that has been observed and calculated for all nucleobases possessing an amino group.[63−66] An important role, especially in introducing asymmetry in the amino group hydrogen displacements, is played by the repulsive electrostatic interaction between the H_b and H_c protons.[61] On the other hand, the formation of hydrogen bonds tends to decrease amino group nonplanarity (see also our study of the structure of the $\pi\kappa$ base pair[60a]). The complicated interplay of base-pair ionization, hydrogen bonding, nonplanar amino groups, buckle, and propeller twist is difficult or even impossible to model by simulations based on empirical force fields. Finally, it should be noted that our structural findings, such as the flexibility of the base pairs with respect to the propeller-twist and buckle deformations, are generally supported by X-ray crystallography.[67] However, any detailed quantitative comparisons with X-ray data are meaningless, because our calculations neglect crystal-structure and sequence-dependent effects.

Interaction Enthalpies. The total, relative, and zero-point energies of individual constituents forming the GC1 and GC2 base pairs are given in Table 4. Except for the MINI-1 method, which strongly underestimates the energy of the enol form of guanine, all other methods provide similar stabilities for isolated rare tautomers as compared to the canonical forms of guanine and cytosine. As expected, the relative stability of canonical bases increases in a polar continuum.

The calculated dissociation enthalpy ($-\Delta H_0^{int}$(CP), Table 4) of the canonical base pair agrees well with the experimental value of 21.0 kcal/mol obtained by mass spectrometry measurements.[68] Because of the energy difference between the GC1 and GC2 base pairs (∼9 kcal/mol), the dissociation enthalpy of the GC1 base pair is about 9 kcal/mol larger than that of the GC2 complex. In polar media, both the dissociation enthalpies of the GC1 and GC2 complexes and their difference are significantly decreased.

Replication Fidelity. The relative energy of the GC1 and GC2 tautomers is the most important quantity for the examination of the role of proton transfer in the GC base pair in the mutation theory. In this study, reliable ab initio HF/6-31G* and MP2/6-31G** calculations were used for prediction of its magnitude. On the basis of the comparison of experimental and calculated tautomeric equilibria in isolated guanine and cytosine, the accuracy of our results can be estimated to be within 2 kcal/mol. Naturally, the relative stability of the GC1 and GC2 tautomers embedded in DNA will differ somewhat from that predicted by us for isolated base pairs. However, the fact that transferred protons are located in the central part of the GC base pair supports the plausibility of our model. In addition, we attempted to estimate the magnitude of solvent effects by using the continuum solvation model. We found that, due to the similar size and direction of GC1 and GC2 dipole moments, polar solvents can be expected to cause only small perturbations in the relative stability of these base pairs. Including this effect, we predict the GC2 tautomer to be 10 ± 2 kcal/mol less stable than the canonical GC1 base pair. Using Boltzmann statistics, this energy difference implies a 10^{-6}–10^{-9} ratio (equilibrium constant) between the GC2 and GC1 tautomers.

For the adenine·thymine base pair, a similar 10^{-7} ratio between the imino−enol (A*T*) and canonical (AT) tautomers was calculated using the minimal basis set (MINI-1).[37] However, in view of the present results and also by comparison with the tautomeric equilibrium in the cyclic formamide dimer,[16,17] the AT ↔ A*T* equilibrium would be significantly shifted toward the canonical base pair if MP2 calculations with more complete basis sets were used. To support this statement, and also upon the request of one of the reviewers, we carried out additional calculations of the AT and A*T* energies at the MP2/6-31G**//HF/6-31G* level, i.e., at the same level as used here for the guanine·cytosine base pair. The obtained 16.6 kcal/mol free energy difference (ΔG_{298}) implies a 10^{-12} probability of the formation of A*T*, which makes double proton transfer in AT a highly improbable cause of spontaneous mutations. The role of the A*T* tautomer in the mutation process is further

Table 4. Relative and Interaction Energies of Canonical and Rare Tautomers of Guanine and Cytosine Forming the GC1 and GC2 Base Pairs

	GC1		GC2	
method[a]	Gua	Cyt	Gua (enol)	Cyt (imino)
Relative Energy[b] (kcal/mol)				
MINI-1	0	0	−7.3	1.2
6-31G*	0	0	1.5	0.5
6-31G**	0	0	−0.2	0.6
MP2	0	0	0.6	1.1
SCRF	0	0	6.1	4.1
Reorganisation Energy[c] (kcal/mol)				
MINI-1	3.2	3.7	3.8	2.9
6-31G*	1.5	1.1	1.3	0.4
Zero-Point Vibrational Energy (unscaled) (kcal/mol)				
MINI-1	74.2	62.5	74.0	63.5
6-31G*	79.7	66.9	79.4	67.6
SCRF	79.5	66.9	79.4	67.6
BSSE (kcal/mol)				
MINI-1	3.0	3.6	2.0	3.8
6-31G*	1.3	1.5	1.3	1.6
6-31G**	1.1	1.5	0.9	1.5
MP2	2.5	3.5	2.4	3.5
ΔH_0^{int}(CP)[d] (kcal/mol)				
MINI-1	−23.4		−17.7	
6-31G*	−20.9		−12.0	
6-31G**	−21.3		−12.5	
MP2	−22.6		−15.0	
SCRF[e]	−10.8		−9.7	
exp[f]	−21.0			

[a] SCRF denotes the SCRF HF/6-31G*//SCRF HF/6-31G* method. For abbreviations of other methods see Table 1. [b] Difference in total energies (electronic energy plus repulsion of nuclei) of the canonical and rare tautomers. Energies of the canonical bases were taken as the reference points. [c] The reorganization energy was evaluated as the difference in the energy of monomer M (M = G, C, G*, C*) calculated in its optimized geometry and in the geometry it assumes in the complex: $\Delta E_{reorg}(M) = E(M)|_{geom=complex} - E(M)|_{geom=monomer} > 0$. For evaluation of the reorganization energy, the monomer-spanned basis set was used. [d] ΔH_0^{int}(CP) denotes the interaction enthalpy at 0 K, corrected for the BSSE. $\Delta H_0^{int}(CP) = E(complex) - E(subsystems) + \Delta E_0^{vib}(complex) - \Delta E_0^{vib}(subsystems) + BSSE$, where E denotes total energy. Total energies and zero-point vibrational energies (ΔE_0^{vib}) were calculated in the fully optimized geometry of the given system. For evaluation of 6-31G** and MP2 interaction enthalpies, ΔE_0^{vib} calculated at the HF/6-31G* level was used. [e] HF/6-31G* BSSE was used. [f] Experimental gas phase interaction enthalpy (ref 68).

(63) DelBene, J. E. *J. Phys. Chem.* **1983**, *87*, 367.
(64) Taylor, R.; Kennard, O. *J. Mol. Struct.* **1982**, *78*, 1.
(65) Florián, J. *J. Mol. Struct.: THEOCHEM* **1992**, *253*, 83.
(66) Florián, J.; Baumruk, V. *J. Phys. Chem.* **1992**, *96*, 9283.
(67) Dickerson, R. E. In *Structure and Methods, Vol. 3: DNA & RNA*; Sarma, R. H., Sarma, M . H., Ed.; Adenine Press: New York, 1990; p 1.
(68) Yanson, I. K.; Teplitsky, A. B.; Sukhodub, L. F. *Biopolymers* **1979**, *18*, 1149.

eliminated by its extremely short lifetime and its decreased stability at larger interbase separations.[37]

In contrast, the GC2 base pair was predicted to be separated from the main GC1 tautomer by a relatively large barrier. At the highest (MP2) level, the energy barrier for the simultaneous double proton transfer from GC2 to GC1 amounts to 5 kcal/mol, whereas GC2 decay by stepwise proton transfer via the GC4 ion pair requires a higher activation energy (7.3 kcal/mol). However, one can assume the ion-pair structure to be somewhat stabilized in the DNA environment, even when direct solvation by water molecules is limited via DNA−protein and DNA−DNA interactions, as in densely packed chromosomal DNA or during DNA replication. Thus, two different reaction mechanisms are, in principle, available for the transition between the GC1 and GC2 tautomers. In both cases, the proton transfer must be preceded by significant shortening of the interbase hydrogen bonds. The actual path, as well as the rate constants, will depend on the interplay of many factors involving contacts with surrounding molecules, initial conditions, and proton tunneling. Using transition-state theory and calculated barrier heights, one can roughly estimate that the rate constant for the GC1 → GC2 reaction at 300 K will fall in the $0-10^4$ s^{-1} range. This rate is high enough for GC2 base pairs to be formed during the lifetime of a cell. The GC2 → GC1 reaction will proceed with a much higher rate constant of 10^7-10^{10} s^{-1}. This rate is lower than the frequency of interbase intermolecular vibrations ($\sim 10^{10}$ s^{-1}). If GC2 is being separated because of the impact of outer factors, such as the action of DNA polymerase, the barrier height for the GC2 → GC1 proton transfer reaction (and, consequently, GC2 lifetime) will increase. (This behavior of the proton transfer barrier is common for all hydrogen-bonded systems having H bond lengths constrained at larger than equilibrium separation.[69,70]) Also, the smaller dissociation enthalpy of the GC2 base pair would assist the formation of spontaneous mutations by facilitating dissociation of the GC2 base pair into the G* and C* rare tautomers during the DNA replication process.

The proton donor/acceptor pattern of the G* and C* tautomers enables the formation of the G*T and C*A mismatches with Watson−Crick pairing geometry. This mispairing results in the formation of a GC → AT transition mutation after the next round of replication. If this were the only possible substitution mutation mechanism, the number of GC base pairs in DNA would decrease on the evolutionary time scale. In reality, this trend is most probably offset by the wobble base pairing, to which the AT base pair is more prone due to its smaller interaction enthalpy.[71−74] However, the observed deficiency in GC content of the DNA of higher organisms,[6] which is generally as low as 0.5 times (0.4 in humans) the AT content, is an impressive consonance.

Competition among different mutation mechanisms also prevents the straightforward comparison of the calculated GC2-to-GC1 ratio with the measured fidelity of DNA replication. However, there is one important difference between template-based and free nucleotide triphosphate-based substitution mutations; namely, nucleotides that were newly added at the growing DNA strand can be excised by DNA polymerases through their multistage proofreading activity,[75] provided the added nucleotide mismatches the template base or exhibits an unusual structure. On the other hand, mutations originating from the tautomerization of the template are hardly recognizable by the polymerase. Also, there is no simple way of correcting them without breaking the template DNA strand. Indeed, the relative amount of rare tautomeric base pairs in DNA predicted by us ($10^{-6}-10^{-9}$) agrees well with the frequency of substitution mutations observed for proofreading polymerases.[76] However, this agreement is only a fortuitous coincidence, since all chemical or genetic studies of DNA replication fidelity use single-stranded DNA templates. As such, they are principally unable to detect substitution mutations originating from base-pair tautomerization that occurs prior to DNA unwinding.

Though a direct experimental proof of the predicted tautomeric equilibrium is missing, we believe that the technique employed for its derivation is established enough to warrant the plausibility of the obtained results. The concept of mutable GC base pairs provides an incentive for the future development of mutagens able to selectively tautomerize the GC base pairs that are parts of defined DNA sequences.

Conclusions

We have demonstrated that the guanine·cytosine base pair is more structurally variable than has been assumed. The ion-pair (GC4) and imino−keto/amino−enol (GC2) forms of this base pair are energetically accessible, though the probability of their formation falls below 10^{-6}. The geometries of the GC2 and GC4 base pairs are significantly nonplanar. The degree of this nonplanarity increases for structures involving negatively charged guanine, whereas guanine protonation at the N7 position increases the rigidity of the base pair.

The calculation of the steepest descent path for the stepwise single-proton-transfer reaction by using the mass-weighted coordinates enabled us to account for dynamic effects during the proton transfer. Consequently, the proton-transfer event could be described as consisting of two parts: first, the interbase distance is decreased by 0.2−0.4 Å compared to the equilibrium distance; second, stepwise proton transfer from the GC1 to the GC2 structure occurs via the GC4 reaction intermediate. The role of the ion-pair complex as a reaction intermediate enables proton transfer to be easily triggered by the interactions of the base pair with protein side groups or transition metals. Even without being assisted by external factors, the transition from the canonical GC1 structure to the GC2 rare tautomer can occur during the lifetime of the cell by the mechanism of simultaneous double proton transfer.

We predicted the equilibrium ratio of the GC2 and GC1 base pairs to lie in the $10^{-6}-10^{-9}$ range, i.e., within the range of measured DNA replication fidelity. Thus, the proton transfer-induced formation of the rare tautomers (G*, C*) in the DNA template prior to replication represents a viable expansion of the more established spontaneous mutation mechanisms.

Acknowledgment. This work was supported in part by NTH Grant No. 332090, ONR Grant No. N00014-95-1-0049, and a contract (DAAH 04-95-C-0008) between the Army Research Office and the University of Minnesota for the Army High Performance Computing Research Center. The Mississippi Center for Supercomputing Research is acknowledged for a generous allotment of computer time. The helpful comments of Michael J. Stewart and Dr. Jiří Šponer are greatly appreciated.

JA951983G

(69) Scheiner, S. In *Proton Transfer in Hydrogen Bonded Systems*; Bountis, T., Ed.; Plenum: New York, 1992; p 29.
(70) Kar, T.; Scheiner, S. *J. Am. Chem. Soc.* **1995**, *117*, 1344.
(71) Hobza, P.; Sandorfy, C. *J. Am. Chem. Soc.* **1987**, *109*, 1302.
(72) Sponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem.* **1996**, *100*, 1965.
(73) Dey, M.; Moritz, F.; Grotemeyer, J.; Schlag, E. W. *J. Am. Chem. Soc.* **1994**, *116*, 9211.
(74) Petruska, J.; Sowers, L. C.; Goodman, M. F. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 1559.
(75) Beckman, R. A.; Loeb, L. A. *Q. Rev. Biophys.* **1993**, *26*, 225.
(76) Fersht, A. R.; Knill-Jones, J. W. *J. Mol. Biol.* **1983**, *165*, 633.