SHORT PAPER

J. Chem. Research (S),
2003, 195-199

Ab initio theoretical studies of relative stabilities and IR spectrum of 5-methylcytosine tautomers†

Lida Ghassemzadeh, Majid Monajjemi* and Karim Zare

Islamic Azad University , Science and Research Branch , P.O.Box 14515-775 , Tehran, Iran

The structure and relative energies of the tautomers of 5-methylcytosine in the gasphase and in different solvents are predicted using MP2 and density functional theory methods.The order of stability for these tautomers is C3>C1>C2>C4>C5>C6 calculated by MP2 and C1>C3>C2>C4>C5>C6 calculated by the B3LYP method. Relative energy calculations are performed in wide range of solvent dielectrics and in all solvents the oxo-amino C1 is predicted as the most stable tautomer. The infrared spectra of two dominant tautomers are calculated in the gas phase using HF and density functional theory. Good agreement between calculated (DFT) and experimental harmonic vibrational frequencies is found.

Keywords: physico-chemical properties, 5-methylcytosine tautomers

Introduction

An understanding of the physico-chemical properties of the purine and pyridine bases of the nucleic acids is of fundamental importance not only in relation to qualitative concepts of chemical binding and physical chemistry, but also in relation to molecular biology. It is well known that geometrical and conformational properties of biomolecules have an important effect on their biochemical behavior. The possible existence of one or more of the DNA bases in unusual tautomeric forms can increase the probability of mispairings of the purines with pyrimidines, and hence may lead to point mutation.[1–6] The importance of the methylation of cytosine has been demonstrated experimentally; for example, methylation of DNA is involved in a wide variety of biochemical events.[7-10] Methylated cytosine is the hot spot for mutation by deamination of 5-methylcytosine to thymine, one of the ordinary DNA bases; this means that once 5-methylcytosine deaminates, a guanine–thymine mismatch appears.[11]

5-methylcytosine is a minor base of DNA. Its percentage with respect to the total content of cytosine varies over a wide range from 0.03% in insects, to 2–8% in mammals, to 50% in the higher plants.[12]

The role of environment in the determination of the order of stability of nucleic acid bases tautomers and the importance of 5-methylation of cytosine in many biological process, justify the present paper.

5-Methylcytosine molecule may exist in various tautomeric forms differing from each other by the position of the hydrogens, which may be bound to either the ring nitrogen atoms or the oxygen atoms (Fig. 1).

Matrix isolation IR studies of 5-methylcytosine[13] indentified the amino-oxo (C1), amino-hydroxy (C3) and imino-oxo (C2) exist simultaneously. Therefore, the spectrum obtained is quite complicated because it is a superposition of the spectra of the three tautomers.

During the last 15 years, considerable progress in the theoretical prediction of IR spectra of big molecules has been achieved. Contrary to the previously used calculations at the HF/3-21G or HF/4-31G level, for which the theoretical descriptions of the out-of-plane vibrations were very inaccurate, the calculations performed at the HF/6-31G(d,p) level provide a satisfactory description of both the in-plane and the out of-plane modes.

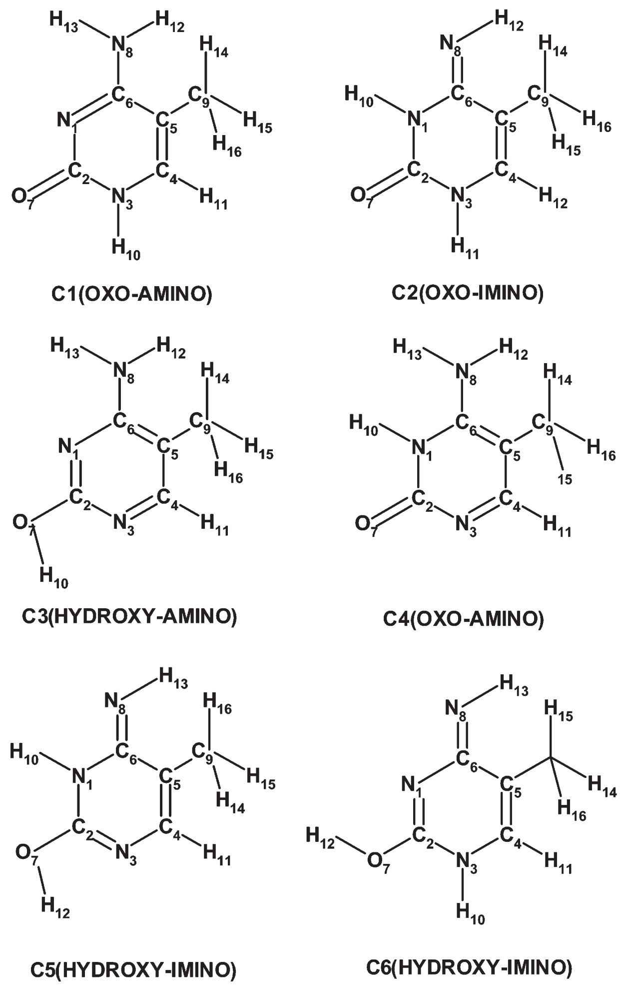

Fig. 1 5-methylcytosine tautomers.

As an experimental reference for the assessments of the theoretical predictions, the IR spectra of the species isolated in weakly interacting argon low-temperature matrices were used. Also, the theoretically predicted intensities are, as a rule, overestimated with respect to the experimental values. The intensities calculated at the HF/6-31G(d,p) level are too high, while at level of taking electron correlation into account [MP2/6-31G(d,p),DFT(B3LYP)/6-31G(d,p)] the intensities are systematically smaller, but still higher than the values determined

*To receive any correspondence. E-mail:m_monajjemi@yahoo.com
† This is a Short Paper, there is therefore no corresponding material in *J Chem. Research (M).*

Table 1 Relative stabilities of 5-methylcytosine in gas phase(kcal/mol)

	MP2/6-31+G(d,p) //MP2/6-31+G(d,p)	MP2/6-311++G(d,p) //MP2/6-31+G(d,p)	B3LYP/6-311+G(2d,p) //MP2/6-31+G(d,p)
C1	2.2749101904	3.1827909350	0.0000000000
C2	4.4053049429	5.1082410848	2.1829800486
C3	0.0000000000	0.0000000000	0.0598644063
C4	10.0119193183	10.7349932694	7.3108621807
C5	14.0441018667	13.7980553917	12.2823689454
C6	22.4676638928	22.0687561036	20.6630720727

Table 2 Relative stabilities of C1 and C3 tautomers in different solvents(kcal/mol)

Solvent	ε	C1	C3
Water	78.39	1.78131121765	1.92363037225
DMSO	46.7	10.27797810050	9.52760224040
Nitro methane	38.2	10.93956137	10.18360067170
Methanol	32.63	0.00000000000	0.00000000000
Ethanol	24.55	2.70519345450	2.61916190205
Acetone	20.7	10.25501125280	9.33909838660
Dichloro ethane	10.36	11.83024835065	10.47620835155
THF	7.58	12.59574718970	10.92374812695
Chloroform	4.9	14.16496019635	11.97106148245
Cyclohexan	2.023	16.39820376	12.33683677

by experiment. It has been well established that solvents with large dielectric constants favour the more polar tautomers. For the tautomerism of hydroxypyridine/pyridone system this means that the equilibria will shift toward the pyridone in more polar solvents because the oxo-form tautomer is usually the more polar species.[14]

Computational details

Calculations were carried out with the GAUSSIAN98[15] program. Geometry optimisation in the gas phase for all six tautomers were performed at Hartree-Fock(HF) and second order Moller-Plesset(MP2) levels with MP2/6-31+G(d,p). The self-consistent isodensity polarized continuum model[16-17] has been used in simulation of the solvent effect.

The calculations of the IR frequencies and intensities were performed with HF/6-31+ G(d,p) and B3LYP/6-311+G(2d,p) basis sets.

Results and discussion

Relative stabilities and geometries: *Gas phase*. Relative stabilities of the six tautomers in gas phase are given in Table 1. The order of stability is as follows: C1>C3>C2>C4>C5>C6 and C3>C1>C2>C4>C5>C6 at B3LYP and MP2 computing levels, respectively. Clearly, both methods show that the C1, C2, and C3 tautomers are more stable than others. B3LYP present the former tautomers very close in energy in a narrow range of 2.18 kcal/mol while MP2 calculations increases this range up to 5.1kcal/mol, now C3 being the most stable tautomer. These results can be related with previous theoretical studies for cytosine presented by Paglieri *et al*.,[18] Kobayashi,[19] Sambrano *et al*.[20] and Leszczynski.[21]

Solvent effects on structure. The oxo-amino(C1) tantomer is predicted as the most stable form by HF methods in all solvents. Our calculations found that the hydroxy-amino(C3) form was considerably destabilised by solvation as shown in Figs 2 and 3.

In the first instant irregular variations were observed concerning relative energy versus dielectric constant where the energy variations result from two levels of regular changes:

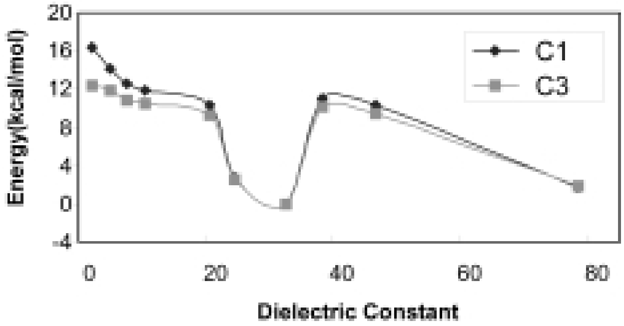

Fig. 2 Variation of energy (kcal/mol) with ε for C1 and C3

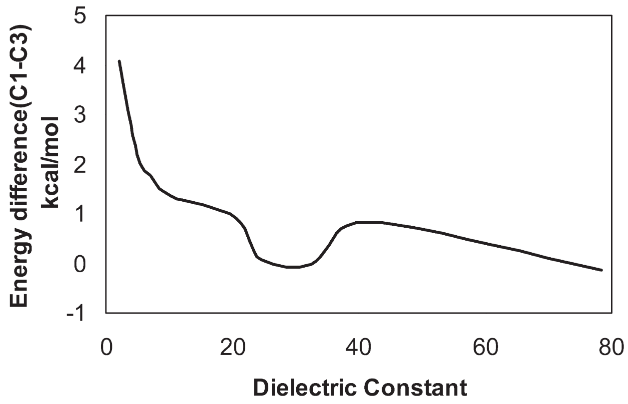

Fig. 3 Variation of energy difference between C1 and C3 with ε.

A. Energy variations with solvents that have no hydrogen bonds to oxygen.

B. Energy variations with solvents that have hydrogen bonds to oxygen.

Like cytosine,[14] energy values increase nonlinearly with decrease in dielectric constant in former case, while future investigations are required to explain the latter. By plotting the relative energy versus ln(ε), the following equation is derived.

C1:E(rel)= –1.8823ln(ε)+12.935

C3:E(rel)= –0.919ln(ε)+12.935

The points that do not match on the line might be due to factors such as polarisability and dipole moment. More accurate results might be obtained if successive values of dielectric constants in a narrower range are chosen.

Variational analysis: In Tables 3 and 4, theoretical vibrational frequencies and IR intensity in the gas phase for C1 and C3 tautomers are presented. The possible tautomers C1, C2, C3 and C4 are very close energetically and coexist in an inert gas matrix.This fact makes the experimental results problematic, since the IR bands obtained in the spectrum sometimes correspond to two or more different tautomers. Moreover, experimental IR studies[13] were developed with the 5-methylcytosine molecule isolated in nobel gas matrix, which is a different environment from the gas phase. Thus we have carried out a comparison with the experimental data taking into account these limitations. The general picture of the 3600–3400cm^{-1} region is similar to that of cytosine.[20-22] Also, the ratio of the experimental intensities in this case is relative higher than similar ratios observed in other pyrimidine derivatives. In the high frequency region of the IR spectra, two bands due to antisymmetric and symmetric stretching vibrations of the amino group are observed in the experimental spectra. The calculations carried out at the levels of theory considered in this paper predicted well the antisymmetric ν NH_2 band at higher frequency and the symmetric νNH_2 band at lower

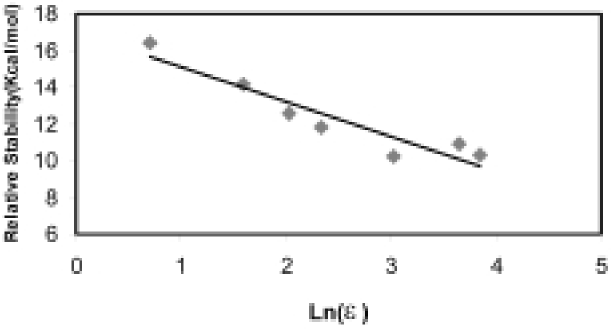

Fig 4 Linear dependence of relative energy to Ln(ε) in C1 for solvents that have no –OH group.

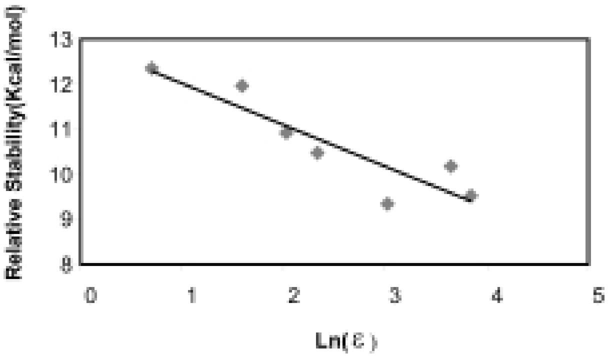

Fig 5 Linear dependence of relative energy to Ln(ε) in C3 for solvents that have no –OH group.

Table 3 Variational frequencies(cm^{-1}) and IR intensities(km/mol) for C1

	HF/6-31+G(d,p)		B3LYP/6-311+G(2d,p)		Experimental data[13]		
No.	Freq	IR inten	Freq	IR inten	Freq	IR inten	Mode*
1	156.3	3.9	112.5	4.3			C_2O_7 b; Me b
2	313.8	1.0	163.8	0.9			Me tors
3	276.4	4.8	223.8	6.7			C_6N_8 b(oopl); C_1C_6,C_1C_2 b
4	340.1	1.2	279.5	3.2			N_8H_2 b(oopl); C_5C_9 b
5	329.7	0.7	293.2	1.2			C_5C_9 b(oopl);C_6N_8 b(oopl)
6	429.6	2.2	375.9	7.7	382		C_6C_8 b;C_2O_7 b
7	493.2	10.6	420.5	12.0			N_8H_2 tors; C_4H_{11} b(oopl)
8	595.6	63.9	438.8	153.6			N_8H_{12} b
9	518.4	29.7	489.1	3.7			ring deformation in ring plan
10	618.3	56.8	548.3	30.3	406	23	N_8H_2 tors; N_3H_{10} b(oopl)
11	640.3	33.6	575.8	62.2			N_3H_{10} b(oopl);N_8H_2 tors
12	775.7	9.2	587.7	11.5			ring deformation
13	734.1	113.3	635.5	107.5	638		N_3H_{10} b(oopl)
14	789.7	164.4	733.5	4.5			C_4C_5 s
15	841.7	3.0	765.2	3.9			N_8C_6 b(oopl); N_3H_{10} b(oopl)
16	849.6	21.8	788.6	0.6	802	18	C_2N_3 b(oopl)
17	894.5	75.4	797.6	33.4	779		Me tors; C_2O_7 b(oopl)
18	1077.3	9.9	915.9	9.9			C_4H_{11} b(oopl)
19	1000.8	12.2	925.3	7.6	877	22	N_3C_2 s;C_5C_6 s
20	1112.1	3.3	1010.9	1.9	998	7	Me b
21	1184.1	0.1	1067.9	0.4	1096	25	Me b
22	1239.7	38.4	1140.9	52.7	1167	151	N_8H_2 b
23	1234.0	77.0	1180.2	13.7			N_3C_4 s;N_1C_2 s
24	1309.2	47.1	1192.2	49.8	1188		N_3H_{10} b;C_4H_{11} b
25	1347.7	7.6	1259.6	6.3			N_1C_2 s
26	1432.4	112.3	1336.4	38.5	1330	24	C_4H_{11} b
27	1569.6	70.3	1414.5	20.1	1439		Me sciss
28	1554.0	159.5	1425.9	56.2	1422	30	N_3H_{11} b
29	1525.9	328.6	1448.4	230.4			N_8C_6 s
30	1610.9	7.7	1482.3	8.6			Me b
31	1635.9	2.4	1501.5	18.4	1524	12	Me b
32	1673.8	173.0	1539.5	125.6	1537	189	C_6N_1 s
33	1770.1	737.2	1630.2	130.9	1595	412	N8H2 sciss
34	1790.7	603.6	1671.6	259.8	1676	404	C_4C_5 s;C_6N_1 s
35	1818.8	222.0	1678.3	821.7	1735	78	C_2O_7 s
36	3124.2	46.0	3028.5	35.0			Me s
37	3175.0	25.9	3077.1	15.3			Me s
38	3209.3	22.8	3111.6	13.5			Me s
39	3306.0	5.3	3189.8	4.8			C_4H_{11} s
40	3655.3	70.4	3565.1	49.5	3443		N_8H_2 s(sym)
41	3696.7	115.8	3596.4	74.7	3464	139	N_3H_{10} s
42	3781.3	54.3	3682.9	37.7	3560		N_8H_2 s(anti-sym)

*b: bonding; s: stretching; oopl: out-of-ring plane; tors: torsion; sciss: scissoring; wagg: wagging; sym: symmetric; anti-sym: antisymmetric.

frequency, as observed in the experimental spectrum. Also, the relative intensities of these two bands are well predicted theoretically, for the oxo-amino tautomer the symmetric νNH_2 band being stronger than the antisymmetric νNH_2 band. For the amino-hydroxy, the band due to the νOH vibration is observed. The theoretically-predicted frequency of this band is considerably overestimated in the calculation performed at the HF level.

The bands due to the normal models of the NH_2 scissoring vibration are placed in the frequency range 1650–1500 cm^{-1}. The experimental pattern of the bands in this region is usually well predicted at the DFT level. Two strong bands due to the ring stretching of the hydroxy form are expected in the 1650–1500 cm^{-1} region. As in the spectra of other pyrimidines with fully aromatic rings (cytosine[22]), their calculated frequencies are underestimated. Bonding motions of the

Table 4 Variational frequencies(cm^{-1}) and IR intensities(km/mol) for C3

	HF/6-31+G(d,p)		B3LYP/6-311+G(2d,p)		Experimental data[13]		
No.	Freq	IR inten	Freq	IR inten	Freq	IR inten	Mode*
1	167.9	0.5	128.5	0.4			C_2O_7 b; Me b
2	306.3	0.1	146.2	0.3			Me tors
3	282.7	4.6	235.3	8.5			C_6N_8 b(oopl); C_1C_6,C_1C_2 b
4	324.2	0.2	285.3	0.3			N_8H_2 b(oopl); C_5C_9 b
5	369.9	0.5	314.9	2.4			C_5C_9 b(oopl); N_8H_2 tors
6	410.9	6.0	355.8	11.0	352	14	C_6C_8 b;C_2O_7 b
7	607.9	57.0	438.5	50.1	450	18	N_8H_2 tors
8	539.8	6.8	474.7	40.5			N_8H_2 wagg; C_4H_{11} b(oopl)
9	530.9	9.1	496.3	7.9			ring deformation in ring plan
10	604.9	10.4	544.2	12.6	499	40	C_2O_7 b
11	638.2	17.8	565.4	40.9			O_7H_{10} tors; N_8H_2 wagg
12	738.7	230.6	574.0	183.1			N_8H_2 wagg; O_7H_{10} tors
13	811.5	116.8	598.6	82.0			ring deformation
14	822.6	26.3	744.9	8.2			C_6N_8 b
15	818.4	2.0	766.4	1.2	777		C_5C_6 s
16	920.0	105.9	781.0	6.6	795	59	C_2O_7 wagg; C_6N_8 wagg
17	842.6	23.5	819.3	35.3			Me tors; C_2O_7 b(oopl)
18	1110.6	10.0	958.6	7.4			C_4H_{11} b(oopl)
19	1046.8	2.2	973.6	8.0	1008	21	N_3C_2 s;C_5C_6 s
20	1102.7	53.8	1018.3	15.1	963	5	Me b
21	1184.3	0.2	1066.2	0.7	1090	144	Me b
22	1242.7	36.7	1109.3	101.0			N_8H_2 b
23	1165.7	40.7	1202.1	13.9			N_3C_4 s; N_1C_2 s; C_5C_9 b
24	1324.4	6.2	1228.3	2.8			C_5C_9 s; N_1C_2 s
25	1363.0	11.6	1264.1	23.3			N_1C_2 s
26	1435.3	28.2	1328.2	110.4	1318	97	C_4H_{11} b
27	1485.2	410.3	1361.8	179.4	1358	126	Me sciss
28	1553.8	38.6	1413.3	14.6	1432	216	N_3H_{11} b
29	1539.3	550.1	1445.0	444.3	1451	442	N_8C_6 s
30	1573.5	47.6	1461.1	32.3	1458	35	Me b
31	1612.3	7.7	1483.7	8.6	1481	106	Me b
32	1636.0	17.4	1504.7	13.5			C_6N_1 s
33	1736.9	367.5	1592.4	228.0	1610	224	N_8H_2 sciss
34	1759.7	206.6	1613.3	118.3	1630	295	C_4C_5 s; C_6N_1 s
35	1817.3	204.6	1642.0	236.2	1586	141	C_2O_7 s
36	3120.9	48.6	3026.1	35.7			Me s
37	3169.6	29.4	3073.9	17.4			Me s
38	3211.7	20.5	3114.0	12.1			Me s
39	3274.2	21.5	3155.6	19.7			C_4H_{11} s
40	3650.4	58.1	3561.3	39.8	3447		N_8H_2 s(sym)
41	3770.0	47.6	3673.0	34.4	3564		N_3H_{10} s
42	3821.9	133.8	3750.4	94.4	3599	227	N_8H_2 s(anti-sym)

methyl group provide the major contributions to the normal modes whose frequencies are in the 1500–1400 cm^{-1} region. 1000–200 cm^{-1} region, for the bands of the out-of-plane vibrations is more complicated.

In the IR spectra of the compounds with an amino group, a medium strong band near 500 cm^{-1} is usually observed. This band is most probably due to some of out-of-plane vibration of the amino group, indicated by its considerable shift toward higher frequencies in the spectra of the compounds isolated in nitrogen matrices. The theoretical calculations, carried in the harmonic approximation, predict the band to be due to the NH_2 twisting vibration, but at all considered levels of theory this band is of low intensity.

Conclusions

Our results can be summarised as follows:

(i) This work presents a self consistent reaction field study of the relative stability of two dominant tautomers of 5-methylcytosine, oxo-amino and hydroxy-amino forms. MP2 level of theory combined with 6-31+G(d,p) basis predicted the energy difference of C1-C3 to be about 2.27 kcal/mol and MP2/6-31+G(d,p) optimised geometry followed by MP2/6-311++G(d,p) single point energy calculation predicted this difference to be about 3.18 in the gas phase. In the liquid phase HF/6-31+G(d,p) shows the oxo-amino C1 tautomer as the stable form.

(ii) Many common freatures appear in the IR spectrum of cytosine and 5-methylcytosine. Comparison of the calculated and experimental spectra led to positive assignments of most of the bands in the spectra of the amino-hydroxy and amino-oxo forms.

(iii) The biological significance of tautomerism of 5-methylcytosine is out of the scope of this paper. We want only to mention that since the biologically active (*e.g.*, in nucleic acids) compound is the N3 derivative, the amino-hydroxy form (which dominates in low- temperature matrix) cannot exist in the biological systems. However, a strong stability of this form, which emerges from both theoretical and spectroscopical studies, seems to support a possibility of an additional bond through the exocyclic oxygen atom, which can considerably enhance the stability of some derivatives.[13]

(iiii) The IR spectra of the bases calculated by the DFT(B3LYP) method agree much better with the recorded experimental spectra than do the spectra predicted at the HF level and also are better than those predicted at the MP2 level[23-25]. The DFT calculations is probably too sensitive and thus unreliable for optimized calculations.

Received 19 August 2002; accepted 21 October 2002
Paper 02/1510

References

1 P.O. Lowdin, *Adv. Quanti Chem*. 1965, **2**, 213.
2 J.D. Watson and F.H.C. Crick, *Nature(London)* 1953, **171,** 1964.
3 V.I. Danilov and G.F. Kventsel, *Electronic Representation in the Point Mutation Theory* (in Russian) (naukova Dumka, kiev, 1971).
4 R. Rein, M. Shibata, R. Garduno-Juarez and T. Kieber-Emmons, in *Structure and Dynamics: Nucleic Acids and Proteins*, eds. E. Clementi and R.H. Sarma, Adenine Press, Guilderland, N.Y. 1983, p.269.
5 W. Saenger; *Principles of nucleic acid structure*, Springer-Welag; New York, 1984.
6 G.M. Blackborn and M.J. Gait. *Nucleic acids in chemistry and biology*, Oxford University Press. Oxford 1996.
7 E.Li, H. Bestor and R. Jaerish, *Cell* 1992, **69**, 915.
8 S.B. Baylin, *Cancer Cells*, 1992, **69**, 91.
9 E.Li, R. Beard and R. Jaenish, *Nature*, 1993, **366**, 362.
10 S.T. Warrant, C.T. Ashley, *Ann. Rev. Neurosci*, 1995, **94**, 2545.
11 R.F. Weaver, P.W. Hedrick, *Basic Genetics*, Wm C. Brown, Dubuque, Iowa, 1991.
12 W. Doerfler, *Annu. Rev. Bio Chem*., 1983, **52**, 93.
13 L. Lapinski, M.J. Nowak, J. Fulara, A. Les and L. Adamowicz, *J. Phys. Chem*., 1990, **94**, 6555.
14 M. Monajjemi, R. Fazaeli, K. Zare, F. Ataherian and J.M. Struct *THEOCHEM* in press.
15 Gaussian98, RevisionA.7, M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E Scuseria, M.A. Robb, J.R. Cheeseman, V.G. Zakrzewski, J.A. Montgomery, Jr., R.E. Stratmann, J.C. Burant, S. Dapprich, J.M. Millam, A.D. Daniels, K.N. Kudin, M.C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G.A. Petersson, P.Y. Ayala, Q. Cui, K. Morokuma, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman J. Cioslowski, J.V. Ortiz, A.G. Baboul, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P.M.W. Gill, B. Johnson, W. Chen, M.W. Wong, J.L. Andres, C. Gonzalez, M. Head-Gordon, E.S. Replogle and J.A. Pople, Gaussian, Inc., Pittsburgh PA, 1998.
16 J. Tomasi and M. Persico, *Chem. Rev.*, 1994, **94**, 2027.
17 J. Tomasi, In: C.J. Cramer and D.G. Truhlar (eds), *Structure and reactivity in Aqueous Solution*, Am. Chem. Soc., Washington, DC, 1994, p.10.
18 L. Paglieni, G. Corongin and D.A. Estrin, *Int. J. Quantum Chem.*, 1995, **56**, 615.
19 R. Kobayashi, *J. Phys. Chem* **A 102** 1998 10813.
20 J.R. Sambrano, A.R. Souza, J.J. Queralt and J. Andres, *Chem. Phys. Lett.*, 2000, **317**, 437
21 J. Lesczynski, In: P.V.R. Schleyer (ed.), *Encyclopedia of Computational Chemistry*, Vol. 5 Wiley, New York, 1999, p.2951.
22 M.J. Nowak, L. Lapinski and J. Fulara, *J. Spectro Chem. Acta*, 1989, **45A**, 229.
23 J.S. Kwiakowski and J. Lesczynski, *J. Mol. Struct.*, 1996, **376**, 325.
24 L. Lapinski, M.J. Nwak, A. Les and L.Adamowicz, *Vibra. Spectrosc*., 1995, **8**, 313.
25 J.S. Wiatkowski and J. Leszczynski, *J. Mol. Struct. (THEOCHEM)*, 1994, **312**, 201.