Theoretical Study of the Electronic Spectrum of Cytosine

M. P. Fülscher and B. O. Roos*

Contribution from the Department of Theoretical Chemistry, Chemical Centre, P.O. Box 124, S-221 00 Lund, Sweden

Received September 15, 1994[®]

Abstract: *Ab initio* quantum chemical methods based on multiconfigurational second order perturbation theory have been employed to compute the electronic excitation spectrum of cytosine in the gas phase. The computed vertical $\pi \rightarrow \pi^*$ transition energies for the four lowest singlet excited valence states are 4.4, 5.4, 6.2, and 6.7 eV, in agreement with experiment (measured energies are 4.6−4.7, 5.2−5.8, 6.1−6.4, and 6.7−7.1 eV, respectively). It is also suggested that the lowest n $\rightarrow$ π^* transition should be seen at energies close to 5.0 eV. The computed polarization directions for the three lowest $\pi \rightarrow \pi^*$ transitions are approximately 60°, −5°, and −40° and assign a large value for the first angle. This is in contradiction to experiment, which gives a small angle. The crystal field effects on the angles can, however, be expected to be large.

Introduction

Detailed knowledge of the electronic structure of nucleic acids and related model compounds is an important prerequisite for our understanding of the static and dynamic properties of polynucleotides. For this purpose, reliable data concerning the excitation energies, oscillator strengths, and transition moment directions in a molecule-fixed reference frame are needed. However, the spectra of the nucleic acids consist of rather braod, strongly overlapping bands. The excitation energies of the strong absorption lines are known experimentally, but information is missing concerning the weak transitions. Moreover, the transition moment directions are difficult to assess unambiguously from experiment. There is thus a marked need for support from theoretical studies. For these reasons we have undertaken a project aimed at the determination of the $\pi \rightarrow \pi^*$ and n $\rightarrow$ π^* excitation energies and transition moments of cytosine. The study is part of a larger project to determine the spectral properties of all the base monomers of DNA and RNA.[1]

Over the years a wealth of information about the electronic spectrum and transition moments of cytosine has been gathered. An extensive review of this work has been published by Callis.[2] There is general agreement concerning the characteristics of the three bands located at 4.5, 5.4, and 6.3 eV, of which the first two are weak oscillators. Because of the broad, overlapping absorption bands and potentially large solvent effects, it is difficult to resolve the 5.0−6.0 eV energy region into unique transitions. By combining polarized reflection spectra and stretched film dichroism data, Zaloudek *et al.*[3] arrived at a model consisting of four transitions. The new assignment puts two bands at 5.3 and 5.7 eV instead of only one at 5.4 eV, and the suggested polarization angles of the three lowest transitions are +6, −46, and −76°. This proposal contrasts with a large body of theoretical evidence which favors a polarization angle of ≈70° for the first transition. Semiempirical calculations,[5−7]

using numerous methods and parametrizations, consistently predict two weak and one strong oscillators with excitation energies less than 6.5 eV. On the other hand, the majority of the *ab initio* calculations[8−11] seem, at first glance, to favor the four transition model or even a larger number of transitions.

In an effort to bring new information to bear on these questions and to alleviate some of the contradictions, we report *ab initio* quantum chemical calculations using two different basis sets and methods based on the multiconfigurational approach, that is, the complete active space self-consistent field (CASSCF)[12] method. Due to its inherent flexibility the CASSCF approximation has proven to be particularly suited to cope with situations where the electronic structure varies strongly and accounts for near-degeneracy effects common to many excited states.

A second motivation stems from the fact that all of the previous *ab initio* calculations neglected the dynamic $\sigma-\pi$ polarization effects, which are known to be important for an accurate prediction of excitation energies. Traditionally, these correlation effects have been obtained by supplementing the CASSCF calculation with an MRCI calculation, which includes all the σ-electrons in the correlation treatment. Such an approach is, however, hardly feasible for molecular systems of the present size. Alternatively, it has been suggested that the remaining dynamic correlation effects can be estimated by means of second-order perturbation theory with a CASSCF wave function constituting the reference state. The approach, called the CASPT2 method,[13,14] has proven its validity in calculating differential correlation effects on excitation energies in a number of applications[15−26] (for a recent review of the applications of

® Abstract published in *Advance ACS Abstracts,* February 1, 1995.
(1) Lorentzon, J.; Fülscher, M. P.; Roos, B. O. To be published.
(2) Callis, P. R. *Annu. Rev. Phys. Chem.* **1983**, *34*, 329.
(3) Zaloudek, F.; Novros, J. S.; Clark, L. B. *J. Am. Chem. Soc.* **1985**, *107*, 7344.
(4) Hug, W.; Tinoco, I., Jr. *J. Am. Chem. Soc.* **1973**, *95*, 2803.
(5) Callis, P. R. *Photochem. Photobiol.* **1986**, *3*, 315.
(6) Fischer-Hjalmars, I.; Henriksson-Enflo, A. *Int. J. Quantum Chem.* **1990**, *37*, 517.
(7) Jensen, H. J. A.; Koch, H.; Jorgensen, P.; Olsen, J. *Chem. Phys.* **1988**, *119*, 297.

(8) Petke, J. D.; Maggiora, G. M.; Christoffersen, R. E. *J. Phys. Chem.* **1992**, *96*, 6992.
(9) Matos, J. M.; Roos, B. O. *J. Am. Chem. Soc.* **1988**, *110*, 7664.
(10) Fülscher, M. P.; Malmqvist, P.-Å.; Roos, B. O. Ab Initio Quantum Chemical Calculations of Excitation Energies and Transition Moments for the Nucleic Acid Base Monomers. In *Time-Resolved Laser Spectroscopy in Biochemistry II*; Lakowicz, J., Ed.; SPIE: Bellingham, WA, 1990; p 1204.
(11) Roos, B. O. In *Advances in Chemical Physics; Ab Initio Methods in Quantum Chemistry - II*; Lawley, K. P., Ed.; Wiley: Chichester, England, 1987; p 399, Chapter 69.
(12) Andersson, K.; Malmqvist, P.-Å.; Roos, B. O.; Sadlej, A. J.; Wolinski, K. *J. Phys. Chem.* **1990**, *94*, 5483.
(13) Andersson, K.; Malmqvist, P.-Å.; Roos, B. O. *J. Chem. Phys.* **1992**, *96*, 1218.
(14) Fülscher, M. P.; Andersson, K.; Roos, B. O. *J. Phys. Chem.* **1992**, *96*, 9204.
(15) Fülscher, M. P.; Roos, B. O. *Theor. Chim. Acta* **1994**, *87*, 403.

the CASPT2 method in spectroscopy, see ref 25). Transition energies obtained with this approach usually agree with experiment to within a few tenths of an electronvolt.

The CASPT2 equations are formulated exclusively in terms of one-, two-, and three-body matrices over a small subset of the one-electron space (the active orbitals). As a result, the computational effort increases only weakly with the number of orbitals, and large basis sets can be used, thus avoiding contamination of the results due to basis set deficiencies. The range of applicability of the CASPT2 method is also larger than that for any method of comparable accuracy, not only because it allows the use of large basis sets but also because the method scales slowly with the number of correlated electrons, as long as the number of active orbitals is kept limited. The CASPT2 method was recently used to study the electronic spectrum of free-base porphin with all 114 valence electrons correlated.[27]

Here we shall report results for the gas phase spectra. Such spectra may not be of direct experimental interest, since environmental effects are important, especially for the transition moments and their directions.[4] From a theoretical standpoint it is, however, important to solve first all problems related to the description of the excited states in the free molecule, before invoking the interaction with a surrounding solvent or crystal. This will be done in a second stage of this project. It should be emphasized that even if gas phase transition moment directions are sensitive to the surrounding, this is less so for the energies and total intensities of the transitions. Gas phase studies will therefore give information, which is of interest for all types of experimental studies.

Our results for cytosine strongly favor the three-transition model, and the absorption lines are predicted to be located at 4.39, 5.36, and 6.16 eV. It is suggested that the lowest n → π^* transition should be seen at energies close to 5.0 eV. We shall review critically earlier *ab initio* calculations in light of the present results. Finally, this report is summarized in the last section.

Details of the Calculations

Geometry and Basis Sets. The geometry of cytosine used in the present calculations is based on the averaged experimental heavy atom bond distance and angles reported by Taylor and Kennard,[28] and the 2-oxo-4-amino form is adopted. The molecule is placed in the *xy*-plane. The NH and CH distances were assumed to be 1.01 and 1.08 Å, respectively, and we use Hug and Tinoco's[5] conventions to label atoms and define angles throughout.

Generally contracted basis sets of atomic natural orbital (ANO) type[29,30] are used. The smaller basis sets, hereafter called BS1, have been derived from van Duijneveldt's[31] (10s,6p) primitive sets and contracted to DZP quality in the valence shell. The second and larger basis sets, hereafter called BS2, include (14s,9p,4d) and (8s,4p) primitives for the first-row atoms and hydrogen, respectively, and are contracted to the following structure: (C,N,O/4s,3p,1d) and (H/2s).

The present ANO-type basis sets are constructed to optimally treat correlation and polarization effects and are large enough[16] to describe the electronic structure of the valence excited states. On the other hand, diffuse functions are missing in the basis sets, and we cannot expect to be able to treat Rydberg states with high precision. However, the first members of the Rydberg series fall into the energy region of interest, and interaction with valence states may lead to spurious states of dubious character. Therefore, we extended BS2 by adding a set of Rydberg-like ANO basis functions, placed at the center of nuclear charge. They are constructed in the following way: A CASSCF wave function for the $1^2A''$ cation was computed using BS2 contracted to a minimal basis and augmented with an uncontracted set of eight primitives of s-, p-, and d-type of which the exponents are selected according to ref.[32] The diffuse, virtual orbitals of each angular type were used to determine the ANO contraction coefficients using the method of density averaging.[30] A set of 1s, 1p and 1d contracted, Rydberg-type orbitals are then included in the final basis set. This approach has yielded well described Rydberg states[21] which interact only weakly with other states.

CASSCF, CASSI, and MRCI Calculations. In continuation of our previous attempts,[11] the first set of CASSCF calculations were performed using BS1. The active space included the 10 π-electrons distributed in all possible ways among the 10 π-orbitals. Preliminary calculations optimizing orbitals for the average energy of the six lowest roots of A' symmetry indicated the presence of two Rydberg states. The corresponding orbitals could be identified and deleted from the molecular orbital (MO) basis. Four final CASSCF calculations were then performed for the averages containing the roots 1−5, 2−5, 1−4, and 2−4, respectively. The CASSCF state interaction (CASSI) method[33] is used to allow the 16 CAS wave functions to interact. Noninteracting wave functions, transition moment directions, and oscillator strengths are obtained by solving the corresponding secular problem. This somewhat involved procedure was followed in order to minimize the bias introduced by the averaging. Notice that it is virtually impossible to attempt an individual optimization of each of the excited states.

To estimate the effect of π-electron correlation not included at the CASSCF level of approximation, we performed additional π-MRCI (multireference CI) calculations. In order to compute transition moments and oscillator strengths, multiroot calculations were performed; that is, the program was asked to compute the five lowest roots simultaneously. The MO basis was constructed by diagonalizing the averaged one-particle density obtained from the CASSI calculations for the five lowest roots using BS1. The reference CSFs were selected such that the

(16) Roos, B. O.; Andersson, K.; Fülscher, M. P. *Chem. Phys. Lett.* **1992**, *192*, 5.

(17) Roos, B. O.; Serrano-Andrés, L.; Merchán, M. *Pure Appl. Chem.* **1993**, *65*, 1693.

(18) Roos, B. O.; Merchán, M.; McDiarmid, R.; Xing, X. *J. Am. Chem. Soc.* **1993**, *116*, 5927.

(19) Rubio, M.; Merchán, M.; Ortí, E.; Roos, B. O. *Chem. Phys.* **1994**, *179*, 395.

(20) Serrano-Andrés, L.; Merchán, M.; Nebot-Gil, I.; Lindh, R.; Roos, B. O. *J. Chem. Phys.* **1993**, *98*, 3151.

(21) Serrano-Andrés, L.; Merchán, M.; Nebot-Gil, I.; Roos, B. O.; Fülscher, M. P. *J. Am. Chem. Soc.* **1993**, *115*, 6184.

(22) Serrano-Andrés, L.; Lindh, R.; Roos, B. O.; Merchán, M. *J. Phys. Chem.* **1993**, *97*, 9360.

(23) Serrano-Andrés, L.; Merchán, M.; Fülscher, M. P.; Roos, B. O. *Chem. Phys. Lett.* **1993**, *211*, 125.

(24) Serrano-Andrés, L.; Roos, B. O.; Merchán, M. *Theor. Chim. Acta* **1994**, *87*, 387.

(25) Roos, B. O.; Fülscher, M. P.; Malmqvist, P.-Å.; Merchán, M.; Serrano-Andrés, L. Theoretical Studies of Electronic Spectra of Organic Molecules. In *Quantum Mechanical Electronic Structure Calculations with Chemical Accuracy*; Langhoff, S. R., Ed.; Kluwer: Dordrecht, Neth., 1994.

(26) Merchán, M.; Ortí, E.; Roos, B. O. *Chem. Phys. Lett.* **1994**, *221*, 136.

(27) Woody, R. W.; Callis, P. R. *Biophys. J.*, **1992**, *61*, A168.

(28) Taylor, R.; Kennard, O. *J. Am. Chem. Soc.* **1982**, *104*, 3209.

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

(30) Widmark, P.-O.; Malmqvist, P.-Å.; Roos, B. O. *Theor. Chim. Acta* **1990**, *77*, 291.

(31) van Duijneveldt, F. B. Gaussian basis sets for the atoms H−Ne for use in molecular calculations. Technical Report RJ945; IBM Research, 1970.

(32) Kaufmann, K.; Baumeister, W.; Jungen, M. *J. Phys. B.* **1989**, *22*, 2223.

(33) Malmqvist, P. Å.; Roos, B. O. *Chem. Phys. Lett.* **1989**, *155*, 189.

subset contained approximately 96% (sum of weights) of the CASSCF state functions. The final CI wave functions included 1 405 964 CSFs. Davidson's correction[34] was applied for the estimation of the contributions from unlinked quadruple excitations.

CASSCF, CASSI, and CASPT2 Calculations. The set of calculations described above was performed in order to explore the effect of π-electron correlation on the relative energies and the interaction with Rydberg excited states. In the second set of calculations described below, the aim is to obtain quantitative results for the energies and transition moments. Thus, larger basis sets are needed, and electron correlation must be fully accounted for (including also the σ-electrons). The calculations were all performed using BS2, and the 1s core orbitals of the first-row atoms were kept frozen in the form determined by the ground state SCF calculation. The minimal choice of the active spaces should include at least the valence π-orbitals and a set of Rydberg orbitals. It is important to include Rydberg-type basis functions into the AO basis set, since otherwise Rydberg states will interact in an artificial way with the valence excited states, using the more diffuse part of the original basis set. Preliminary CASSCF calculations correlating eight electrons were performed in order to optimize the Rydberg orbitals (the most stable π-electron orbital was left inactive). Since we are primarily interested in the valence excited states, the Rydberg orbitals were deleted from the MO basis in subsequent calculations. The preliminary calculations take into account any possible real interaction between Rydberg and valence states by a modification of the valence orbitals. An explicit treatment of all the Rydberg states is not of immediate interest, since they will not appear in spectra taken in solution or in crystals.

To explore the energies of the singlet excited $\pi \rightarrow \pi^*$ valence excited states, the active space was again extended to include all 10 π-electrons and the number of active orbitals was increased, starting with 10 orbitals, until convergence of the excitation energies, transition dipole moment directions, and the weight, ω, of the reference function. The final results are obtained with an active space including 12 orbitals. To estimate the energies of the singlet excited n $\rightarrow \pi^*$ transitions, we allowed the electrons of the σ-orbital highest in energy to propagate into the π-space, by also making this orbital active. Finally, the CASSI method is used to compute the transition dipole moment matrix elements, which are used together with the CASPT2 energy differences to obtain the oscillator strengths.

The CASPT2 method computes the first-order wave function and the second-order energy in the full CI space without any further approximation, with a CASSCF wave function constituting the reference function. The zeroth order Hamiltonian is defined as a Fock-type one-electron operator and is constructed such that a Möller−Plesset type perturbation theory is obtained in the closed-shell single determinant case. Two different formulations are possible: one which utilizes only the diagonal part of the Fock matrix (PT2D) and one which includes also the off-diagonal elements (PT2F). The first choice is computationally simpler and leads, in most cases, to results close to those obtained with the PT2F method. It should be emphasized, however, that it is only the nondiagonal approach which is invariant to rotations of the molecular orbitals. Here, we shall only present the results obtained by the PT2D approach.

The CASPT2 program calculates the weight, ω, of the CASSCF reference in the first-order wave function, which is a measure of how large a fraction of the wave function is treated variationally. The relative weight ω in different states then gives a measure of how balanced a calculation is. Normally, one

Table 1. Ground State Properties of Cytosine: Total Energies, Dipole Moments, and Quadrupole Moments

wave function	energy[a]	μ[b]	Q_{xx}[c]	Q_{yy}	Q_{zz}
CASSCF(BS1)[d]	−392.8031	7.20	−9.96	16.25	−6.29
CASSCF(BS2)	−392.8199	7.35	−10.10	16.04	−5.94
MRCI(BS1)[e]	−392.8299	7.60	−10.02	17.03	−7.00

[a] Energies in hartree. [b] Dipole moments in debye. [c] The x-, y-, and z-components of the diagonal quadrupole moment tensor in atomic units. [d] The CASSCF results are obtained from individual ground state calculations. [e] These results are extracted from the multiroot-CI calculations.

requires ω to be about the same for the ground and the excited states in order for the calculation to be balanced with respect to the treatment of electron correlation. In some cases interference with nearby Rydberg-type states not included in the reference CI space deteriorates this balance for some of the excited states. Such a situation can be tolerated in cases where the interaction is weak, and consequently the effect on the second-order energy is small. A large coefficient for the interfering state may then appear in the first-order wave function with only a small effect on the energy.

All calculations which made use of BS1 were performed on an IBM 3090 computer (Model 600) using a prerelease of the MOLCAS-2 quantum chemistry software. The remaining computations were carried out within the production release of MOLCAS Version 2[35] installed on an IBM RS/6000 workstation (Model 550).

Results and Discussion

Ground State Properties. The ground state properties for various wave functions are collected in Table I. The magnitude of the computed dipole moment, $\approx$7.5 D, is in accord with various experimental estimates, 7.0[36] and 6.0−8.0 D[37,38] for cytosine embedded in an aqueous solution or in a crystalline environment, respectively. The dipole moment vector makes an angle of $\approx$10° with the C_2-O_2 axis. A similar value was found by analyzing the electron density obtained from neutron diffraction data.[36] The results indicate that the electron density is well-described at the CASSCF level of approximation, and additional correlation effects appear to be of minor importance. The two different basis sets give similar results for these properties.

The CASSI/MRCI Results. From previous experience[39] we may expect the valence excited states of cytosine to be interspersed with Rydberg states. Basis sets lacking diffuse functions are obviously quite inadequate to describe such states and may lead to an artificial mixing of Rydberg and valence states due to erratic near-degeneracies occurring when the Rydberg states are described with a bad basis set. The problem can be overcome by systematic improvements of the basis sets. Therefore, to explain the contradiction between the first CASSCF calculations on cytosine published by Matos and Roos[10] and those reported later by Fülscher and Roos,[11] we extended the basis sets used in the later work by adding d-type functions (BS1).

(34) Davidson, E. R. *J. Chem. Phys.* **1975**, *62*, 400.

(35) Andersson, K.; Blomberg, M. R. A.; Fülscher, M. P.; Kellö, V.; Lindh, R.; Malmqvist, P.-Å.; Noga, J.; Olsen, J.; Roos, B. O.; Sadlej, A. J.; Siegbahn, P. E. M.; Urban, M.; Widmark, P.-O. *MOLCAS Version 2*; Department of Theoretical Chemistry Center, University of Lund: Lund, Sweden 1992.

(36) Weber, H.-P.; Craven, B. M. *Acta Crystallogr.* **1990**, *B46*, 532.

(37) Kulakowska, I.; Geller, M.; Lesyng, B.; Bolewska, Wierchowski, K. L. *Biochim. Biophys. Acta* **1970**, *407*, 420.

(38) Palmer, M. H.; Wheeler, J. R.; Kwiatkowski, J. S.; Lesyng, B. J. *J. Mol. Struct.* **1983**, *92*, 283.

(39) Malmqvist, P.-Å.; Roos, B. O.; Fülscher, M. P.; Rendell, A. *Chem. Phys.* **1992**, *162*, 359.

Table 2. Singlet Excited States of Cytosine[a]

state	CASSI			MRCI		
	ΔE	f	θ	ΔE	f	θ
$\pi-\pi^*$ Transitions						
2A′	4.95	0.076	76.7	4.96	0.069	69.8
3A′	6.41	0.139	0.1	6.22	0.201	−5.2
4A′	7.46	0.877	−39.1	7.41	0.989	−41.4
5A′	8.01	0.139	13.1	7.94	0.091	17.1

[a] Excitation energies (ΔE) the eV, oscillator strengths (f), and transition directions (θ) in degrees computed at the CASSI/MRCI level of approximation (BS1).

Table 3. Singlet Excited States of Cytosine[b]

state	ΔE(CASSCF)	ΔE(CASPT2)[a]	f	θ	μ
$\pi-\pi^*$ Transitions					
2A′	5.18	4.39 (0.73)	0.061	60.6	4.7
3A′	6.31	5.36 (0.72)	0.108	−1.5	7.0
4A′	7.30	6.16 (0.70)	0.863	−39.7	6.2
5A′	7.82	6.74 (0.29)	0.147	14.9	9.3
6A′	9.13	7.61 (0.40)	0.239	45.4	5.0
$n-\pi^*$ Transitions					
1A″	5.13	5.00 (0.73)	.005		4.7
2A″	7.14	6.53 (0.68)	.001		6.4

[a] The weight of the CASSCF reference function within parentheses. [b] Excitation energies (ΔE) in eV, oscillator strengths (f), transition directions (θ) in degrees, and dipole moment (μ) in debye, computed at the CASSCF/CASPT2 level of approximation (BS2).

Two-state average CASSCF calculations, including the roots 1−6 and 2−6, were first performed. In these wave functions the third and the fifth states were dominated by single excited CSFs involving excitations from the π-orbitals highest in energy into a diffuse Rydberg-like orbital with a second electric moment of ≈60 au, which was almost unoccupied in the other states. We therefore performed a new set of state average CASSCF calculations with the Rydberg orbital removed from the MO basis. A second orbital of Rydberg character was then identified. Thus, the state average CASSCF calculations had to be repeated once more, now with both Rydberg orbitals deleted from the MO basis and including the roots 1−5 and 2−5, respectively. Finally, these wave functions were allowed to interact by solving the corresponding secular problem using the CASSI approach. The results are shown in Table 2.

To estimate the remaining correlation effects in the π-orbital space and to ensure that the removal of the orbitals with Rydberg character have not significantly perturbed the system, multiroot π-CI calculations were performed. The results are included in Table 2. The excitation energies computed at the π-CI level of approximation differ at most by 0.2 eV from those obtained in the CASSI calculations. The oscillator strengths of the first transition are virtually identical for both sets of calculations, and the transition moment directions changed at most by 7°. The similarity of the results is further confirmed by the weights of the references function, 96−92% which indicate that the CASSCF wave function covers all essential correlation effects within the π-electron space.

We find the present results to be in agreement with the data presented by Matos and Roos.[10] By matching excitation energies and transition directions, we can correlate the first excited state reported by Fülscher and Roos[11] with the 2A′ state. The oscillator strengths indicate a correspondence of the 3A′ states. By the same token, however, root 4A′ appears to be related to root 5A′ computed previously. Thus, excluding the old state 4A′, the overall agreement of the transition directions amounts to ≈10° and is only slightly bigger than the differences observed between the present CASSI and MRCI calculations. In conclusion, the old 4A′ state is identified as a quasi-Rydberg state with an incorrect mixing with valence excited states due to basis set deficiencies. This reanalysis of our earlier results thus yields a consistent picture, which is in agreement with the present result: the energy region below 6.5 eV contains three $\pi \rightarrow \pi^*$ excited states.

The CASSCF/CASPT2 Results. In the previous section we have given arguments and an example showing the importance of basis sets suitable to describe Rydberg-type orbitals. As a consequence the CASSCF/CASPT2 calculations have been performed with BS2, which was augmented with basis functions designed to describe Rydberg states. Table 3 summarizes the results obtained in this second set of calculations.

A comparison of Table 3 with Table 2 shows that the CASSCF energies reproduce the MRCI results with an accuracy of ≈0.2 eV. Differential dynamic correlation effects, which

now fully include the σ-electrons and the $\sigma-\pi$ interactions, lower the excitation energies substantially. It is worth emphasizing that this is the first set of calculations on cytosine in which these effects are included. It is well-known that they are essential for obtaining excitation energies that are in agreement with experiment. It is thus only these results that can be used for a conclusive assignment of the experimental transition energies. For example, the transition energy for the first excited state is lowered by 0.79 eV. The order of the states is, however, preserved. The transition directions (obtained at the CASSCF level of theory), which are independent of the excitation energies, do not differ by more than 10° from those obtained with the smaller basis set. In contrast, a comparison of the oscillator strengths exposes substantial differences. The intensities of the second and the third transitions are reduced by 50% and 13%, respectively, whereas the higher excitations gain intensity.

The lower half of Table 3 includes the two lowest n → π^* transitions. They arise from excitations of a nitrogen lone pair electron into the 6a″ and higher orbitals. The lowest singlet excited n → π^* state is computed (CASPT2) to occur at 5.0 eV. Its location is almost unaffected by differential correlation effects. One might expect a second state of A″ symmetry at slightly higher energies corresponding to excitation out of the in-plane oxygen lone-pair orbital. Since the active space contains only one σ-orbital, this state is not included in the calculation.

The weights, ω, of the reference state in the first-order wave function are included in Table 3. For the 5A′ and 6A′ state we observe that interferences with a nearby diffuse state not included in the reference function deteriorate the balance. Such situations may occur in cases where the active space is not large enough to cover all nearly degenerate states. It can be tolerated in cases where the effect is small and the contributions to the second order energy can be neglected.[15,16] The large coefficients in the first-order wave function are due to a Rydberg-type excitation, which was not included in the CAS CI space. It interacts only weakly with the valence excited states, and the contribution to the excitation energy is therefore small. It is difficult to avoid these types of intruder states for the higher excited states, since it is virtually impossible to delete all Rydberg character from the basis set. The problem could have been solved by extending the active space, but the number of active orbitals would then have exceeded 12 in one symmetry, which is more than the current implementation of the program can handle.

It is important to note that the similarity of the CASSCF results obtained with either BS1 or BS2 supports the assumptions made in the first set of calculations. On the one hand, diffuse orbitals were associated with states of Rydberg character. The orbitals were therefore deleted from the basis set. On the

Table 4. Excited States of Cytosine: A Summary of Experimental Data[a]

	I	II	III	IV	V
Absorption Spectra					
Voet *et al.*,[46] aqueous solutions					
ΔE	4.6		5.4		6.3
ϵ	6100		4300		22500
Morita and Nagakura,[47] aqueous solutions					
ΔE	4.6		5.5		6.3
f	0.11				0.68
Clark and Tinoco,[42] solution of trimethyl phosphate					
ΔE	4.5		5.2	6.1	6.7
ϵ	7500		3500	11900	12200
Yamada and Fukutome,[48] sublimed films					
ΔE	4.6				6.2
Raksányi *et al.*,[49] sublimed films					
ΔE	4.6		5.4	6.1	6.7
f	0.06		0.07	0.115	0.07
Circular Dichroism					
Miles *et al.*,[40] the study includes many cytosine derivatives					
ΔE	4.6	5.0	5.6	6.4	
Sprecher and Johnson[41]					
ΔE	4.6	5.3	5.8	6.2	7.1
Magnetic Circular Dichroism					
Voelter *et al.*[50]					
ΔE	4.6	5.3			
Linear Dichroism					
Bott and Kurucsev[43]					
θ	(35 ± 14°, −28 ± 7°)		(15 ± 20°, −6 ± 16°)		
Matsuoka and Nordén[44]					
ΔE	4.6		5.2		
θ	(25 ± 3°, −46 ± 4°)		(6 ± 4°, −27 ± 3°)		
Fucaloro and Forster[51]					
ΔE	4.6		5.2		
θ	(9°)		(7.5°, 10.5°)	(94°, −76°)	
Polarized Absorption of Crystals of Cytosine Monohydrate					
Lewis and Eaton[52]					
ΔE	4.4		5.5		
θ	(14°, 48°)		(−5 ± 3°)		
Callis and Simpson[45]					
ΔE	4.7		(−11°)		
Zaloudek *et al.*[3]					
ΔE	4.6	5.3	5.8	6.3	
ΔE	0.14	0.03	0.13	0.36	
θ	(6°, 54°)		(76°, −17°)	(86°, −27°)	

[a] Experimental excitation energies (ΔE) in eV, oscillator strengths (*f*), extinction coefficients (ϵ), and transition directions (θ) in degrees.

other hand, BS2 is especially designed to properly describe the low lying Rydberg states, and their interaction with the valence states is reduced and reflects a true mixing, which is not an artifact of the basis set choice. Thus, the approach taken in the first set of calculations appears to be justified.

Comparison with Experiment. The experimental spectra of cytosine can be divided into two classes, one which partitions the energy range extending from 4.5 to 6.5 eV into three bands, and the other giving evidence for a fourth transition close to 5.2 eV in energy. Four spectra were published reporting transition energies higher than 6.5 eV. In total, there is information on the positions of at most six bands. A summary of the most relevant data is shown in Table 4.

From the results given in Table 4 it can be seen that, overall, the solution spectra correlate well with the crystal and film measurements for the bands I, III, and IV. Therefore, the discussion has concentrated on the position and origin of band

II. The crystal reflection spectrum for light polarized perpendicular to the molecular plane shows a slight maximum at 5.3 eV which suggests that it may involve an n → π* transition. Zaloudek *et al.*[3] were, however, not able to make a definitive assignment. Further support for an n → π* transition can be derived from CD spectroscopy; Miles *et al.*[40] studied a number of cytosine derivatives of which some molecules show Cotton effects in the energy region around 5.0 eV. On the other hand, the observed solvent shifts (up to 5 nm) are somewhat too small. In contrast to these authors, Sprecher and Johnson found no hint pointing to an n → π* transition.[41]

The solvation effects measured by Miles *et al.*[40] are of approximately the same size as the band shifs derived from comparison of spectra of cytosine films sublimed onto a carrier with the data for cytosine embedded in an aqueous or a crystalline environment. There appears to be a contradiction with the data presented by Clark and Tinoco.[42] By correlating the cytosine spectrum with that of pyrimidine, they argued that the intense bands exposed at 6.1 and 6.7 eV in trimethyl phosphate move together in water, giving rise to a weaker band at 5.8 eV. The disparity can be resolved if the second and following absorption lines are assigned to bands III−V. Furthermore, the realignments allow an interpretation of the weak features observed in CD spectra at energies around 5.0 eV to be weak n → π* transitions, without conflicting with the absorption measurements. As a consequence band V is given a somewhat large extinction coefficient attributable to solvent effects.

Solvent effects on the computed excitation energies can be expected to be on the order of a few tenths of an electronvolt. The dipole moment of the ground state is about 7.5 eV. We can therefore expect a blue shift of the first π → π* transition, which has a considerably smaller dipole moment (cf. Table 3). This seems to be in accordance with experiment since the calculated excitation energy is 0.0−0.3 eV too small. The two next states have moments close to those of the ground state, and only small shifts are expected. The n → π* states will undergo a blue shift in solution, since the excitation breaks strong solute−solvent interactions involving the lone-pair electrons. This blue shift can be estimated to be 0.3−0.5 eV, which places the corresponding band under the second π → π* transition. More detailed studies of the solvent shift are under way, using a reaction field Hamiltonian in a dielectric continuum model.

The polarization of the first few transitions has been measured for cytosine embedded in stretched poly(vinyl alcohol) films and crystalline environments. The studies included cytosine monohydrate and 1-methylcytosine. At a first glance the experimental data appear to disagree on the polarization directions. Zaloudek *et al.*[3] argued that by rotating the molecular axis system by −34° in the work of Bott and Kurucsev[43] and by −19° in the work of Matsuoka and Norden,[44] the data can be brought into coincidence. Such a large rotation can be explained by the actual experimental difficulties in determining the average molecular orientation in stretched films. On the basis of consistency arguments, Callis and Simpson[45] recom-

(40) Miles, D. W.; Robins, M. J.; Robins, R. K.; Winkley, M. W.; Eyring, H. *J. Am. Chem. Soc.* **1969**, *91*, 831.
(41) Sprecher, C. A.; Johnson, W. C., Jr. *Biopolymers* **1977**, *16*, 2243.
(42) Clark, L. B.; Tinoco, I., Jr. *J. Am. Chem. Soc.* **1965**, *87*, 11.
(43) Bott, C. C.; Kurucsev, T. *Spectrosc. Lett.* **1977**, *10*, 495.
(44) Matsuoka, Y.; Nordén, B. *J. Phys. Chem.* **1982**, *86*, 1378.
(45) Callis, P. R.; Simpson, W. T. *J. Am. Chem. Soc.* **1970**, *92*, 3593.
(46) Voet, D.; Gratzer, W. B.; Cox, R. A.; Doty, P. *Biopolymers* **1963**, *1*, 193.
(47) Morita, H.; Nagakura, S. *Theor. Chim. Acta* **1968**, *11*, 279.
(48) Yamada, T.; Fukutome, H. *Biopolymers* **1968**, *6*, 43.

Table 5. Singlet Excited $\pi-\pi^*$ States of Cytosine: A Summary of *ab Initio* Calculations

Matos and Roos,[10] CASSCF				
energy[a]	5.0	6.6		8.2
θ[b]	67.0	−9.8		−30.6
f[c]	0.04	0.12		1.02
Fülscher and Roos,[11] CASSCF				
energy	4.81	6.68	6.98	7.92
θ	86	6	−27	−25
f	0.06	0.11	0.03	0.91
Petke et al.,[9] MRCI1 (extended basis sets, 4472 configurations)				
energy	5.6	7.0	8.9	9.2
θ	64	−5	−13	−47
f	0.07	0.27	0.24	0.61
Petke et al.,[9] MRCI2 (extended basis sets, 12 603 configurations)				
energy	5.5	6.8	6.8	7.3
θ	68	−3	28	−34
f	0.08	0.12	0.03	0.25
Petke et al.,[9] MRCI3 (small basis sets, 1627 configurations)				
energy	5.4	6.7		7.9
θ	68	3		−39
f	0.07	0.1		1.279
Petke et al.,[9] RPA (extended basis sets)				
energy	6.19	7.40	7.54	7.83
θ	41	−40	−8	−43
f	0.13	0.03	0.30	0.44
Jensen et al.,[8] RPA				
energy	5.85	7.14	7.34	7.46
θ	50	−18	−52	−44
f	0.11	0.43	0.09	0.41

[a] Excitation energy in eV. [b] Transition direction. [c] Oscillator strength.

mended a value of $\theta_I = 6°$ as the correct choice. Subsequently, owing to the lack of conclusive experimental data this viewpoint was adapted and led to the following preferences: $\theta_I = 6°$, $\theta_{II} = -46°$, $\theta_{III} = 76°$ and $\theta_{IV} = -27°$ or $86°$. The present gas phase results cannot give any conclusive confirmation of the experimental data, since the crystal field effects are large.[4] Except for the first state, the differences between computed and measured angles are small.

The excitation energies and transition moments (except for transition I) computed at the CASSCF/CASPT2 level of approximation match the experimental data if the following assumptions are made: (1) the absorption band II is due to an n → π^* transition; (2) the experimental excitation energies are then reproduced with an error of 0.2 eV or smaller. This accuracy is within the error bars associated with numerous applications of the CASSCF/CASPT2 method. The computed polarization directions, approximately −5° and −40°, are within ±15° of the values reported by Zaloudek et al.[3] for the bands III and IV. The observed intensity distribution is reproduced at a qualitative level.

Comparison to Previous *ab initio* Calculations. Over the years, four *ab initio* studies were reported on the singlet excited states of cytosine. The most recent study has been published by Petke et al.[9] The authors applied MRCI and RPA techniques which can be regarded as an extension of the RPA calculations reported by Jensen et al.[8] Both groups applied Huzinaga's basis sets but differ in the way the latter have been augmented with polarization and diffuse functions. In contrast, Matos and Roos[10], as well as the present authors[11] investigated the spectrum

of cytosine using the CASSCF method combined with ANO-type basis sets. A summary of the various calculations is given in Table 5.

As pointed out above, the CASSCF calculations of Matos and Roos and Fülscher et al. are consistent with the present results and clearly favor the three-transition model. In contrast, the RPA calculations of Petke et al. and Jensen et al. result in a large number of valence excited singlet states with excitation energies smaller than 6.5 eV. In the discussion above on the importance of diffuse functions, it was shown that basis set deficiencies can lead to spurious states due to incorrect Rydberg−valence state mixing. A close inspection of the results reported by Petke et al. suggests that their calculations suffer from the same defects as the calculations by Fülscher and Roos did. First, we observe that the smallest MRCI calculations using the original basis sets contracted to double-zeta quality (lower part of Table 8 in ref 9) produce results virtually identical to the CASSCF and MRCI results presented here. When the basis sets are extended (upper part of Table 8 in ref 9), the second electric moments of the states 3A′−7A′ dramatically increase due to large contributions from excitations (Table 10 in ref 9) to diffuse virtual orbitals (Table 2 in ref 9). Owing to the lack of enough flexibility in the basis sets, valence excited states are seen to mix with the diffuse Rydberg-like states, e.g., the roots 3A′ and 4A′ are found to be nearly degenerate—the splitting amounts to less than 200 cm^{-1}. Their transition dipole moments are, however, not perpendicular as expected. Similar arguments may be used to explain the spreading of the intensities.

In spite of the severe valence−Rydberg state mixing observed for the MRCI calculations of Petke et al., it is possible to arrive at a description of the spectrum which is consistent with the present results. If all diffuse valence excited singlet states of the "medium-sized" MRCI calculations (Table 7 in ref 9) are ignored, we arrive at a model that includes the excited states S_1, S_5, and S_{19}. (The states S_{19} and S_{20} also seem to mix with diffuse orbitals, and hence, one may assume that they would split into one valence and one Rydberg state if the basis sets are further improved.) In addition, we observe that the calculations by Petke et al. predict that the lowest n → π^* transitions occur at energies between those of the first two π → π^* bands.

Since electric moments cannot be computed by RPA calculations, we cannot analyze the character of the states in the terms used above. We note, however, that the RPA method predicts the lowest state to be π → π^*, followed by a n → π^* transition. This is consistent with our calculations. The first π → π^* valence excited singlet state obtained by this method is found at ≈6.0 eV with a polarization angle of about 40°.

Summary

In this paper we have presented results on the properties of the excited states of the cytosine molecule. The calculations included a number of vertical valence excited singlet states. On the one hand, we reported results obtained from CASSCF and π-MRCI calculations using basis sets of modest size. On the other hand, we also applied the CASSCF/CASPT2 method in combination with high-quality basis sets which were extended to describe Rydberg orbitals. These studies included the necessary correlation contributions from the interaction between the σ- and the π-electrons. It is the first calculation from which we can expect to obtain quantitative agreement between the computed and the measured excitation energies. The calculations predict the five lowest excited states in the free cytosine molecule to be located at 4.4, 5.0, 5.4, 6.2, and 6.7 eV above

(49) Raksányi, K.; Földváry, I.; Fidy, J.; Kittler, L. *Biopolymers* **1978**, *17*, 887.

(50) Voelter, W.; Records, R.; Bunnenberg, E.; Djerassi, C. *J. Am Chem. Soc.* **1968**, *90*, 6163.

(51) Fucaloro, A. F.; Forster, L. S. *J. Am. Chem. Soc.* **1971**, *93*, 6443.

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

the ground state. The second state is of n → π* character, the other four are π → π* states. The energies should be correct to within 0.3 eV. There is a possibility that one more n → π* state may appear with a slightly larger energy. Obviously, for the free molecule, Rydberg states will appear in the spectrum. They were excluded from the present treatment by deleting the corresponding part of the one-electron space.

One aim of the present study was to obtain a detailed understanding of the electronic spectrum of cytosine, including both the excitation energies and the transition moments and their directions. We believe that the goal has been accomplished, and numerous arguments are given in favor of the three-transition model. In agreement with previous *ab initio* and semiempirical calculations we obtain a gas phase value of the polarization angle for band I of ≈60°. We suggest that band II is due to a n → π* transition. The computed excitation energies reproduce the experimental line positions with an accuracy of 0.2 eV or better, and the polarization directions match within ±15° of the measured data for all but the first state, indicating that the crystal field effects for the other states are more limited.

By comparing the results obtained from the different methods and applying various basis sets, we have illustrated that the discrepancies between the present and earlier results for cytosine can be attributed to basis set problems and a too limited treatment of the electron correlation effects. It is only when all valence electrons are correlated that results of quantitative accuracy can be obtained for the transition energies. In order to avoid an erroneous mixing of valence and Rydberg excited states, it is imperative to have a good representation of the Rydberg orbitals in the basis set. Once they have been defined, they can be deleted in order to concentrate the study to the valence excited states. Results of the same accuracy will later be presented for the other base monomers of DNA and RNA.[1] These studies will include solvent and crystal field effects on the spectrum.

JA943047R