THEORETICA CHIMICA ACTA

Theoret. Chim. Acta (Berl.) 54, 251–258 (1980)

Non-Empirical SCF MO Studies on the Protonation of Biopolymer Constituents

II. Protonation of Adenine, Guanine and Their Tautomeric Forms

Paul G. Mezey

Laboratory of the National Foundation for Cancer Research at the Chair of Theoretical Chemistry, University of Erlangen-Nürnberg, D-8520 Erlangen, Federal Republic of Germany, and Department of Chemistry and Chemical Engineering, University of Saskatchewan, Saskatoon, Canada S7N 0W0*

János J. Ladík and Michael Barry

Chair of Theoretical Chemistry at the Friedrich-Alexander-University Erlangen-Nürnberg, D-8520 Erlangen, Federal Republic of Germany, and Laboratory of the National Foundation for Cancer Research at the Chair of Theoretical Chemistry, University of Erlangen-Nürnberg, D-8520 Erlangen, Federal Republic of Germany

Proton affinities of a large number of tautomeric adenine and guanine structures have been calculated using the *ab initio* Hartree–Fock–Roothaan SCF method. For guanine, and to a lesser extent for adenine, it is likely that several protonated forms may co-exist in acidic solutions. Protonation at the "free" adenine and guanine positions in DNA may introduce effective acceptor levels between the energy bands of the polymer that may cause fundamental changes in the conduction properties of DNA.

Key words: Biopolymer constituents, protonation of ~ – Purine bases

1. Introduction

Protonation of nucleotide bases at various electronegative centres plays an important role in certain biochemical–biophysical processes [1]. A protonation followed by deprotonation at a different site leads to an actual tautomerization of the nucleotide base. While direct tautomerization of nucleotide bases as a cause of spontaneous mutation in DNA is probably less important [2, 3] than it has been thought previously, several tautomeric forms may have important roles in the replication process.

There are many possible tautomeric forms of both adenine and guanine and

* Permanent address

0040-5744/80/0054/0251/$01.60

protonation further increases the number of distinct structures. For many of the possible protonated forms experimental information is scarce or non-existent, although fluorescence studies carried out on protonated adenine in acidic solution indicated the presence of several tautomers [4]. While it is difficult to assess the solvent effect on various protonated forms properly, this finding suggests that the total energy difference between some of the most stable protonated adenine structures is probably small. Due to the experimental difficulties in observing and studying a large number of possible protonated tautomeric forms, theoretical techniques appear as the natural choice for studying the problem.

Protonation may also cause fundamental changes in the conductance properties of biopolymers. Protonated units in the polymer may introduce new energy levels in the band gap and these may become effective acceptor levels. The actual location of such new energy levels depends on the electronic structure of the biopolymer as a whole that may "readjust" to accommodate the extra positive charge. Nevertheless, the most important determining factor is the electronic structure of the protonated monomeric unit. Consequently, by calculating the energy level shifts of the monomeric unit due to protonation, an approximate model may be constructed for the analysis of the new acceptor levels in protonated biopolymers [5]. Using this simple model it was suggested that isomerization or conformational change in the side chains of a series of amino acids. (Arg, His, Ser, Tyr, Lys) do not alter the energy band structure of a polypeptide significantly. In contrast, side chain protonation may introduce effective acceptor levels according to earlier *ab initio* calculations using the above mentioned model [5].

In the present study a similar technique is used to investigate the effects of protonation of purine bases adenine and guanine, and of their various tautomeric forms. In addition, gas phase proton affinities are calculated for a large number of protonation reactions involving various tautomeric forms of adenine and guanine.

The protonation of purine bases in DNA is considered more important than protonation of pyrimidine bases since in the latter all locations that are primary targets for protonation are occupied either by hydrogen bonds or by the sugar group.

2. Method

Throughout this study the one-determinant Hartree–Fock–Roothaan LCAO SCF MO method [6] was used with the STO-3G (minimum) basis set [7]. Due to the large size of the protonated adenine and guanine molecules (protonated adenine and guanine have 16 and 17 atoms, respectively), geometry optimization was not attempted and standard or designed geometries were used throughout. With the exception of the immediate neighbourhood of the extra proton, the geometries were kept the same as in an earlier study on purine base tautomerization [3].

For protonated N or O moieties the same geometry changes were adopted as in an earlier study on amino acid side chain protonation [5]. In order to avoid reference to an excessive number of figures the following convention is used when referring to

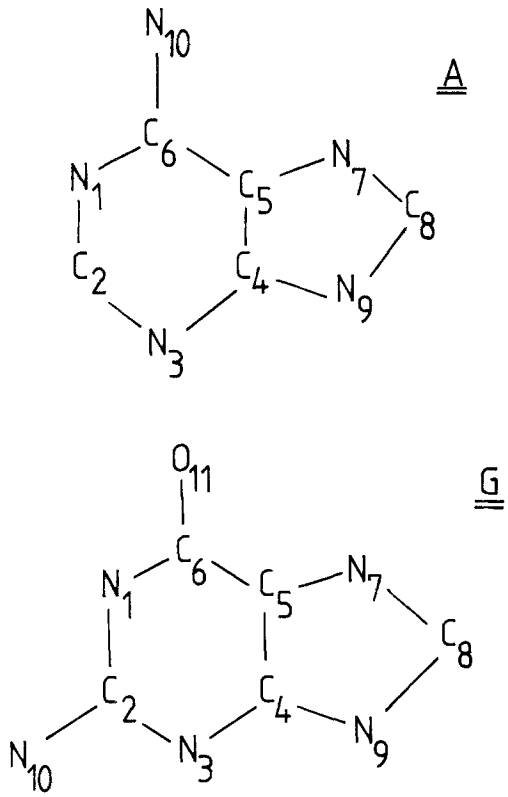

Fig. 1. A) Adenine skeleton; G) Guanine skeleton

various protonated purine bases: The atoms in the purine ring skeleton are numbered according to the usual convention by 1, . . . , 9 and numbers 10 and 11 are assigned to heavy atoms N and O outside the ring system, as shown in Fig. 1. These atoms form the adenine and guanine skeletons.

When referring to a given protonated tautomeric form, the serial numbers of the hydrogen bearing skeleton atoms are listed. Thus, the most common tautomeric (neutral) form of adenine is referred to as Adenine (2, 8, 9, 10, 10) or A(2, 8, 9, 10, 10). This tautomer is the same as form A1 of Ref. [3].

It should be mentioned that solvent interactions usually play a more important role for charged species than for neutral molecules and an error estimate that one may try to apply for the results of such minimum basis *ab initio* calculations is even less rigorous for positive ions than for the neutral tautomeric forms. We arbitrarily set 10 kcal/mole calculated energy difference between tautomeric protonated forms as the lower limit of significant difference, similarly to the limit applied in an earlier tautomerization study [3]. That is, no attempt has been made to draw chemical conclusion based on a calculated energy difference between protonated forms which is less than 10 kcal/mole. However, for the gas phase energy difference between a

protonated form and the neutral parent tautomer the above "limit of significance" appears somewhat excessive and smaller energy differences have also been subjects of cautious interpretations.

3. Result and Discussion

The calculated total molecular energies (in atomic units), the relative energies and protonation energies (in kcal/mole) together with HOMO and LUMO orbital energy levels (in eV) of various protonated forms of adenine and guanine are listed in Tables 1 and 2, respectively. The relative energies are given with respect to the most stable protonated form, that is, form A(1, 2, 8, 9, 10, 10) for adenine and G(7, 8, 9, 10, 10, 11) for guanine. A given protonated form may be obtained by protonation from several neutral tautomeric forms as reflected by multiple entries under $E_{\text{prot.}}$. The reference to the actual neutral "parent" tautomer is given in parentheses after each protonation energy value. For the neutral forms these numbers represent proton affinities and in each row the proton affinity of the most stable neutral tautomer is listed first. If the protonation of a less stable neutral form results in the same ion, the proton affinity is larger, as shown by subsequent numbers.

Form A(1, 2, 8, 9, 10, 10) is clearly preferred among the protonated adenines, although the relative energies of four different forms listed subsequently in Table 1 are only a few kcal/mole above the arbitrarily set 10 kcal/mole "limit of significant difference". All the other eight protonated adenines are of much higher relative energy and are not expected to be present in acidic solutions in any significant concentrations. In general, forms with $-NH_3$ and $=NH$ groups are the least stable, while protonation at one of the nitrogen atoms of the six numbered rings is preferred. The calculated energy values and the uncertainties involved in their calculation suggest that in addition to the most stable form, forms

A(2, 3, 8, 9, 10, 10)

A(2, 3, 7, 8, 10, 10)

A(1, 2, 7, 8, 10, 10)

and

A(2, 7, 8, 9, 10, 10)

may be sufficiently stabilized by favourable solvent effects to have non-negligible concentrations in a biochemical system.

In contrast to the adenine results, there is no clearly preferred protonated guanine structure, although protonation at the oxygen atom is preferred. According to the calculations the most stable form is

G(7, 8, 9, 10, 10, 11),

Table 1. Total energies (E_t), relative energies[a] (E_{rel}), HOMO and LUMO energies and protonation energies[b] (E_{prot}) of various adenine structures

Protonated form	E_t(a.u.)	E_{rel}(kcal/mole)	E_{HOMO}(eV)	E_{LUMO}(eV)	E_{prot}(kcal/mole)		
A(1, 2, 8, 9, 10, 10)	−459.06484	0.00	−11.939	0.059	286.62(2, 8, 9, 10, 10) [−5.499[c], −6.323[d]]	316.47(1, 2, 8, 10, 10) [−5.924, −5.314]	324.33(1, 2, 8, 9, 10) [−6.446, −5.677]
A(2, 3, 8, 9, 10, 10)	−459.04322	13.57	−12.208	−0.170	273.05(2, 8, 9, 10, 10) [−5.769, −6.552]	299.25(2, 3, 8, 10, 10) [−6.056, −4.949]	321.36(2, 3, 8, 9, 10) [−6.784, −6.346]
A(2, 7, 8, 9, 10, 10)	−459.03640	17.85	−11.709	−0.204	268.77(2, 8, 9, 10, 10) [−5.270, −6.586]	280.48(2, 7, 8, 10, 10) [−4.979, −6.226]	
A(1, 2, 7, 8, 10, 10)	−459.04089	15.03	−12.643	−0.481	283.30(2, 7, 8, 10, 10) [−5.913, −6.503]	301.43(1, 2, 8, 10, 10) [−6.628, −5.854]	308.31(1, 2, 3, 8, 10) [−6.948, −6.342]
A(2, 3, 7, 8, 10, 10)	−459.04154	14.63	−12.835	−0.330	283.71(2, 7, 8, 10, 10) [−6.105, −6.352]	298.19(2, 3, 8, 10, 10) [−6.683, −5.109]	310.25(2, 3, 7, 8, 10) [−7.322, −6.344]
A(2, 7, 8, 10, 10, 10)	−458.96837	60.56	−12.242	−0.001	237.78(2, 7, 8, 10, 10) [−5.512, −6.023]		
A(1, 2, 3, 8, 10, 10)	−458.99509	43.78	−11.348	−0.885	269.04(2, 3, 8, 10, 10) [−5.196, −5.664]	272.68(1, 2, 8, 10, 10) [−5.333, −6.258]	
A(1, 2, 8, 10, 10, 10)	−458.90651	99.39	−11.026	−1.166	217.08(1, 2, 8, 10, 10) [−5.011, −6.539]		
A(2, 3, 8, 10, 10, 10)	−458.94174	77.27	−11.268	−1.248	235.55(2, 3, 8, 10, 10) [−5.116, −6.027]		
A(1, 2, 3, 8, 9, 10)	−458.99645	42.93	−11.102	−1.015	281.40(1, 2, 8, 9, 10) [−5.609, −6.751]	292.00(2, 3, 8, 9, 10) [−5.678, −7.191]	
A(2, 3, 7, 8, 9, 10)	−458.97509	56.34	−10.313	0.169	268.53(2, 3, 7, 8, 10) [−4.800, −5.845]	278.59(2, 3, 8, 9, 10) [−4.889, −6.007]	
A(1, 2, 3, 7, 8, 10)	−458.99539	43.59	−11.173	−0.736	279.75(1, 2, 7, 8, 10) [−5.478, −6.597]	281.28(2, 3, 7, 8, 10) [−5.660, −6.750]	
A(1, 2, 7, 8, 9, 10)	−458.99334	44.88	−10.514	0.275	278.46(1, 2, 7, 8, 10) [−4.819, −5.586]	279.45(1, 2, 8, 9, 10) [−5.021, −5.461]	

[a] Relative to the most stable form. [b] With respect to neutral tautomeric form specified in parentheses.
[c] The first number given in square brackets under each protonation energy is the change of HOMO energy level (in eV) with respect to the neutral form.
[d] The second number given in square bracket is the change of LUMO energy level (in eV).

Table 2. Total energies (E_t), relative energies[a] (E_{rel}), HOMO and LUMO energies and protonation energies[b] (E_{prot}) of various protonated guanine structures

Protonated form	E_t(a.u.)	E_{rel}(kcal/mole)	E_{HOMO}(eV)	E_{LUMO}(eV)
G(1, 7, 8, 9, 10, 10)	−532.89876	5.62	−10.590	0.787
G(1, 3, 7, 8, 10, 10)	−532.88494	14.30	−11.484	0.866
G(1, 8, 9, 10, 10, 11)	−532.90039	4.60	−11.315	0.374
G(3, 8, 9, 10, 10, 11)	−532.88809	12.32	−11.976	0.105
G(7, 8, 9, 10, 10, 11)	−532.90772	0.00	−11.247	0.093
G(8, 9, 10, 10, 11, 11)	−532.81268	59.66	−9.057	0.672
G(1, 7, 8, 10, 10, 11)	−532.87742	19.02	−11.811	0.757
G(3, 7, 8, 10, 10, 11)	−532.89886	5.56	−12.410	0.274
G(7, 8, 10, 10, 11, 11)	−532.78068	79.74	−10.775	0.460
G(1, 3, 8, 9, 10, 11)	−532.84565	38.96	−11.310	0.644
G(1, 3, 7, 8, 10, 11)	−532.83209	47.47	−11.565	−1.024

[a] Relative to the most stable form.
[b] With respect to neutral tautomeric form specified in parentheses.
[c] The first number given in square brackets under each protonation energy is the change of HOMO energy level (in eV) with respect to the neutral form.
[d] The second number given in square brackets is the change of LUMO energy level (in eV).

however, there are three other forms,

G(1, 8, 9, 10, 10, 11)
G(3, 7, 8, 10, 10, 11)

and

G(1, 7, 8, 9, 10, 10)

within a 10 kcal/mole range of the most stable form, and three further forms within a 20 kcal/mole range:

G(3, 8, 9, 10, 10, 11)
G(1, 3, 7, 8, 10, 10)

and

G(1, 7, 8, 10, 10, 11).

E_{prot}(kcal/mole)			
293.23(1, 8, 9, 10, 10)	297.20(1, 7, 8, 10, 10)		
[−5.286[c], −6.017[d]]	[−5.022, −5.706]		
288.53(1, 7, 8, 10, 10)	295.21(1, 3, 7, 8, 10)	303.94(3, 7, 8, 10, 10)	
[−5.916, −5.627]	[−4.995, −5.137]	[−5.799, −5.712]	
282.55(8, 9, 10, 10, 11)	294.26(1, 8, 9, 10, 10)	321.89(1, 8, 10, 10, 11)	342.65(1, 8, 9, 10, 11)
[−5.241, −5.837]	[−6.011, −6.430]	[−5.832, −4.610]	[−6.440, −4.500]
274.83(8, 9, 10, 10, 11)	297.85(3, 8, 10, 10, 11)	312.85(3, 8, 9, 10, 11)	315.16(3, 8, 9, 10, 10)
[−5.902, −6.106]	[−5.929, −4.843]	[−6.523, −5.754]	[−6.432, −7.103]
287.15(8, 9, 10, 10, 11)	292.57(7, 8, 10, 10, 11)		
[−5.173, −6.118]	[−4.876, −5.708]		
227.49(8, 9, 10, 10, 11)			
[−2.983, −5.539]			
273.55(7, 8, 10, 10, 11)	283.81(1, 7, 8, 10, 10)	307.47(1, 8, 10, 10, 11)	335.59(1, 7, 8, 10, 11)
[−5.440, −5.044]	[−6.243, −5.736]	[−6.328, −4.227]	[−6.934, −3.754]
287.01(7, 8, 10, 10, 11)	304.61(3, 8, 10, 10, 11)	312.67(3, 7, 8, 10, 10)	314.75(3, 7, 8, 10, 11)
[−6.039, −5.527]	[−6.363, −4.674]	[−6.725, −6.304]	[−7.003, −4.994]
212.83(7, 8, 10, 10, 11)			
[−4.404, 5.341]			
274.64(1, 3, 8, 9, 10)	286.21(3, 8, 9, 10, 11)	295.48(1, 3, 8, 10, 11)	308.29(1, 8, 9, 10, 11)
[−4.968, −6.154)	[−5.857, −5.215]	[−5.457, −3.790]	[−6.435, −4.230]
262.04(1, 3, 7, 8, 10)	272.83(3, 7, 8, 10, 11)	286.96(1, 3, 8, 10, 11)	307.13(1, 7, 8, 10, 11)
[−5.076, −7.027]	[−6.158, −6.292]	[−5.712, −5.458]	[−6.688, −5.535]

These results suggest that in actual acidic biochemical systems there is likely to coexist a group of diverse forms of protonated guanines.

In general, protonated guanine structures containing —OH_2 or =NH groups are the least stable. For both adenine and guanine positive ions structures with hydrogen atoms in both 1 and 3 positions of the six-membered ring are also somewhat disfavoured.

The calculated highest occupied and lowest unoccupied (HOMO, LUMO) orbital energy levels are also listed in Tables 1 and 2. Both the E_{HOMO} and E_{LUMO} values show rather large variations from tautomer to tautomer.

Protonation energies for several protonation processes and the accompanying changes in the HOMO and LUMO orbital energy levels have been calculated. The relevant data for the neutral tautomeric forms (specified in parentheses after each protonation energy value in Tables 1 and 2) have been taken from an earlier study [3]. As expected, the most stable protonated forms may be obtained by direct protonation from the most stable neutral tautomeric forms. All neutral forms of adenine, and nine neutral forms of guanine, may be converted into one of the first four most stable positive ions by direct protonation, i.e. without an additional tautomerization process. The calculated proton affinity values for these most likely protonation processes spread over a rather wide range. It is expected that this

interval could be reduced somewhat by full geometry optimization for all the structures involved, however, the general trends are not expected to change.

Also listed in Tables 1 and 2 are the HOMO and LUMO energy level changes associated with the most likely protonation processes. Most shifts in these MO levels are of the order of 0.2 a.u. $\sim$ 5–6 eV. Particularly important are those protonation processes that correspond to protonation in positions 3 or 7. These positions are "free" for both adenine and guanine, built in a DNA chain. Protonation in positions 1, 10 and 11 would, on the other hand, interfere with the existing H-bond system. The 5–6 eV value indicates that, assuming no major readjustment of the band structure of the polymer, a protonation process may introduce new, discrete energy levels in the band gap, several eV above the upper edge of the valence band. In earlier studies [8] with a different basis on the energy band structure of homopolynucleotides, polyadenine and polyguanine, the upper edge of the valence band was calculated as -8.767 eV and -7.398 eV, respectively, while the lower edge of the conduction band was obtained as 2.815 eV and 3.679 eV, respectively. The 5–6 eV lowering of the LUMO levels in the protonated forms, according to the present results, would generate new levels approximately half way within the band gap. These new energy levels may serve as effective acceptor levels in DNA. The reorganization of the band structure may possibly lead to semiconductive properties by producing holes in the valence band.

Acknowledgement. The authors express their thanks to the "Fond der Chemischen Industrie" and to the Natural Sciences and Engineering Research Council of Canada for continuing financial support.

References

1. Ulbricht, T. L. V.: Purines, pyrimidines and nucleotides and the chemistry of nucleic acids. New York: Pergamon Press 1964
2. Goddard, J. D., Mezey, P. G., Csizmadia, I. G.: Theoret. Chim. Acta (Berl.) **39**, 1 (1975)
3. Mezey, P. G., Ladik, J. J.: Theoret. Chim. Acta (Berl.), in press
4. Boerresen, H. C.: Acta Chim. Scand. **21**, 2463 (1967)
5. Mezey, P. G., Ladik, J. J., Suhai, S.: Theoret. Chim. Acta (Berl.) **51**, 323 (1979)
6. Roothaan, C. C. J.: Rev. Mod. Phys. **23**, 69 (1951)
7. Hehre, W., Stewart, R. F., Pople, J. A.: J. Chem. Phys. **51**, 2657 (1969)
8. Ladik, J. J.: Proceedings of the Ciba Symposium on "Electronic Theory of Cancer", London, 1978 (in press)

Received August 27, 1979