Protonation of Thymine, Cytosine, Adenine, and Guanine DNA Nucleic Acid Bases: Theoretical Investigation into the Framework of Density Functional Theory

NINO RUSSO,[1] MARIROSA TOSCANO,[1] ANDRÉ GRAND,[2]* FRANCK JOLIBOIS[2]†

[1]*Dipartimento di Chimica, Universitá della Calabria, I-87030 Arcavacata di Rende (CS), Italy*
[2]*Département de Recherche Fondamentale sur la Matière Condensée, Service de Chimie Inorganique et Biologique, Grenoble, France*

Received 15 October 1997; accepted 21 January 1998

ABSTRACT: Gradient-corrected density functional computations with triple-zeta-type basis sets were performed to determine the preferred protonation site and the absolute gas-phase proton affinities of the most stable tautomer of the DNA bases thymine (T), cytosine (C), adenine (A), and guanine (G). Charge distribution, bond orders, and molecular electrostatic potentials were considered to rationalize the obtained results. The vibrational frequencies and the contribution of the zero-point energies were also computed. Significant geometrical changes in bond lengths and angles near the protonation sites were found. At 298 K, proton affinities values of 208.8 (T), 229.1 (C), 225.8 (A), and 230.3 (G) kcal/mol were obtained in agreement with experimental results.
 J Comput Chem 19: 989–1000, 1998

Keywords: gas-phase proton affinity and basicity; density functional computations; protonation of DNA acid bases

* *Present affiliation:* Laboratoire de Chimie de Coordination (Unité de Recherche Associée au CNRS No. 1194)
† *Present affiliation:* Laboratoire de Léisions des Acides Nucléiques
Correspondence to: M. Toscano
Contract/grant sponsors: CNR; MURST; CINECA

 CCC 0192-8651 / 98 / 090989-12

Introduction

Protonation of nucleic acid bases plays a crucial role in many biochemical (i.e., enzymatic reactions, stabilization of triplex structures) and mutagenic processes.[1,2] For this reason the acid–base equilibria involving the nucleic acid components have been studied extensively in both gas and condensed phases by theoretical[3–8] and experimental[9–12] means. The development of mass spectrometry techniques has favored gas-phase investigation in this field.[13] Data from these techniques give different indications with respect to those obtained by experimental studies in solution. For example, on the basis of pK_a values in aqueous solution,[14] cytosine is predicted to be protonated, but not guanine, whereas a recent fast-atom-bombardment tandem mass spectrometry study appeared to show different results.[13] This fact is not surprising because the protonation energetics depend on the chemical surrounding, so theoretical investigations represent a practicable way for obtaining information about the energetics of proton attachment to the different sites of thymine, cytosine, adenine, and guanine and to establish which of these are favored.

Nucleic acid bases may, in fact, be protonated at the different nitrogens or oxygen atoms, but experimentally, even using the more sophisticated techniques, it is not possible to discriminate between the various possibilities.

The first theoretical study[15] on the protonation processes of nucleic acid bases was performed using the molecular electrostatic potential derived from *ab initio* wave functions. This study provides indications of the protonation sites but does not yield proton affinity (PA) values. PAs for all bases were subsequently calculated by the *ab initio* method employing minimal STO-3G[16] and 4-31G[17,18] basis sets. For thymine and cytosine the calculations were also performed with the more extended 6-31G basis set.[19] To our knowledge, the PA values obtained by the high-level *ab initio*-correlated method [MP4/6-311 + + G(d,p)//MP2/6-31G(d)] are available only for cytosine and guanine.[3]

In this article, we report a systematic study of the protonation of the four DNA nucleic acid bases considering all the different attachment sites by using the Gaussian density functional (DF) method. It is well known that this approach, which uses density as the basic variable, demands less computational effort than the *ab initio*-correlated molecular orbital calculations. The use of DF methods has grown considerably in the last few years in many fields of chemistry. Recently, several DF studies, at different levels of sophistication, have also been devoted to the prediction of proton affinity in a series of nitrogen- and oxygen-containing compounds.[20–22] Satisfactory agreement between calculated and measured PA values is reached especially when the gradient-corrected exchange and correlation functionals are employed.

In our computations, the DF lower energy tautomer of each of the DNA bases has been considered in the protonation process. Although in one case i.e., cytosine[3,6]) there is no correspondence between DF and *ab initio*-correlated methods in the prediction of tautomeric relative stabilities, we believe that our choice is the better one from a biological point of view. In fact, the most stable DF tautomer is always the one with the hydrogen bonded to the nitrogen to which the sugar moiety of nucleosides is attached. On the other hand, in the case of cytosine, the energy differences between the most stable tautomer proposed by the *ab initio* study[3] and that coming from the previous DF computation[6] is about 1 kcal/mol and does not significantly affect the PA values.

Computational Procedure

Local spin density (LSD) full geometry optimization for both protonated and neutral species of each system were performed with the density functional code DGauss,[23] which calculates variational self-consistent solutions to the Kohn–Sham equations in the framework of the local Vosko–Wilk–Nusair spin-density approximation of the exchange/correlation potential.[24] The nonlocal corrections were added using the gradient-corrected functionals of Becke for the exchange[25] and Perdew[26] for the correlation part (hereafter denoted NLSD). For C, N, O, and H triple-zeta (TZVP) orbital types with corresponding C, N, O (8/4/4), and H (4/1), auxiliary basis sets[27] were employed.

Vibrational frequencies were computed using the analytical second derivatives. From inspection of the frequencies of the normal mode of vibration all the protonated and neutral species were minima.

Gas-phase proton affinity was assumed as the negative of the enthalpy for the process:

$$B + H^+ \rightarrow BH^+$$

and calculated as:

$$\mathrm{H} = E(BH^+) - E(B) - E(H^+) + \Delta(PV)$$
$$PA = -[E_{\mathrm{SCF}}(BH^+) - E_{\mathrm{SCF}}(B) + (E_{\mathrm{vib}}(BH^+) - E_{\mathrm{vib}}(B))] + 5/2RT$$

where E_{SCF} valuse were obtained from the SCF calculations, E_{vib} includes zero-point energy and temperature corrections to the vibrational enthalpy, and $5/2RT$ includes the translational energy of the proton and $\Delta(PV)$ term.

The absolute gas-phase basicity was calculated as the negative of the standard free energy, ΔG:

$$\Delta G = \Delta H - T\Delta S$$

The entropy contribution is given by:

$$-T\Delta S = -T[S(BH^+) - S(B) - S(H^+)]$$

at 298 K the $TS(H^+)$ term has a value of 7.76 kcal/mol.[28]

The BH^+ and B entropies were obtained from thermochemical calculations by using the NLSD theoretical harmonic frequencies and the equilibrium geometries.

The electrostatic potential at point r is defined as an electrostatic interaction energy of a probe charge of 1.0 a.u. with all nuclei and electrons of the molecule[23]:

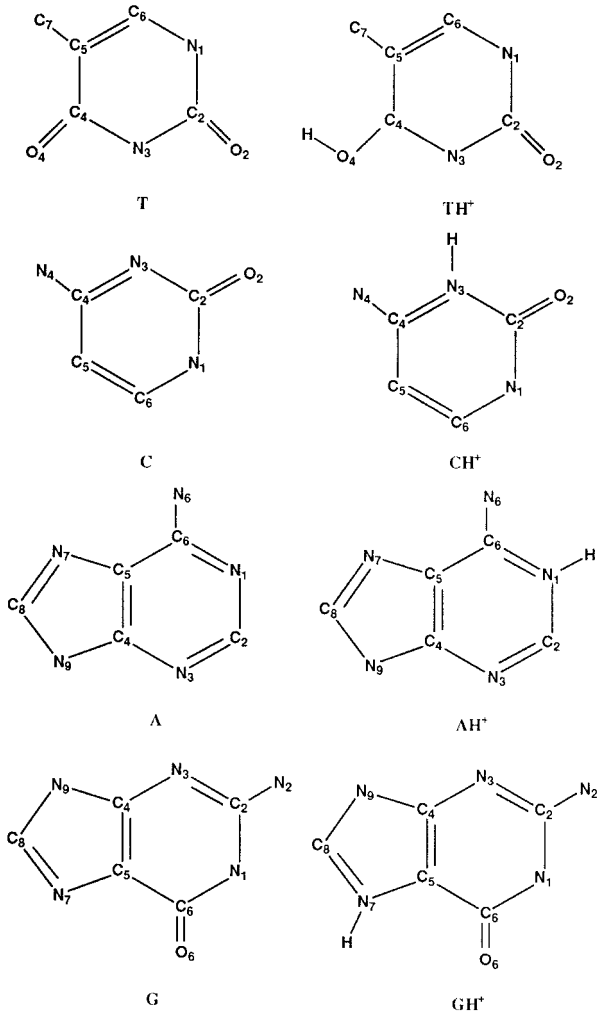
$$\mathbf{V}(r) = \sum_A \frac{Z_A}{|R_A - r|} - \int \frac{\rho(r')}{|r - r'|}\, dr'$$

and calculated in connection with a grid defined as: $x\ \mathrm{min} = \mathrm{min}(x_n) - \Delta x$ and $x\ \mathrm{max} = \mathrm{max}(x_n) + \Delta x$; $y\ \mathrm{min} = \mathrm{min}(y_n) - \Delta y$ and $y\ \mathrm{max} = \mathrm{max}(y_n) + \Delta y$; $z\ \mathrm{min} = \mathrm{min}(z_n) - \Delta z$ and $z\ \mathrm{max} = \mathrm{max}(z_n) + \Delta z$. with $\Delta x = \Delta y = \Delta z = 3.0$ Å; $\mathrm{min}(x_n)$ and $\mathrm{max}(x_n)$ are, respectively, the smallest and the largest x-coordinate of any atom in the molecule. Thirty grid points in each direction (x, y, z) were used.

Results

STRUCTURAL FEATURES

Geometrical parameters for the most stable neutral and protonated tautomer of thymine, cytosine, guanine, and adenine (see Scheme 1) are reported in Tables I–IV together with theoretical MP2[30] and experimental data coming from two statistical surveys of the x-ray structures in the *Cambridge Structural Database*.[31,32] In all neutral species, the ring conformations are found to be planar, whereas the amine groups of cytosine and guanine deviate from the planarity. Similar behaviors have been found at the MP2[30] level and with previous DF computations on cytosine.[6] In the protonated molecules the amine groups are nearly planar. As a general feature we note that our bond lengths and valence angles agree well with the experimental data and are consistent with MP2-optimized structures. For the neutral systems the average deviations between our bond lengths and the corresponding x-ray bond lengths are 0.008, 0.016, 0.007, and 0.015 Å for thymine, cytosine, adenine,



SCHEME 1.

TABLE I.
Theoretical and Experimental Geometrical Parameters for *T* and *TH*$^+$ Most Stable Protonated Form (Distances in Angstroms and Angles in Degrees).

Parameters	NLSD *T*	MP2 *T*[a]	Exp *T*[b]	NLSD *TH*$^+$
C4—C5	1.453	1.460	1.445	1.403
C5—C6	1.349	1.354	1.339	1.370
C6—N1	1.366	1.380	1.378	1.343
N1—C2	1.379	1.385	1.376	1.392
C2—N3	1.374	1.386	1.373	1.401
N3—C4	1.396	1.403	1.382	1.339
C5—C7	1.483	1.496	1.473	1.484
C2—O2	1.220	1.224	1.220	1.202
C4—O4	1.225	1.230	1.228	1.306
C4—C5—C6	118.0	118.4	118.0	115.4
C5—C6—N1	122.7	122.3	123.7	122.4
C6—N1—C2	123.8	124.1	121.3	124.6
N1—C2—N3	112.6	112.2	114.6	111.6
C2—N3—C4	128.1	128.6	127.2	125.2
N3—C4—O4	120.0	120.7	119.9	115.0
N1—C2—O2	124.2	123.6	123.1	123.7
C4—C5—C7	117.8	117.7	119.0	119.8

[a] From ref. 29. [b] From ref. 30.

and guanine, respectively. The maximum deviation (0.03 Å) occurs for the C6–C1 of guanine. For the valence angles the deviations are in the same order, 1.0°, 1.6°, 1.0°, and 1.2°. In this case, the maximum deviation occurs for the C6—N1—C2 angle of cytosine. Similar behavior can be verified for the protonated species (see Tables I–IV).

The protonation process produces, as expected, significant changes in the structural parameters of neutral systems, especially near the protonation

TABLE II.
Theoretical and Experimental Geometrical Parameters for *C* and *CH*$^+$ Most Stable Protonated Form (Distances in Angstroms and Angles in Degrees).

Parameters	NLSD *C*	MP2 *C*[a]	Exp. *C*[b]	Exp. *C*[c]	NLSD *CH*$^+$	Exp. *CH*$^+$[b]	Exp. *CH*$^+$[c]
C4—C5	1.427	1.438	1.447	1.425	1.407	1.413	1.413
C5—C6	1.354	1.358	1.337	1.339	1.359	1.341	1.346
C6—N1	1.344	1.357	1.364	1.367	1.345	1.362	1.365
N1—C2	1.417	1.419	1.399	1.397	1.387	1.381	1.381
C2—N3	1.358	1.380	1.356	1.353	1.401	1.387	1.384
N3—C4	1.317	1.320	1.334	1.335	1.351	1.352	1.353
N4—C4	1.350	1.358	1.337	1.335	1.324	1.313	1.315
C2—O2	1.222	1.226	1.237	1.240	1.204	1.211	1.212
C4—C5—C6	116.0	116.0	117.6	117.4	118.1	118.5	118.4
C5—C6—N1	119.9	119.6	121.0	121.0	121.9	122.5	122.2
C6—N1—C2	123.5	124.0	120.6	120.3	123.6	121.5	121.7
N1—C2—N3	116.1	116.0	118.9	119.2	112.7	114.9	114.7
C2—N3—C4	120.5	119.8	120.0	119.9	125.9	125.1	125.3
N3—C4—N4	116.8	116.7	117.9	118.0	119.9	119.5	119.5
N1—C2—O2	117.9	118.7	119.2	118.9	124.5	123.5	123.4

[a] From ref. 29. [b] From ref. 31. [c] From ref. 30.

TABLE III.
Theoretical and Experimental Geometrical Parameters for *A* and AH^+ Most Stable Protonated Form (Distances in Angstroms and Angles in Degrees).

Parameters	NLSD *A*	MP2 A^a	Exp. A^b	Exp. A^c	NLSD AH^+	Exp. AH^{+b}	Exp. AH^{+c}
C4—C5	1.395	1.398	1.382	1.383	1.396	1.385	1.378
C5—C6	1.402	1.410	1.409	1.406	1.397	1.405	1.403
C6—N1	1.337	1.342	1.349	1.351	1.362	1.359	1.360
C6—N6	1.342	1.353	1.337	1.335	1.323	1.322	1.320
N1—C2	1.335	1.352	1.338	1.339	1.374	1.362	1.357
C2—N3	1.332	1.339	1.332	1.331	1.297	1.306	1.305
N3—C4	1.330	1.343	1.342	1.344	1.339	1.354	1.356
C4—N9	1.368	1.378	1.376	1.374	1.357	1.366	1.365
N9—C8	1.370	1.372	1.367	1.373	1.396	1.378	1.373
C8—N7	1.310	1.326	1.312	1.311	1.289	1.316	1.312
N7—C5	1.370	1.381	1.385	1.388	1.356	1.378	1.379
C4—C5—C6	115.9	116.0	116.9	117.0	118.1	117.7	117.9
C5—C6—N1	118.5	118.9	117.6	117.7	113.7	114.3	114.0
C6—N1—C2	119.1	118.2	118.8	118.6	123.7	123.2	123.3
N1—C2—N3	128.4	129.1	129.0	129.3	124.2	125.5	125.7
C2—N3—C4	113.3	110.7	110.8	110.6	113.5	112.0	111.6
C4—N9—C8	107.1	106.9	105.9	105.8	106.5	106.3	105.9
N9—C8—N7	112.8	113.6	113.8	113.8	112.5	113.0	113.5
C8—N7—C5	104.4	103.2	103.9	103.9	105.0	104.1	103.7
N7—C5—C4	112.2	112.1	110.7	110.7	111.7	111.0	111.0
N1—C6—N6	119.1	119.2	119.0	118.6	123.1	120.2	120.2

[a] From ref. 29. [b] From ref. 31. [c] From ref. 30.

site. For instance, the C2—N3 bond of cytosine becomes longer by about 0.05 Å and the C2—N3—C4 angle is opened by about 5°. In the neutral molecule, the N1—C2—O2 angle is 117.9° and becomes 124.5° upon the protonation on N1 atom. This effect is easily comprehensible if we consider that now the lone-pair(on N3)–lone-pair(on O2) interaction is replaced by the hydrogen lone-pair interaction. As a result the geometry appears more symmetrical.

PREFERRED PROTONATION SITES

For the systems investigated different protonation sites are possible. On the basis of previous experimental and theoretical indications, as well as on the basis of simple chemical considerations, the preferred positions are O4 and O2 for thymine; N3 and O2 for cytosine; N1, N3, and N7 for adenine, and N7, N3, and O6 for guanine. Total and relative energies of neutral and protonated nucleic acid bases are reported in Table V. In the case of thymine, protonation occurs preferentially on the O4 site. Protonation on the O_2 is unfavored by 9.44 kcal/mol (8.94 kcal/mol is the value obtained taking into account the zero-point energy correction [ZPEC]). The situation for cytosine is not as well defined because the energy difference between protonation at N3 and O2 is only 0.19 kcal/mol (0.79 kcal/mol with the ZPEC). A similar small energy difference (1.12 and 2.22 kcal/mol without and with ZPEC, respectively) is found when the N1 and N3 sites of adenine are considered. The attachment of the proton at the N7 position of adenine yields a tautomer that lies at 7.38 kcal/mol (7.58 kcal/mol with ZPEC). Finally, in the case of guanine, we found that the preferred protonation sites is N7 with O6 and N3 at 6.8 and 18.49 kcal/mol, respectively (the counterpart values with ZPEC are 6.5 and 17.69 kcal/mol). The rationalization of the results on the basis of the molecular electrostatic potential, or charge distribution, or the analysis of the HOMO composition reveals itself to be inadequate.[15,17] In fact, charge transfer, polarization, and structural changes that occur during the protonation process must be considered. In our study all these factors are taken into account.

TABLE IV.
Theoretical and Experimental Geometrical Parameters for *G* and *GH*$^+$ Most Stable Protonated Form (Distances in Angstroms and Angles in Degrees).

Parameters	NLSD *G*	MP2 *G*[a]	EXP *G*[b]	EXP *G*[c]	NLSD *GH*$^+$
C4—C5	1.393	1.394	1.377	1.379	1.383
C5—C6	1.429	1.440	1.415	1.419	1.429
C6—N1	1.424	1.431	1.393	1.391	1.410
C6—O6	1.220	1.225	1.239	1.237	1.215
N1—C2	1.362	1.374	1.375	1.373	1.371
C2—N2	1.365	1.363	1.341	1.341	1.326
N3—C4	1.345	1.364	1.355	1.350	1.334
C2—N3	1.309	1.314	1.327	1.323	1.326
C4—N9	1.362	1.370	1.377	1.375	1.381
N9—C8	1.375	1.376	1.374	1.374	1.342
C8—N7	1.306	1.323	1.304	1.305	1.327
N7—C5	1.368	1.378	1.389	1.388	1.367
C4—C5—C6	118.3	118.8	119.1	118.8	120.8
C5—C6—N1	109.9	109.1	111.7	111.5	108.8
C6—N1—C2	126.7	127.0	124.9	125.1	126.4
N1—C2—N3	123.1	123.9	124.0	123.9	123.0
N1—C2—N2	117.2	116.8	116.3	116.2	118.2
C4—N9—C8	106.9	107.0	106.0	106.4	109.5
N9—C8—N7	112.3	112.9	113.5	113.1	108.2
C8—N7—C5	105.1	103.8	104.2	104.3	109.3
N7—C5—C4	110.7	111.7	110.8	110.8	107.5
C5—C6—O6	131.0	131.3	128.3	128.6	128.7

[a] From ref. 29. [b] From ref. 31. [c] From ref. 30.

It is important to note that the proton is a hard species and that the protonation process can be considered essentially in terms of hardness and softness. In a more traditional language the protonation process is an acid–base reaction; therefore, it seems more appropriate to consider the charge distribution rather than the HOMO composition.

Because of the variety of effects derived by the protonation process, a description will be presented for each molecule individually.

TABLE V.
Total and Relative Energies and Zero-Point Corrections for Neutral and Protonated Nucleic Acid Bases.

System	*E* (a.u.)	ΔE (kcal / mol)	ΔE_{corr} (kcal / mol)	ZPE (kcal / mol)
A	−467.493530	/	/	68.7
AH$^+$ (N1)	−467.863983	0.0	0.0	76.8
AH$^+$ (N3)	−467.862201	1.12	2.22	77.9
AH$^+$ (N7)	−467.852227	7.38	7.58	77.0
G	−542.764346	/	/	71.8
GH$^+$ (N7)	−543.141781	0.0	0.0	79.8
GH$^+$ (O6)	−543.130899	6.80	6.50	79.5
GH$^+$ (N3)	−543.112313	18.49	17.69	79.0
T	−454.308547	/	/	69.9
TH$^+$ (O4)	−454.650341	0.0	0.0	77.7
TH$^+$ (O2)	−454.635288	9.44	8.94	77.2
C	−395.084514	/	/	60.6
CH$^+$ (N3)	−395.458405	0.0	0.0	67.5
CH$^+$ (O2)	−395.458102	0.19	0.79	68.1

Thymine

On the basis of HOMO composition, thymine offers three protonation possibilities, located on O2, O4, and N3 atoms. The proton attachment at N3 implies a significant geometrical distortion of the ring.

The molecular electrostatic potential (see Fig. 1) clearly shows that O4 and O2 are the preferred sites for the protonation process. Further information is not possible on the basis of this property only, because it is predicated on the assumption that the interaction is purely electrostatic. From the total energy values obtained by geometry optimization (see Table V) of the two protonated species, and in agreement with previous *ab initio* studies,[17,19] we found that the O4 is the favorable protonation site. The higher stability of the O4 protonated tautomer is due to two different reasons. First, the attachment of the proton to the O4 site, with an H—O4—C4—N3 *anti* conformation, is favored from a steric point of view due to less repulsion between the added proton and the hydrogens of the methyl group (which has a slow rotational barrier and can easily rearrange two hydrogens out of the plane of the ring) with respect to imide hydrogen (in this case the steric rearrangements might cause a more energetically expensive distortion of the ring). In addition (see net charges and bond orders in Fig. 2), in the case of protonation on O4, the charge is better delocalized. In fact, in the first case, N1, C6, C5, and C4 atoms all become more positive than in the neutral molecule. In the O2 protonation the incoming charge is localized practically only on the C2 atom due to the presence of C4—O4 carbonyl that works as an electron-withdrawal group. The variation of the bond orders is consistent with the trends of the charge distributions obtained by Mulliken population analysis.

Cytosine

N1, N3, and O2 lone pairs contribute mainly to the HOMO of cytosine. On the other hand, the molecular electrostatic potential indicates that protonation is possible only at N3 and O2 sites (see Fig. 3). As shown in Table V, the N3 protonation site is preferred over the O2 one by only 0.19 kcal/mol. Previous electrostatic potential and

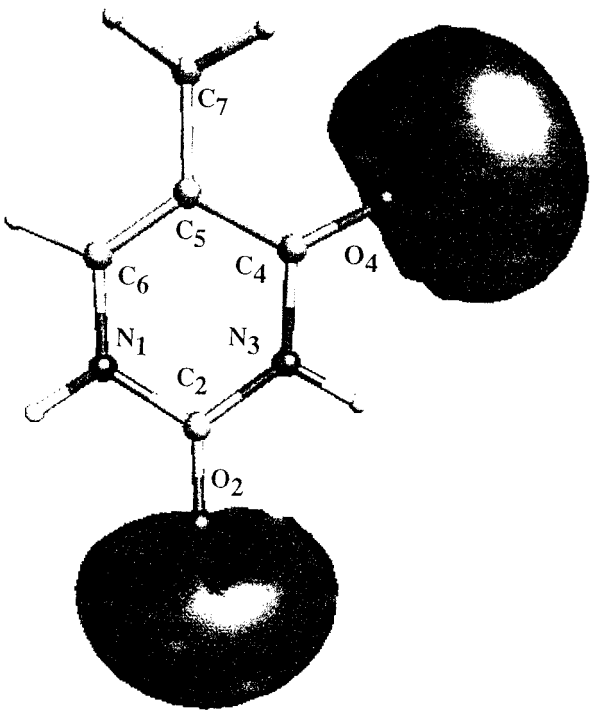


FIGURE 1. Electrostatic molecular potential energy map for thymine.

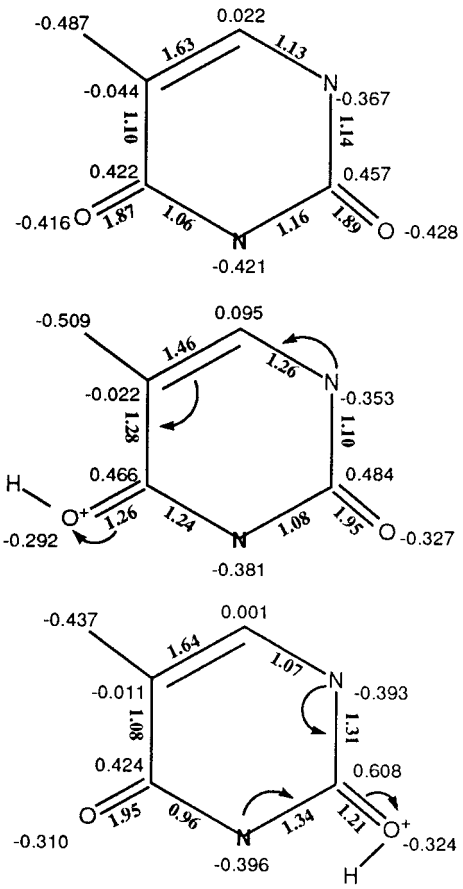


FIGURE 2. Net charges and bond orders (in bold) for neutral thymine and its protonated tautomers.

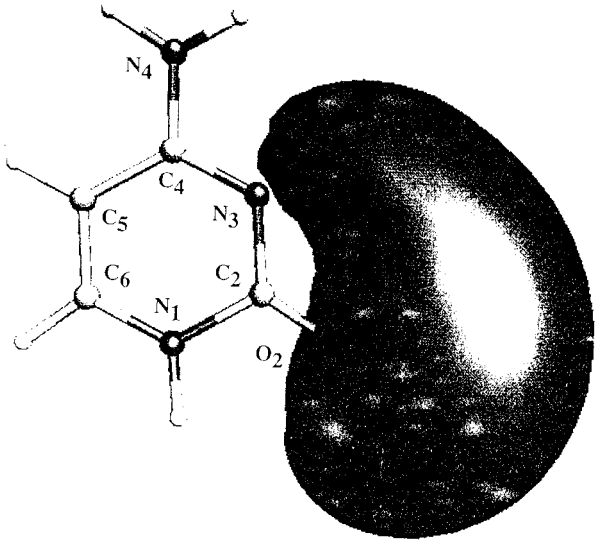


FIGURE 3. Electrostatic molecular potential energy map for cytosine.

STO-3G studies performed without geometry optimization[14,15] favor the N3 site by about 20 and 2 kcal/mol, respectively. The optimization of geometries of both neutral and protonated species with minimal STO-3G basis sets predicts that the O2 protonation energy is 4 kcal/mol greater than the corresponding N3 protonation energy. A recent *ab initio* MP4/6-311 + + G(d,p)//MP2/6-31G(d) study[3] found the two protonated forms to have similar stability, the enol form being preferred by only 0.4 kcal/mol. A clear explanation for the preference of one site rather than the other is difficult due to the very small energy difference and the variety of effects. Looking at Figure 4 we can see that the protonation on O2 seems to favor charge delocalization. Actually, the N3 atom has a more negative charge than in the neutral cytosine. This means that the induced delocalization is not practical there. On the other hand, we must also consider that the presence of hydrogen on N1 forces the orientation of hydroxyl hydrogen toward the N3 lone pair.

In the case of protonation on N3 the net charges of all other atoms become more positive. A better distribution of electronic densities between the bonds is now in place.

Adenine

HOMO of adenine is formed essentially by the lone pairs of all nitrogens. The orbital of the amine group contributes with the higher coefficient, and

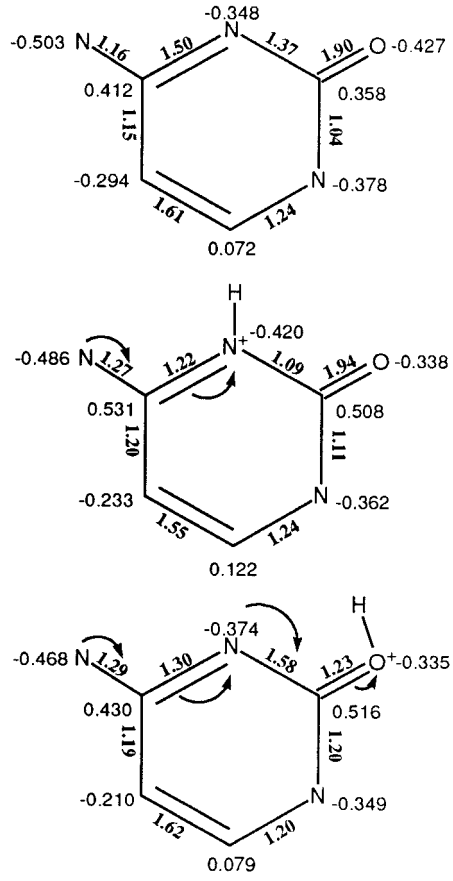


FIGURE 4. Net charges and bond orders (in bold) for neutral cytosine and its protonated tautomers.

that of N1 with the lower one. On the basis of this information, N1 protonation might be the less favorable process. On the contrary, the molecular electrostatic potential map (Fig. 5) indicates that N1, N3, and N7 are the better candidates for proton attachment. In particular, the N1 and N3 regions appear to be favored over N7. Experimentally, the preferred protonation sites of adenine are N1 and N3 with the former slightly favored.[18] Our computation favors protonation at N1 over those at N3 and N7 by 1.12 and 7.38 kcal/mol, respectively. This trend is in agreement with previous theoretical computations.[17] From bond orders analysis (see Fig. 6), and due to the small energy difference it is possible to view the N1 and N3 protonated forms (in N1 the double bonds occur in C4—C5 and in the C2—N3 positions, whereas, in the second, in the C4—C5 and N1—C2 positions) as two resonance hybrids of a unique structure with a consistent charge delocalization. This observation allows better insight into the reason for which the protonation on the five-membered ring

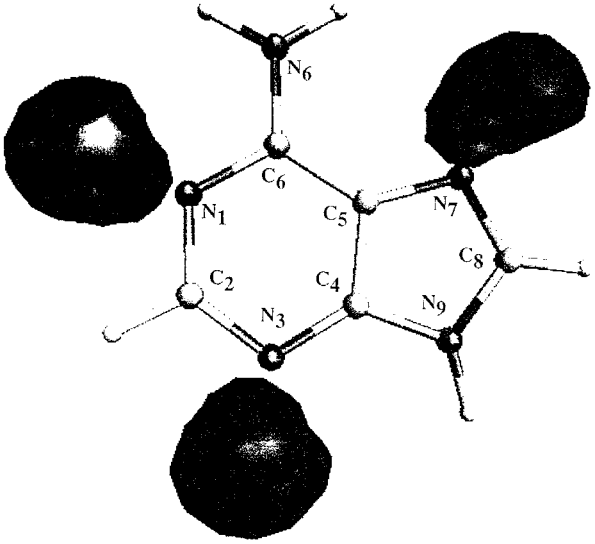


FIGURE 5. Electrostatic molecular potential energy map for adenine.

is unfavorable. In fact, as shown in Figure 6, the positive charge appears to be distributed only on the C8 and N9 atoms, because it is evident that no interactions exist between the two rings upon protonation.

Guanine

As in the case of adenine, information about preferred protonation sites based only on HOMO composition, appear to also be incorrect for guanine. The greater orbital coefficients are those of N3, O6, C5, C8, and N2. No contribution arises from the N7 atom. Electrostatic potential assessment (Fig. 7) indicates that a wide region between N7 and O6 is a good place for the proton, which can also attach to the N3 site, albeit with minor probability.

Experimental data[2,11,14] and previous *ab initio* computations[3,17] suggest that N7 protonation is the most favorable process, followed by the O6 process. Our NLSD computations agree with these findings. The N7 protonated form is the most stable (see Table V). At 6.5 kcal/mol above the absolute minimum we found the form protonated on O_6, MP4/6-311 + + G(d,p)//MP2/6-31G(d) study[3] establishes an energy difference (ΔE) of 5.6 kcal/mol between the two forms. The corresponding B3LYP 6-311 + + G(d,p)//MP2/6-31G(d) ΔE value is 6.4 kcal/mol.[3] Finally, in agreement with the suggestions of molecular electrostatic potential computations, the N3 protonated tautomer is less

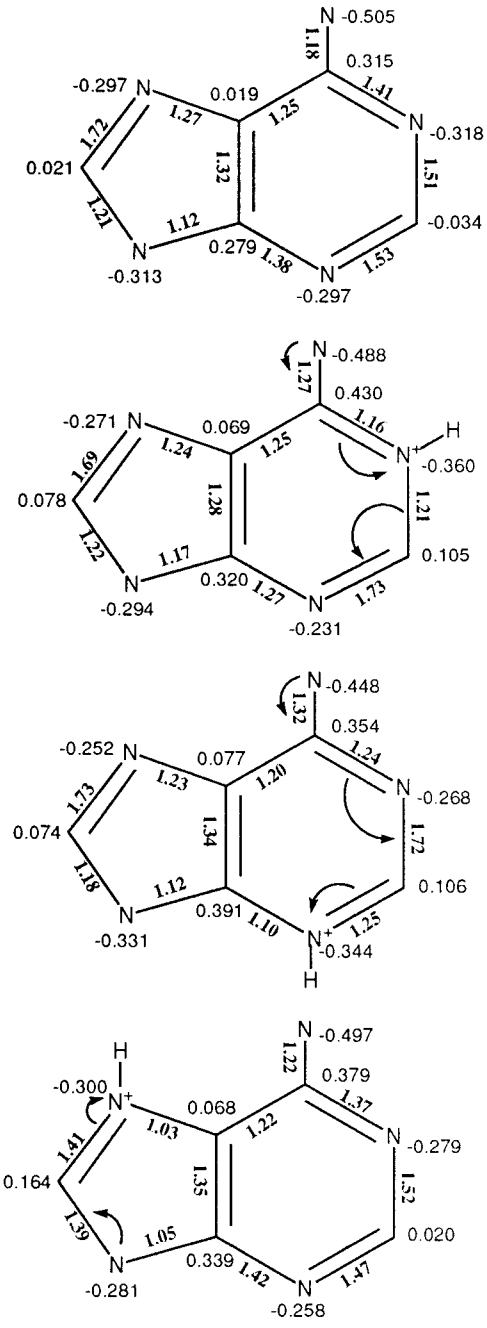


FIGURE 6. Net charges and bond orders (in bold) for neutral adenine and its protonated tautomers.

stable and lies at 17.69 kcal/mol above the absolute minimum.

Contrary to the adenine case, in guanine, the most stable protonated site is on the five-membered ring. This is not surprising because of the different chemical nature of the two six-membered rings in the two molecules. N7 protonation is accompanied by a major redistribution of the elec-

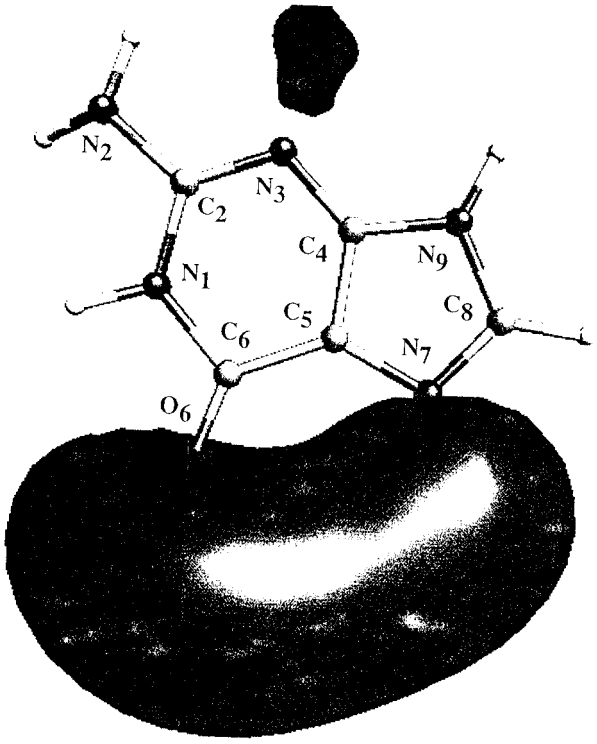
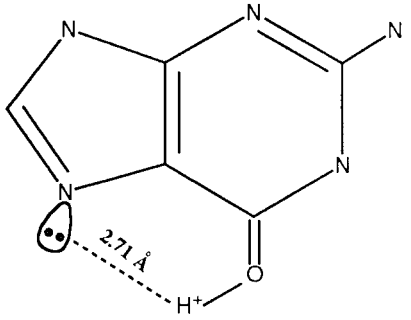


FIGURE 7. Electrostatic molecular potential energy map for guanine.

tron density, as shown by Mulliken population and Mayer bond order analyses (see Fig. 8). O6 protonation does not allow a similar charge distribution, but the resulting product is stabilized by the presence of an intramolecular hydrogen bond of 2.71 Å (see Scheme 2) giving rise to a further five-membered cycle. The stabilizing effect of the H-bond formation is confirmed by the fact that O6 protonation with the hydrogen toward the N1 atom yields to a product less stable by 8.23 kcal/mol.



SCHEME 2.

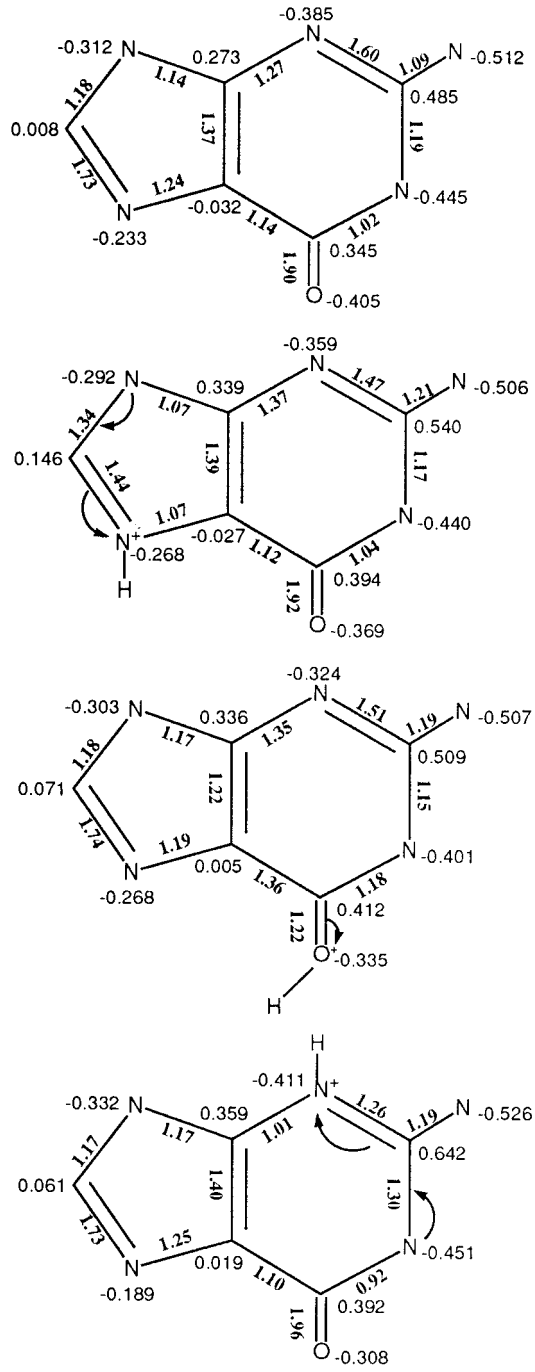


FIGURE 8. Net charges and bond orders (in bold) for neutral guanine and its protonated tautomers.

PROTON AFFINITIES AND GAS PHASE BASICITIES

Theoretical proton affinities (PA) are generally well reproduced with the Gaussian-2 method[33] with a target accuracy of 2 kcal/mol.[34] Unfortunately, these calculations are limited to small sys-

tems because they require high computational cost. On the other hand, DFT-based methods give similar reliable predictions of PA[20–22] with much less computational demand.

NLSD proton affinities and gas-phase basicities (GB) are given in Table VI, together with previously available theoretical[3,19] and the most recent experimental values.[13]

Results show that the NLSD PA order is the same as that proposed by a mass spectrometric experimentation[13] in the framework of the kinetic method.[35] A different trend is given by Lias et al.[12] for which cytosine (223.8 kcal/mol), adenine (223.5 kcal/mol), and guanine (223.0 kcal/mol) have almost the same PA value, whereas that of thymine (208.8 kcal/mol) remains the lowest of the series. Our individual absolute PA values are very close to the experimental data of Greco et al.[13] The maximum deviations (3.2 and 2.9 kcal/mol) occur in the case of cytosine and guanine, respectively. A possible source of error can be due to the existence, in a narrow range of energy, of other neutral tautomers for both these systems and, consequently, of other protonated forms.

However, the average error falls in the range of experimental inaccuracy.[13] Furthermore, we emphasize that our values are corrected for the ZPE and thermal contributions at 298 K, whereas, the temperature of the experimental measurements is not well defined. HF 4-31G[17] and 6-31G[19] results are sensibly overestimated. MP4//MP2[3] calculation for cytosine and guanine seems to reproduce accurately the experimental data, but, like the other *ab initio* values, does not include ZPE and (5/2)RT corrections. These contributions (see Table V for ZPE) are significant for a correct prediction of thermochemical properties including proton affinities as well as gas-phase basicities.

Starting from the calculated PA, we have computed, for the first time, the GB of DNA nucleic acid bases (see Table VI) at room temperature. The order of basicity in the gas phase is found to be:

$$\text{guanine} > \text{cytosine} > \text{adenine} \gg \text{thymine}$$

and differs from that in polar solvent.[8,11,14]

Conclusions

Density functional computations were performed to determine preferred protonation sits of DNA nucleic bases to evaluate their proton affinities and gas-phase basicities. Results support the following conclusions:

- Protonation occurs preferentially at O4 in thymine, N3 in cytosine, N1 in adenine, and N7 in guanine.
- The most favorable final product is always that in which a better redistribution of electron density is present.
- A provision about the preferred protonation site is not possible on the basis of individual relationships regarding the nature of HOMO, molecular electrostatic potential, or charge distribution. All these factors are necessary to explain the effects that stabilize the most stable protonated forms.
- The presence of intramolecular hydrogen bonds contributes to the stabilization of some protonated forms.
- The calculated gas-phase proton affinities are in good agreement with the most recent experimental data and have the same levels of accuracy as those generally obtained by the

TABLE VI.
Theoretical and Experimental Proton Affinity Values (kcal / mol) for DNA Bases.

	Method				
Molecule	NLSD[a]	HF / 4-31G[b]	6-31G[c]	MP4 / / MP2[d]	Exp.[e]
Thymine	208.8 (201.3)	211.8	209.0	/	209.0
Cytosine	229.1 (221.3)	249.2	238.9	227.0	225.9
Adenine	225.8 (217.9)	240.0	/	/	224.2
Guanine	230.3 (222.7)	244.7	/	225.8	227.4

[a] NLSD values are calculated at 298 K. Gas-phase basicity (kcal / mol) predictions are shown in parentheses.
[b] From ref. 17. [c] From ref. 18. [d] From ref. 3. [e] From ref. 12. Values from these references are not corrected for ZPE and thermal contribution.

more expansive Gaussian-2 method for smaller systems.

- The gas-phase basicity order is predicted theoretically for the first time.

Acknowledgment

The authors are grateful to the Direction des Sciences de la Matière/CEA–Grenoble for Cray C94 computer time.

References

1. S. M. Mirkin, *Annu. Rev. Biophys. Biomol. Struct.*, **24**, 319 (1995).
2. R. R. Sinden, *DNA Structure and Function*, Academic Press, San Diego, CA, 1994.
3. C. Colominas, F. J. Luque, and M. Orozco, *J. Am. Chem. Soc.*, **118**, 6811 (1996).
4. G. H. Roehrig, N. A. Oyler, and L. Adamowicz, *J. Phys. Chem.*, **99**, 14285 (1995).
5. A. Holmen and A. Broo, *Int. J. Quantum Chem. (Quantum Biol. Symp.)*, **22**, 113 (1995).
6. (a) L. Paglieri, G. Corongiu, and D. A. Estrin, *Int. J. Quantum Chem.*, **56**, 615 (1995); (b) D. A. Estrin, L. Paglieri, and G. Corongiu, *J. Phys. Chem.*, **98**, 5635 (1994).
7. I. R. Gould, N. A. Burton, R. J. Hall, and I. H. Hillier, *J. Mol. Struct. (Theochem)*, **331**, 147 (1995).
8. A. R. Katritzki and M. Karelson, *J. Am. Chem. Soc.*, **113**, 1561 (1991).
9. N. C. Gonnella, H. Nakanishi, J. B. Holtwick, D. S. Horowitz, K. Kanamori, N. J. Leonard, and J. D. Roberts, *J. Am. Chem. Soc.*, **105**, 2050 (1983).
10. R. D. Brown, P. D. Godfrey, D. McNaughton, and P. Pierlot, *Chem. Phys. Lett.*, **156**, 61 (1989), and references therein.
11. W. Saenger, *Principles of Nucleic Acid Structure*, Springer, New York, 1984.
12. S. G. Lias, J. F. Liebmann, and R. D. Levin, *J. Phys. Chem. Ref. Data*, **13**, 659 (1984).
13. F. Greco, A. Liguori, G. Sindona, and N. Uccella, *J. Am. Chem. Soc.*, **112**, 9092 (1990).
14. J. J. Christensen, J. H. Rytting, and R. M. Izatt, *J. Phys. Chem.*, **71**, 2700 (1967).
15. (a) R. Bonaccorsi, A. Pullman, E. Scrocco, and J. Tomasi, *Theor. Chim. Acta*, **24**, 51 (1972); (b) R. Bonaccorsi, E. Scrocco, J. Tomasi, and A. Pullman, *Theor. Chim. Acta*, **36**, 339 (1975).
16. (a) A. Pullman and M. Armbruster, *Theor. Chim. Acta*, **45**, 249 (1977); (b) P. G. Mezey, J. J. Ladik, and M. Barry, *Theor. Chim. Acta*, **54**, 251 (1984).
17. J. E. Del Bene, *J. Phys. Chem.*, **87**, 367 (1983).
18. J. J. Christensen, J. H. Rytting, and R. M. Izatt, *Biochemistry*, **9**, 4907 (1970), and references therein.
19. A. O. Colson, B. Besler, D. M. Close, and M. D. Sevilla, *J. Phys. Chem.*, **96**, 661 (1992).
20. N. Russo, Y. Abashkin, P. Calaminici, T. Mineva, E. Sicilia, and M. Toscano, *Recent Advances in Density Functional Methods*, P. D. Chong, Ed., World Scientific, Singapore, 1996, p. 335.
21. A. K. Chandra and A. Goursot, *J. Phys. Chem.*, **100**, 11596 (1996).
22. A. M. Schmedekamp, I. A. Topol, and C. J. Michejda, *Theor. Chim. Acta*, **92**, 83 (1995).
23. J. Andzelm and E. Wimmer, *J. Chem. Phys.*, **96**, 508 (1992).
24. S. H. Vosko, L. Wilk, and M. Nusair, *Can. J. Phys.*, **58**, 1200 (1980).
25. A. D. Becke, *Phys. Rev. A*, **38**, 3098 (1988).
26. J. P. Perdew, *Phys. Rev. B*, **33**, 8822 (1986).
27. N. Godbout, D. R. Salahub, J. Andzelm, and E. Wimmer, *Can. J. Chem.*, **70**, 560 (1992).
28. S. W. Benson, *Thermochemical Kinetics*, John Wiley & Sons, New York, 1976.
29. UNICHEM 3.0 (1994) Cray Research Inc., 2360 Pilot Knob Road, Mendota Heights, MN.
30. E. L. Stewart, C. K. Foley, N. L. Allinger, and J. P. Bowen, *J. Am. Chem. Soc.*, **116**, 7282 (1994).
31. L. Clowney, S. C. Shri, A. R. Srinivasan, J. Westbrook, W. K. Olson, and H. M. Berman, *J. Am. Chem. Soc.*, **118**, 509 (1996).
32. R. Taylor and O. Kennard, *J. Mol. Struct.*, **78**, 1 (1982).
33. L. A. Curtiss, K. Raghavachari, G. W. Trucks, and J. A. Pople, *J. Chem. Phys.*, **94**, 7221 (1991).
34. B. J. Smith and L. Radom, *Chem. Phys. Lett.*, **231**, 345 (1994).
35. R. G. Cooks, J. S. Patrick, T. Kotiaho, and S. A. McLuckey, *Mass Spectrom. Rev.*, **13**, 287 (1994).