

## A Non-Empirical Molecular Orbital Study on the Relative Stabilities of Adenine and Guanine Tautomers

Paul G. Mezey

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

Janos J. Ladik

Chair of Theoretical Chemistry, University of Erlangen-Nürnberg, Germany, and  
Laboratory of the National Foundation for Cancer Research  
at the Chair of Theoretical Chemistry, University of Erlangen-Nürnberg,  
Egerlandstraße 3, D-8520 Erlangen, Federal Republic of Germany

The relative stabilities of a series of adenine and guanine tautomers have been calculated using an *ab initio* Hartree-Fock-Roothaan SCF MO method. The calculated relative stabilities agree in general with the results of earlier semi-empirical studies. According to the present study, tautomeric forms with regular Kekulé structure for the six-membered purine ring are the most stable. The amine-imine tautomerization of purine bases is not likely to be responsible for spontaneous mutations in DNA.

**Key words:** Adenine tautomers, relative stabilities of ~ – Guanine tautomers, relative stabilities of ~

### 1. Introduction

The electronic structure and relative stabilities of biological purines and pyrimidines play an important role in determining the physical properties and chemical reactions of some of the most fundamental biopolymers. The base components of RNA and DNA may participate in various weak bonding interactions as well as in hydrogen exchange reactions, that may alter the properties of the biopolymer significantly [1]. One of the simplest of such hydrogen transfer reactions is tautomerization of the nucleotide base. While the number of possible tautomeric forms for pyrimidine bases (not including forms where hydrogen atoms attached to a carbon are involved in the tautomerization) is relatively small, six for both cytosine and thymine [2], this number is generally larger for purines such as adenine (8 forms) and guanine (15 forms). Some of the tautomerization processes possible for the free base may not occur directly in the biopolymer without disruption of the molecular chain; nevertheless, the relative stabilities of various possible tautomeric forms of purines and pyrimidines are clearly of some interest.

The importance of quantum chemical techniques in studying the molecular and electronic structures of nucleotide bases has been recognized for some time [3]. Due to the large number of possible tautomeric forms, and the experimental difficulties in observing and studying "rare" forms, theoretical methods appear particularly suitable for studying the problem.

"Rare" tautomeric forms are postulated to have particular importance in biological mutation processes. The rate of spontaneous biological mutation is estimated as about  $10^{-5}$  per nucleotide base per generation [4], that corresponds to ca. 6 kcal/mole free energy difference between the "normal" and "mutant" structures. According to the Crick-Watson theory [5] the spontaneous biological mutations may be attributed to the presence in DNA of nucleotide bases in "abnormal" tautomeric forms. This could result in irregularities in the replication process, e.g. the imine form of adenine would form a hydrogen-bonded pair with cytosine instead of the normal coupling with thymine. The estimated ca. 6 kcal/mole difference quoted above may very well correspond to stability differences of the tautomeric forms.

Most IR [6, 7] and UV [7–11] spectroscopic studies, as well as X-ray crystallographic investigations [12, 13] indicate that nucleotide base components exist predominantly in amino and keto forms, however, various other forms have also been suggested on the basis of experimental (spectroscopic) investigations [14–16].

The electronic absorption spectra of cytosine and thymine have been recorded in various solvents and an attempt has been made to correlate these experimental results with the theoretical spectra calculated using semiempirical PPP CI MO methods [17, 18]. Based on these studies one could not identify unambiguously the tautomer present in the solution, nor could one assign relative abundances to the various tautomeric forms. A more recent study on the photoelectron spectra of biological pyrimidines, however, indicated little direct evidence for the presence of more than one tautomeric form [19].

The 1H-purine-6-amine tautomeric form of adenine and the 2-amino-1,7-dihydro-6H-purine-6-on form of guanine (these forms are denoted A1 and G1 in the present paper, respectively) are considered the most stable tautomers in most solvents. Nevertheless, a comparison of the electronic spectra of the purine bases in water solution [10] and in the vapour phase [20] did not exclude the possibility of the presence of more than one tautomeric form in the solution. Fluorescence studies carried out in acidic solution on protonated adenine did indicate the presence of more than one tautomeric form [21].

The electronic structure of biological purines and pyrimidines has been the subject of several theoretical studies, using both semiempirical [3, 17, 18, 19a, 19b, 22–32] and *ab initio* [2, 33–37] molecular orbital methods. Following the suggestion of Szent-Györgyi [38] that the cancerous state is a reversal to a more primitive state of the cell, and this reversal is due to a change in the conduction properties of proteins and possibly of DNA, detailed theoretical studies have been carried out on the energy band structure and conduction properties of polynucleotides [24, 32, 36, 37, 39–41] and polypeptides [32].

Starting from the first Hückel calculations of B. and A. Pullman [19a] several semiempirical studies have given some insight into the problem of relative stabilities of rare tautomeric forms of nucleotide bases [19a, 19b, 23, 25, 27, 29], but *ab initio* investigations have been carried out only for the tautomerism of pyrimidine bases cytosine and thymine [2].

In the present paper we report the first non-empirical MO study on the tautomerism of two purine bases: adenine and guanine.

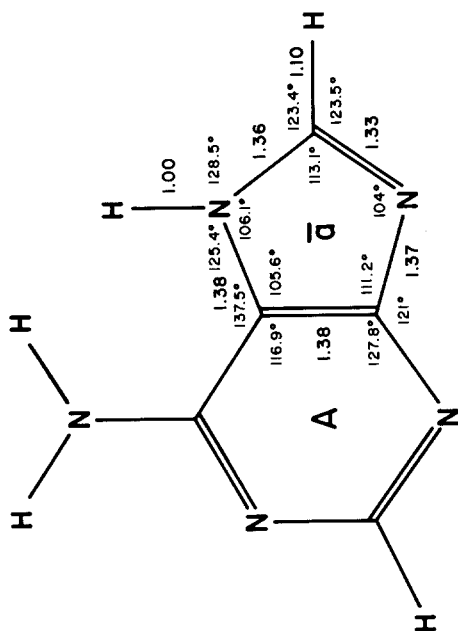
## 2. Method

Throughout this study Roothaan's restricted Hartree-Fock method was applied [42], using a modified version of the Gaussian 70 program system [43] and a set of small  $6^s3^p$  Gaussian basis functions contracted to a minimum basis [44]. Because of the rather large size of the molecules (adenine: 15 atoms, 70 electrons; guanine: 16 atoms, 78 electrons), geometry optimization was clearly impractical. With the exception of the tautomeric forms A1 and G1 (observed in crystal lattice), where experimental X-ray geometries were used [13], the geometries were designed on the basis of chemical analogies and standard geometrical parameters. The assumed geometrical features are discussed in some detail in the first part of the study on nucleotide base tautomerism [2], where a similar technique was used in designing geometries for pyrimidine base tautomers.

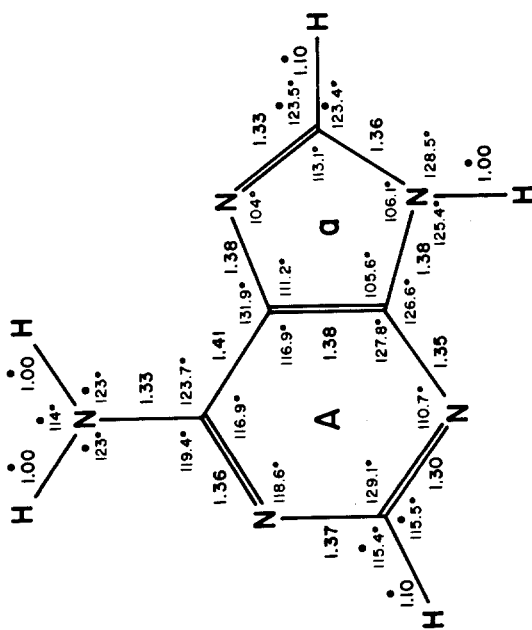
Any attempt for the estimation of the error of energy differences obtained in such calculations necessarily contains some arbitrary elements. In the present study we were concerned with three main sources of possible errors:

- 1) Neglect of interactions with the solvent. (The degree of interaction is likely to vary with the tautomer present.)
- 2) The lack of geometry optimization.
- 3) Inherent inaccuracy of the single determinant,  $6^s3^p$  basis *ab initio* SCF MO calculations.

We assume that variations in the solvent effect from tautomer to tautomer do not exceed the energy of two hydrogen bonds. The choice of geometry for the "rare" forms seems to be the most arbitrary element of the present study. It is noteworthy, however, that according to a preliminary study on formamide tautomerism the designed, chemically intuitive structures and the fully geometry optimized ones gave an energy difference constant to within approximately 1 kcal/mole [2]. Assuming a comparable success in estimating the individual geometrical parameters in the present case, one may expect an increase of the error roughly proportional to the size of the molecules. The third type of error is the most fundamental and difficult to estimate, however, past experience with similar calculations indicates that this error is of the same order of magnitude as the first two types. Based on the above points, but still arbitrarily, we set 10 kcal/mole as the lower limit of significant difference between calculated energies, and no attempt was made to draw chemical conclusions based on a calculated energy difference less than 10 kcal/mole. The tautomeric forms considered for adenine (A1...A8) and guanine (G1...G17) are shown in Figs. 1 to 6, where all relevant geometrical



A2



A1

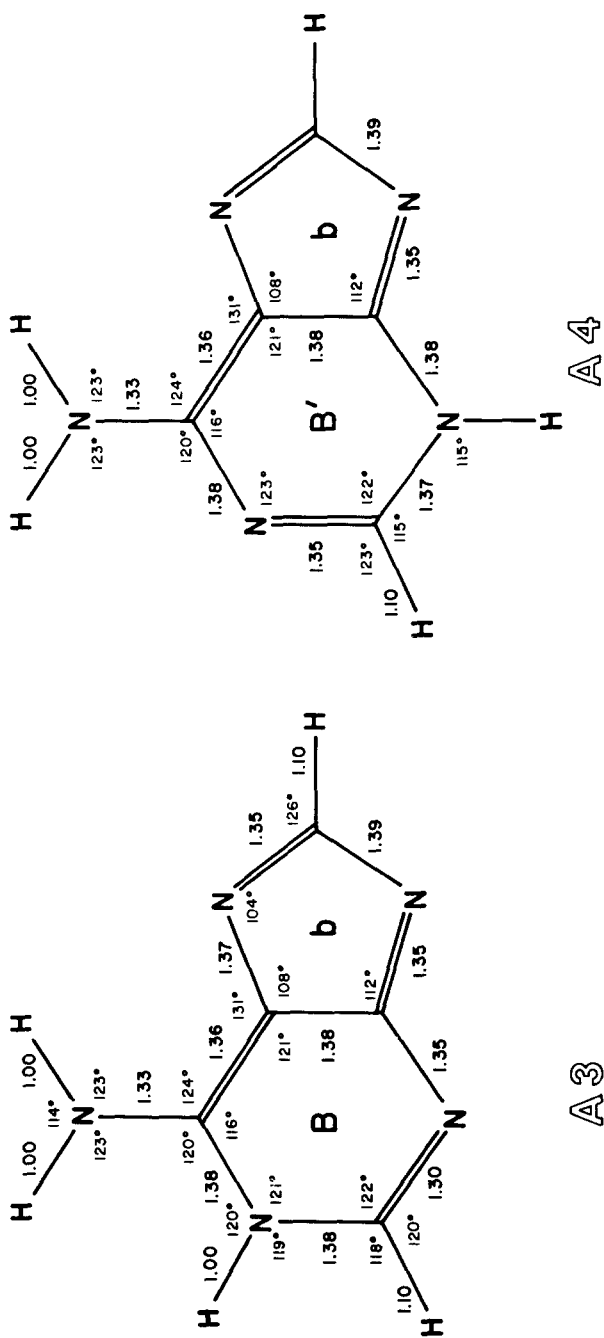
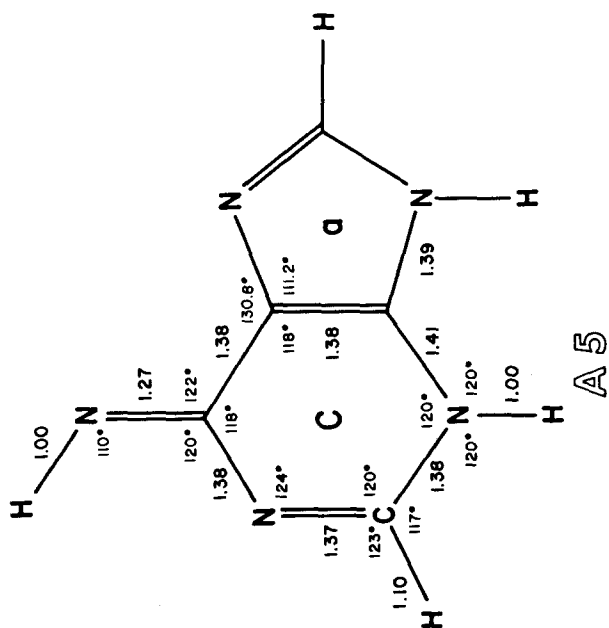
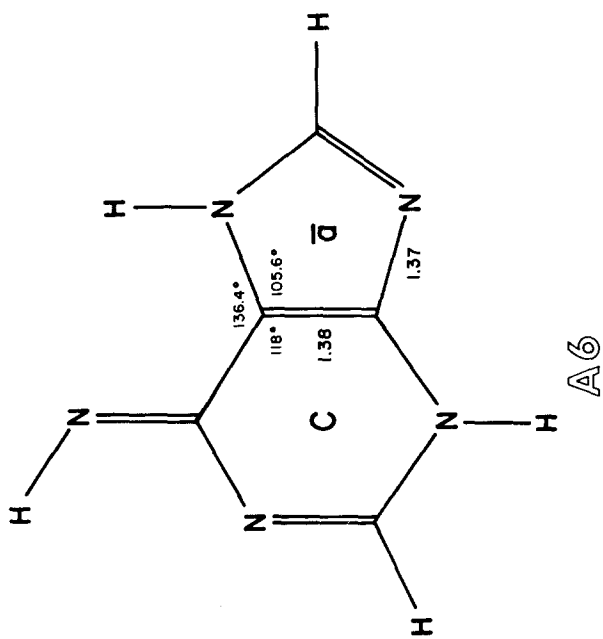


Fig. 1. Geometries of adenine tautomeric forms A1...A4 (bond lengths in Å, bond angles in degrees)



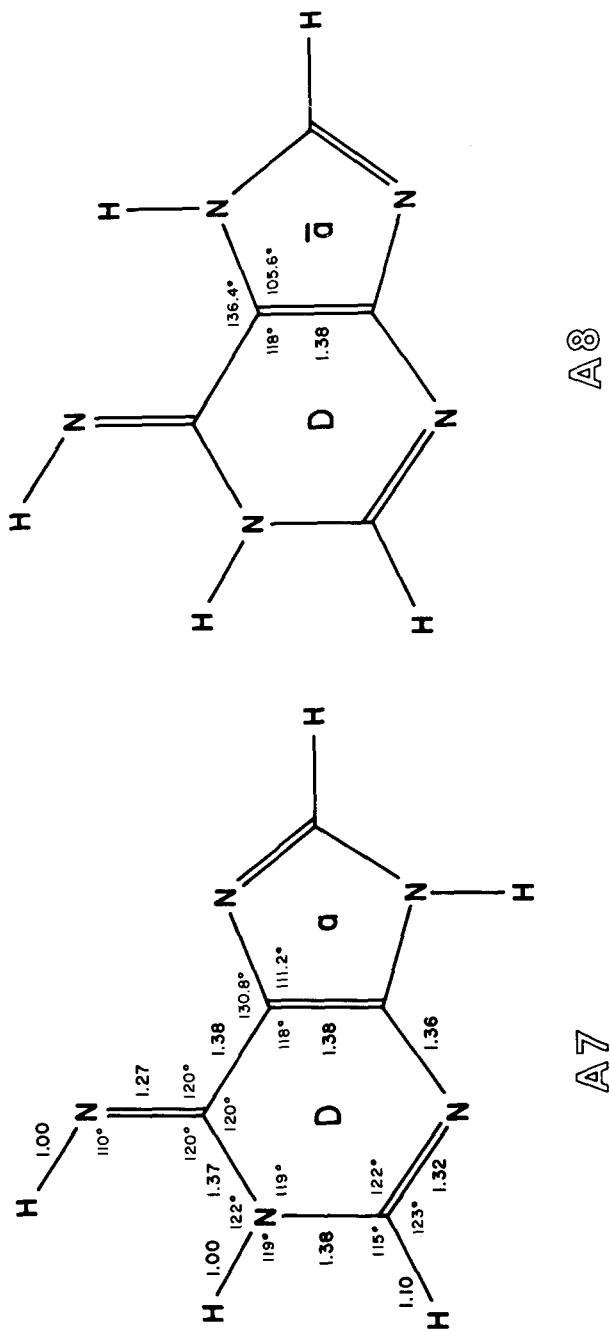


Fig. 2. Geometries of adenine tautomeric forms A5 ··· A8 (bond lengths in Å, bond angles in degrees)

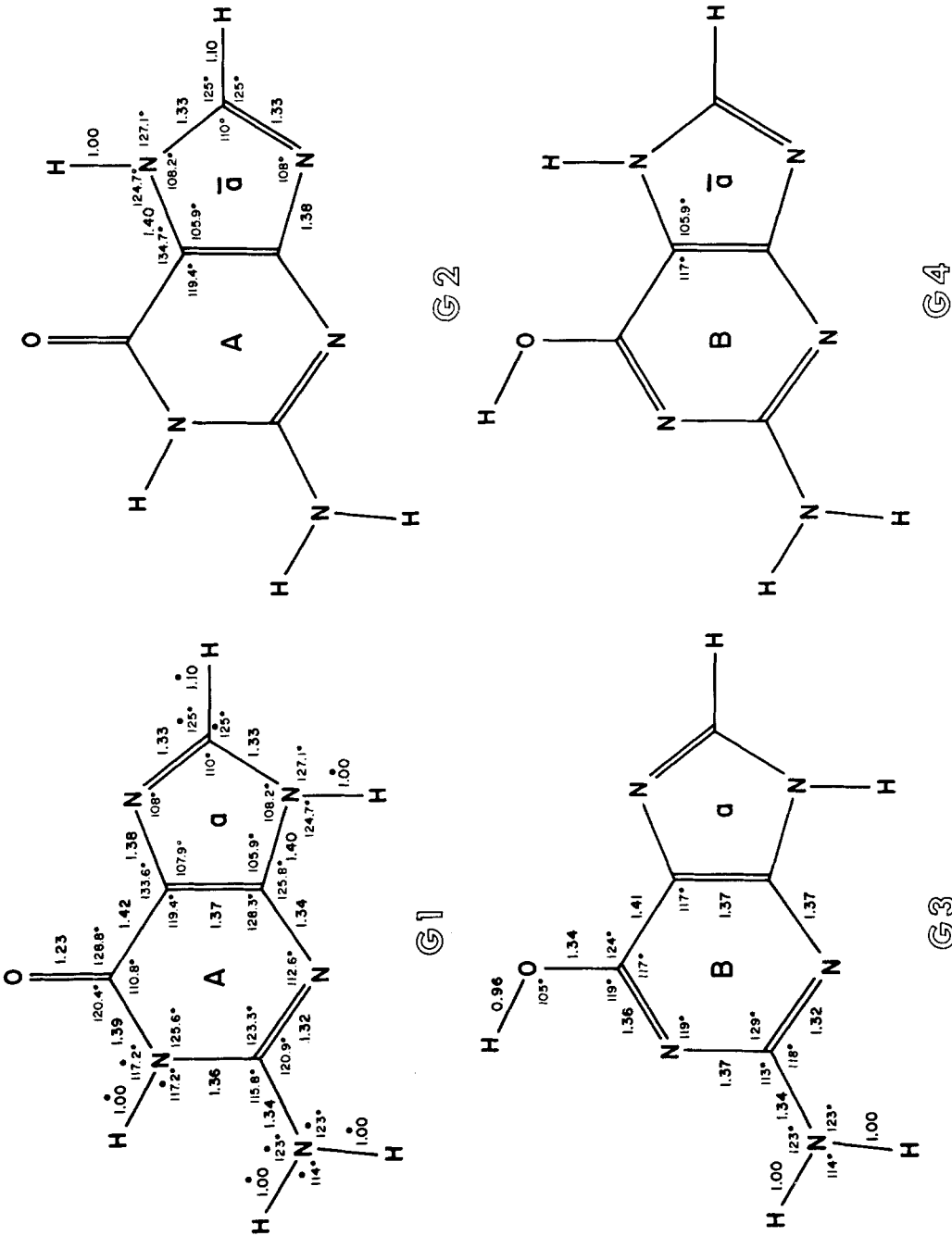


Fig. 3. Geometries of guanine tautomeric forms G1...G4 (bond lengths in Å, bond angles in degrees)



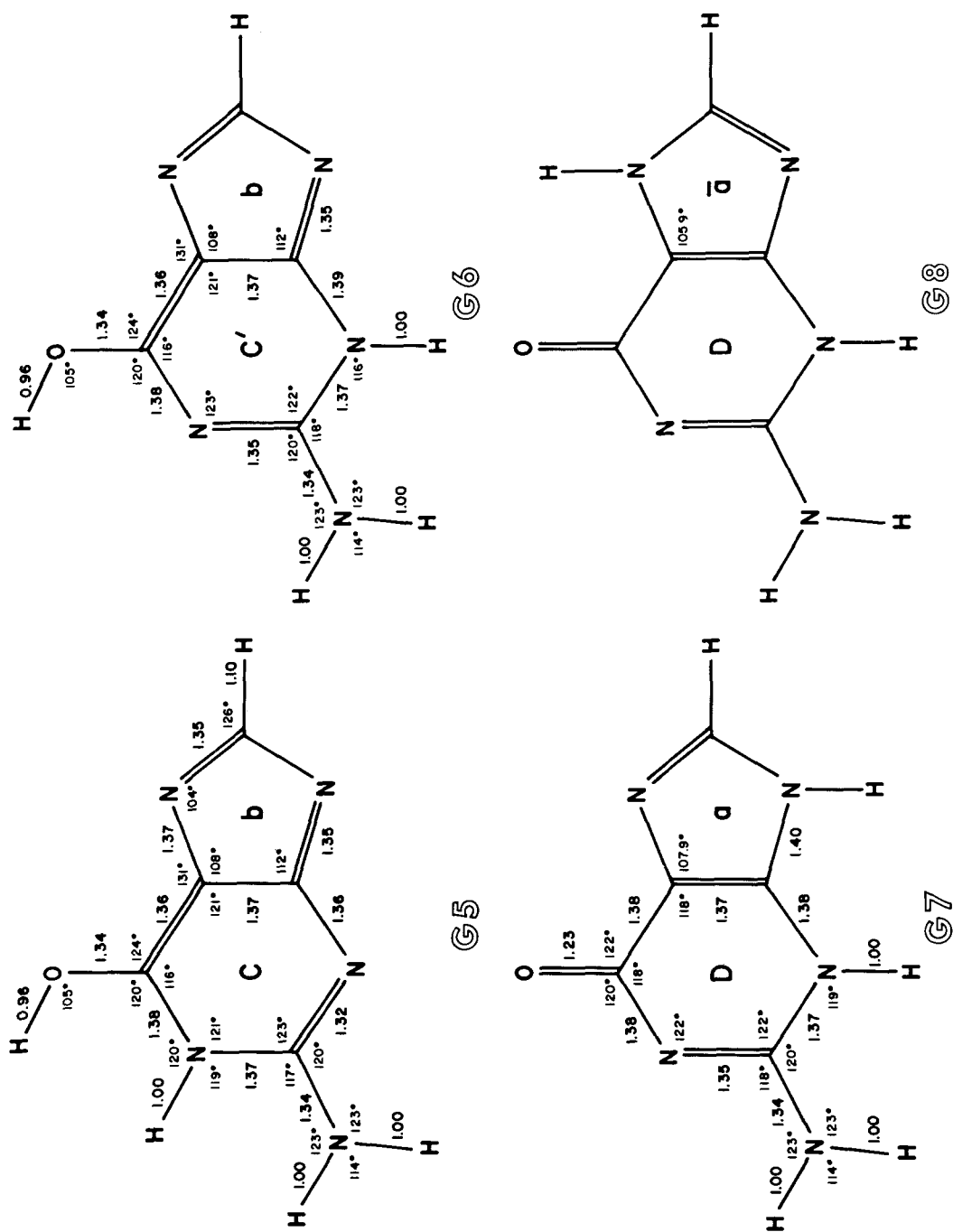


Fig. 4. Geometries of guanine tautomeric forms G5...G8 (bond lengths in Å, bond angles in degrees)

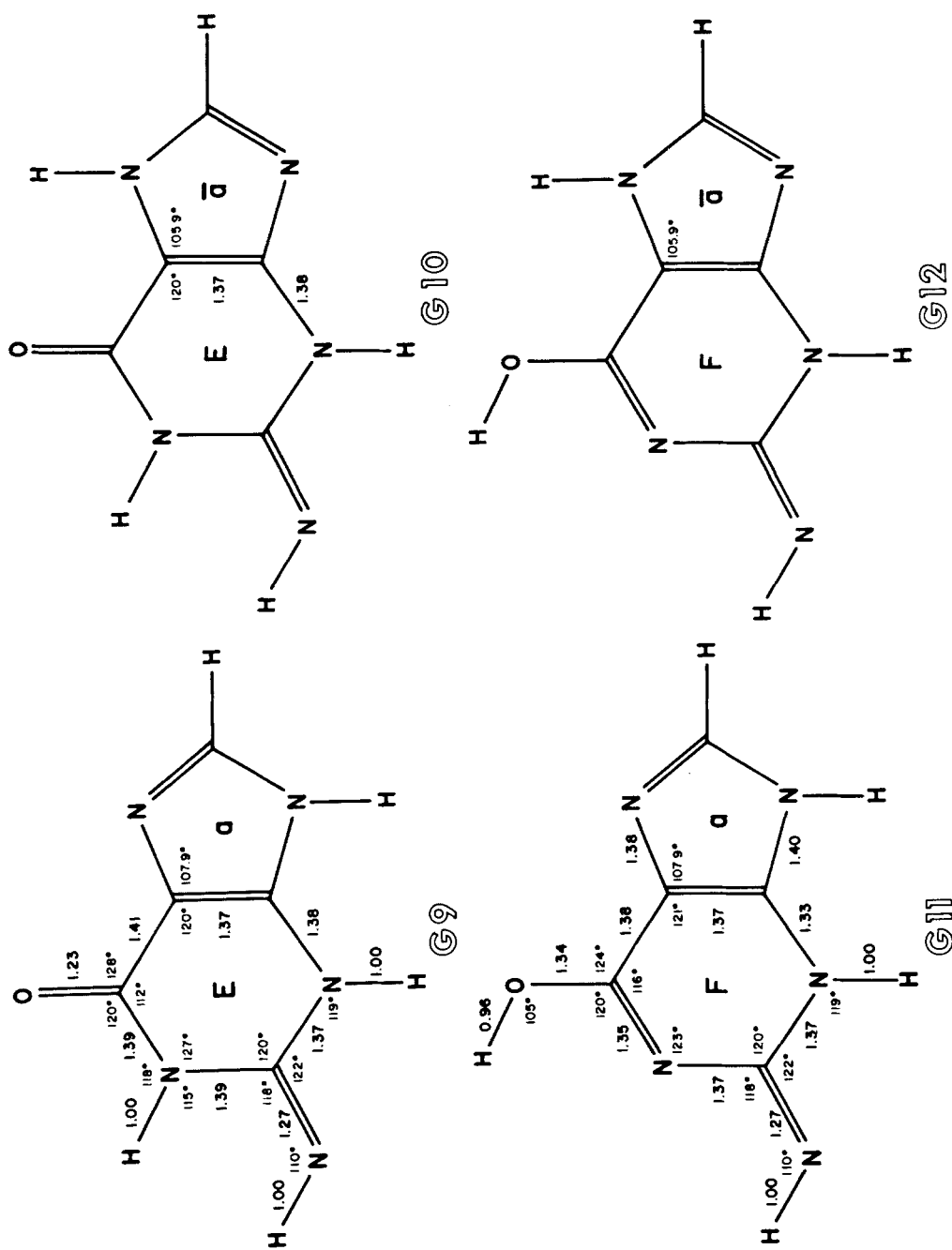
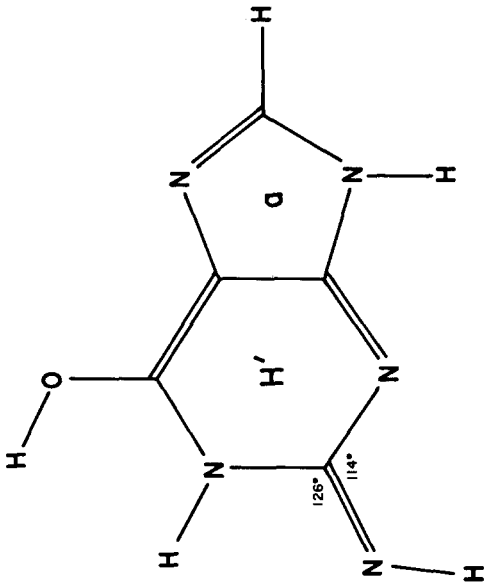
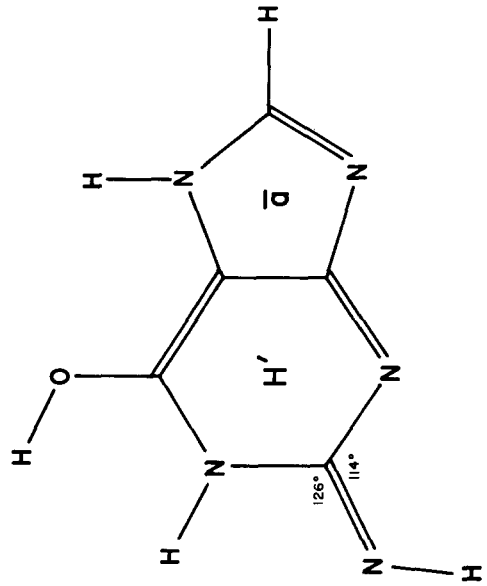


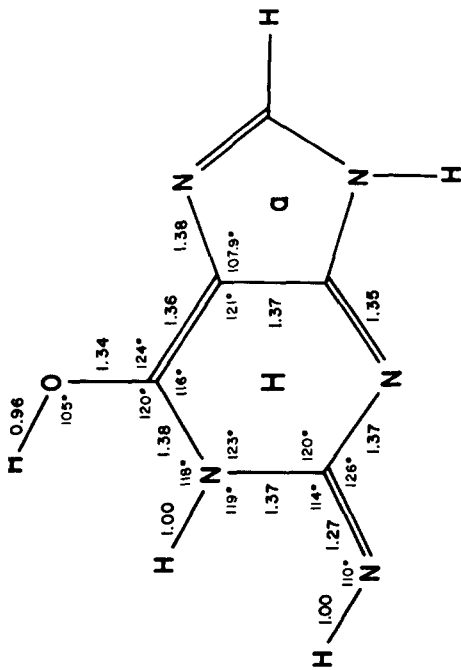
Fig. 5. Geometries of guanine tautomeric forms G9–G12 (bond lengths in Å, bond angles in degrees)



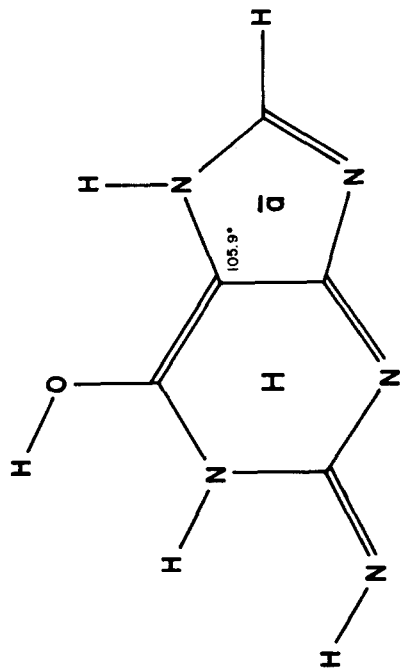
G14



G16



G13



G15

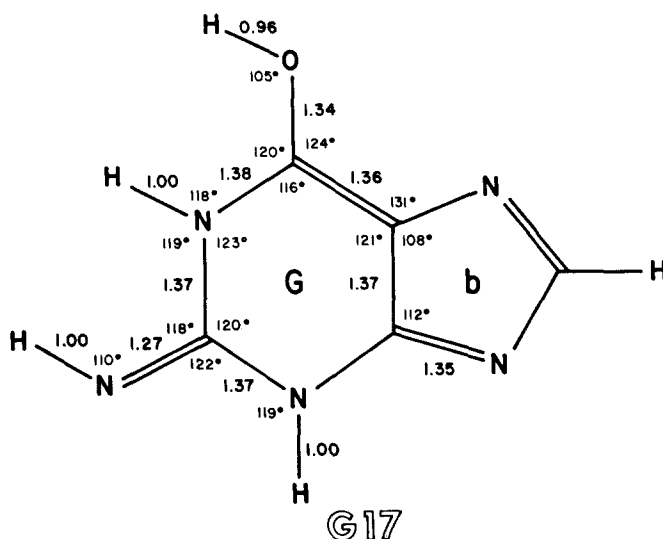


Fig. 6. Geometries of guanine tautomeric forms G13–G17 (bond lengths in Å, bond angles in degrees)

parameters are also given. No tautomers involving tautomerization reactions of C—H bonds were considered. Even for tautomeric forms A1 and G1, which correspond to those observed in the crystalline state [13], the experimental information had to be augmented with assumed geometrical parameters. The dots above the parameters in structures A1 and G1 denote these assumed values. The six-membered rings are denoted by capital letters A · · · H while the five-membered rings are denoted by lower case letters *a* and *b*. The bar in symbol  $\bar{a}$  means “inverted *a* ring”, i.e. tautomeric forms where the N(7) and N(9) moieties are interchanged with respect to forms containing rings of the *a* type. Forms G14 and G16 are not separate tautomeric forms, rather conformers of tautomeric forms G13 and G15, respectively, derived from the latter ones by 180° rotation or in-plane inversion of the imine group.

### 3. Results and Discussion

The calculated total electronic, nuclear repulsion, and total energies in hartree, the tautomerization energies relative to the most stable tautomer in kcal/mole and the first ionization potentials in eV, as calculated via Koopmans’ theorem are presented in Tables 1 and 2 for adenine and guanine, respectively.

According to the present calculations tautomeric form A1 of adenine has the lowest total energy, in agreement with most experimental results. Tautomeric form A2, with a hydrogen in the N(7) position instead of the N(9) position, has the second-lowest total energy, the tautomerization energy being 12 kcal/mole. It is

noteworthy that the total energy of the most stable of the four imine forms considered, A8, is about 37 kcal/mole higher than that of form A1. Although complete geometry optimization for these tautomeric forms could decrease the energy difference considerably, it is significant that the difference is much larger than 6 kcal/mole, the estimated free energy change associated with spontaneous mutation. This indicates that an amine-imine tautomerization is not the likely process primarily responsible for spontaneous mutations.

The total energy of the third most stable adenine tautomer, A4, is 26 kcal/mole higher than that of A1, and all other forms are of much higher energy. This

**Table 1.** Energies and ionization potentials of adenine tautomers

Tautomer	Total electronic energy (a.u.)	Nuclear repulsion energy (a.u.)	Total energy (a.u.)	Relative energy (kcal/mole)	Ionization potential (eV)
A1	-961.24743	502.63921	-458.60822	0.00	6.439
A2	-961.88580	503.29623	-458.58957	11.70	6.731
A3	-962.66721	504.10653	-458.56068	29.83	6.016
A4	-960.32514	501.75865	-458.56649	26.81	6.152
A5	-959.71675	501.18548	-458.53127	48.29	5.425
A6	-960.35454	501.80726	-458.54729	38.23	5.514
A7	-963.53313	504.98498	-458.54815	37.69	5.493
A8	-964.15108	505.60136	-458.54972	36.71	5.695

indicates that form A1 is predominant in most circumstances and only form A2 may have any significant additional role, assuming favourable solvent effects or other intermolecular interactions.

The results for guanine tautomers are less clear-cut than the adenine results. There are three, possibly four, tautomeric forms that are within a relatively narrow energy range of the most stable tautomeric form. According to the present results, form G1, considered most commonly as the "normal" form of guanine, is not the most stable one, its total energy is 12 kcal/mole above the most stable form, G3, that is an "enol" form of guanine. This energy difference is slightly above the value set as "limit of significance". It is of some interest that for both adenine and guanine the tautomeric forms with the "fully aromatic" six-membered rings, i.e. the ones with regular Kekulé structures appear to be the most stable. In this context it is also remarkable that earlier semiempirical CNDO/2 studies also indicated enhanced stability for the "enol" form of guanine as compared to the "keto" form G1, even though a rather different geometry was assumed [23]. This suggests that the prediction of the "enol" form as the most stable one in gaseous state is not likely to be an artifact of the design of geometries. In addition, the previous *ab initio* study on cytosine and thymine [2], as well as CNDO/2 studies on the same molecules [25], also indicated greater stability for "enol" forms than for "keto" forms. This apparent agreement of all the above theoretical studies suggests that in the absence

of intermolecular interactions (i.e. in vacuum, or in low pressure gas) enol forms of nucleotide bases are the most stable tautomers. In the present study the calculated relative energy of a different "enol" form, G4, is 5 kcal/mole, also less than the 12 kcal/mole value of the keto form G1. The fourth form with reasonably small relative energy ( $E_{\text{rel}} = 16$  kcal/mole) is "keto" form G2 and all other tautomeric forms are of much higher relative energy.

**Table 2.** Energies and ionization potentials of guanine tautomers

Tautomer	Total electronic energy (a.u.)	Nuclear repulsion energy (a.u.)	Total energy (a.u.)	Relative energy (kcal/mole)	Ionization potential (eV)
G1	-1130.75985	598.32824	-532.43161	11.71	5.304
G2	-1130.98234	598.55705	-532.42529	15.67	5.568
G3	-1127.18355	594.73329	-532.45026	0.00	6.074
G4	-1127.42495	594.98332	-532.44163	5.42	6.371
G5	-1129.72521	597.33762	-532.38759	39.33	5.483
G6	-1127.25470	594.84111	-532.41359	23.01	6.048
G7	-1129.38663	597.00062	-532.38601	40.32	5.544
G8	-1129.63095	597.23022	-532.40074	31.08	5.686
G9	-1131.27728	598.86915	-532.40813	26.44	6.343
G10	-1131.50466	599.09002	-532.41464	22.36	6.489
G11	-1131.32945	598.93975	-532.38969	38.01	5.453
G12	-1131.68477	599.28733	-532.39744	33.15	5.407
G13	-1130.88594	598.53143	-532.35451	60.09	4.875
G14	-1130.59702	598.23329	-532.36374	54.29	4.700
G15	-1131.28230	598.93950	-532.34280	67.43	4.877
G16	-1130.96308	598.61023	-532.35284	61.13	4.674
G17	-1132.73375	600.35882	-532.37493	47.27	5.854

For a set of similar molecules the nuclear repulsion energy may be regarded as a crude "measure" of the compactness (or "crowdedness") of the molecules [45]. A larger value, i.e. larger Coulomb repulsion between the nuclei indicates a more compact structure. It is noteworthy that among the guanine tautomers the most stable form, G3, has the smallest nuclear repulsion energy, i.e. this tautomer is the "least compact". Indeed, G3 is one of the most "spread out" tautomers, and the hydrogen atoms are distributed fairly evenly along the perimeter of the purine ring system. Normally, hydrogen atoms contribute little to the variations in total nuclear repulsion energy due to their small nuclear charge. In tautomerization problems, however, their contribution becomes important as a consequence of large changes in their positions. For adenine a similar tendency is not apparent, although the nuclear repulsion energy for A1 is in the lower one-third of the interval of all values, spreading from 501.185 a.u. (A5) to 505.601 a.u. (A8).

It is remarkable that the variations in the nuclear repulsion energy and total electronic energy are about 100 times larger than variations in the total energy values of the tautomeric forms.

Denoting nuclear and electronic energy differences (between tautomeric forms  $i$  and  $j$ ) by  $\Delta E_n^{ij}$  and  $\Delta E_e^{ij}$ , respectively, and the corresponding total energy difference by  $\Delta E_t^{ij}$ , the relation

$$|\Delta E_n^{ij}| + |\Delta E_e^{ij}| > |\Delta E_t^{ij}| \quad (1)$$

holds within both the adenine (A1  $\cdots$  A8,  $1 \leq i < j \leq 8$ ) and guanine (G1  $\cdots$  G17,  $1 \leq i < j \leq 17$ ) tautomeric series, for any choice of  $i$  and  $j$ . Relation [1] is a special form (applicable for a set of *discrete* geometrical structures) of a general criterion for the approximate conservation of the molecular total energy along a reaction path [46]. Consequently, the validity of relation (1) indicates that the total energy is approximately maintained in all tautomeric rearrangements in spite of the very large ( $\sim 2000$  kcal/mole) variations in the nuclear repulsion energy. Indeed, the variations in the internuclear repulsion, associated with a given tautomeric rearrangement, are approximately 99% compensated for by variations of the opposite sign in the total electronic energy. The fact that the electronic energy shows such large differences from tautomer to tautomer indicates that there are major variations in the electronic structures, in spite of the rather similar geometries assumed for the purine ring structures in all tautomers.

The ionization potentials, calculated via Koopmans' theorem, are given in the last columns of Tables 1 and 2. The calculated values are very similar to those obtained for the tautomers of pyrimidine bases cytosine and thymine [2], indicating that the extension of the ring system by a five-membered ring does not change appreciably the first ionization potential of these nucleotide bases.

In structures G14 and G16 the orientation of the imine hydrogen is different from that in tautomeric forms G13 and G15, respectively. The total energy differences, associated with these two in-plane inversion processes are 5.80 and 6.30 kcal/mole, respectively. Energy differences of this order are insufficient to change the order of stability of the tautomers substantially and imine hydrogen inversion has no effect on the most stable tautomers. Nevertheless, the in plane inversion of the imine group and rotation of the OH group in the "enol" forms may play an important role in biological systems. In a forthcoming study we shall consider the influence of the above two factors on the stability of various nucleotide bases.

#### 4. Conclusions

The present *ab initio* study on purine bases indicates that in the gas phase the tautomeric forms with fully aromatic Kekulé structures for the six-membered ring are the most stable. Relatively few, only two of the 8 tautomeric forms of adenine and four of the 15 forms of guanine, may play any significant role in real biological systems, all other tautomeric forms are of much higher relative energy. Energy differences between amine and imine forms suggest that an amine  $\rightarrow$  imine tautomerization is not likely to be responsible for spontaneous mutations in DNA.

*Acknowledgement.* The authors express their thanks to the National Foundation for Cancer Research and the Natural Sciences and Engineering Research Council, Canada, for financial support.

## References

1. Jordan, D. O.: The chemistry of nucleic acids. London: Butterworths 1960; 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. Hoffman, T. A., Ladik, J.: Advan. Chem. Phys. **7**, 84 (1964); Ladik, J., Appel, K.: Theoret. Chim. Acta (Berl.) **4**, 132 (1966)
4. Strickberger, M. W.: Genetics. New York: Macmillan and Co. 1968
5. Watson, J. D.: The molecular biology of the gene, 2nd ed. New York: Benjamin 1970
6. Angell, C. L.: J. Chem. Soc. **504** (1961)
7. Marshall, J. R., Walker, J.: J. Chem. Soc. **1004** (1951)
8. Shugar, D., Fox, J. J.: Biochim. Biophys. Acta **9**, 199 (1952)
9. Brown, D. J., Lyall, J. M.: Aust. J. Chem. **15**, 851 (1962)
10. Clark, L. B., Tinoco, I.: J. Am. Chem. Soc. **87**, 11 (1965)
11. Maevskii, A. A.: Biofizika **20**, 957 (1975)
12. Spencer, M.: Acta Cryst. **12**, 59 (1959)
13. Arnott, S., Dover, S. D., Wonacott, A. J.: Acta Cryst. **B25**, 2192 (1969)
14. Blout, E. R., Fields, M.: J. Am. Chem. Soc. **72**, 479 (1950)
15. Kokks, J. P., Goldstein, J. H., Mandell, L.: J. Am. Chem. Soc. **83**, 2909 (1961)
16. Austin, J. E.: J. Am. Chem. Soc. **56**, 2141 (1934)
17. Tanaka, M., Nagakura, S.: Theoret. Chim. Acta (Berl.) **6**, 320 (1966)
18. Morita, H., Nagakura, S.: Theor. Chim. Acta (Berl.) **11**, 279 (1968)
19. Pullman, B., Pullman, A.: Quantum biochemistry. New York-London: Academic Press 1963; Yu, C., Peng, S., Akiyama, I., Lin, J., LeBreton, P. R.: J. Am. Chem. Soc. **100**, 2303 (1978)
20. Clark, L. B., Peschel, G. G., Tinoco, I., Jr.: J. Phys. Chem. **69**, 3615 (1965)
21. Boerresen, H. C.: Acta Chim. Scand. **21**, 2463 (1967)
22. Berthod, H., Giessner-Prettre, C., Pullman, A.: Theoret. Chim. Acta (Berl.) **5**, 53 (1966)
23. Fujita, H., Imamura, A., Nagata, C.: Bull. Chem. Soc. Jap. **42**, 1467 (1969)
24. Ladik, J., Biczó, G., Rédy, J.: Phys. Rev. **188**, 710 (1969)
25. Breen, D. L., Flurry, R. L.: Theoret. Chim. Acta (Berl.) **23**, 138 (1971)
26. Giessner-Prettre, C., Pullman, A.: Theoret. Chim. Acta (Berl.) **9**, 242 (1968); Pullman, A., Kochanski, E., Gilbert, M., Denis, A.: Theoret. Chim. Acta (Berl.) **10**, 231 (1968)
27. Pullman, B., Berthod, H., Dreyfuß, M.: Theoret. Chim. Acta (Berl.) **15**, 265 (1969); Pullman, B., Berthod, H., Bergmann, F., Neiman, Z., Weiler-Felchenfeld, H., Bergmann, E. D.: Tetrahedron **26**, 1483 (1970)
28. Kwiatkowski, J. S.: Theoret. Chim. Acta (Berl.) **13**, 149 (1969)
29. Bodor, N., Dewar, M. J. S., Harget, A. J.: J. Am. Chem. Soc. **92**, 2929 (1970)
30. Hug, W., Tinoco, I., Jr.: J. Am. Chem. Soc. **95**, 2803 (1973)
31. Hug, W., Tinoco, I., Jr.: J. Am. Chem. Soc. **96**, 665 (1974)
32. Ladik, J., in: Electronic structure of polymers and molecular crystals, André, J.-M. and Ladik, J., eds., p. 663. New York: Plenum Press 1975
33. Clementi, E., André, J.-M., André, M.-Cl., Klint, D., Hahn, D.: Acta Phys. Hungarian Acad. Sci. **27**, 493 (1969)
34. Clementi, E., Mehl, J., Niessen, W.: J. Chem. Phys. **54**, 508 (1971)
35. Mely, B., Pullman, A.: Theoret. Chim. Acta (Berl.) **13**, 278 (1969)
36. Ladik, J., Suhai, S., Otto, P., Collins, T. C.: Intern. J. Quantum Chem. Quant. Biol. Symp. **4**, 55 (1977)
37. Suhai, S., Merkel, Ch., Ladik, J.: Phys. Letters **61A**, 487 (1977)
38. Szent-Györgyi, A.: Bioelectronics **4**, 535 (1973); Szent-Györgyi, A.: Acta Biochem. Biophys. Acad. Sci. Hung. **8**, 117 (1973); Szent-Györgyi, A.: Life Sci. **15**, 863 (1977)
39. Ladik, J.: Intern. J. Quantum Chem. Quantum Biol. Symp. **1**, 65 (1974)
40. Laki, J.: Intern. J. Quantum Chem. Quantum Biol. Symp. **2**, 133 (1975)
41. Laki, K., Ladik, J.: Intern. J. Quantum Chem. Quantum Biol. Symp. **3**, 51 (1976)



42. Roothaan, C. C. J.: *Rev. Mod. Phys.* **23**, 69 (1951)
43. Hehre, W. J., Lathan, W. A., Ditchfield, R., Newton, M. D., Pople, J. A.: QCPE program 236, Indiana University, Bloomington, Indiana
44. Hehre, W. J., Stewart, R. F., Pople, J. A.: *J. Chem. Phys.* **51**, 2657 (1969)
45. Kapur, A., Steer, R. P., Mezey, P. G.: *J. Chem. Phys.* **69**, 968 (1978)
46. Mezey, P. G.: *Chem. Phys. Letters* **47**, 70 (1977)

*Received November 21, 1978/December 14, 1978*