Quantum-Chemical Investigation of the Structures and Electronic Spectra of the Nucleic Acid Bases at the Coupled Cluster CC2 Level

Timo Fleig* and Stefan Knecht

Institute of Theoretical and Computational Chemistry, Heinrich Heine University Düsseldorf, Universitätsstrasse 1, D-40591 Düsseldorf, Germany

Christof Hättig

Lehrstuhl für Theoretische Chemie, Ruhr-Universität Bochum, Universitätsstrasse 150, D-44801 Bochum, Germany

Received: October 23, 2006; In Final Form: March 8, 2007

We study the ground-state structures and singlet- and triplet-excited states of the nucleic acid bases by applying the coupled cluster model CC2 in combination with a resolution-of-the-identity approximation for electron interaction integrals. Both basis set effects and the influence of dynamic electron correlation on the molecular structures are elucidated; the latter by comparing CC2 with Hartree−Fock and Møller−Plesset perturbation theory to second order. Furthermore, we investigate basis set and electron correlation effects on the vertical excitation energies and compare our highest-level results with experiment and other theoretical approaches. It is shown that small basis sets are insuffient for obtaining accurate results for excited states of these molecules and that the CC2 approach to dynamic electron correlation is a reliable and efficient tool for electronic structure calculations on medium-sized molecules.

1. Introduction

The nucleic acid (NA) bases adenine, thymine, guanine, cytosine, and uracil are essential building blocks of DNA and RNA. A profound understanding of their electronic structure and dynamics is of great interest, as the NA bases are remarkably stable with respect to damaging UV irradiation.[1] In recent works, rapid decay pathways for electronic energy in excited states of adenine and also the other NA bases have been investigated,[2−8] indicating the importance of high-accuracy data for excited singlet and possibly also triplet states of the NA bases.

With the impressive advance of quantum-chemical methodology and computational power in the past decade, ab initio calculations of molecules of the size of the NA bases and also their dimers have become possible using large basis sets and including electron correlation by coupled cluster methods.[9,10] A consistent study of the NA bases using the same high-level treatment for both full geometry optimizations and the calculation of excited states, however, has not been available to the date.

Apart from applications of various density functional theory and semiempirical models, only a few ab initio approaches have been applied to the calculation of excited states of the NA bases. Dynamic electron correlation has either been treated by configuration interaction (CI) theory (e.g., in refs 11 and 12) or multireference perturbation theory via the complete active space perturbation theory to second order (CASPT2, as in refs 13 and 14). Shifted CI singles is not a rigorous method, as electron correlation is merely accounted for in an implicit fashion. Moreover, CI approaches generally suffer from the low compactness of the truncated CI wave function. CASPT2 based on CASSCF wave functions has been the most rigorous approach to the electronic spectra of the NA bases so far. However, this method becomes difficult to apply when the electronic structure of a species requires the use of an extensive complete active orbital space, which for larger molecules quite frequently is the case. In coupled cluster theory, on the other hand, the leading higher excitations representing the multireference space of the CASPT2 calculations are contained in the cluster expansion of the wave function.

Using the approximate coupled cluster model CC2[15] throughout, we pursue the following purposes with this investigation: (1) Full ground-state geometry optimization of all 5 NA bases without molecular symmetry, thus allowing for relaxation to the energetically most favorable nonplanar structures. We use a series of one-particle basis sets and systematically elucidate basis set effects and the influence of electron correlation treated at different levels of sophistication on molecular geometries. (2) Vertical excited-state calculation of singlet and triplet states of the NA bases using the optimized geometries. We here specifically focus on basis set and electron correlation effects on excitation energies and the character of excited states in terms of contributing orbitals/excitations. (3) Assessment of a level of high accuracy and a profound understanding of the subunits for subsequent investigation of NA base dimers.[16]

In the following section (2), we summarize the employed computational approaches. Section 3 comprises the main body of the paper with all results and their discussion, and in Section 4, we draw conclusions from our studies, with an emphasis on future work concerning larger related systems.

2. Computational Method and Basis Sets

The CC2 equations are an approximation to the coupled cluster singles and doubles (CCSD) equations, where the singles equations are retained in the original form and the doubles

---

* Corresponding author. E-mail: timo@theochem.uni-duesseldorf.de.

10.1021/jp0669409 CCC: \$37.00

equations are truncated to first order in the fluctuation potential.[15] We used the implementation[17−19] in the TURBOMOLE[20] quantum chemistry program package. A resolution-of-the-identity (RI) approximation is employed for molecular orbital two-particle integrals. The errors made within this approximation are, with optimized auxiliary basis sets, in general negligible as compared to errors due to the one-electron basis set incompleteness.[17,18] All electrons were correlated, i.e., 78 in guanine, 70 in adenine, 66 in thymine, and 58 in cytosine and uracil. This computational level was retained for both the geometry optimizations and the excited-state calculations via coupled cluster response theory. For a few molecules and basis sets, we also checked for the effect of forming a frozen core from the 1s electrons on structures and excitation energies. All calculations have been carried out in C_1 point-group symmetry.

We used a sequence of standard Gaussian basis sets, the polarized split-valence SV(P) (C, N, O: 7s4p1d/3s2p1d; H: 4s1p/2s1p),[21] and the augmented correlation-consistent sets aug-cc-pVDZ (C, N, O: 10s5p2d/4s3p2d; H: 5s2p/3s2p), cc-pVTZ (C, N, O: 10s5p2d1f/4s3p2d1f; H: 5s2p1d/3s2p1d), aug-cc-pVTZ (C, N, O: 11s6p3d2f/5s4p3d2f; H: 6s3p2d/4s3p2d), aug-cc-pVQZ (C, N, O: 13s7p4d3f2g/6s5p4d3f2g; H: 7s4p3d2f/5s4p3d2f).[22,23] The corresponding auxiliary basis sets for the RI approximation are documented in refs 24,25.

For analyzing the character of excited states, we have used the MOLDEN program suite version 4.4.[26]

3. Results and Discussion

A. Geometry Optimizations. The geometries of the nucleic acid bases are sensitive to both the employed basis sets and the treatment of electron correlation. For consistency in our study, the geometry at which the single-point calculations of vertical excitation energies were carried out are optimized using the same electronic structure method. We therefore optimize all geometries with the CC2 approach and compare the obtained structures employing different basis sets. Moreover, we elucidate the influence of electron correlation by comparing structures obtained with the same basis sets but different approaches to electron correlation.

1. Basis Set Effects. The optimized geometries of all five nucleic acid bases can be found in the Supporting Information (figures S1−S5) to this article. The following discussion is based on the CC2 results. Because of the high computational demand, only uracil was treated with the set of quadruple-ζ quality. For the series aug-cc-pVXZ ($X = D,T,Q$), we observe that bonds contract systematically by about 0.01 Å for DZ to TZ and by about 0.005 Å for TZ to QZ. The small SV(P) set, which is comparable to the commonly used 6-31G* basis set in quality, overestimates bond lengths in the range of 0.01 to 0.02 Å except for C═O bonds, where bond lengths resemble those obtained with the aug-cc-pVQZ set. The effect of diffuse augmenting functions, by comparison with the results for the cc-pVTZ for uracil, is to stretch all bonds by roughly 0.005 Å. We have added these functions for an appropriate description of Rydberg-type excited states and especially preparing for a further study on the NA base dimers, where they are required for modeling excited states of charge-transfer type and the hydrogen bonds in many dimer structures. We confirm the trend of the CC2 method to give slightly longer bond lengths than MP2 (cc-pVTZ, ref 27), which has been established earlier for a test set of small molecules.[19,28] The effect of augmenting basis functions becomes particularly interesting when considering the cc-pVDZ geometries of uracil, which are the same up to 0.001 Å as those with the cc-pVTZ basis set. The contraction from increasing the cardinal number is only observed with augmenting functions. In total, the elongation due to augmenting basis functions and this contraction happen just to level out so that cc-pVDZ and aug-cc-pVQZ geometries of uracil are almost exactly the same.

TABLE 1: Sum of Bond Angles C−N−H$_1$, C−N−H$_2$, H$_1$−N−H$_2$, in deg at the Amino Group with Different Basis Sets and Levels of Electron Correlation[a]

method/basis set	adenine	guanine	cytosine
RI-CC2/SV(P)	353.80	343.39	352.72
RI-CC2/aug-cc-pVDZ	353.69	338.98	352.58
RI-CC2/aug-cc-pVTZ	355.77	341.69	354.53
RI-CC2/aug-cc-pVTZ (fc)		342.97	
MP2/aug-cc-pVTZ	353.74	340.96	353.77
HF/aug-cc-pVTZ	358.45	348.79	359.64

[a] (fc) denotes an optimization with all 1s electrons forming a frozen core. The deviation from 360° is a measure of the nonplanarity of the amino group.

In view of the very small changes from going to the QZ basis set and the general fact that the error from the limited treatment of electron correlation surpasses the basis set error already at the aug-cc-pVTZ level,[29] we consider our results converged with respect to the extent of the one-particle basis set.

Regarding the nonplanarity of amino groups in adenine, guanine, and cytosine, we confirm earlier findings.[27,30,31] Table 1 gives the sum of the three optimized bond angles around the amino nitrogen atom in adenine, guanine, and cytosine. The individual bond angles can be found in the Supporting Information for this article. The largest nonplanarity is found for guanine. The nonplanarity decreases with increasing size of the basis sets, where the smallest SV(P) does not follow this trend. It is interesting to note that, for a given approach to electron correlation, a basis set of increasing size leads to a more planar structure. This finding is in line with the results of a study on the planarity of formamide by Fogarasi et al.[32]

2. Effects of Electron Correlation. For obtaining information on the effect of electron correlation on the molecular geometries, we have carried out Hartree−Fock geometry optimizations and, furthermore, compare with the MP2 results by Wang et al.[27]

Commencing with the effect on nonplanarity, the inclusion of electron correlation leads to a considerable increase for the amino groups. However, the perturbative optimizations reveal that MP2 nonplanarities are slightly larger than those obtained with CC2, as can be seen in Table 1. This indicates that the partial sp^3 character of the amino nitrogen atoms is decreased by including single excitations in the wavefunction at the CC2 level.

As expected, due to the shift of electron density into antibonding orbitals, bonds are stretched upon accounting for electron correlation (Figures S1−S5 in the Supporting Information). This increase ranges from 0.004 to 0.04 Å, with the exception of one C−C ring bond per system, which is slightly contracted. The contracted bond is always the longest ring bond with the largest σ-bonding character. The elongation of C═O bonds is most pronounced compared to other elongations and at the upper limit of the range of C═O bond lengths in smaller molecules.[28]

We have furthermore investigated the correlation contributions due to the 1s electrons by optimizing geometries using a respective frozen core. The effects on bond lengths are very small, with elongation of bonds due to the frozen core approximation in the order of 0.005 Å. For guanine and the aug-cc-pVTZ basis set, we observe the largest effects.

B. Excited States. Tables 2 through 6 contain the vertical excitation spectra of the five molecules, a qualitative charac-

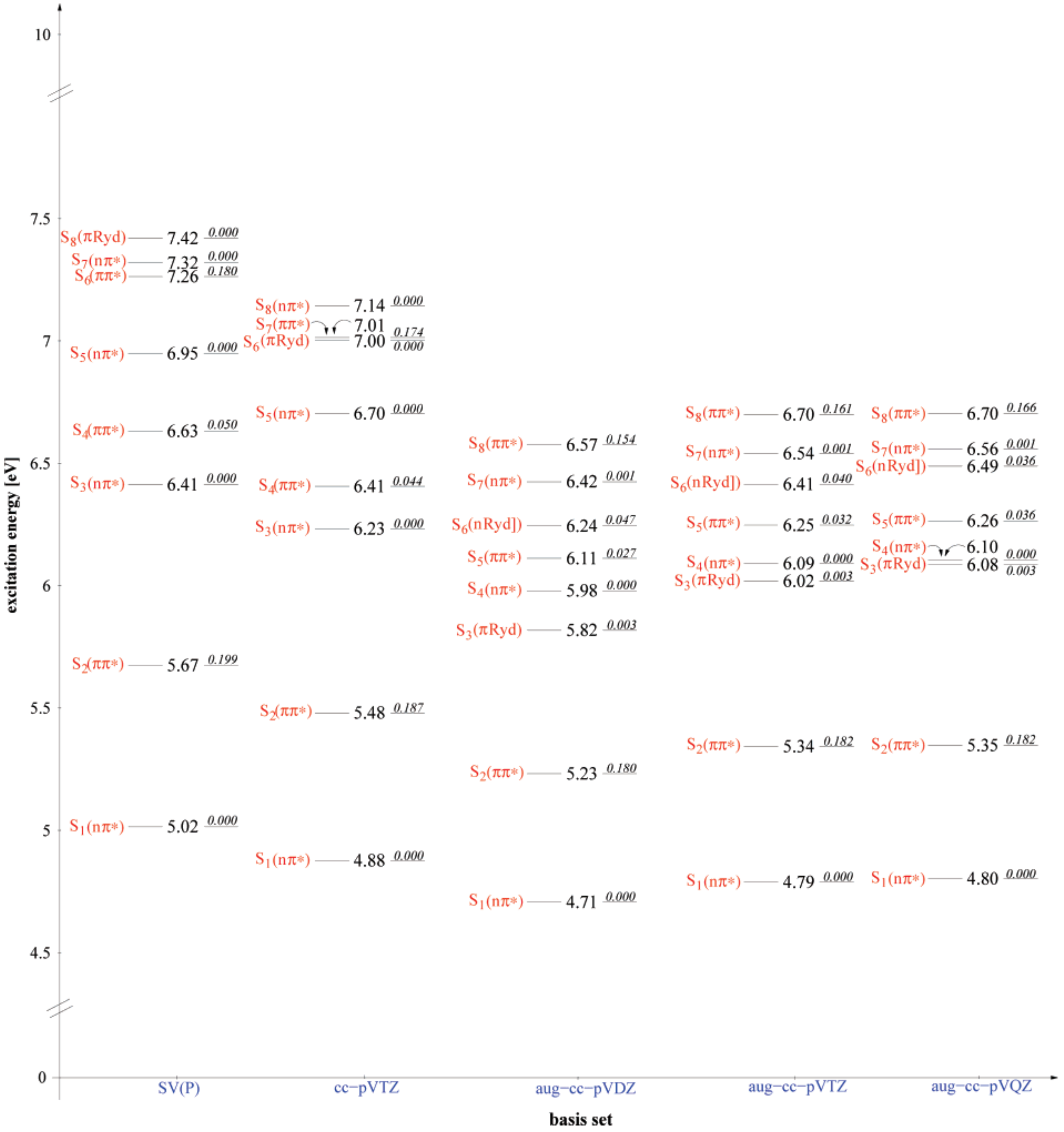

Figure 1. Uracil: CC2 vertical singlet excitation energies with different basis sets, at corresponding ground-state optimized geometries; oscillator strengths of the transitions are given in italics.

terization of the electronic transitions and the involved molecular orbitals, and the oscillator strengths for singlet states, obtained with the most extensive of the investigated basis sets, respectively. The comparison of results with various basis sets has been carried out for all five molecules and is displayed graphically in Figure 1 for uracil as a representative example.[49]

All results have been obtained from calculations at the optimized geometry with the respective basis set.

1. Basis Set Effects. For uracil, it was computationally feasible to increase the basis set cardinal number up to 4 and beside SV(P) and cc-pVTZ compare the series aug-cc-pVXZ (X = D,T,Q). The most striking observation is the inadequacy of the very small split-valence basis set for giving accurate vertical excitation energies in CC2 calculations. The deviations may become as large as 1.0 eV when comparing to the aug-cc-pVDZ set. Both the ordering and the character of the excited states, on the other hand, remain essentially unchanged. Excitation energies increase slightly with increasing cardinal number, but the observed changes are well within the expected error margins from the incomplete treatment of dynamic electron correlation due to the truncated coupled cluster expansion. A favorable error compensation, i.e., the increasing excitation energies through increasing cardinal number combined with decreasing energies when the inherent correlation error is accounted for, lets us anticipate that the aug-cc-pVDZ level provides results closest to experimental values. This statement finds support in the direct comparison with experimental results in the following subsection (3.B.2).

The general trends for uracil are also observed for the other NA bases and triplet excited states. The inadequacy of small basis sets may be of particular relevance for the reliability of detailed studies (e.g., deactivation and other photophysical processes) also with other approaches to dynamic electron correlation when such small basis sets are employed, quite

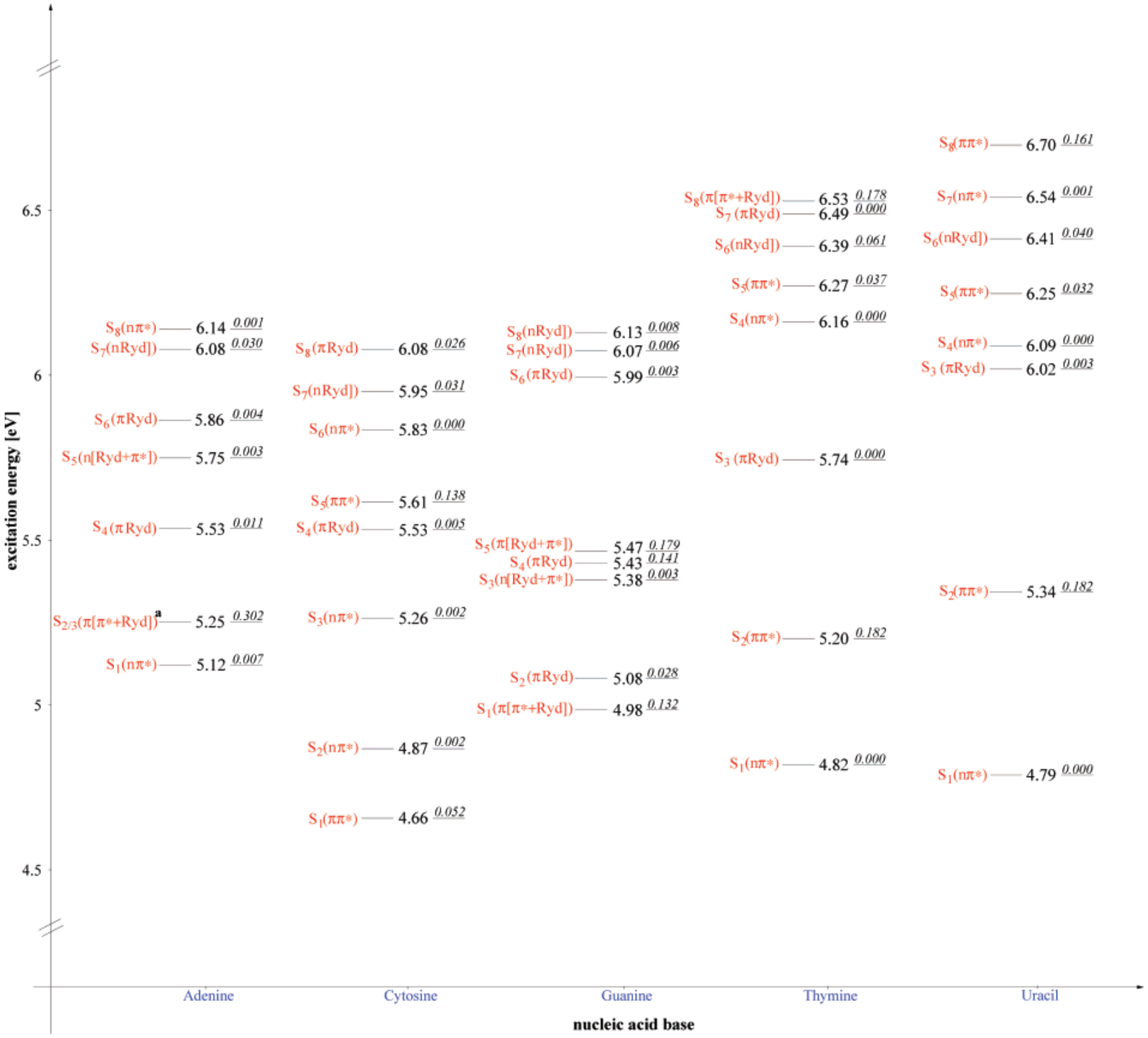

Figure 2. CC2 vertical singlet excitation energies with the aug-cc-pVTZ basis set, at corresponding ground-state optimized geometry; oscillator strengths of the transitions are given in italics. (a) Two excited states are obtained as a conjugated pair of degenerate roots with complex eigenvalues (see text for details).

frequently 6-31G(*) (e.g., in refs 3, 33). Moreover, the inclusion of augmenting diffuse basis functions (aug-cc-pVTZ compared with cc-pVTZ) not only opens for the description of Rydberg-type states but also improves the excitation energies of valence-excited states significantly (Figure 1). We ascribe this finding to the fact that the one-particle basis sets we are discussing have been optimized for atomic ground states and that the additional diffuse basis functions play a non-negligible role in the description of molecular excited states, even if these are valence-excited states like in the present case.

A specific observation of Neiss et al. for the uracil molecule,[34] namely that $n\pi^*$ transitions are supposed to be rather insensitive to the size of the basis set is not supported by our study. As shown in Figure 1, we find the same trend in the same order of magnitude for singlet excitation energies with varying basis sets, irrespective of the character of the transition.

Conclusively, we consider aug-cc-pVDZ a reliable basis set quality for the study of larger systems involving the NA base monomers. From systematic investigations, it is known that the errors due to incomplete treatment of electron correlation typically surpass the basis set error at the aug-cc-pVTZ level,[18,35,36] which underlines our judgment.

2. Excitation Energies and Character of Excited States. Vertical excitation energies of all five NA bases, both for singlet and triplet states, and using the aug-cc-pVTZ basis set throughout (the largest common set for all studies), are displayed in Figures 2 and 3. In the following, we discuss the NA bases separately, but refrain from repeating all reference values reported in the literature, as a very extensive overview has been given by Crespo-Hernández et al.[1] Instead, we compare our results selectively with available experimental values and some recent studies from the theoretical literature. An overview is given in Table 7.

Adenine. Despite the fact that adenine has received considerable attention by both theoreticians and experimentalists, there does not seem to be a consensus on the character of the lowest singlet-excited state. Our most sophisticated calculation (CC2/aug-cc-pVTZ, Table 2) predicts $n\pi^*$ as the S_1 state, with two close-lying $\pi\pi^*$ states 0.13 eV above the S_1 state. We obtain the two low-lying $\pi\pi^*$ states as degenerate complex eigenvectors within response theory, an artifact which may be attributed to the nonsymmetric Jacobi matrix for CC2, leading in the respective region of the potential hypersurface to a conjugated pair of degenerate roots with complex eigenvalues[37] instead of

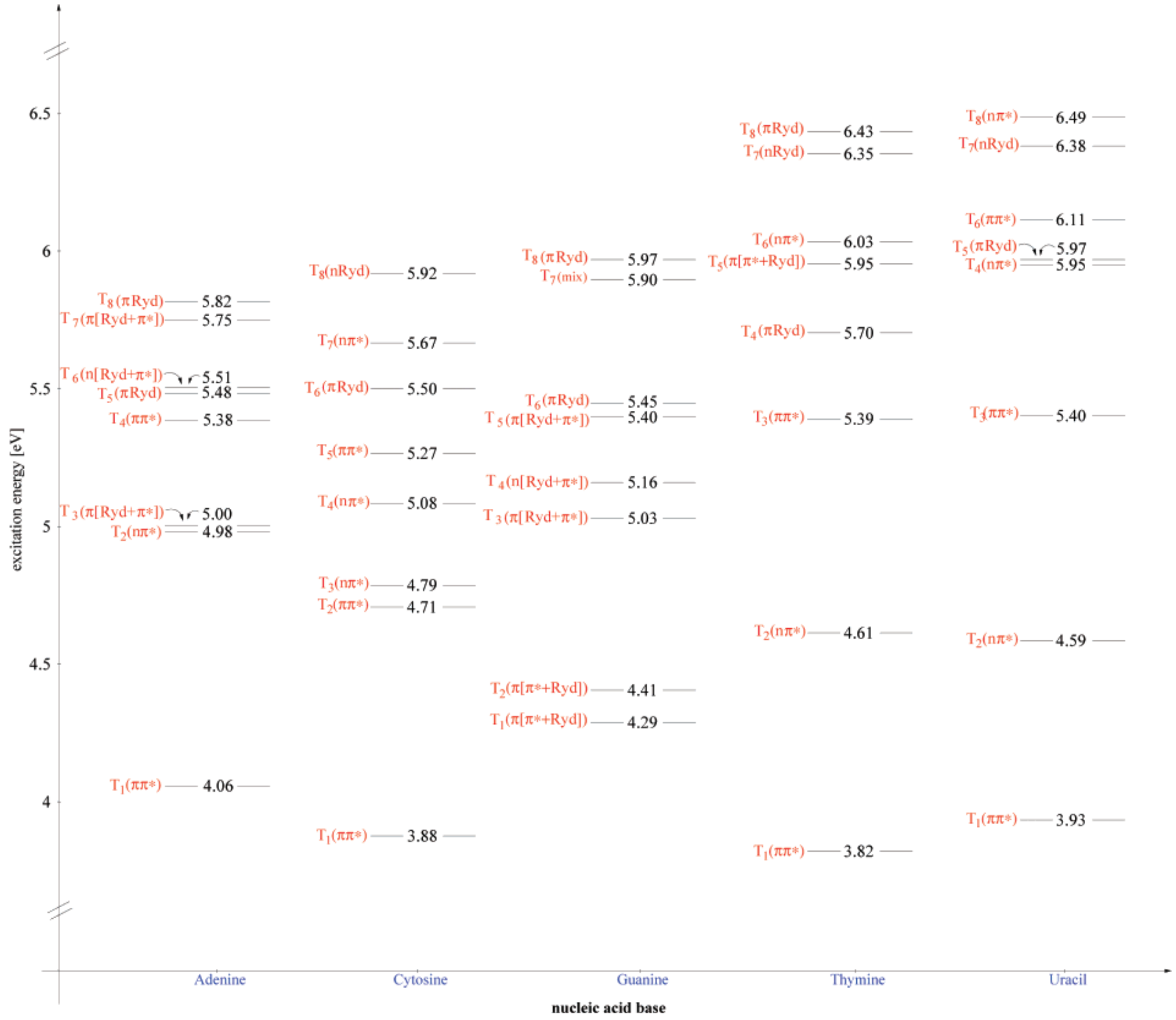

Figure 3. CC2 vertical triplet excitation energies with the aug-cc-pVTZ basis set, at corresponding ground-state optimized geometry.

two almost degenerate real eigenvalues. We are confident, however, that the two states correspond to the lowest $\pi\pi^*$ states identified previously both by experiment and theory (e.g., refs 4,33,38), as the orbital character and the oscillator strengths we obtain are in agreement with earlier findings.

From the work of Clark et al.,[39] which allows for direct comparison with most theoretical work, it cannot be deduced whether S_1 is of $\pi\pi^*$ (at 4.98 eV) or $n\pi^*$ character. Our prediction is, however, in general agreement with the experimental work of Kim et al.,[38] from which vibration-corrected (E_{0-0}) excitation energies of 4.40 eV for $n\pi^*$ and 4.48 eV for $\pi\pi^*$, respectively, can be deduced. Vibronic coupling between the respective electronic states is known to be large,[40] but the proximity effect would lead to a further lowering of the $n\pi^*$ state relative to the $\pi\pi^*$ excited states if considered in our calculations. For adenine, we have also carried out a geometry optimization in the lowest ($\pi-\pi^*$) excited state at the CC2/aug-cc-pVDZ level, yielding an adiabatic excitation energy of 4.47 eV (ls electrons frozen). This value is in perfect agreement with the vibration-corrected experimental result of Kim et al.[38]

CC2 excitation energies are typically up to a few tenths of an eV too high,[17] and this is rather systematic for all excited states. Therefore, we cannot support an $n\pi^*$ excitation energy in adenine of 6.15 eV as obtained by Fülscher et al. with CASSCF/CASPT2.[13] In a more recent study with CASSCF/CASPT2,[2] a value of 4.96 eV is reported that agrees well with our result. The defect of the earlier calculation was there explained by the use of an active orbital space in ref 13 insufficient for describing the $n\pi^*$ excitation properly. Two quite recent LDA/TDDFT studies on the NA bases and also their dimers report the $\pi\pi^*$ excited states at lower energies, and this applies to all NA bases as compared to our results. Tsokalidis et al.[41] find the lowest $\pi\pi^*$ states at 4.52 and 4.95 eV, and Varsano et al.[42] at 4.51 and 4.88 eV, respectively. Their results are generally in good agreement with experiment, but deviations may be in both directions, in contrast to our values where the excitation energies are systematically overestimated.

For higher excited states, we observe considerable Rydberg-valence mixing when the augmented basis sets are used. This applies also to the triplet-excited states of adenine in Table 2, for which there is fairly good agreement in the lower part of the spectrum with earlier DFT/MRCI calculations.[4]

Thymine. Owing to the double keto group in this NA base, we find a low-lying $n\pi^*$ singlet state involving the nonbonding oxygen electrons, shown in Table 3. The lowest excited states are well separated energetically, which also applies to the triplet-excited states in Table 3 and in Figure 3. The excitation energies are in general agreement with those obtained with other methods,[1] in particular, the TD-DFT(B3LYP) calculations of Crespo-Hernández et al. (in ref 1). Our best value of 5.20 eV for the lowest $\pi\pi^*$ excitation energy exhibits about the same deviation from experimental results (obtained in gas phase and

TABLE 2: Excited Singlet and Triplet States of Adenine and Their Character (aug-cc-pVTZ); Oscillator Strengths f for Transitions[a]

state	transition	ΔE (eV)	$t_1/\|\|t\|\|$ (%)	f
S_1	$n \rightarrow \pi^*$	5.121	0.704 (49.6)	0.00672
			0.498 (24.8)	
S_2	$\pi \rightarrow \pi^*$	5.250	−0.453 (20.5)	−0.01737
	$\pi \rightarrow$ Ryd		−0.404 (16.3)	
	$\pi \rightarrow \pi^*$		0.336 (11.3)	
S_3	$\pi \rightarrow \pi^*$	5.250	0.534 (28.5)	0.30215
			0.444 (19.7)	
	$\pi \rightarrow$ Ryd		−0.337 (11.4)	
S_4	$\pi \rightarrow$ Ryd	5.534	0.685 (46.9)	0.01057
	$\pi \rightarrow$ Ryd		−0.400 (16.0)	
S_5	n $\rightarrow$ Ryd	5.749	−0.533 (28.4)	0.00264
			−0.519 (27.0)	
	$n \rightarrow \pi^*$		0.448 (20.1)	
S_6	$\pi \rightarrow$ Ryd	5.863	0.668 (44.6)	0.00372
	$\pi \rightarrow$ Ryd		0.478 (22.8)	
S_7	$n \rightarrow$ Ryd	6.078	0.762 (58.1)	0.03023
S_8	$n \rightarrow \pi^*$	6.139	0.682 (46.6)	0.00121
			0.499 (24.9)	
T_1	$\pi \rightarrow \pi^*$	4.058	0.702 (49.3)	
			0.471 (22.2)	
T_2	$n \rightarrow \pi^*$	4.979	−0.694 (48.2)	
			−0.486 (23.6)	
T_3	$\pi \rightarrow$ Ryd	5.002	−0.539 (29.0)	
			−0.484 (23.4)	
	$\pi \rightarrow \pi^*$		0.423 (17.9)	
T_4	$\pi \rightarrow \pi^*$	5.385	0.476 (22.7)	
			−0.403 (16.2)	
T_5	$\pi \rightarrow$ Ryd	5.483	0.563 (31.7)	
	$\pi \rightarrow$ Ryd		−0.412 (16.9)	
T_6	$n \rightarrow$ Ryd	5.505	−0.461 (21.2)	
			−0.459 (21.1)	
	$n \rightarrow \pi^*$		0.363 (13.2)	
T_7	$\pi \rightarrow$ Ryd	5.747	−0.475 (22.6)	
	$\pi \rightarrow \pi^*$		0.441 (19.4)	
	$\pi \rightarrow$ Ryd		−0.376 (14.2)	
	$\pi \rightarrow \pi^*$		0.320 (10.2)	
T_8	$\pi \rightarrow$ Ryd	5.816	0.425 (18.0)	
	$\pi \rightarrow$ Ryd		0.374 (14.0)	

[a] $t_1/||t||$ denotes the relative weight of the single-excitation amplitudes (t_1) in the CC expansion.

TABLE 3: Excited Singlet and Triplet States of Thymine and Their Character (aug-cc-pVTZ); Oscillator Strengths f for Transitions[a]

state	transition	ΔE (eV)	$t_1/\|\|t\|\|$ (%)	f
S_1	$n \rightarrow \pi^*$	4.818	0.606 (36.7)	0.00002
			0.573 (32.9)	
S_2	$\pi \rightarrow \pi^*$	5.201	0.629 (39.6)	0.18168
			0.539 (29.1)	
S_3	$\pi \rightarrow$ Ryd	5.743	0.827 (68.4)	0.00038
			−0.320 (10.2)	
S_4	$n \rightarrow \pi^*$	6.162	0.409 (16.7)	0.00007
			0.400 (16.0)	
			0.387 (14.9)	
S_5	$\pi \rightarrow \pi^*$	6.271	0.561 (31.5)	0.03686
			0.556 (30.9)	
S_6	$n \rightarrow$ Ryd	6.391	0.718 (51.5)	0.06121
			−0.395 (15.60)	
S_7	$\pi \rightarrow$ Ryd	6.489	0.565 (31.9)	0.00042
			0.475 (22.6)	
			0.360 (13.0)	
S_8	$\pi \rightarrow \pi^*$ + Ryd	6.528	−0.591 (34.9)	0.17755
			0.458 (21.0)	
			−0.360 (13.0)	
T_1	$\pi \rightarrow \pi^*$	3.820	0.645 (41.6)	
			0.559 (31.2)	
T_2	$n \rightarrow \pi^*$	4.614	0.599 (35.9)	
			0.560 (31.4)	
T_3	$\pi \rightarrow \pi^*$	5.390	0.530 (28.1)	
			0.483 (23.3)	
			0.334 (11.2)	
T_4	$\pi \rightarrow$ Ryd	5.703	0.816 (66.5)	
			−0.326 (10.6)	
T_5	$\pi \rightarrow \pi^*$ + Ryd	5.954	−0.491 (24.1)	
			0.398 (15.8)	
			−0.363 (13.2)	
T_6	$n \rightarrow \pi^*$	6.034	0.425 (18.0)	
			0.352 (12.4)	
			0.327 (10.7)	
T_7	$n \rightarrow$ Ryd	6.354	−0.723 (52.3)	
			0.466 (21.7)	
T_8	$\pi \rightarrow$ Ryd	6.435	0.608 (36.9)	
			0.397 (15.7)	
			0.354 (12.5)	

[a] $t_1/||t||$ denotes the relative weight of the single-excitation amplitudes (t_1) in the CC expansion.

solution, as shown in Table 7) as those for the other NA bases. For both the singlet- and the triplet-excited states, Rydberg-valence mixing is far less pronounced than in adenine.

Guanine. Guanine comprises an exception in our series of calculations, in the sense that it is the only molecule where the ordering of lower excited states varies with the extent of the basis set. Mainly for this reason, we include more details on this molecule, which are shown in Figure 4. When using the aug-cc-pVTZ basis and a frozen-core (fc) approximation for the 1s electrons, the S_1 state has the expected $\pi\pi^*$ character, and the second $\pi\pi^*$ as well as the $n\pi^*$ states reported in the literature[1] are found at significantly higher energies. The identification of these states is, however, unambiguous due to the oscillator strengths, which are in good agreement, e.g., with the values of Fülscher et al. with CASSCF/CASPT2.[13] We obtain a low-lying additional state, though, which we ascribe strong Rydberg character and which does not appear when using a basis set without diffuse functions (Figure 4). In comparison with experiment (in ref 13), our excitation energies appear to be somewhat too high, in this case with deviations up to 0.5 eV for the low-lying $\pi\pi^*$ state. Our oscillator strength of 0.132 from the vertical calculation for this state, however, agrees quite well with the experimental value of 0.16 (in ref 13). Most of the excited states are mixed both with respect to Rydberg contributions as well as the approximate designation in terms of angular momentum projection (σ, π, etc.). The Rydberg-valence mixing is especially pronounced when all electrons are correlated, in which case we find strong contributions of Rydberg-type excitations to all excited states (aug-cc-pVTZ). For some of the higher triplet states in Figure 3, this mixing does not even allow a precise assignment. Beside the oscillator strengths, which should be largest for states of $\pi\pi^*$ character, we used $\langle r^2\rangle$ expectation values for the involved orbitals to identify valence and Rydberg characters of excited states. This led us to the assignments as given in Figure 4. A pronounced feature of the guanine spectrum when larger basis sets are used is the occurrence of groups of excited states separated by significant energy gaps. The groups arise through changes of the participating virtual orbitals, whereas the occupied orbitals from which the excitations are performed remain essentially the same (Table 4).

Cytosine. For the singlet-excited states of cytosine, we can support the consensus[1] that S_1 is of $\pi\pi^*$ and S_2 of $n\pi^*$ character, with the two states located quite close in energy (Table 5). With respect to the oscillator strengths of the respective transitions, however, our results agree very well with the CASSCF/CASPT2 values of Merchan et al.[43] and Fülscher et al.,[44] whereas the result for the lowest $\pi\pi^*$ transition obtained with a shifted CIS calculation[11] deviates from these by a factor of 2. Most

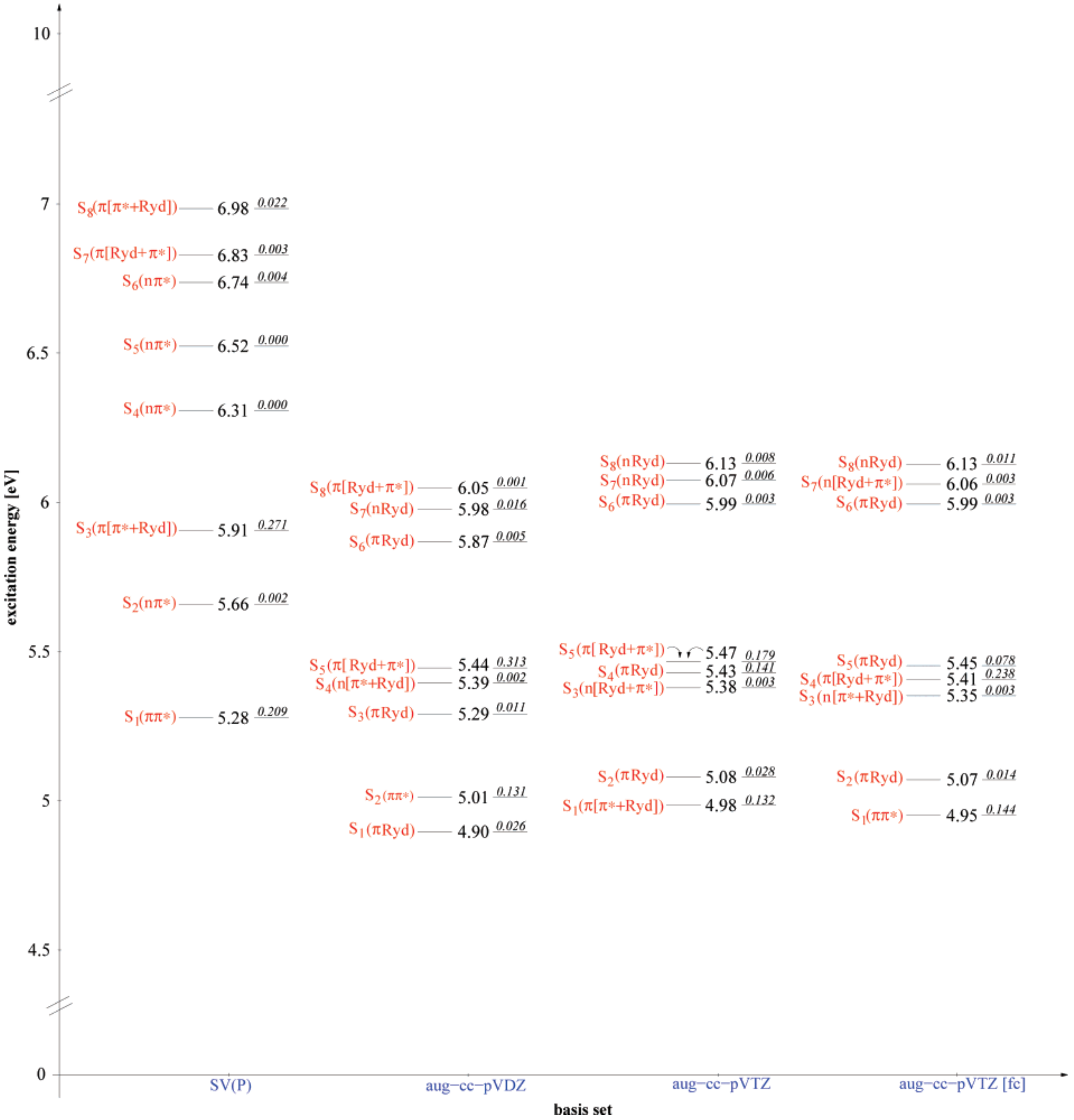

Figure 4. Guanine: CC2 vertical singlet excitation energies with different basis sets, at corresponding ground-state optimized geometry; oscillator strengths of the transitions are given in italics. [fc] denotes a calculation with a frozen core of 1s electrons.

experimental values for the S_1 excitation energy (in ref 44) are around 4.6 eV, in perfect agreement with our highest-level value (CC2/aug-cc-pVTZ). The gas-phase result of Clark et al.,[39] however, with a maximum at 4.28 eV, is significantly lower.

Particularly interesting is the comparison of our triplet excitation energies for cytosine with those of Nguyen et al.[45] For the T_1 state, Nguyen et al. obtain 3.63 eV with DFT/B3LYP (largest basis set reported), 3.83 eV with CCSD, and 3.75 eV with CCSD(T). Our RI-CC2/aug-cc-pVTZ value of 3.88 eV (Table 5) agrees quite well with their CC results, but the DFT value appears to be somewhat too low. In the following subsection on uracil, we will look more closely into electron correlation errors arising from truncated CC expansions.

Uracil. As shown in Table 6 and Figure 1, two low-lying excited singlet states are found for uracil, the lower identified as an $n\pi^*$ transition, the higher as $\pi\pi^*$, in accord with most previous studies.[1,14,45] The excitation energy of 5.35 eV for the $\pi\pi^*$ state agrees fairly well with the gas-phase result of Clark et al.[39] (5.08 eV), and the corresponding oscillator strengths of the two states closely resemble those obtained by Lorentzon et al.[14] The lowest $\pi\pi^*$ excitation energy of Varsano et al.[42] from LDA/TDDFT at 4.69 eV appears to be somewhat too low. All excitations involve electrons from occupied molecular orbitals with strong oxygen participation, and the energy gap in the spectrum arises from the character of virtual orbitals, which are higher lying for the upper part of the spectrum. Especially the lower-energy part of the calculated spectrum closely resembles that obtained for thymine, both with respect to the ordering/character of the excited states and the excitation energies. Our triplet excitation spectrum in Table 6 agrees well with that obtained earlier with the DFT/MRCI method.[46]

Finally, we have been able to assess electron correlation errors through calculations of higher accuracy on uracil. In Table 8, we report coupled cluster excitation energies obtained with the smallest basis set used here, SV(P), and the DALTON[47] program package. For very small model systems, the errors for excitation

TABLE 4: Excited Singlet and Triplet States of Guanine and Their Character (aug-cc-pVTZ); Oscillator Strengths f for Transitions[a]

state	transition	ΔE (eV)	$t_1/\|\|t\|\|$ (%)	f
S_1	$\pi \rightarrow \pi^* +$ Ryd	4.984	−0.447 (20.0)	0.13151
			0.400 (16.0)	
			0.361 (13.0)	
			0.317 (10.0)	
S_2	$\pi \rightarrow$ Ryd	5.079	0.695 (48.4)	0.02776
S_3	$n \rightarrow$ Ryd $+ \pi^*$	5.379	−0.368 (13.5)	0.00281
			−0.336 (11.3)	
S_4	$\pi \rightarrow$ Ryd	5.429	0.646 (41.7)	0.14133
S_5	$\pi \rightarrow$ Ryd $+ \pi^*$	5.466	0.595 (35.3)	0.17925
			−0.335 (11.2)	
S_6	$\pi \rightarrow$ Ryd	5.993	0.755 (57.0)	0.00335
S_7	$n \rightarrow$ Ryd	6.073	0.444 (19.8)	0.00587
			0.331 (11.0)	
S_8	$n \rightarrow$ Ryd	6.130	0.361 (13.0)	0.00770
T_1	$\pi \rightarrow \pi^* +$ Ryd	4.289	0.520 (27.0)	
			−0.411 (16.9)	
			−0.402 (16.2)	
			0.333 (11.1)	
T_2	$\pi \rightarrow \pi^*$	4.405	0.470 (22.1)	
	$\pi \rightarrow$ Ryd		0.376 (14.1)	
T_3	$\pi \rightarrow$ Ryd $+ \pi^*$	5.030	0.789 (62.2)	
T_4	$n \rightarrow$ Ryd $+ \pi^*$	5.161	−0.377 (14.2)	
			−0.326 (10.6)	
T_5	$\pi \rightarrow$ Ryd	5.397	−0.518 (26.8)	
	$\pi \rightarrow \pi^*$		0.351 (12.3)	
			−0.332 (11.0)	
T_6	$\pi \rightarrow$ Ryd	5.447	0.684 (46.7)	
T_7	mix	5.896		
T_8	$\pi \rightarrow$ Ryd	5.970	0.703 (49.4)	

[a] $t_1/||t||$ denotes the relative weight of the single-excitation amplitudes (t_1) in the CC expansion.

TABLE 5: Excited Singlet and Triplet States of Cytosine and Their Character (aug-cc-pVTZ); Oscillator Strengths f for Transitions[a]

state	transition	ΔE [eV]	$t_1/\|\|t\|\|$ (%)	f
S_1	$\pi \rightarrow \pi^*$	4.657	0.562 (31.5)	0.05180
			0.436 (19.0)	
			−0.353 (12.5)	
S_2	$n \rightarrow \pi^*$	4.868	0.561 (31.4)	0.00188
			0.441 (19.4)	
			−0.363 (13.2)	
S_3	$n \rightarrow \pi^*$	5.262	0.547 (29.9)	0.00160
			0.447 (20.0)	
			−0.362 (13.1)	
S_4	$\pi \rightarrow$ Ryd	5.530	0.800 (64.0)	0.00511
S_5	$\pi \rightarrow \pi^*$	5.615	0.541 (29.2)	0.13808
			0.401 (16.0)	
			−0.381 (14.5)	
S_6	$n \rightarrow \pi^*$	5.834	0.475 (22.6)	0.00002
			−0.389 (15.1)	
			0.328 (10.7)	
S_7	$n \rightarrow$ Ryd	5.950	0.657 (43.2)	0.03084
S_8	$\pi \rightarrow$ Ryd	6.079	0.552 (30.4)	0.02553
			0.490 (24.0)	
T_1	$\pi \rightarrow \pi^*$	3.876	0.622 (38.7)	
			0.375 (14.0)	
			−0.360 (12.9)	
T_2	$\pi \rightarrow \pi^*$	4.708	−0.509 (25.9)	
			−0.419 (17.5)	
			0.323 (10.4)	
T_3	$n \rightarrow \pi^*$	4.787	0.528 (27.9)	
			0.415 (17.2)	
			−0.342 (11.7)	
T_4	$n \rightarrow \pi^*$	5.085	0.559 (31.2)	
			0.461 (21.2)	
			−0.367 (13.5)	
T_5	$\pi \rightarrow \pi^*$	5.266	0.543 (29.5)	
			−0.447 (20.0)	
			−0.328 (10.7)	
T_6	$\pi \rightarrow$ Ryd	5.500	0.792 (62.7)	
			0.321 (10.3)	
T_7	$n \rightarrow \pi^*$	5.667	0.455 (20.7)	
			−0.375 (14.1)	
			0.347 (12.1)	
T_8	$n \rightarrow$ Ryd	5.919	−0.625 (39.0)	

[a] $t_1/||t||$ denotes the relative weight of the single-excitation amplitudes (t_1) in the CC expansion.

energies are known to decrease along the series CC2-CCSD-CC3, where CC3 provides a very high level of accuracy close to the exact basis set energies from full configuration interaction (FCI) calculations.[48] For more realistic systems such as benzene, however, CCSD excitation energies tend to overshoot for many low-lying states of $\pi-\pi^*$ character,[35] and CC2 excitation energies are in fact closer to the exact vertical electronic excitation energies. We make precisely this finding for two selected excited states of uracil, from which the CC2 correlation errors are seen to be around +0.1 eV. This is also in line with earlier findings, that CC2 errors are small if the double replacement character of a given excited state is well below 10%,[35] which is the case here. It was not possible to carry out the same study using a larger basis set due to the steep increase in computation time. The results in Table 8 also substantiate that errors from the RI approximation are negligible.

We gain confidence that the obtained vertical RI-CC2 excitation energies also for the other NA bases, which have similar electronic structure, are of high accuracy. As far as the present studies are concerned, we recommend the computational level RI-CC2/aug-cc-pVDZ as the most beneficial CC electronic structure level for the calculation of vertical electronic excitation energies of medium-sized molecules.

4. Concluding Remarks

Our systematic series of geometry optimizations for the NA bases suggests that an accuracy of up to 0.01 Å for equilibrium bond lengths can only be obtained with basis sets of at least triple-ζ quality. Such basis sets can be used for the monomer studies here, but for larger systems (like, e.g., the corresponding dimers), one has to compromise. Furthermore, we show the trends of electron correlation on molecular geometries and establish the expected behavior of MP2 as compared to our CC2 results, where the latter typically come out slightly too long and vice versa for the former. However, the deviations are quite small overall, making both MP2 and CC2 geometries reliable starting points for further studies of vertical excitation spectra.

For both singlet and triplet vertical excitation energies, we obtain reliable results with the CC2/response method. The description of various excited states is balanced and we assign errors in a range of 0.1−0.3 eV, the obtained energies overshooting the expected experimental results. Upon investigating electron correlation errors and regarding basis set errors, we observe a favorable error cancellation from basis set limitations and limited correlation treatments at the CC2/aug-cc-pVDZ level. In these studies, the use of a high-level approach to electron correlation like the CC2 method and the use of extensive basis sets becomes a major issue in obtaining accurate results. In particular, small basis sets like the split-valence set with polarization functions (SV(P)) in combination with the CC2 approach is clearly insufficient in describing the excited states of NA bases satisfactorily. We consider it rather important to note that the addition of diffuse functions (aug) to one-particle basis sets yields significantly lower and more precise excitation energies than calculations without such functions. For geometry

TABLE 6: Excited Singlet and Triplet States of Uracil and Their Character (aug-cc-pVQZ) and Oscillator Strengths f for Transitions[a]

state	transition	ΔE [eV]	$t_1/\|\|t\|\|$ (%)	f
S_1	$n \rightarrow \pi^*$	4.802	0.612 (37.4)	0.00000
			−0.509 (25.9)	
			0.367 (13.5)	
S_2	$\pi \rightarrow \pi^*$	5.347	0.645 (41.6)	0.18226
			−0.456 (20.7)	
			0.405 (16.4)	
S_3	$\pi \rightarrow$ Ryd	6.085	0.779 (60.7)	0.00250
			−0.393 (15.4)	
S_4	$n \rightarrow \pi^*$	6.103	−0.566 (32.0)	0.00011
			0.391 (15.3)	
S_5	$\pi \rightarrow \pi^*$	6.265	0.630 (39.7)	0.03593
			−0.496 (24.6)	
			0.418 (17.5)	
S_6	$n \rightarrow$ Ryd	6.487	0.744 (55.4)	0.03605
			−0.400 (16.0)	
S_7	$n \rightarrow \pi^*$	6.558	0.595 (35.5)	0.00094
			0.506 (25.6)	
			−0.363 (13.2)	
S_8	$\pi \rightarrow \pi^*$	6.702	0.818 (67.0)	0.16623
T_1	$\pi \rightarrow \pi^*$	3.945	0.665 (44.2)	
			−0.465 (21.6)	
			0.408 (16.6)	
T_2	$n \rightarrow \pi^*$	4.605	0.601 (36.1)	
			−0.501 (25.1)	
			0.352 (12.4)	
T_3	$\pi \rightarrow \pi^*$	5.421	0.552 (30.5)	
			−0.417 (17.4)	
			−0.408 (16.7)	
			0.339 (11.5)	
T_4	$n \rightarrow \pi^*$	5.967	−0.583 (34.0)	
			−0.370 (13.7)	
			0.368 (13.6)	
T_5	$\pi \rightarrow$ Ryd	6.038	0.789 (62.3)	
			−0.408 (16.7)	
T_6	$\pi \rightarrow \pi^*$	6.125	0.682 (46.4)	
			0.401 (16.1)	
T_7	$n \rightarrow$ Ryd	6.459	0.732 (53.6)	
			−0.436 (19.0)	
T_8	$n \rightarrow \pi^*$	6.506	0.555 (30.8)	
			0.531 (28.2)	
			−0.384 (14.7)	

[a] $t_1/||t||$ denotes the relative weight of the single-excitation amplitudes (t_1) in the CC exapansion.

TABLE 7: Selected Vertical Excitation Energies of the NA Bases from Present CC2 Calculations (aug-cc-pVTZ Basis Set, aug-cc-pVQZ for Uracil) and Comparison with Experimental Data

NA base	state (transition)	RI-CC2 T_v [eV]	experimental	
adenine	S_1 ($n-\pi^*$)	5.12	4.40[a]	
				4.98[b]
	S_2 ($\pi-\pi^*$)	5.25	4.48[a]	
thymine	S_1 ($n-\pi^*$)	4.82		
	S_2 ($\pi-\pi^*$)	5.20	4.8[c]	4.7[d]
guanine	S_1 ($\pi-\pi^*$)	4.98	4.4−4.5[e]	
cytosine	S_1 ($\pi-\pi^*$)	4.66	4.28[b]	
uracil	S_1 ($n-\pi^*$)	4.80		
	S_2 ($\pi-\pi^*$)	5.35	5.08[b]	

[a] E_{0-0}, vibration corrected.[38] [b] Vapor phase.[39] [c] 1,3-Dimethyluracil, gas phase.[39] [d] In water and TMP.[39] [e] From ref 13.

optimizations in both ground and excited states, on the other hand, one-particle basis sets without diffuse functions may be more favorable as basis set superposition errors are smaller.

For thymine (SV(P), aug-cc-pVDZ, and aug-cc-pVTZ), guanine (SV(P) and aug-cc-pVTZ), and uracil (SV(P), aug-cc-pVDZ, and aug-cc-pVTZ), we compared our calculations with frozen-core (all 1s electrons) approximated ones and found very small deviations for both geometries (max of 0.002 Å) and excitation energies (max of 0.01 eV, 0.03 eV for guanine) which are much smaller than the other errors discussed here.

TABLE 8: Lowest $\pi-\pi^*$ Excitation Energies of Uracil, SV(P) Basis Set[a]

state	RI-CC2	CC2	CCSD	CC3
S_2	5.67	5.67	5.81	5.61
S_4	6.63	6.63	6.91	6.51

[a] RI-CC2 from TURBOMOLE, all others from DALTON.

In ongoing studies, we exploit the obtained knowledge and investigate particularly charge-transfer states of the NA base dimers at the CC2/aug-cc-pVDZ level. Here TDDFT is no longer applicable with standard density functionals and the application of the CASSCF/CASPT2 approach is hampered by the computational demand arising especially from the requirement for extensive complete active orbital spaces. We expect that the CC2 method is a significant and accurate tool for the investigation of excited singlet and triplet states of systems of increasing size like the mentioned NA base dimers or caged compounds.

Acknowledgment. T.F. and S.K. acknowledge support from the Collaborative Research Center (SFB) 663, “Molecular Response after Electronic Excitation” (Düsseldorf). We thank the reviewers for a number of helpful comments and suggestions.

Supporting Information Available: Detailed tables and additional figures. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Crespo-Hernández, C. E.; Cohen, B.; Hare, P. M.; Kohler, B. *Chem. Rev.* **2004**, *104*, 1977.

(2) Serrano-Andrés, L.; Merchán, M.; Borin, A. C. *Chem.—Eur. J.* **2006**, *12*, 6559.

(3) Perun, S.; Sobolewski, A. L.; Domcke, W. *J. Am. Chem. Soc.* **2005**, *127*, 6257.

(4) Marian, C. M. *J. Chem. Phys.* **2005**, *122*, 104314.

(5) Canuel, C.; Elhanine, M.; Mons, M.; Piuzzi, F.; Tardivel, B.; Dimicoli, I. *Phys. Chem. Chem. Phys.* **2006**, *8*, 3978.

(6) Chen, H.; Li, S. *J. Chem. Phys.* **2006**, *124*, 154315.

(7) Blancafort, L. *J. Am. Chem. Soc.* **2006**, *128*, 210.

(8) Matsika, S. *J. Phys. Chem. A* **2005**, *109*, 7538.

(9) Fogarasi, G. *J. Phys. Chem. A* **2002**, *106*, 1381.

(10) Dabkowska, I.; Jurečka, P.; Hobza, P. *J. Phys. Chem. A* **2002**, *106*, 1381.

(11) Shukla, M. K.; Mishra, P. C. *Chem. Phys.* **1999**, *240*, 319.

(12) Mennucci, B.; Toniolo, A.; Tomasi, J. *J. Phys. Chem. A* **2001**, *105*, 7126.

(13) Fülscher, M.; Serrano-Andrés, L.; Roos, B. O. *J. Am. Chem. Soc.* **1997**, *119*, 6168.

(14) Lorentzon, J.; Fülscher, M.; Roos, B. O. *J. Am. Chem. Soc.* **1995**, *117*, 9265.

(15) Christiansen, O.; Koch, H.; Jørgensen, P. *Chem. Phys. Lett.* **1995**, *243*, 409.

(16) Fleig, T. **2006**, unpublished.

(17) Hättig, C.; Weigend, F. *J. Chem. Phys.* **2000**, *113*, 5154.

(18) Hättig, C.; Köhn, A. *J. Chem. Phys.* **2002**, *117*, 6939.

(19) Hättig, C. *J. Chem. Phys.* **2003**, *118*, 7751.

(20) Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H. Kölmel, C. *Chem. Phys. Lett.* **1989**, *162*, 165.

(21) Schäfer, A.; Horn, H.; Ahlrichs, R. *J. Chem. Phys.* **1992**, *97*, 2571.

(22) Dunning, T. H., Jr. *J. Chem. Phys.* **1989**, *90*, 1007.

(23) Kendall, R. A.; Dunning, T. H., Jr.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796.

(24) Weigend, F.; Köhn, A.; Hättig, C. *J. Chem. Phys.* **2002**, *116*, 3175.

(25) Weigend, F.; Häser, M.; Patzelt, H.; Ahlrichs, R. *Chem. Phys. Lett.* **1998**, *294*, 143.

(26) Schaftenaar, G.; Noordik, J. H. *J. Comput. Aided Mol. Des.* **2000**, *14*, 123.

(27) Wang, S.; Schaefer, H. F., III. *J. Chem. Phys*. **2006**, *124*, 044303.

(28) Helgaker, T.; Jørgensen, P.; Olsen, J. *Molecular Electronic Structure Theory*; Wiley & Sons: Chichester, UK, 2000.

(29) Helgaker, T.; Ruden, T. A.; Jørgensen, P.; Olsen, J.; Klopper, W. *J. Phys. Org. Chem.* **2004**, *17*, 913.

(30) Sponer, J.; Hobza, P. *J. Mol. Struct.* **1994**, *304*, 35.

(31) Sponer, J.; Hobza, P. *J. Phys. Chem.* **1994**, *98*, 3161.

(32) Fogarasi, G.; Szalay, P. G. *J. Phys. Chem. A* **1997**, *101*, 1400.

(33) Blancafort, L. *J. Am. Chem. Soc.* **2006**, *128*, 210.

(34) Neiss, C.; Saalfrank, P.; Parac, M.; Grimme, S. *J. Phys. Chem. A* **2003**, *107*, 140.

(35) Christiansen, O.; Koch, H.; Halkier, A.; Jørgensen, P.; Helgaker, T.; Sánchez de Merás, A. *J. Chem. Phys.* **1996**, *105*, 6921.

(36) Christiansen, O.; Stanton, J.; Gauss, J. *J. Chem. Phys*. **1998**, *108*, 3987.

(37) Hättig, C. *Adv. Quantum Chem.* **2005**, *50*, 37.

(38) Kim, N. J.; Jeong, G.; Kim, Y. S.; Sung, J.; Kim, S. K.; Park, Y. D. *J. Chem. Phys.* **2000**, *113*, 10051.

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

(40) Kang, H.; Lee, K. T.; Jung, B.; Ko, Y. J.; Kim, S. K. *J. Am. Chem. Soc*. **2002**, *124*, 12958.

(41) Tsolakidis, A.; Kaxiras, E. *J. Phys. Chem. A* **2005**, *109*, 2373.

(42) Varsano, D.; Felice, R. D.; Marques, M. A. L.; Rubio, A. *J. Phys. Chem. A* **2006**, *110*, 7129.

(43) Merchán, M.; Serrano-Andrés, L. *J. Am. Chem. Soc.* **2003**, *125*, 8108.

(44) Fülscher, M.; Roos, B. O. *J. Am. Chem. Soc*. **1995**, *117*, 2089.

(45) Nguyen, M. T.; Zhang, R.; Nam, P.-C.; Ceulemans, A. *J. Phys. Chem. A* **2004**, *108*, 6554.

(46) Marian, C. M.; Schneider, F.; Kleinschmidt, M.; Tatchen, J. *Eur. Phys. J. D* **2002**, *20*, 357.

(47) Helgaker, T.; Jensen, H. J. Aa.; Jørgensen, P.; Olsen, J.; Ruud, K.; Ågren, H.; Auer, A. A.; Bak, K. L.; Bakken, V.; Christiansen, O.; Coriani, S.; Dahle, P.; Dalskov, E. K.; Enevoldsen, T.; Fernandez, B.; Hättig, C.; Hald, K.; Halkier, A.; Heiberg, H.; Hettema, H.; Jonsson, D.; Kirpekar, S.; Kobayashi, R.; Koch, H.; Mikkelsen, K. V.; Norman, P.; Packer, M. J.; Pedersen, T. B.; Ruden, T. A.; Sanchez, A.; Saue, T.; Sauer, S. P. A.; Schimmelpfenning, B.; Sylvester-Hvid, K. O.; Taylor, P. R.; Vahtras, O. *DALTON, A Molecular Electronic Structure Program*, release 2.0; Oslo University: Oslo, Norway, 2005; http://www.kjemi.uio.no/software/dalton/dalton.html.

(48) Christiansen, O.; Koch, H.; Jørgensen, P.; Olsen, J. *Chem. Phys. Lett*. **1996**, *256*, 185.

(49) Corresponding data for the other systems may be provided upon request.