J. Am. Chem. Soc. 1988, 110, 7664-7671

Ab Initio Quantum Chemical Study of the π-Electron Spectrum of the Cytosine Molecule

J. Mauricio O. Matos[†] **and Björn O. Roos***

Contribution from the Department of Theoretical Chemistry, Chemical Centre, University of Lund, P.O. Box 124, S-221 00 Lund, Sweden. *Received February 22, 1988*

Abstract: Complete active space (CAS) SCF calculations have been performed for the ground state and three π-electron excited states of the cytosine molecule, using basis sets of the ANO type, including 3d-type polarization functions on all heavy atom centers. In order to check the methods used, excited states of the formamide molecule were also studied using the same computational scheme. The CASSCF transition energies for the lowest $^1n\pi^*$, $^3n\pi^*$, $^1\pi\pi^*$, and $^3\pi\pi^*$ states were 6.2, 5.9, 8.6, and 6.2 eV, respectively. Contracted CI calculations (with CEPA corrections) gave the energies 6.0 (5.7), 5.7 (5.3), 7.7 (7.3), and 6.0 eV, respectively, with the experimental energies within parentheses. The oscillator strength for the $^1\pi\pi^*$ state was computed to be 0.25 (0.37) with the direction 21.0° (16.6°). The spectrum of cytosine in the energy interval 4.5–6.5 eV was found to contain three $^1\pi\pi^*$ transitions. The best computed energies were 5.6, 6.7, and 8.3 eV, being between 1 and 2 eV larger than experiment. The calculations led to the following estimates of the oscillator strengths and transition moment directions: 0.05 (68 ± 10°), 0.15 (−8 ± 10°), and 1.00 (−40 ± 10°), respectively. The results rule out the possibility that cytosine has four $^1\pi\pi^*$ states in the energy interval studied.

1. Introduction

There are several reasons for undertaking a theoretical study of the electronic structure and spectral properties of the nucleic acid monomers. These molecules form the building blocks of the genetic material, and knowledge and understanding of their electronic structure are therefore of outmost importance.

During the last two decades much effort has been made to understand the electronic properties of nucleic acids,[1-3] but to achieve this goal a knowledge of the properties of their constituent monomers is required. Of special importance is information concerning the spectral properties of the monomers, since they are fundamental for the understanding of the optical properties of the nucleic acids and the polynucleotides.[4,5]

Several experiments aimed to give transition moment data for the purine and pyrimidine bases have been carried out.[2,4] The interpretation of these experiments are, however, difficult. It is, for example, impossible to determine unequivocally the direction of the transition moment for a given electronic transition. In experiments on crystals (absorption or reflection), there are for a given transition two possible directions of the transition moment which are consistent with the measured dichroic ratio.[4,6-8] The same kind of difficulty is found in experiments on stretched films, where also two possibilities for the transition moment direction have to be considered.[9,10] Attempts have been made to solve this ambiguity by comparing the results from different types of experiments. However, such comparisons are difficult to perform and the results rely on a number of more or less founded assumptions. Moreover, the absorbtion spectra of the nucleic acid monomers present broad diffuse and overlapping bands, making the resolution of the spectra into individual electronic transitions difficult.[11-13]

There is thus a strong need for theoretical support in these studies. Quantum chemical calculations, using semiempirical methods, have also been applied in studies of the spectroscopic properties of the purine and pyrimidine bases.[14] Even though these studies have made possible a general classification of the spectra, there is a large spread in the results of different studies and no consistent agreement with experiment.[2]

In this paper we shall present results from an ab initio study of the low-lying π excited states of one of the nucleic acid monomers: the cytosine molecule. Cytosine has been chosen, partly because it is the smallest of the bases, but also because of some recent experimental studies of the electronic spectrum, with the purpose to determine especially the transition moments for the π–π^* excited states.[4,6,7] Earlier semiempirical calculations have given results which show large differences from experiment, and a more definite theoretical interpretation is urgently needed. A recent RPA calculation[15] also gives results for cytosine transition moments that disagree both with the earlier semiempirical results and with experimentally deduced values.

The Complete active space (CAS) SCF method[16-18] has been used in these calculations. This method has earlier proved to give reliable results for transition moments in benzene and naphthalene.[19,20] The method will also be checked here by a study of some excited states of formamide, which is a natural candidate for testing the computational procedure, since it is the smallest molecule that contains the atoms of cytosine. The testing included both the basis set and the active space used in the CASSCF calculations.

The CGTO basis set used was of the ANO type[22] and included both 3d-type polarization functions and diffuse 2p-type functions. For cytosine, CASSCF calculations have been performed with three different active spaces and two basis sets both with and without polarization functions. The ground state was studied as

---

[†] On leave from the Department of Physics of the University of Ceará, 60000 Fortaleza, Ceará, Brazil.

(1) Bloomfield, V. A.; Crothers, D. M.; Tinoco, I. Jr. *Physical Chemistry of Nuclei Acids*; Harper & Row: New York, 1974.

(2) Callis, P. R. *Annu. Rev. Phys. Chem.* **1983**, *34*, 329 and references therein.

(3) Rizzo, V.; Schellman, J. A.; *Biopolymers* **1984**, *23*, 435.

(4) Zaloudek, F.; Novros, J. S.; Clark, L. B. *J. Am. Chem. Soc.* **1985**, *107*, 7344 and references therein.

(5) Sprecher, C. A.; Johnson, W. C., Jr. *Biopolymers* **1977**, *16*, 2243.

(6) Callis, P. R.; Simpson, W. T. *J. Am. Chem. Soc.* **1970**, *92*, 3593.

(7) Lewis, T. P.; Eaton, W. A. *J. Am. Chem. Soc.* **1971**, *93*, 2054.

(8) Chen, H. H.; Clark, L. B. *J. Chem. Phys.* **1973**, *58*, 2593.

(9) Bott, C. C.; Kurucsev, T. *Spectrosc. Lett.* **1977**, *10*, 495.

(10) (a) Matsuoka, Y.; Nordén, B. *J. Phys. Chem.* **1982**, *86*, 1378. (b) Fucaloro, A. F.; Forster, L. S. *J. Am. Chem. Soc.* **1971**, *93*, 6443.

(11) Voet, D.; Gratzer, W. B.; Cox, R. A.; Doty, P. *Biopolymers* **1963**, *1*, 193.

(12) Clark, L. B.; Tinoco, I., Jr., *J. Am. Chem. Soc.* **1965**, *87*, 11.

(13) Clark, L. B.; Peshel, G. G.; Tinoco, I., Jr., *J. Phys. Chem.* **1965**, *69*, 3615.

(14) Pullman, A.; Pullman, B. In *Advances in Quantum Chemistry*; Löwdin, P.-O., Ed.; Academic Press: New York, 1968; Vol. 4.

(15) Jensen, H. J. Aa.; Koch, H.; Jørgensen, P.; Olsen, J. *Chem. Phys.* **1988**, *119*, 297.

(16) Roos, B. O.; Taylor, P. R.; Siegbahn, P. E. M. *Chem. Phys.* **1980**, *48*, 157.

(17) Siegbahn, P. E. M.; Heiberg, A.; Roos, B. O.; Levy, B. *Phys. Scr.* **1980**, *21*, 323.

(18) Siegbahn, P. E. M.; Almlöf, J.; Heiberg, A.; Roos, B. O. *J. Chem. Phys.* **1981**, *74*, 2384.

(19) Matos, J. M. O.; Roos, B. O.; Malmqvist, P.-Å. *J. Chem. Phys.* **1987**, *86*, 1458.

(20) Matos, J. M. O.; Roos, B. O. *Theor. Chim. Acta*, in press.

0002-7863/88/1510-7664$01.50/0

well as the three lowest singlet π–π^* excited states. Transition moments for excitation to these states have been computed from the CASSCF wave functions, and also the dipole and quadrupole moments for each state. The effect of dynamic electron correlation, not included in the CASSCF wave function, was studied for the ground-state properties by complete fourth-order MBPT calculations and compared with the CASSCF results.

The spectroscopic results were obtained without inclusion of dynamic electron correlation. It was shown in the earlier work on benzene[19] that the σ–π correlation has a profound effect on the computed excitation energies, while good agreement for the transition moments could be obtained without including these effects. We can thus expect rather large errors in the energies also for cytosine, while the results for transition moments can be assumed to be reliable. In order to obtain an estimate of the errors in computed energies we have also performed CCI and CEPA[21] calculations for the singlet and triplet π–π^* and n–π^* excited states of formamide.

2. Details of the Calculation

a. The Basis Sets. Contracted Gaussian functions of the ANO type[22] were used as atomic basis functions in all calculations. The largest basis set used in the formamide calculation was derived from Huzinaga's (9s,5p) primitive set[23] for carbon, nitrogen, and oxygen. The s and p subsets were augmented with one diffuse function each with exponents: C ($\alpha_s = 0.04736$, $\alpha_p = 0.03654$), N ($\alpha_s = 0.06499$, $\alpha_p = 0.05147$), and O ($\alpha_s = 0.08619$, $\alpha_p = 0.06371$), respectively. The exponents of the 3d-type polarization functions for the same atoms were obtained from a series of CASSCF calculations for the atomic ground states, starting with the single exponents proposed by Roos and Siegbahn,[24] adding one larger and one smaller exponent with a fixed scale factor. Optimization of the scale factor and the original single exponent led to the following set of d-type functions: C (1.33, 0.53, 0.21), N (2.55, 0.79, 0.25), and O (3.88, 1.16, 0.34). The CASSCF calculations, using the active space 2s, 2p, 3p, and 3d for C(^{3}P), N(^{4}S), and O(^{3}P), furnished the contraction coefficients following the method proposed by Almlöf and Taylor.[22] The final ANO basis is (10s, 6p, 3d/3s, 3p, 1d) for carbon, nitrogen, and oxygen.

The hydrogen basis set (6s, 3p/2s, 1p) is also of the ANO type. The (6s) primitive set was taken from the compilation of Duijneveldt,[25] and the p exponents were chosen as (2.0, 0.8, 0.32). The contraction coefficients were obtained from CASSCF calculations on the H_2 molecule.[26] The basis set described above will be labeled A.

An SCF calculation was performed on formamide as a first test of the basis set. The calculated SCF energy was −168.9830 au to be compared with the earlier result of Stenkamp and Davidson: −168.9857 au,[27] which was obtained with a much larger segmented basis set. It comprised (11s, 7p/6s, 4p) for the heavier atoms and (6s/3s) for hydrogen. Polarization functions were represented by uncontracted lobe functions around the axis between each pair of bonded atoms and in the oxygen lone pair region. The experimental geometry of Kitano and Kuchitsu[28] was employed in both calculations. The experimental total energy for formamide has been estimated to be −169.896 au.[27]

Because of size and the low symmetry of cytosine, the basis set described above is prohibitively large (169 CGTO's). Therefore, two smaller basis sets were constructed for the cytosine calculations. One basis set (B) was obtained by dropping the polarization functions. It can be written as (10s, 6p/3s, 3p) for the heavier atoms and (6s/2s) for hydrogen. The other set (C) was obtained by reducing the p-type subset for C, N, and O to two functions and deleting the p-type function for hydrogen. The contracted d-type polarization functions were replaced by a single primitive Gaussian, using the exponents proposed by Roos and Siegbahn.[24] This basis set then has the following structure: C, N, O (10s, 6p, 1d/3s, 2p, 1d) and H (6s/2s). For cytosine basis B comprises 106 CGTO's and basis C 130 CGTO's.

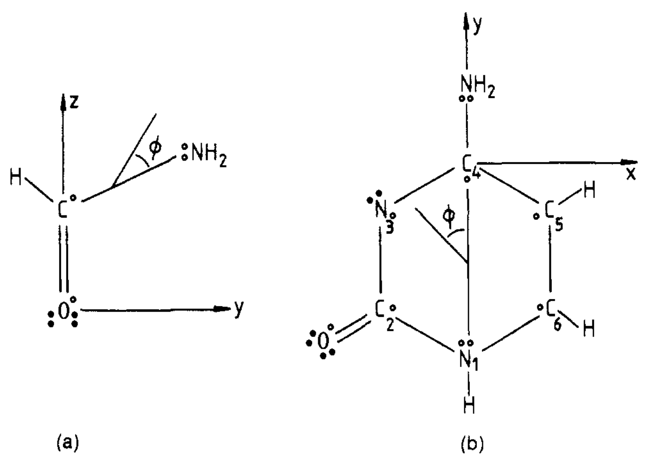

Figure 1. Molecular coordinate system for formamide (a) and cytosine (b): (●) lone-pair electrons, (○) the π electrons.

The SCF energies of cytosine obtained with these basis sets are −392.4930 au (B) and −392.6660 au (C). The former value compares favorably with a recent SCF result reported by Jensen et al.,[15] −392.4926 au, where a basis set of segmented contracted Gaussian functions, (5s, 3p/3s) was used. The geometry used in the present calculations on cytosine is the same as the one used by Jensen et al.,[15] where details can be found. The heavy atom geometry was obtained from the experimentally deduced distances and angles as reported by Taylor and Kennard.[29] The NH and CH distances were assumed to be 1.01 and 1.08 Å, respectively.

b. The Wave Functions. The main purpose of this work is to compute the transition moments of π–π^* excitations in the cytosine molecule. In order to verify the dependence of those quantities on basis set and CASSCF active orbitals, a study of the properties and electronic spectrum of the formamide molecule has first been carried out. this is a reasonably small molecule, which has some features similar to cytosine. It contains the same type of atoms (C, N, O, H). There are nonbonding electrons, and the number of π electrons is larger than the number of heavy atoms. Therefore, the conclusions drawn from a study of formamide can be useful in the analysis of the results obtained for cytosine. Complete active space (CAS) SCF[16-18] calculations have been carried out for formamide, and the effect of dynamic electron correlation was studied by multireference contracted CI (CCI) and CEPA calculations.[21] This latter study is impossible for cytosine. Some conclusions regarding the effect of dynamic correlation on the spectral properties of cytosine can also be inferred from our earlier study of the benzene molecule.[19]

The CASSCF method has been described elsewhere[16-18] and has been applied in many molecular electronic structure calculations.[30] Three sets of CASSCF calculations have been performed for the formamide molecule. Basis set A, described above, was used in the two first studies and basis B in the third.

A representation of the formamide molecule is given in Figure 1a, showing also the lone-pair electrons (•) and the four π electrons (∘). The two calculations with basis A used as active orbitals (2σ, 5π) and 6σ, 3π), respectively. In the first case two lone-pair electrons and the four π electrons were correlated, while the second calculation correlated six σ electrons and the four π electrons, at the expense of using a smaller correlating orbital space for the π electrons. Finally with the smaller basis set B, the active orbital

---

(21) (a) Siegbahn, P. E. M. In *Current Aspects of Quantum Chemistry: Proceedings of the International Congress, Barcelona, Spain, 1981*; Carbo, R., Ed.; Elsevier: Amsterdam, 1981. (b) Meyer, W. *J. Chem. Phys.* **1973**, *58*, 1017.

(22) Almlöf, J.; Taylor, P. R. *J. Chem. Phys.* **1987**, *86*, 4070.

(23) Huzinaga, S. *J. Chem. Phys.* **1965**, *42*, 1293.

(24) Roos, B.; Siegbahn, P. *Theor. Chim Acta* **1970**, *17*, 199.

(25) van Duijneveldt, F. B. *IBM Res. Rept.* **1971**, RJ945.

(26) Widmark, P. O., private communication.

(27) Stenkamp, L. Z.; Davidsson, E. R. *Theor. Chim. Acta* **1977**, *44*, 405.

(28) Kitano, M. Kuchitsu, K. *Bull. Chem. Soc. Jpn.* **1974**, *47*, 67.

(29) Taylor, R.; Kennard, O. *J. Mol. Struct.* **1982**, *78*, 1.

(30) Roos, B. O. In *Ab initio Methods in Quantum Chemistry II*; Lawley, K. P., Ed.; Wiley: New York, 1987.

Table I. Formamide CASSCF Ground, $^{1,3}n\pi^*$, and $^{1,3}\pi\pi^*$ Excited States: Total Energy, and Dipole and Quadrupole Moments Computed with Different Basis Sets and Active Spaces

state	E_T[d]	$\lvert\mu\rvert$	ϕ_{CN}	Q_{xx}	Q_{yy}	Q_{zz}	θ
CASSCF I[a]							
GS	−169.0641	3.95	38.5°	−2.06	2.89	−0.83	14.5°
$^1n\pi^*$	−168.8360	2.01	32.7°	−3.71	2.98	0.74	14.8°
$^3n\pi^*$	−168.8473	1.88	32.8°	−3.81	3.13	0.68	17.1°
$^1\pi\pi^*$	−168.7487	5.28	27.7°	−2.21	2.56	−0.35	−6.6°
$^3\pi\pi^*$	−168.8366	4.36	18.3°	−3.06	2.19	0.86	−26.9°
CASSCF II[b]							
GS	−169.1247	3.95	39.2°	−1.93	2.66	−0.72	14.5°
$^1n\pi^*$	−168.8607	1.86	32.6°	−3.64	2.87	0.77	16.0°
$^3n\pi^*$	−168.8726	1.70	32.0°	−3.77	2.79	0.97	18.7°
$^1\pi\pi^*$	−168.7843	5.52	25.3°	−2.13	2.34	−0.21	−10.6°
$^3\pi\pi^*$	−168.8718	4.63	18.1°	−2.98	2.18	0.80	−27.9°
CASSCF III[c]							
GS	−168.9325	4.16	40.1°	−1.86	3.22	−1.36	15.3°
$^1\pi\pi^*$	−168.6152	6.57	29.9°	−2.59	3.37	−0.78	2.4°
$^3\pi\pi^*$	−168.7293	3.57	21.8°	−2.68	2.35	0.33	−8.1°
Experimental							
GS		3.71 ± 0.06[e]	39.6°[e]	−2.3 ± 0.6[f]	2.5 ± 0.3[f]	−0.2 ± 0.4[f]	

[a] Basis set A, active space: 2σ, 5π orbitals. [b] Basis set A, active space: 6σ, 3π orbitals. [c] Basis set B, active space: 5π orbitals. [d] E_T, total energy in atomic units; $\lvert\mu\rvert$, dipole moment in Debye; ϕ_{CN}, the angle between the dipole vector and the CN bond (see Figure 1a); $Q_{xx,yy,zz}$, quadrupole moment components in the principal axis system. θ is the counterclockwise angle this system has with the coordinate system given in Figure 1a. [e] Reference 33. [f] Reference 34.

space consisted of four π electrons only with no correlation of the lone-pair electrons. This calculation is similar to one of the calculations in cytosine. In all cases the ground state and the lowest $^{3,1}n\pi^*$ and $^{3,1}\pi\pi^*$ states were studied.

The best CASSCF results for excitation energies and transition moments were obtained using the active space (2σ, 5π). Based on the corresponding wave functions, CCI and CEPA[21] calculations were performed for all states. The 1s core orbitals were kept frozen and the remaining 18 valence electrons were correlated. The reference configurations were in each case selected as the configurations in the CASSCF wave function having CI coefficients larger than 0.05.

CASSCF calculations for cytosine, using basis set B, were performed with two different active subspaces comprising 8 and 10 π orbitals, respectively, and correlating the 10 π electrons (see Figure 1b). They will be labeled CASSCF1 and -2, respectively. In CASSCF1 the wave function comprised 1176 configuration state functions (CSF's). The corresponding number for CASSCF2 was 19404. With basis set C we performed three sets of CASSCF studies. The first active space consisted again of 10 π electrons (CASSCF3). This space includes two orbitals for each pair of electrons (one strongly and one weakly occupied). However, since the basis B study showed only minor differences in results obtained with the 8π and the 10π active spaces, two more studies were performed using the active space (σ, σ^*, 8π) and correlating 12 electrons (2σ, 10π). Different results could be obtained in this case depending on the choice of starting orbitals for σ and σ^*. In the first calculation (CASSCF4), the wave function converged to a minimum where σ and σ^* were localized at the nitrogen σ lone-pair region, while in the second case (CASSCF5) the oxygen lone pair was described by these two orbitals. Obviously it would have been perferable to include both lone pairs in the same study. This would, however, lead to a prohibitively large active subspace (4σ, 8π). The calculations will at any rate give an indication of the importance of the most pertinent σ–π interactions for the spectral properties of the π–π^* excitations in cytosine. Calculations were performed for the ground state and the three lowest π–π^* excited states.

The CASSCF3 wave function comprises the same number of CSF's as CASSCF2, while for CASSCF4 and -5 the number of CSF's was 7140. Transition moments were computed from the CASSCF wave functions using the method developed by Malmqvist.[31] Care was taken to show that the wave functions for the different excited states were orthogonal to each other, since this is not guaranteed in the CASSCF approach. In no case was the overlap integral larger than 0.04.

The ground-state dipole moment and polarizability were also computed with the full fourth-order MBPT method[32] in order to study the effect of different correlation effects on these properties. The calculations were performed with the computer codes generated by Hubac et al.[32d] and implemented in Lund by Kellö and Sadlej.[32e] Three sets of calculations were performed. The first included excitations for the σ lone pairs and π electrons into σ^* and π^* orbitals, the second excitations for the 10 π electrons into σ^* and π^* orbitals, and finally the third excitations among the 10 π electrons into all π^* orbitals. Dipole and quadrupole moments were at the CASSCF level of theory computed as average values of the corresponding operators, while MBPT dipole moments and polarizabilities were obtained using the finite field method with the electric field strengths: $F_{x,y} = \pm 0.002$ au. All calculations reported have been performed on an IBM 3090-150VF computer.

3. Results and Discussion

a. The Formamide Molecule: Ground, $^{1,3}n\pi^*$, and $^{1,3}\pi\pi^*$ Excited States. Calculated molecular properties for the ground state and the lowest singlet and triplet $n\pi^*$ and $\pi\pi^*$ excited states are presented in Table I, which gives the total energies and dipole and quadrupole moments obtained from three different CASCF calculations, labeled CASSCF I, CASSCF II, and CASSCF III, respectively. The CASSCF I and II calculations were carried out with basis set A and active spaces comprising (2σ, 5π) and (6σ, 3π) orbitals, respectively. Basis set B was used in the CASSCF III calculations and the active space was limited to five π orbitals.

The CASSCF II calculations were performed in order to test the importance of including more σ orbitals into the wave function. The changes in computed properties are not very large. Interestingly enough the most conspcious effects are found for the $\pi\pi^*$ excited states (cf. Table II). It is also here that we find the largest occupation numbers for the weakly occupied orbitals. Thus for the $^1\pi\pi^*$ state the occupation numbers for the six σ orbitals are 1.99, 1.98, 1.96, 0.04, 0.02, and 0.01. The fourth orbital is involved in a number of singly, doubly, and triply excited configurations,

---

(31) Malmqvist, P.-Å. *Int. J. Quantum Chem.* **1986**, *30*, 479.

(32) (a) Bartlett, R. J.; Dykstra, C. E.; Paldus, J. In *Advanced Theories and Computational Approaches for the Electronic Structure of Molecules*; Dykstra, C. E., Ed.; Reidel: Dordrecht, 1984. (b) Kucharski, S.; Bartlett, R. J. *Adv. Quantum Chem.* **1986**, *18*, 281. (c) Bartlett, R. J. *Annu. Rev. Phys. Chem.* **1981**, *32*, 359. (d) Urban, M.; Hubac, I.; Kellö, V.; Noga, J. *J. Chem. Phys.* **1980**, *72*, 3378. (e) Kellö, V.; Sadlej, A. J., private communication.

Table II. Formamide Computed Excitation Energies, Transition Moments, and Oscillator Strengths of the $\pi\pi^*$ Transition

	CASSCF I[a]	CASSCF II[b]	CASSCF III[c]	CCI[d]	DAV[e]	CEPA[f]	exptl
			$^1n\pi^*$				
ΔE	6.2	7.2		6.0	5.8	6.0	5.7[g]
			$^3n\pi^*$				
ΔE	5.9	6.9		5.7	5.5	5.7	5.3[h]
			$^1\pi\pi^*$				
ΔE	8.6	9.3	8.6	8.2	7.8	7.7	7.3[g]
$\lvert m\rvert$[k]	1.18	1.36	1.58				1.44[i]
ϕ_{CN}[l]	21.0°	22.6°	23.3°				16.6°[j]
f[m]	0.25	0.33	0.45				0.37[i]
			$^3\pi\pi^*$				
ΔE	6.2	6.9	5.5	6.0	5.9	6.0	(6.6)[h]

[a-c] See corresponding footnotes in Table I. [d] Contracted CI results with reference configurations and orbitals from the CASSCF I calculation. [e] As in footnote *d* with Davidson correction. [f] Multireference CEPA results. [g] Reference 36. [h] Reference 37; the value of 6.6 eV for $^3\pi\pi^*$ is very uncertain. [i] Reference 38. [j] Reference 39, angle determined in myristamide. [k] $|m|$: magnitude of the transition moment. [l] See footnote *d* in Table I. [m] Computed using the experimental excitation energy.

Table III. Cytosine Ground-State CASSCF Total Energy and Dipole and Quadrupole Moments

	E_T[f]	μ_x	μ_y	$\lvert\mu\rvert$	ϕ	Q_{xx}	Q_{yy}	Q_{zz}	θ
CASSCF1[a]	−392.5711	2.8717	0.6868	7.5	104.2°	−9.36	14.22	−4.87	−13.8°
CASSCF2[b]	−392.5922	2.8948	0.7116	7.6	104.5°	−9.50	14.39	−4.89	−13.8°
CASSCF3[c]	−392.7645	2.7552	0.8303	7.3	107.5°	−9.32	14.86	−5.54	−13.5°
CASSCF4[d]	−392.7766	2.6900	0.7611	7.1	106.5°	−9.01	14.39	−5.39	−13.8°
CASSCF5[e]	−392.7828	2.6720	0.7669	7.1	106.7°	−8.86	14.41	−5.56	−13.2°

[a] Basis set B, 8π orbitals active. [b] Basis set B, 10π orbitals active. [c] Basis set C, 10π orbitals active. [d] Basis set C, n(N), σ^*, 8π orbitals active. [e] Basis set C, n(O), σ^*, 8π orbitals active. [f] E_T, total energy in atomic units; $\mu_{x,y}$, components of the dipole moment in atomic units; $|\mu|$, dipole moment in Debye; ϕ, angle that $|\mu|$ makes with the line N_1–C_4 (Devoe and Tinoco convention, ref 41, see Figure 1b); $Q_{xx,yy,zz}$, components of the quadrupole moment in atomic units in the principal axis system. θ is the counterclockwise angle this system has with the coordinate system given in Figure 1b.

with a summed weight of around 3% in the CAS wave function.

The CASSCF III calculation is with these results in mind clearly less accurate, since it does not involve any active σ orbitals. The calculated dipole moments also deviate somewhat from the presumably more accurate results obtained with the CASSCF I and II wave functions. We notice, however, that the variations in the computed angle of the dipole moments for the different states are only a few degrees.

The computed ground-state dipole moment is in good agreement with the experimental value,[33] both with respect to the magnitude and the angle (ϕ_{CN}) it makes to the CN bond. This is true for all three sets of calculations, even if the error in the magnitude obtained from CASSCF III is slightly larger. A recent fourth-order MBPT calculation, with all valence electrons correlated, gave a dipole moment of 3.74 D with $\phi_{CN} = 40.1°$, in perfect agreement with experiment. A large polarized basis set of the type (C, N, O: 10s, 6p, 4d/5s, 3p, 2d)/(H: 6s, 4p/3s, 2p) was used in these calculations.[34] This study together with the experimental values confirm that the CASSCF III calculation (being of the same size as the study of cytosine) reproduces reasonably well the ground-state dipole moment of formamide. Also the ground-state quadrupole moment compares well with the rather uncertain experimental estimate.[35] The best results are obtained with the CASSCF II have function (cf. Table I). Dipole and quadrupole moments are also given for the excited states. No experimental data are available for comparison.

The computed excitation energies from the CASSCF and the MR-CCI and CEPA calculations based on the CASSCF I reference configurations are displayed in Table II. It also includes the transition moment and the oscillator strength for the $^1\pi\pi^*$ transition computed at the CASSCF level of theory. The best CASSCF results are obtained in calculation I. Errors in computed transition energies vary between 0.5 and 1.3 eV, the error being largest for the $^1\pi\pi^*$ transition. The same error is obtained in calculation III. We notice that the inclusion of σ orbitals yields smaller transition moments, while there is only a minor effect on the direction. The error in this angle varies between 4 and 6° in the three calculations.

Inclusion of dynamic correlation through MR-CCI and CEPA calculations, diminishes the errors in computed excitation energies to around 0.4 eV. The experimental value of 6.6 eV for the $^3\pi\pi^*$ transition is very uncertain. It seems that the trapped electron spectrum failed to properly account for this electronic state.[37]

As the results of Table II show, the value of the transition moment increases in going from CASSCF I to CASSCF III. Also the angle ϕ_{CN} increases somewhat. The largest difference from the angle measured in myristamide[39] is 6.7°. The experiment is made in the solid phase and gives two possible angles: 16.6 and 52.4°. Thus the present CASSCF calculations confirm the assignment of Peterson and Simpson: an angle of 9.1° with the N–O bond towards the C–N bond, which gives an angle of 16.6° relative to the C–N bond. A recent CNDO/CI calculation[40] on myristamide yields the angle, $\phi_{CN} = 31.2°$ with an oscillator strength 0.123, which is one-third of the experimental value.

The conclusion to be drawn from these test calculations on formamide is that it should be possible to reproduce the transition moment directions in cytosine with an error not larger than 5–10°, and oscillator strengths with errors of the order of 20–25%. It is interesting to notice that the CASSCF III value for the oscillator strength is around 20% larger than the experimental value. A similar error is found for the strongest $\pi\pi^*$ transition in cytosine. The computed transition energies are expected to be larger than experiment with errors ranging from a few tenths of an eV to between 1 and 2 eV. These conclusions are fully consistent with results obtained in previous studies of the π electron spectra of benzene and naphthalene.[19,20]

b. Cytosine Ground-State Properties. The results of the five different CASSCF calculations of the ground-state properties of cytosine are presented in Table III. The lowest energy is obtained

---

(33) Kurland, R. J.; Wilson, E. B., Jr., *J. Chem. Phys.* **1957**, *27*, 585.

(34) Matos, J. M. O.; Sadlej, A. J.; Roos, B. O., to be published.

(35) Flygare, W. H.; Benson, R. C. *Mol. Phys.* **1971**, *20*, 225.

(36) Robin, M. R. *Higher Excited States of Polyatomic Molecules*; Academic Press: New York, 1975; Vol. II.

(37) Staley, R. H.; Harding, L. B.; Goddard, W. A., III; Beauchamp, J. L. *Chem. Phys. Lett.* **1975**, *36*, 589.

(38) Bash, H.; Robin, M. B.; Kuebler, N. A. *J. Chem. Phys.* **1968**, *49*, 5007.

(39) Peterson, D. L.; Simpson, W. T. *J. Am. Chem. Soc.* **1955**, *77*, 3929, **1957**, *79*, 2375.

(40) Volosov, A. *J. Chem. Phys.* **1987**, *87*, 6653.

Table IV. Cytosine Ground-State MBPT Total and Correlation Energy, Dipole Moment, and Polarizability

	$E_T{}^a$	E_{corr}	μ_x	μ_y	$\lvert\mu\rvert$	ϕ	α_{xx}	α_{yy}
SCF	−392.4931		3.3642	0.8100	8.8	104.2	68.75	80.00
			Excitations $\pi \rightarrow \pi^*$					
MBPT2[b]	−392.5685	−0.0754	3.1900	0.8171	8.4	105.1	72.20	82.80
MBPT3	−392.5907	−0.0976	3.1900	0.8159	8.1	105.4	74.23	84.30
MBPT4	−392.5997	−0.1066	3.0524	0.8093	8.0	105.5	75.50	85.00
			Excitations: $\pi \rightarrow \sigma^*,\pi^*$					
MBPT2	−392.5800	−0.0869	3.1790	0.9032	8.4	106.6	72.45	83.03
MBPT3	−392.6029	−0.1098	3.0892	0.9464	8.2	107.7	74.25	84.50
MBPT4	−392.6117	−0.1186	3.0424	0.9662	8.1	108.3	75.50	85.25
			Excitations: $2n \rightarrow \sigma^*,\pi^*$; $\pi \rightarrow \sigma^*,\pi^*$					
MBPT2	−392.6608	−0.1677	3.0729	0.7956	8.1	105.2	72.66	84.41
MBPT3	−392.6842	−0.1911	3.0160	0.8166	7.9	105.8	73.50	85.25
MBPT4	−392.6969	−0.2038	2.9624	0.8166	7.8	106.1	74.50	86.25

[a] See footnote *f* in Table III; E_{corr}, correlation energy in atomic units; $\alpha_{xx,yy}$, in-plane components of the polarizability. [b] MBPT*X* gives the results obtained in order *X* of the perturbation.

with the CASSCF5 wave function, showing that larger correlation effects are obtained (with two σ orbitals active) for the oxygen lone pair than for the nitrogen lone pair.

The magnitude of the dipole moment is around 7.5 D with basis B. With basis C, which includes polarization functions, the value drops to 7.3 D when only π electrons are correlated, while correlating also one σ lone pair gives a slightly smaller value. As was the case in formamide, the direction of the dipole moment remains almost the same at the different level of approximation. The direction is here given in the convention of Devoe and Tinoco[41] as the counterclockwise angle it makes with the line connecting N_1 and C_4 (see Figure 1b).

The fourth-order MBPT calculations were performed using basis set B, and the results are presented in Table IV. The first set of results given in the table is obtained by including only excitations among the π orbitals into the intermediate states of the MBPT expansion. These results can be directly compared to the CASSCF1 and -2 results in Table III. The small decrease in energy obtained at the MBPT4 level of theory is more a reflection of the limitations in the basis set than a measure of the accuracy of the CASSCF results. Nevertheless, it is interesting to note that the CASSCF calculations almost are at the limit of the basis set usef for the π orbitals. The effect on the density matrix is, however, larger as can be seen from the change of the dipole moment, where the CASSCF2 calculations give 7.6 D while MBPT4 gives 8.0 D. Again, the effect on the direction of the dipole moment is small.

The second set of results given in Table IV is obtained by allowing also excitations from the occupied π orbitals into empty σ orbitals. As can be seen from the table, these excitations influence the results to a very small extent. More important is the inclusion of excitations from the two σ lone pairs into empty σ and π orbitals. The effect is to decrease the magnitude of the dipole moment with 0.2 D. Exactly the same decrease was obtained when going from CASSCF3 to CASSCF5. Larger MBPT calculations than those presented in Table IV could not be performed due to program limitations. These calculations are anyway among the largest full fourth-order MBPT calculations performed to date.

The MBPT4 calculations were performed with basis B. The extension of the basis set to include polarization functions was shown to decrease the dipole moment with around 0.3 D, increasing the direction of the angle with 3° (compare CASSCF2 and -3 in Table III). From these results we therefore predict the dipole moment of cytosine to be 7.5 ± 0.5 D, where the error bars are due to basis set limitations and correlation effects, mainly among the σ orbitals. The angle of direction is predicted to be109 ± 3°.

The diagonal in plane polarizability tensor components are also given in Table IV. We notice that the correlation effects recovered in the MBPT4 calculations are small and almost entirely due to π electron excitations. The out-of-plane component was not calculated, since the basis set is too small to allow a reasonable estimate to be made.

The quadrupole moment tensor was computed on the CASSCF level of theory. the results are presented in Table III. The different choices of active space have only a very small effect on computed values for the tensor components, while the enlargement of the basis set (from B to C) leads to a small increase of the magnitude of the Q_{zz} component. The values obtained from the CASSCF5 wave function are probably rather close to the limits obtainable with basis set C.

c. π-Electron Excited States in Cytosine. The three lowest π–π^* excited singlet states have been investigated at the CASSCF level of theory. The same active spaces and basis sets have been used as already described for the ground state. The results of the five different calculations are presented in Table V, which gives the computed excitation energies, transition moment components and directions (using the convention of Devoe and Tinoco[41], and the oscillator strength for each of the three transitions. Experimental results from solution,[42] crystal,[4,6],[7] and stretched film[9,10] spectra are also displayed.

The cytosine spectrum in solution shows different features depending on the solvent. The spectrum in water has three broad, diffuse, and overlapping bands with maxima around 4.6, 5.4, and 6.3 eV. In addition a shoulder is found at 5.9 eV.[13] On the other hand, Morita and Nagakura[42] find only two maxima (4.6 and 6.3 eV) and one shoulder at 5.4 eV, using the same solvent. Common to all spectra is a strong overlap between the different bands which makes a resolution into individual components difficult. A number of n–π^* transitions are probably also located to the same energy region[15,43] (see also below). Even if they normally are considerably weaker than the π–π^* transitions, it cannot be excluded that they give rise to some minor features appearing in the absorbtion band. Attempts to record the electronic spectrum of cytosine in the gas phase have failed owing to decomposition of the molecule.[13]

The small basis set B calculations show that the effect of increasing the π active space from 8 to 10 orbitals (CASSCF1 and -2 in Table V) has only minor effects on the computed properties. Adding polarization functions (basis C) using again 10 active π orbitals (CASSCF3) changes the properties somewhat more, especially the directions of the transitions. The CASSCF4 calculations shows the effect of correlating the nitrogen lone pair. Here the only noticeable effect is on the transition moment for the third transition. The angle decreases 9.4° and the oscillator strength increases with 0.24. The effects are somewhat smaller when the oxygen lone pair is correlated. In this case the comparison is made between the CASSCF3 and -5 calculations. The largest effect is here on the transition moment of the first transition, where the angle is decreased by 5.4° while the oscillator strength by increases from 0.04 to 0.10. CASSCF5 also leads

---

(41) Devoe, H.; Tinoco, 1. *J. Mol. Biol.* **1962**, *4*, 518.

(42) Morita, M.; Nagakura, S. *Theor. Chim. Acta* **1986**, *11*, 279.

(43) Rich, A.; Kasha, M. *J. Am. Chem. Soc.* **1960**, *89*, 6197.

Table V. Cytosine Lower π–π^* Transitions: CASSCF Excitation Energy, Transition Moment Components, and Direction and Oscillator Strength, Compared with Experimental Results

	transition I		transition II	transition III
			CASSCF1[a]	
ΔE[f]	4.8		6.5	7.9
m_x	−0.6266		0.0955	−1.0881
m_y	0.1865		0.8384	−1.9601
ϕ	74.1°		−5.6	−28.3°
f	0.01		0.11	0.97
			CASSCF2[b]	
ΔE	4.8		6.5	7.7
m_x	−0.6389		0.0702	−1.1255
m_y	0.1614		0.8250	−1.9759
ϕ	76.5°		−4.2°	−29.0°
f	0.05		0.11	0.98
			CASSCF3[c]	
ΔE	5.0		6.6	8.2
m_x	−0.5542		0.1600	−1.1695
m_y	0.2437		0.8630	−1.9257
ϕ	67.0°		−9.8°	−30.6°
f	0.04		0.12	1.02
			CASSCF4[d]	
ΔE	5.0		6.8	8.5
m_x	−0.5431		0.2066	−1.5999
m_y	0.2278		1.0079	−1.8633
ϕ	67.9°		−10.9°	−40.0°
f	0.04		0.18	1.26
			CASSCF5[e]	
ΔE	5.6		6.7	8.3
m_x	−0.5842		0.1662	−1.1496
m_y	0.3127		0.9412	−1.9556
ϕ	62.5°		−9.3°	−29.7°
f	0.10		0.15	0.99
			Experimental[f]	
ΔE	4.6[g]	5.3[g]	5.8[g]	6.3[g,k]
	4.7[h]		5.4[h]	
	4.4[i]		5.5[i]	
	4.6[j]		5.2[j]	
	4.7[k]		5.4[k]	
ϕ	(6°,54°)[g]	(−46°)[g]	(76°,−17°)[g]	(86°,−27°)[g]
	(9°,51°)[h]		(−11°)[h]	
	(10°,52°)[i]		(−5±3°)[i]	
	(25 ± 3°,−46 ± 4°)[j]		(6±4°,−27 ± 3°)[j]	
	(35 ± 14°,−28 ± 7°)[l]		(15±20°,−6 ± 16°)[l]	
	(9°)[m]		(7.5°,10.5°)[m]	(94°,−76°)[m]
f	0.14[g]	0.03[g]	0.13[g]	0.36[g]
	0.13[i]			0.68[k]
	0.11[k]			

[a–e] See corresponding footnotes in Table III. [f] ΔE, excitation energy in eV; $m_{x,y}$, components of the transition moment vector in atomic units; ϕ, see footnote in Table III; f, oscillator strength. [g] Crystal polarized reflection spectra, ref 4. [h] Crystal polarized fluorescence and reflection spectra, ref 6. [i] Crystal polarized absorption spectra, ref 7. [j] Stretched film, UV linear dichroism spectra, ref 10a. [k] Solution (water) absorption spectrum, ref 42. [l] Same as j, ref 9. [m] Same as j, ref 10b.

to a general increase, varying between 0.1 and 0.6 eV, of the transition energies.

It is tempting to add the changes in computed properties arising from the three changes in the active space: (1) increase of the active π orbitals from 8 to 10, (2) correlating the nitrogen (N_3) lone pair, and (3) correlating the oxygen (O_2) lone pair. The additional σ and π lone pairs correlated, when the active space is changed, starting with 8π, are well separated (N_1 and N_4 π lone pairs and N_3 and O_2 σ lone pairs), and it is reasonable to assume an additive effect in introducing a correlating orbital for each of them. Adding the corrections to the excitation energies leads to the values 5.6, 6.9, and 8.1 eV, respectively, for the three transitions. This result is almost identical with that obtained from the CASSCF5 calculations. Additivity was assumed for the changes in the components m_x and m_y of the transition moments, and new directions and oscillator strengths were computed from the corrected values. The results for the oscillator strengths are 0.06, 0.19, and 1.31, respectively. These values have been obtained using the transition energies given above. The error is here between 1 and 2 eV. If we use instead the experimental estimates of the energies, the values of the oscillator strengths are reduced somewhat: 0.05, 0.15, and 1.03. The largest effect is on the oscillator strength for the third transition. The corresponding transition moments are the quantities which are affected most profoundly by the correlation of the σ lone pairs. Adding up the corrections to the transition moment components leads to the following angles: 68.1, −8.0, −39.7°. Again this is not very different from the results of the CASSCF5 calculations.

It is well known that π electron transition energies computed at the present level of theory are between 0 and 2 eV too large.[19,20] This was also illustrated above for the formamide molecule. There are three major reasons for the error. (1) Basis set effects. Larger basis sets are generally needed to describe the more diffuse molecular orbitals of the excited states. Thus the error is positive. With the present basis C we estimate this error to be less than 0.5 eV. (2) Dynamic correlation between the π electrons. This effect is normally also positive, due to the increase of the ionic character of the wave function for the excited state. From earlier experience we expect this effect to be a few tenth of an eV. (3) Dynamic correlation between the σ and π electrons. This effect is important especially for excited states with a high degree of ionic character in the wave function. The earlier study of benzene[19] showed that it could be as large as around 1 eV in some cases. The added effect of reasons 2 and 3 was in the formamide molecule found to be 0.9 eV (compare CASSCF I and CEPA in Table II).

The experimental estimates of the energy for the first transition varies between 4.4 and 4.7 eV.[4,6,7,10,42] The theoretical estimate is 5.6 eV. Thus the error is around 1 eV. The second π–π^* transition appears only as a shoulder in some of the spectra. Energies given vary between 5.2 and 5.4 eV.[4,6,7,10,42] The present estimate is 6.9 eV, which is around 1.5 eV too large. Finally the third and most intense transition has the maximum in the range 6.3–6.5 eV.[4,42] The corresponding theoretical estimate is 8.1 eV. Thus all errors are in the expected range. It should, however, be remembered that all experiments are from solution, crystal, or stretched film measurements, while the theoretical values have been derived for a single isolated molecule. Intermolecular interactions will affect both transition energies and transition moments.

Zaloudek et al.[4] have interpreted the absorbtion spectrum of cytosine dissolved in water and the reflection spectrum of the monohydrate crystal as composed of four different π–π^* transitions in the energy range 4.3–6.5 eV, with peaks at 4.6, 5.3, 5.8, and 6.4 eV. This interpretation is not consistent with the present results; the intense band around 6.5 eV clearly corresponds to the third calculated transition at 8.1 eV. Computed intensities also show that the bands at 4.6 and 5.8 eV correspond to the two lower of the excited states. In order to explain a fourth weak transition ($f \sim 0.03$), we would therefore have to assume that the computed energy is in error with more than 3 eV, which is highly unlikely. Earlier studies of π electron spectra in ethylene, fomaldehyde, benzene, naphthalene, and formamide at the same level of theory have in no case resulted in errors larger than 2 eV (ref 19 and 20 and unpublished results). It is not unlikely that the features in the absorption band found by Zaloudek et al.[4] are due to unsymmetrical shapes of the individual absorption bands or/and underlying n–π^* transitions, which borrows in plane intensity through vibronic coupling to the π–π^* transitions.

In order to check this further, we have also calculated the electronic state corresponding to an excitation from the oxygen lone pair to a π orbital. Basis set C was used and the same active space as was used in the CASSCF5 calculation. The calculated transition energy was 5.97 eV, which place it between the first and the second of the π electron states. The computed intensity was very small ($f \sim 5 \times 10^{-4}$). Assuming the error in the energy to be the same as in formamide, ~0.5 eV, places this state very close to the second $\pi\pi^*$ singlet, indicating a possibility of vibronic coupling between the two states.

It may be argued that within the nonlinear MCSCF method it might be difficult to decide whether all states within a given energy interval have been determined. The normal way of controlling this would be to compute the number of negative eigenvalues of the Hessian matrix, a possibility which does not exist with the CASSCF program used in the present work. In order to check that we have not bypassed an energy level in the CASSCF calculation, we have performed a calculation where the same orbitals were used for all states. The orbitals were determined in a CASSCF5 calculation by optimizing the average energy of the *five* lowest states of $^1A'$ symmetry (the ground state and *four* $\pi\pi^*$ excited states). The excitation energies computed with these orbitals were 5.2, 6.5, 8.1, and 8.4 eV. The similarity between these energies and those obtained in the independent CASSCF5 calculations (cf. Table V) shows clearly that no state has been missed. We note that the fourth excited state has an energy 0.3 eV above the third state. The computed, oscillator strength is 0.14 ($\theta = -87°$). The corresponding absorption band has been found at around 180 nm, in the TMP solution spectrum recorded by Clark and Tinoco.[12]

The oscillator strengths for the three first excited states obtained with the average orbitals are 0.06, 0.07, and 0.49 (θ = 66, 2, and -44°). A comparison with the CASSCF5 results in Table V shows that orbital optimization for each state independently has a larger effect on the transition moment directions than on the excitation energies. The differences are, however, not very large. An average orbital CASSCF calculation is therefore of interest for a first scan over the excited states. The orbitals should also be good as starting orbitals in the state specific calculations, where convergence problems often are encountered.

d. Transition Moments in Cytosine. The most important aspect of the present work is the calculation of the transition moments for the π–π^* transitions. Much experimental work has been devoted to the determination of the transition moment directions in nucleic acids and polynucleotides, since these quantities are necessary for an understanding of conformational properties of the nucleic acids.[1,2,4,10,44] As a first step towards understanding these properties, attempts have been made to determine them for the purine and pyrimidine bases. The most common methods used are polarized reflection spectra on single-crystal faces[2,4,6,7] and linear dichroism (LD) studies on stretched films.[9,10] A basic difficulty with both these methods is that two different directions are consistent with the experimental data for each transition. The strong overlap between the different bands also results in considerable ambiguity in the interpretation. There is thus a strong need for trustworthy theoretical work which can be used to help interpret the experimental data. This work is aimed at presenting such results.

While the CASSCF method, as discussed above, yields transition energies which are too large, the corresponding transition moments are generally in good agreement with experiment.[19,20] This has been illustrated in the present work by the results obtained for formamide as well as in earlier studies of excited states in benzene and naphthalene.[19,20] In formamide the direction of the moment was found to be stable with respect to changes in the basis set and active space. The variation was only 2° and the difference from experiment was around 6°.

In cytosine we find for the first transition an angle varying between 67.9 and 62.5°, when basis set C is used. The smaller basis B yields a slightly larger angle, ~75°. Our final estimate is 68°, based on the analysis of the contributions to the transition moment from the different correlation effects, as discussed in the previous section. Supported by earlier experience we estimate the error in this quantity to be not larger than ±10°. The LD studies of Matsuoka and Nordên[10a] on stretched film yields two possible values: 25 ± 3 or -46 ± 4°. Two different values, +9 or +51°, were inferred from the reflection spectra of the (001) face of cytosine monohydrate by Callis and Simpson.[6] Based on consistency arguments with data from 1-methylcytosine, they chose the angle +9° with an estimated accuracy of ±10°. Lewis and Eaton[7] used direct absorption measurements and obtained the angles +10 and +52°. Finally Zaloudek et al.[14] derived the values +6 and +54° from their polarized reflection spectra. In all cases the small angle was preferred. This interpretation is in conflict with the present result: considering the difference between an isolated molecule and the monohydrate crystal, the agreement between the larger angle and the theoretical value 68 ± 10° must be considered satisfactory. It is highly unlikely that the error in the computed value should be as large as ~60°.

Turning now to the second transition, it can be seen that the angle varies between -9.3 and -10.9° for the three CASSCF studies made with basis set C (CASSCF3–5). The smaller basis set yields an angle which is ~4° larger. Our final estimate for this angle is -8 ± 10°. The crystal data of Zaloudek et al.[4] yields two possible angles: +76 and -17° (note: this is their transition III; transition II does not correspond to a π electron state, as discussed earlier in this work). Based on a comparison with the LD data of Matsuoka and Nordên[10] and Bott and Kurucsev,[9] the large angle is selected. This again disagrees with the theoretical value. The negative angle agrees, however, well with the present result. It is highly unlikely that the theoretical transition moment is perpendicular to the true moment of the free molecule, and the large positive angle can therefore safely be discarded.

The moment of the third transition has only been studied in the monohydrate crystal.[4] The two possible angles are -27 or +86°. The present estimate is -39°, but CASSCF5 gives -30° while CASSCF3 gives -40°. Thus the effect of correlating the nitrogen σ lone pair is to decrease the angle with around 10°. The agreement with the experimental angle (-27°) must be considered as satisfactory, and the other angle (+86°) is clearly inconsistent with the present results.

Summarizing, the computed angles 68 ± 10, -8 ± 10, and -40 ± 10° are clearly consistent with the experimental values of Zaloudek et al. if the experimental angles chosen are 54, -17, and -27°. The agreement with the crystal spectra of ref 6 for the first two transitions is also acceptable. Our result can also be brought into agreement with the stretched film experiment of Matsuoka and Nordên,[10a] provided that their orientation angle is chosen as around +20° instead of -11°. Using the same orientation angle also brings the results of Bott and Kuruscev[9] into reasonable agreement with the present theoretical angles. Thus a consistent picutre emerges, where both crystal and stretched film experiments and theory agree on the transition moment directions in the cytosine molecule.

Zaloudek et al.[4] have also presented values for the oscillator strengths derived for the free molecule. The data are somewhat obscured by their interpretation of the spectrum as consisting of four bands, but if the intensities of the transitions II and III are added, the following values are obtained: 0.11, 0.14, and 0.38.

The theoretical estimates, based on the experimental transition energies, are 0.05, 0.15, and 1.28. The first two of these values are in as good agreement with experiment as can be expected considering the uncertainties in the calculations and the differing experimental conditions. As can be seen in Table V the variations of the oscillator strengths between the five different calculations are much larger than the corresponding variations in the directions. The intensity of the third transition is computed to be three times as large as the value given by Zaloudek et al.[4] Based on our earlier results for benzene and naphthalene[19,20] and the present study of formamide, we expect a somewhat smaller error and deduce that the free molecule value derived by Zaloudek et al. is probably too small.

e. Comparison with Other Theoretical Studies. A number of semiempirical studies of the spectral properties of the purine and pyrimidine bases have appeared in the literature.[45-49] The early

---

(44) For a review see: R. W. Woody, *J. Polym. Sci.* **1977**, *12*, 181.

(45) Berthod, H.; Prettre, C. G. *Theor. Chim. Acta* **1966**, *5*, 53.

(46) Hug, W.; Tinoco, I., Jr., *J. Am. Chem. Soc.* **1973**, *95*, 2803.

(47) Ito, H.; I'Haya, Y. J. *Bull. Chem. Soc. Jpn.* **1976**, *49*, 3466.

(48) Danilov, V. I.; Pechenaya, V. I.; Zheltovsky, N. V. *Int. J. Quantum Chem.* **1980**, *17*, 307.

Table VI. CASSCF Excitation Energy (ΔE) in eV, Oscillator Strength (f), and Transition Moment Directions for the π–π^* Transitions of Cytosine Compared with Previous Theoretical Work

	transition I			transition II			transition III		
semiempirical	ΔE	f	ϕ[a]	ΔE	f	ϕ[c]	ΔE	f	ϕ[a]
PPP–SCF[b]	4.2	0.1	85°	5.3	0.4	−16°	6.0	0.5	−47°
CNDO–CI[c]	4.7	0.14	18°	5.8	0.16	−24°	6.6	0.30	−45°
RPA[d]	4.6	0.16	45°	6.0	0.17	−13°	6.5	0.71	−43°
CNDO/S–CI[e]	4.6	0.15	40°	5.6	0.14	−13°	6.4	0.14	−40°
VE–PPP[f]	4.2	0.15	65°	5.3	0.12	−56°	6.3	1.14	−32°
ab initio									
RPA[g]	5.9	0.11	50°	7.3	0.09	−52°	7.1	0.43	−18°
CASSCF[h]	5.6	0.05	68 ± 10°	6.9	0.15	−8 ± 10°	8.1	1.03	−40 ± 10°

[a] See corresponding footnote in Table III. [b] Reference 45. [c] Reference 46. [d] Reference 47. [e] Reference 48. [f] Reference 49. [g] Reference 15. [h] Estimated from the results of the five calculations.

work was reviewed in 1966.[14] These studies employ various variants of the Pariser–Parr–Pople, the CNDO, and the random phase approximations. Recently an ab initio RPA study of cytosine has been published by Jensen et al;[15] which is made using a basis set similar to basis B of this study. The results can therefore be directly compared to the CASSCF1 and -2 calculations.

It might be of some interest to compare these earlier theoretical studies with the results obtained in the present work. The results obtained here can be viewed as a link between the semiempirical studies and the experiments, since a direct comparison is difficult due to the ambiquities involved in the interpretation of the experimental data. We also believe that the estimates made in this work are correct within the error bars given in the preceding discussion.

The semiempirical work gives a consistent description of the transition energies in good overall agreement with experiment (cf. Table VI). All these studies[45–49] predict the existence of two weak (f = 0.1–0.2) and one strong (f = 0.3–1.1) transitions in the energy interval of the absorption band found for cytosine (4.5–6.5 eV), in complete agreement with the present work. The weakness of the semiempirical work is, however, clearly illustrated in the results obtained for the transition moment directions.

Thus for transition I, different semiempirical estimates yield values varying between 85 (PPP-SCF) and 18° (CNDO–CI). More consistent results have been obtained for transition II with a predicted angle varying between −13 and −24° in all but one of the semiempirical calculations (cf. Table VI). These results are also in reasonable good agreement with the present ab initio result, −8°. The angles reported for the third transition are rather close to the ab initio value and do not vary much between the different semiempirical methods used. In all, these semiempirical estimates of the transition moments in cytosine give support to the present interpretation of the experimental studies.

It is difficult to have a decided conception about the sources of the errors encountered in semiempirical calculations of transition moment directions. One important factor is probably a change in the σ electron distributions, which is not accounted for in the PPP method. For example, the CASSCF5 calculations shows that the third $\pi\pi^*$ excitation leads to a loss of 0.15 σ electron on N_4 and a corresponding increase on other atoms (mainly C_2 and C_4). Similar redistributions occur for the other excitations. The polarization of the σ electrons thus has a profound influence on the computed transition moments, and a method that does not include this effect can in general not be expected to yield reliable results.

It is not equally straightforward to analyze the contrasting results obtained with the CNDO–CI method.[46,48] However, the minimal basis set and the severe approximations involved in the integral evaluation can easily destroy the balance needed for a correct evaluation of the transition moments.

The ab initio RPA calculation[15] gives results which are neither consistent with the present results nor with the earlier semiempirical RPA study.[47] The calculated intensities indicate a reversal of the order of the second and the third transition. A fourth π–π^* transition was reported at 7.5 eV with an oscillator strength of 0.41 and a direction $\phi = -44°$. More interesting is maybe that RPA calculation also yields five σ–π and π–σ transitions in the energy range 6.4–7.0 eV, with a summed intensity corresponding to f = 0.02. If this result is correct it could explain the occurrence of a fourth band in some of the solution spectra for cytosine, which then should be composed of a number of σ–π and π–σ transitions.

4. Conclusions

The present work is the first complete ab initio study of the properties of cytosine in its ground state and the three lowest π–π^* excited states. The calculations have been performed with an extended and well-balanced AO basis set including polarization functions on the heavy center. The most important correlation effects have been included into the wave functions through the use of the complete active space SCF method. The error estimates of computed properties, especially the transition moments, are based on earlier studies of aromatic systems and on the present test study of the formamide molecule.

There is in *no* case full agreement between computed transition moment directions and the predictions based on experimental spectra. Our conclusion is that the interpretations made on the basis of experimental data are incorrect, since we believe that the error estimates made on the theoretical data are well founded. This theoretical study therefore ought to be of value for future experimental studies of the optical properties of cytosine. Our aim is to continue this line of research and extend the investigations to include also other pyrimidine and purine bases.

Acknowledgment. We thank Professor A. J. Sadlej for helping with the MBPT calculations, and Professor Bengt Nordén and docent G. Karlström for fruitful discussions. The research reported in this communication has been supported by a grant from the Swedish Natural Science Research Council (NFR) and by IBM Sweden under a joint study contract. J.M.O.M. acknowledges support from the Brazilian Agency CAPES.

Registry No. Cytosine, 71-30-7.

---

(49) Srivastava, S. K.; Mishra, P. C. *Int. J. Quantum Chem.* **1980**, *18*, 827.