JOURNAL

OF THE AMERICAN CHEMICAL SOCIETY

VOLUME 110, NUMBER 8 APRIL 13, 1988

Contributions from Electron Correlation to the Relative Stabilities of the Tautomers of Nucleic Acid Bases

Józef S. Kwiatkowski,[†] Rodney J. Bartlett,*[‡] and Willis B. Person*

Contribution from the Department of Chemistry and the Quantum Theory Project, University of Florida, Gainesville, Florida 32611. *Received July 11, 1986*

Abstract: The contribution of electron correlation to the relative stabilities of tautomeric forms of nucleic acid bases is assessed. Different levels of many-body perturbation theory (MBPT) up to full fourth-order and the coupled-cluster single and double excitation model including the effects of triple excitation contributions are used to study the relative stabilities of models of tautomers of uracil and cytosine, with MBPT(2) being used for the complete molecule. The general phenomena of lactim–lactam and amino–imino tautomerization is analyzed with reference to formamide, formamidine, formamidic acid, and the lactim–lactam tautomers of 2-oxopyridine. The order of the relative electron correlation energies depends upon the systems considered as well as on the method used for calculation. However, the results indicate that in the case of lactim–lactam tautomeric pairs of nucleic acid bases, the electron correlation contributions favor the lactim tautomer, and are a significant correction for the relative energies of the lactim–lactam pair of cytosine, although for uracil it is unimportant. In the case of amino–imino tautomers of the bases, the electron correlation contributions are greater for the amino forms. These correlation effects make a contribution to the relative stability of cytosine tautomers that is comparable to the ΔE_{SCF} differences.

A knowledge of relative stabilities of tautomeric forms of N-heterocyclic molecules as well as of the tautomeric conversion from one tautomeric form to another is important from the point of view of structural chemistry, and in relation to spontaneous point mutations as a consequence of mispairing by rare tautomeric forms of purinic and pyrimidinic nucleic acid bases.[1-3] During recent years a large amount of experimental and theoretical work has been performed in order to understand the phenomenon of tautomerism, and to estimate the relative stabilities of tautomeric forms of nucleic acid bases and related model systems.[4-7] The relative energies and spectra of tautomers are potentially amenable to ab initio theoretical predictions, yet the importance of electron correlation in such studies has not yet been assessed.

The theoretical prediction of the tautomeric stability of a molecule involves a determination of differences between the total SCF + correlated electronic energies of the tautomers including any potential differences in their zero-point vibrational energies (ZPE) to obtain the relative internal energy ΔE°_0 at 0 K. Yet, the correct prediction of the relative total electronic energies of two tautomers is difficult, since this difference corresponds to the difference between two large and nearly equal electronic energies. One hopes in this approach that several errors involved in the estimation of electronic energies (due to the numerous uncertainties from approximations used in the calculations) will cancel when relative energies are considered. However, in general this expectation is probably too optimistic.

Previous calculations of the relative stability of tautomers have been made by semiempirical methods or by ab initio methods at the Hartree–Fock self-consistent field (SCF) level. In this approach electron correlation contributions to the stability of each tautomer are considered to be of the same magnitude, so they are generally omitted from consideration. In fact, calculations with some semiempirical methods presumably include some electron correlation effects, but to an unpredictable extent. On the other hand, electron correlation is well defined in ab initio approaches, but is invariably subject to basis set limitations. Consequently, in only a few cases have correlation effects been estimated for tautomeric equilibria. Examples include the ground states of lactam–lactim and keto–enol tautomeric forms of molecules (acetaldehyde–vinyl alcohol,[8] formamide–formamidic acid, and lactim–lactam pairs of 2- and 4-oxopyridine[9]). The scarcity of such studies is explained partly by the general expectation that

[†] Permanent address: Institute of Physics, N. Copernicus University, 87-100, Torun, Poland.

[‡] Guggenheim Fellow.

(1) Löwdin, P. O. *Adv. Quantum Chem.* **1965**, *2*, 213.
(2) Cooper, W. G. *Int. J. Quantum Chem.* **1978**, *14*, 71.
(3) Rein, R.; Shibata, M.; Garduno-Juarez, R.; Kieber-Emmons, T. In *Structure and Dynamics: Nucleic Acids and Proteins*; Clementi, E., Sarma, R. H., Eds.; Adenine Press: Guilderland, N.Y., 1983; p 269.
(4) Kwiatkowski, J. S.; Pullman, B. *Adv. Heterocycl. Chem.* **1975**, *18*, 199.
(5) Kwiatkowski, J. S.; Zielinski, T. J.; Rein, R. *Adv. Quantum Chem.* **1986**, *18*, 85.
(6) Berndt, M.; Kwiatkowski, J. S. In *Theoretical Chemistry of Biological Systems*; Nâray-Szabó, G., Ed. (Vol. 41 of *Studies in Theoretical and Physical Chemistry*); Elsevier: Amsterdam, 1986; p 349.
(7) For general discussion of tautomerism of heterocycles including biological important molecules, see: Katritzky, A. R.; Lagowski, J. M. *Adv. Heterocycl. Chem.* **196.**, *1*, 311, 339; **1969**, *2*, 1. Elguero, J.; Marzin, C.; Katritzky, A. R.; Linda, P. *Ibid.* **1976**, Suppl. No. 1.
(8) Rodwell, W. R.; Bouma, W. J.; Radom, L. *Int. J. Quantum Chem.* **1980**, *28*, 107.
(9) Schlegel, H. B.; Gund, P.; Fluder, E. M. *J. Am. Chem. Soc.* **1982**, *104*, 5347.

0002-7863/88/1510-2353$01.50/0

the relative electron correlation contributions are less important (although contradicted in the examples cited), and partly because the calculation of these contributions for several electron molecules is not an easy problem.

In the present study we present calculations for the electron correlation contributions for the ground states of tautomers of nucleic acid bases and of certain model systems. We use many-body perturbation theory (MBPT)[10-12] at second-order for cytosine, uracil, and oxypyridine and selected tautomeric forms, while using full fourth-order (MBPT(4)) and couple-cluster (CC) methods,[10,13-15] including some triple excitation effects, for the smaller model systems. Such "many-body" methods are most appropriate for large molecules because of their size-extensive nature[10]. Although the calculations presented here are clearly the most accurate for these systems to date, we cannot expect that extremely small energy differences between tautomeric forms of the large molecules in question can be predicted definitively by even a very high level inclusion of electron correlation contributions for the tautomers due to basis set limitations, However, the present work demonstrates that for certain kinds of bonding patterns electron correlation effects on calculated relative electronic energies of nucleic acid tautomers are essential, and some trends are evident.

Statement of the Problem

Before presenting the results from our study, we shall devote a few words to the discussion of the general problem of the calculation of relative energies of tautomers.

The total internal energy of a molecule can be written as the sum of the energy calculated at the Hartree–Fock level (E^{SCF}), the electron correlation energy (E^{corr}), and the zero-point correction, ZPE. Hence, the difference between the internal energies of two tautomers (for simplicity we consider a molecule existing in only two tautomeric forms A or B) is represented as

$$\Delta E^0_{AB} = \Delta E^{SCF(eq)}_{AB} + \Delta E^{corr(eq)}_{AB} + \Delta ZPE_{AB} \qquad (1)$$

The superscript (eq) means that the differences should be calculated at the global energetic minima of the tautomers; theoretically that means that the full geometry optimization should be carried out for both tautomers A and B with inclusion of electron correlation effects. In practice, the calculations for the optimal structures of complex systems, such as those considered here, are performed at the HF–SCF level only and, as a matter of fact, usually with a relatively small basis set (bs). Such geometries tend often to benefit from cancellation of errors, with 3-21G SCF usually providing structures in better agreement with experiment than 6-31G** SCF (e.g., ref 16). Additional single-point energy calculations with a better basis set (bs') may be performed at the optimal geometries obtained with the bs basis set in order to estimate an approximate value of $\Delta E^{SCF(eq)}_{AB}$ (instead of the exact value; the difference defined in eq 1 is replaced by the approximate value designed as $\Delta E^{SCF(bs'//bs)}_{AB}$). The latter procedure is justified by the observation that the basis set dependence is usually less important for geometry prediction than it is for energies.

On the other hand, the electron correlation contributions are scarcely meaningful in the absence of polarization functions in the basis, so they are calculated with the better basis set bs' at the optimized geometry obtained with basis set bs ($\Delta E^{corr(eq)}_{AB} \approx \Delta E^{corr(bs'//bs)}_{AB}$). Thus instead of evaluating the exact value by eq 1, we obtain an approximate estimate for the relative internal energy difference:

$$\Delta E^0_{AB} \approx \Delta E^{SCF(bs'//bs)}_{AB} + \Delta E^{corr(bs'//bs)}_{AB} + \Delta ZPE_{AB} \qquad (2)$$

The zero-point correction can be obtained as a by-product of the geometry optimization using the usual empirical scale factors.[16] The second and third terms in this expression may often be unimportant if the value of the first term is relatively large. However, we should keep in mind that the value of $\Delta E^{SCF(bs'//bs)}_{AB}$ depends upon the size of the basis set bs' as well as on the geometry assumed for the calculation. A change of the basis set bs' changes the relative energies calculated at the Hartree–Fock level;[5] however, we can expect that improving the basis set bs' may lead to

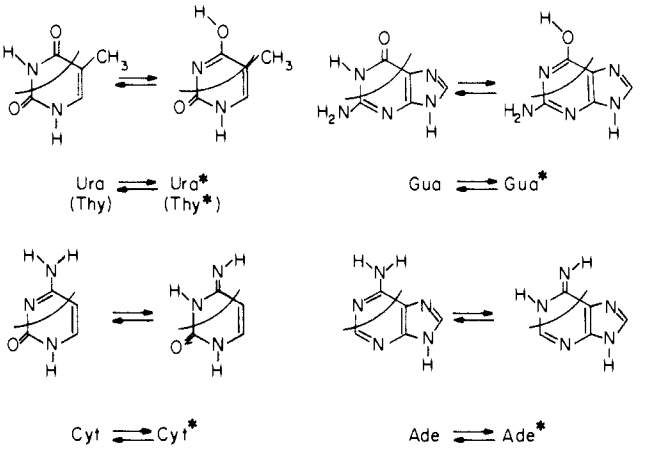

Figure 1. Nucleic acid bases: normal (lactam, amino) and rare (starred; lactim, imino) tautomers. Solid arcs denote the regions of the tautomers involved with tautomeric rearrangements (see Figure 2). Three-letter abbreviations for nucleic acid bases are used throughout the text according to the recommendation of the International Union of Pure and Applied Chemistry and of the International Union of Biochemistry (see also: Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984).

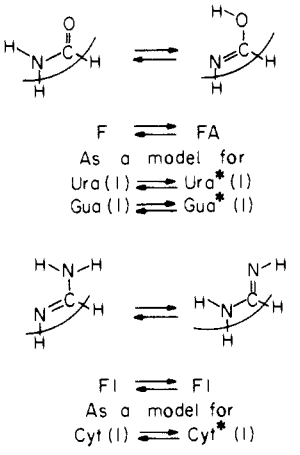

Figure 2. The simplest model systems [formamide (F), formamidic acid (FA), and formamidine (FI)] simulating lactam–lactim [Ura(1), Ura*(1), Gua(1), Gua*(1)] and amino–imino [Cyt(1), Cyt*(1)] tautomeric pairs of nucleic acid bases when the geometries are appropriately adjusted. See text for explanation.

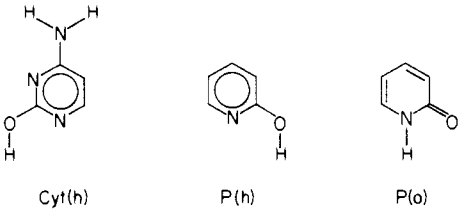

Figure 3. Lactim or hydroxy form C(h) of cytosine and the lactim [P(h) or hydroxy)] and lactam [P(o) or oxo)] tautomeric forms of 2-oxopyridine.

(10) Bartlett, R. J. *Annu. Rev. Phys. Chem.* **1981**, *32*, 359.
(11) Bartlett, R. J.; Shavitt, I. *Chem. Phys. Lett.* **1977**, *50*, 190.
(12) Bartlett, R. J.; Silver, D. M. *J. Chem. Phys.* **1975**, *62*, 3258; **1976**, *64*, 4578.
(13) Bartlett, R. J.; Purvis, G. D. *Int. J. Quantum Chem.* **1978**, *14*, 561. Bartlett, R. J.; Purvis, G. D. *Ann. N. Y. Acad. Sci.* **1981**, *367*, 62.
(14) Lee, Y. S.; Kucharski, S. A.; Bartlett, R. J. *J. Chem. Phys.* **1984**, *81*, 5906.
(15) Urban, M.; Noga, J.; Cole, S. J.; Bartlett, R. J. *J. Chem. Phys.* **1985**, *83*, 4041.
(16) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab. Initio Molecular Orbital Theory*; Wiley: New York, 1986.

Table I. Lactam–Lactim Tautomeric Pairs of Formamide–Formamidic Acid Model Systems: Electron Correlation Energies (in au) and Their Relative Values (in kJ mol^{-1})[a]

methods[b]		systems[c] lactam: F lactim: FA	Ura(1) Ura*(1)	Gua(1) Gua*(1)
MBPT(2)	full	−0.474 559 (0.57)	−0.476 177 (1.68)	−0.477 257 (5.15)
		−0.474 774 (ref)	−0.476 815 (ref)	−0.479 215 (ref)
	fc	−0.463 557 (0.96)	−0.465 285 (1.95)	−0.466 801 (4.52)
		−0.463 921 (ref)	−0.466 028 (ref)	−0.468 521 (ref)
SDQ-MBPT(4)	full	−0.491 604 (4.77)	−0.493 223 (5.23)	−0.494 271 (8.34)
		−0.493 420 (ref)	−0.495 214 (ref)	−0.497 445 (ref)
	fc	−0.481 123 (5.22)	−0.482 848 (5.57)	−0.484 236 (8.02)
		−0.483 109 (ref)	−0.484 966 (ref)	−0.487 286 (ref)
SDTQ-MBPT(4)	fc	−0.496 611 (2.76)	−0.498 706 (3.40)	−0.500 466 (6.35)
		−0.497 662 (ref)	−0.500 000 (ref)	−0.502 882 (ref)
CCSD + T(CCSD)	fc	−0.495 686 (6.15)	−0.497 732 (6.88)	−0.499 406 (9.93)
		−0.498 027 (ref)	−0.500 351 (ref)	−0.503 185 (ref)

[a] The upper number for each pair if E^{corr} for F, Ura(1), and Gua(1); the lower is for FA, Ura*(1), and Gua*(1). The values of electron correlation energies of F relative to FA, Ura(1) to Ura*(1), and Gua(1) to Gua*(1) are given in parentheses. SCF energies E^{SCF} (in au = hartrees) calculated with 6-31G* basis set are −168.929 912 (for F), −168.905 808 (FA), −168.925 958 (Ura(1)), −168.900 665 (Ura*(1)), −168.920 885 (Gua(1)), −168.893 795 (Gua*(1)). The relative SCF energies ($\Delta E^{SCF} = E_F^{SCF} - E_{FA}^{SCF}$) for F–FA, Ura(1)–Ura*(1) and Gua(1)–Gua*(1) pairs are −63.4, −66.5, and −69.1 kJ mol^{-1}, respectively, favoring the lactam forms. [b] Full, all MO's included; fc, "frozen core" approximation. See ref 10 for other definitions. [c] See Figures 1 and 2 for structures of these model systems.

relative energy values closer to those estimated at the Hartree–Fock limit. At any rate, the correlation term in eq 2 can be important in a case when the difference in the relative energies of the two tautomers calculated near the SCF limit is small. Clearly, ΔZPE_{AB} can also be important for sufficiently small differences.

We evaluate electron correlation contributions to the relative stabilities of tautomers for the nuclei acid bases uracil and cytosine and for the model systems (formamide, formamidine). Tautomerism in 2-oxopyridine is also discussed.

Choice of Systems. Figures 1–3 show the systems we have considered. We study first of all the tautomeric pairs of nucleic acid bases which are important from the viewpoint of the theory of point mutations, namely, the normal (lactam, amino) and rare (lactim, imino) tautomeric forms of uracil (Ura, Ura*), guanine (Gua, Gua*), and cytosine (Cyt, Cyt*). We shall also discuss the amino–lactim forms of cytosine as well as the lactim–lactam tautomeric pair for 2-oxopyridine (Figure 3).

Optimization of the geometries of the normal and rare tautomers of these bases has been previously studied with a 3-21G split valence shell basis set (except for adenine).[17,18] In all of our calculations we have used the optimized structures reported there.[17,18] It is worthwhile to note that this optimization leads to relatively good agreement with the experimental geometry for normal tautomeric forms found for the nucleic acids (particularly good agreement is obtained for the calculated geometry of uracil itself compared with results from a recent electron diffraction study of this molecule[19]). The geometries of 2-oxopyridine tautomers have also been taken from a previous 3-21G optimization study by Schlegel et al.[9]

In order to see whether it is possible to replace a complex system like a nucleic acid base by a simple model system showing the same type of tautomerism for the purpose of estimating the relative electron correlation contribution, we consider a model system that shows lactam–lactim and amino–imino types of tautomerism. These simplest model systems are formamide–formamidic acid (F-FA) and the formamidine–formamidine (Fi–FI) pairs (Figure 2). Electron correlation contributions for these pairs are calculated (a) using the optimized geometries found in this study with a 3-21G basis set, and (b) for the same systems but at *geometries constrained to the corresponding fragments of the nucleic acid bases* from the prior optimizations using the same basis set.[17,18] These F, FA, and FI systems at the appropriate geometries are designated in Figure 2 as Ura(1), Ura*(1), Gua(1), Gua*(1), Cyt(1), and Cyt*(1) (the N–H and C–H bond lengths in these model systems are taken to be those calculated at the optimized geometries of F, FA, and FI, but the HNC or HCN bond angles in the model are assumed to be the same as the CNC angles calculated for the corresponding nucleic acid bases[17,18]).

The choice of geometry for these calculations deserves some comment. All molecules in question (formamide, formamidic acid, formamidine, and the nucleic acid bases uracil and cytosine) have been assumed to be planar. This assumption may not, however, be correct for all molecules. The equilibrium structure for the formamide molecule is probably slightly nonplanar,[20] although experimental studies of its structure have led to a number of conflicting conclusions.[21–23] Amino tautomers of the cytosine molecule are probably also nonplanar.[20,24–26] However, the deviations from nonplanarity in all cases are small, and it seems that consideration of nonplanarity for the systems in question would not significantly change the conclusions regarding the relative energies.

Results of the Calculation and Discussion

Electron correlation contributions for the model systems in question have been calculated using a 6-31G* split valence basis set at different levels of many-body perturbation theory up to full fourth-order [MBPT(2), SDQ–MBPT(4), SDTQ–MBPT(4)] with all MO's (full) and subject to the 1s core electrons not being correlated, i.e., "frozen core" (fc). We also report results from the coupled-cluster single and double excitation (CCSD) model including the effects of triple excitation contributions obtained by a single triple excitation iteration using converged CCSD coefficients.[15] This model is termed CCSD + T(CCSD). Calculations of energies at the latter level have been shown to be very close (within 5 kJ mol^{-1})[14,15] to the full CI results (i.e., to the basis set limit correlation results) for several molecules.[27]

Model Systems. The results of the present calculations for small model systems are summarized in Tables I and II. The relative electron correlation energy for the F–FA pair was estimated previously by Schelgel et al.[9] at the MBPT(2) level in the fc approximation to be 0.5 or 1.0 kJ mol^{-1} in favor of the FA form when the calculation was made with 6-31G or 6-31G* basis sets,[16]

(17) Scanlan, M. J.; Hillier, I. H. *J. Am. Chem. Soc.* **1984**, *106*, 3737.
(18) Latajka, Z.; Person, W. B.; Morokuma, K. *THEOCHEM* **1986**, *135*, 253.
(19) Ferenczy, G.; Harsányi, L.; Rozsondai, B.; Hargittai, I. *J. Mol. Struct.* **1986**, *140*, 71.
(20) Boggs, J. E.; Niu, Z. *J. Comput. Chem.* **1985**, *6*, 46.
(21) Kurland, R. J.; Wilson, E. B. *J. Chem. Phys.* **1955**, *27*, 585.
(22) Costain, C. C.; Dowling, J. M. *J. Chem. Phys.* **1960**, *32*, 158.
(23) Hirota, E.; Sugisaki, R.; Vielsen, C. J.; Sørensen, O. *J. Mol. Spectrosc.* **1974**, *49*, 251.
(24) Kwiatkowski, J. S.; Lesyng, B.; Palmer, M. H.; Saenger, W. *Z. Naturforsch. C* **1982**, *37*, 937.
(25) Palmer, M. H.; Wheeler, J. R.; Kwiatkowski, J. S.; Lesyng, B. *THEOCHEM* **1983**, *92*, 283.
(26) Sygula, A.; Buda, A. *THEOCHEM* **1985**, *121*, 133.
(27) Cole, S. J.; Bartlett, R. J. *J. Chem. Phys.* **1987**, *86*, 873.

Table II. Amino–Imino Tautomeric Pairs of Formamidine Type Model Systems: Electron Correlation Energies (in au) and Their Relative Values (in kJ mol^{-1})[a]

methods[b]		systems[c] amino: FI	imino: FI	imino: Cyt(1)[c] Cyt*(1)
MBPT(2)	full	−0.459 932 (ref)		−0.462 822 (ref)
		−0.459 932 (0.00)		−0.461 226 (4.20)
	fc	−0.448 255 (ref)		−0.451 207 (ref)
		−0.448 255 (0.00)		−0.449 623 (4.16)
SDQ–MBPT(4)	full	−0.483 374 (ref)		−0.485 849 (ref)
		−0.483 374 (0.00)		−0.484 693 (3.04)
		−0.483 374 (0.00)		−0.484 693 (3.04)
	fc	−0.472 224 (ref)		−0.474 762 (ref)
		−0.472 224 (ref)		−0.473 617 (3.01)
SDTQ–MBPT-(4)	fc	−0.486 509 (ref)		−0.489 635 (ref)
		−0.486 509 (0.00)		−0.488 137 (3.94)
		−0.486 509 (0.00)		−0.488 137 (3.94)
CCSD + T(CCSD)	fc	−0.487 460 (ref)		−0.490 556 (ref)
		−0.487 460 (0.00)		−0.489 098 (3.83)

[a] Compare footnote *a* of Table I. Relative values ($\Delta E^{corr} = E^{corr}_{imino} - E^{corr}_{amino}$) are given in parentheses. SCF energies (in au) calculated with 6-31G* basis set are −149.074 400 (for FI), −149.066 902 (Cyt(1)), and −149.070 273 hartrees (Cyt*(1)). The relative SCF energy ($\Delta E^{SCF} = E^{SCF}_{Cyt(1)} - E^{SCF}_{Cyt*(1)}$ for the Cyt(1)–Cyt*(1) pair is 8.9 kJ mol^{-1} in favor of the Cyt*(1) form. [b] Full, all MO's included; fc, "frozen core" approximation. [c] See Figures 1 and 2 for structures of these model systems.

respectively, at the 3-12G optimized geometry. Such small values can hardly be considered to be significantly different from zero, however. A fourth-order perturbation theory approximation to CISD (CISD$_4$) gave 7.3 kJ mol^{-1} for ΔE^{corr}, also in favor of the FA form[9] (however, this nonsize extensive calculation is probably less reliable than the MBPT(2) (6-31G*//6-31G) result[9]).

Both the results from the calculation by Schlegel et al.[9] and those presented here for the F, FA pair (including the very accurate CCSD + T(CCSD) results) in Table I are similar, in that the electron correlation contribution for the lactim form (FA) of formamide is greater (i.e., more negative) than that for the lactam (F) form. At this point we would like to note that for the F, FA pair the ΔE^{corr} values range from −0.6 to −6.2 kJ mol^{-1}, depending on the method used in the calculation, while the contribution from the relative zero-point vibrational energy $\Delta ZPE_{F,FA}$ is +1.2 or −2.9 kJ mol^{-1} for calculations done with a 3-21G basis set[9] or with a 6-31G** basis (our unpublished results), respectively. Thus contributions from both ΔE^{corr} and ΔZPE are small compared to the relative electronic energy calculated at the SCF level ($\Delta E^{SCF(6\text{-}31G^{**}//3\text{-}21G)}_{F,FA} = -63.4$ kJ mol^{-1}; see Table I).

However, there are two interesting points in Table I that we should like to emphasize. First, improving the accuracy of the calculation of electron correlation increases the magnitude of the total electron correlation energy E^{corr} for each tautomer and also increases the magnitude of ΔE^{corr}_{AB} for all these systems. Secondly, a similar trend is observed in the changes of E^{corr} and ΔE^{corr}_{AB} when the assumed geometries of corresponding tautomers are changed in the series F → Ura(1) → Gua(1) or for the lactims FA → Ura*(1) → Gua*(1). In the first series the most significant change in geometry is the increase of the N–C bond distance from 1.35 Å in F, through 1.396 Å in Ura(1) to 1.427 Å in Gua(1), with some changes also in the bond angles (e.g., ∠NCO 125.3°, 120.7°, 118.8° for F, Ura(1), and Gua(1)). In the second series the most significant changes are also connected with the increasing N–C bond distance from 1.245 Å (FA), 1.280 Å (Ura*(1)) to 1.314 Å (Gua*(1)) and in the changes of some bond angles (e.g., ∠NCH 128.2° (FA), 124.6° (Ura*(1)), and 120.1° (Gua*(1)).

Table II contains results of calculations for the formamidine molecule at two different geometries. When the 3-21G optimized geometry is used for the formamidine molecule, the relative contribution of electron correlation to the amino–imino tautomeric equilibrium is, of course, zero, because both systems are identical in the FI ⇄ FI equilibrium. On the other hand, when the assumed geometries of the formamidine forms are taken from the corresponding cytosine tautomers, the relative correlation contribution ΔE^{corr}_{AB} is of the order of 3–4 kJ mol^{-1} in favor of the form simulating the amino tautomer. The calculations of electron correlation contributions for the amino–imino cytosine tautomeric pair Cyt–Cyt* (see later) shows the same order of magnitude for the $\Delta E^{corr}_{A(amino)B(imino)}$ value.

The relative electronic energy ΔE^{SCF}_{AB} calculated for the Cyt(1)–Cyt*(1) pair is about 9 kJ mol^{-1} in favor of Cyt*(1), so in this case the electron correlation contribution (3–4 kJ mol^{-1}) is likely to be an important contribution to the relative total energies of the pair (in contrast with the example of the Ura(1)–Ura*(1) pair).

2-Oxopyridine Tautomers. The electron correlation contributions to the relative stability of tautomeric forms of 2-oxopyridine have been calculated previously[9] at the MBPT(2) level with the fc approximation using a 6-31G basis set at 3-21G optimized geometries, giving a contribution of 3.4 kJ mol^{-1} for $\Delta E^{corr}_{P(h),P(o)}$ in favor of the lactam tautomer (see Table III). The MBPT(2) calculation with the 6-31G* basis set increases the electron correlation energies for both tautomers, and the value of $\Delta E^{corr}_{P(h),P(o)}$ is increased to 6.5 kJ mol^{-1} in favor of the lactim form. The 6-31G* basis set used here is better than 6-31G[9] for the prediction of the electron correlation energy due to the polarization functions on the heavy atoms. However, polarization functions on the hydrogen atom (as, for example, in a 6-31G** basis set[16]) might be expected to have additional important effects for such hydrogen atom rearrangements.

The stabilities of the tautomeric pairs for 2-oxopyridine have been discussed many times from experimental and theoretical points of view.[5] Since the main aim of the present work is to estimate the electron correlation contribution to the stabilities of the tautomers, we note only that the relative electronic energy calculated at the SCF level depends on the basis set used in the calculation, and it seems that this point remains to be studied more exactly. As to the contribution of the zero-point vibrational energies to the stabilities of the P(h), P(o) tautomeric pair, the ΔZPE value has been estimated experimentally by Beak[28] to be

Table III. Total SCF Energies (E^{SCF} in au), Electron Correlation Energies Calculated by Means of MBPT(2) with Frozen Core ($E^{MBPT(2)}$ in au) for Lactim P(h) and Lactam P(o) Tautomers of 2-Oxopyridine, and Relative Energies (in kJ mol^{-1}) for Tautomeric Pairs

	lactim P(h)	lactam P(o)	lactim–lactam[a]	approx rel total energy[d]
E^{SCF}			ΔE^{SCF}	
3-21G//3-21G	−319.768 150[b]	−319.770 807[b]	7.0 (7.4)[c]	
6-31G//3-21G	−321.430 43[b]	−321.433 71[b]	8.6	
6-31G*//3-21	−321.565 623[d]	−321.566 511[d]	2.3	
6-31G**//3-21G	−321.578 38[c]	−321.576 80[c]	−4.2	
$E^{MBPT(2)}$			$\Delta E^{MBPT(2)}$	$\Delta E^{(eq)}_{AB}$[e]
6-31G//3-21G[b]	−0.651 38	−0.652 69	3.4	
6-31G*//3-21G[d]	−0.961 649	−0.959 185	−6.5	−4.2

[a] Relative energy: $E_{P(h)} - E_{P(o)}$. [b] From Schlegel et al.[9] [c] From Scanlan et al.[30] (these authors used a slightly softer gradient condition for optimization). [d] This work. [e] Defined in eq 2.

Table IV. Total SCF Energies (E^{SCF} in au), Electron Correlation Energies Calculated by Means of MBPT(2) with Frozen Core ($E^{MBPT(2)}$ in au) for Tautomeric Forms of Uracil and Cytosine, and Relative Energies (in kJ mol^{-1}) for Tautomeric Pairs

	basis			
systems	3-21G//3-21G[a]	6-31G*//3-21G[b]	6-31G*//3-21G[b]	
	E^{SCF}	E^{SCF}	$E^{MBPT(2)}$	
Ura	-410.163 10	-412.467 826	-1.162 070	
Ura*	-410.131 95	-412.444 049	-1.162 288	
	ΔE^{SCF}	ΔE^{SCF}	$\Delta E^{MBPT(2)}$	$\Delta E^{(eq)}$[d]
Ura-Ura*[c]	-81.9	-62.5	0.6	-61.9
	E^{SCF}	E^{SCF}	$E^{MBPT(2)}$	
Cyt	-390.416 17	-392.614 149	-1.147 667	
Cyt*	-390.415 52	-392.613 118	-1.146 493	
Cyt(h)	-390.410 11	-392.613 196	-1.148 220	
	ΔE^{SCF}	ΔE^{SCF}	$\Delta E^{MBPT(2)}$	$\Delta E^{(eq)}$[d]
Cyt-Cyt*[c]	-1.7	-2.7	-3.1	-5.8
Cyt-Cyt(h)[c]	-15.9	-2.5	+1.5	-1.0

[a] All data for $E^{SCF(3\text{-}21G//3\text{-}21G)}$ taken from Scanlan and Hillier.[17] Zielinski,[31] using the same basis set, calculated ΔE^{SCF} for Ura-Ura* to be equal to -82.3 kJ mol^{-1} (she used stronger gradient convergence criteria than was used in ref 17). [b] This work. [c] $\Delta E^{SCF} = E^{SCF}_{Ura} - E^{SCF}_{Ura^*}$ (or $E^{SCF}_{Cyt} - E^{SCF}_{Cyt^*}$, etc.). [d] Approximate relative total energy (eq 2).

only 0.4 kJ mol^{-1}. On the other hand, semiempirical estimations of these contributions for hydroxy–oxo tautomeric pairs of oxopyridines and oxopyrimidines are of the order 2.5–3.1 kJ mol^{-1} when the MINDO/3 method[29] has been used or 0.1–1.5 kJ mol^{-1} using the MNDO method[29] (for details see ref 5). Comparing these values with the $\Delta E^{corr}_{P(h),P(o)}$ contribution as calculated here (-6.5 kJ mol^{-1}; see Table III), we see that in this case the ΔE^{corr} contribution is estimated to be even more important than the contributions from the zero-point vibrational energies of the tautomers.

From these values in Table III, the calculated relative total electronic energy for the P(h), P(o) pair is either -4 or -11 kJ mol^{-1} depending upon whether the relative SCF energies were calculated with 6-31G* or 6-31G** basis sets respectively. Since the contribution from vibrational modes is small (see the experimental value quoted above), the predicted relative internal energy is in qualitative agreement with the experimental estimate of $\Delta H^{\circ}_{P(h),P(o)}$ at 132 °C of about -1.3 kJ mol^{-1} with its large uncertainty of 10.5 kJ mol^{-1}.[28] The corresponding estimate of the change in standard free energy $\Delta G^{\circ}_{P(h),P(o)}$ ranges from -2.4 to -3.7 kJ mol^{-1} (for a review of experimental data, see ref 5).

Uracil and Cytosine Tautomers. The tautomerism of pyrimidinic nucleic acid bases has been discussed many times in several papers and has also been summarized in a recent review.[5] Therefore, we restrict our discussion here to only a few points.

Table IV presents the results from SCF calculations with 3-21G and 6-31G* basis sets at 3-21G optimized geometries as well as results at the MBPT(2) level, using the fc approximation, for selected tautomers of uracil and cytosine molecules. It is evident that although the electron correlation contribution for the tautomeric pair of uracil molecules is not important, this conclusion is certainly not true for the cytosine tautomers.

It should be noted that the calculation for the entire cytosine molecule at the SCF level (6-31G*//3-21G) predicts that the imino tautomer Cyt* is less stable than the amino form Cyt by 2.7 kJ mol^{-1}, in contrast with the prediction shown in Table II for the cytosine model compounds (Cyt*(1) more stable than Cyt(1) by 9 kJ mol^{-1}). This difference shows the effect of going from the very simple Cyt(1) models (Figure 1) to the complete molecule. However, the results from the model calculations in Table II certainly indicate that the energies of the rare tautomers of cytosine (Cyt* and Cyt(h)) can be expected to be very close to the energy of the regular amino form (Cyt), in agreement with the more complete calculations in Table IV, and with the experimental results.[34]

Calculations at the SCF/3-21G level indicate that the energy of the 2-hydroxyuracil tautomer[17] is 9.7 kJ mol^{-1} lower in the gas phase than is that of the 4-hydroxy form studied here. However, for uracil, the $\Delta E^{SCF}_{Ura,Ura^*}$ values for either 4- or 2-hydroxy forms are of the order of 60–80 kJ mol^{-1} while the value of $\Delta E^{corr}_{Ura,Ura^*}$ is two orders of magnitude smaller. Hence, $\Delta E^{corr}_{Ura,Ura^*}$ may affect the prediction of relative stabilities of the two hydroxy tautomers, but it has no effect on the prediction of relative stabilities of hydroxy to oxo forms. The ΔZPE contribution calculated by the MINDO/3 method is also small[32] (-0.8 kJ mol^{-1}), and it is of opposite sign to the $\Delta E^{corr}_{Ura,Ura^*}$ value. Thus in the case of Ura–Ura* pair, the relative internal energy for the 4-hydroxy form is the same as the relative energy calculated at the SCF level, $\Delta E^{SCF(6\text{-}31G^*//3\text{-}21G)}_{Ura,Ura^*} + \Delta E^{corr(MBPT(2))}_{Ura,Ura^*} + \Delta ZPE_{Ura,Ura^*} = ((-62.5 + 0.6\text{-}0.8)$ kJ mol^{-1} = -62.7 kJ kJ mol^{-1}). This estimate is in good agreement with the recent experimental estimation of the relative enthalpies for the Ura–Ura* pair in the vapor phase (-79 ± 25 kJ mol^{-1}) by Beak and White.[33] Clearly the correlation effects are unimportant in uracil tautomerization. Assuming that the value of ΔE^{SCF} between the 2- and 4-hydroxyuracil forms calculated with the 3-21G basis[17] is approximately correct, we estimate that isomerization energy from the 2-hydroxy to the oxo form of uracil would be ~-53 kJ/mol.

Until very recently there has been a lack of appropriate experimental data for isolated cytosine tautomers. Very recent matrix isolation studies[34] of vibrational spectra of cytosine and its methylated derivatives in argon and nitrogen matrices indicate that the stable forms of the isolated cytosine molecule are the lactam–amino form Cyt (Figure 1) and lactim–amino form Cyt(h) (Figure 3) which exist in comparable concentrations in the matrix (the tautomeric equilibrium constant K_t = [Cyt/Cyt(h)] is about 0.5, implying that ΔE is nearly zero). The presence of the imino tautomer Cyt* was not detected in these experiments, although the presence of a small amount of this form (on the order of 1–5%) cannot be excluded.

The calculated results for cytosine tautomers listed in Table IV show two important points in connection with the tautomeric equilibrium of cytosine: (a) the use of the larger 6-31G* basis set changes the prediction of relative stabilities of the three tautomers of cytosine calculated in the Hartree–Fock level compared with the case where the 3-21G basis set is used; (b) the electron correlation contributions are comparable in size to the SCF differences and should be included in any proper analysis of the energy difference. In this case our results with correlation included appear to provide the correct order of predicted stabilities for the tautomers. Although the predictions can change significantly with changes in the basis set, at present, the total difference in electronic energies ($\Delta E^{SCF(6\text{-}31G^*//3\text{-}21G)} + \Delta E^{corr(MBPT(2))}$ presented in Table IV for Cyt–Cyt* tautomeric pairs (Cyt* less stable than Cyt by 5,8 kJ mol^{-1}) and for Cyt–Cyt(h) (Cyt(h) less stable than Cyt by 1.0 kJ mol^{-1}) are the best theoretical estimates of relative stabilities for these tautomers. They also appear to be in agreement with the experimental results. It is notable that very recently the 3-21G calculation of the infrared spectra of the cytosine tautomers Cyt and Cyt(h) predicted[35] the $\Delta ZPE_{Cyt,Cyt(h)}$ value to be 2 kJ mol^{-1}. Thus after correction for the zero-point vibrational energies, the relative internal energy of the Cyt(h) tautomer is predicted

(28) Beak, P. *Acc. Chem. Res.* **1977**, *10*, 186.
(29) Sadlej, J. *Semi-Empirical Methods of Quantum Chemistry - CNDO, INDO, NDDO*; Polish Scientific Publishers/Ellis Horwood; Warszawa/Chichester, 1985.
(30) Scanlan, M. J.; Hillier, I. H.; McDowell, A. A. *J. Am. Chem. Soc.* **1983**, *105*, 3568.
(31) Zielinski, T. J. *J. Comput. Chem.* **1983**, *4*, 345.
(32) Shibata, M.; Zielinski, T. J.; Rein, R. *Int. J. Quantum Chem.* **1980**, *18*, 323.
(33) Beak, P.; White, J. M. *J. Am. Chem. Soc.* **1982**, *104*, 7073.
(34) Szczesniak, M. Ph.D. Thesis, Institute of Physics, Polish Academy of Sciences, Warsaw, 1985. Szczesniak, M.; Nowak, M. J.; Szczepaniak, K. *J. Mol. Struct.* **1984**, *115*, 221. For argon matrix isolation study of cytosine tautomers, see also: Radczenko, Y. D.; Plokhotnichenko, A. M.; Sheina, G. G.; Blagoi, Yu. P. *Biofizika* **1983**, *28*, 559. Radczenko, E. D.; Sheina, G. G.; Smorygo, N. A.; Blagoi, Yu. P. *J. Mol. Struct.* **1984**, *116*, 387.
(35) Kwiatkowski, J. S.; Person, W. B.; Szczepaniak, K.; Szczesniak, M. *Acta Biochim. Polon.* **1987**, *34*, 165.

to be *lower* than that of Cyt by 1 kJ mol^{-1}, perhaps fortuitously in excellent agreement with the experiment. However, from the viewpoint of definitive predictions, the level of basis, the extent of correlation included, and the potential geometry changes with correlation to reflect the possible nonplanarity of cytosine may certainly introduce errors of the order of several kilojoules per mole for the calculated relative energies of tautomers of such nucleic acid bases.

Conclusions

Recently a few papers have appeared, devoted to the calculation of electron correlation contributions for large systems. An ab initio localized bond coupled-cluster and MBPT model introduced be Bartlett and co-workers[36] was applied in coupled-cluster calculations for nucleic acid bases.[37] However, the present results are the first to study the details of important electron correlation effects that contribute to the relative stabilities of nucleic acid bases. Comparison of the electron correlation contributions to the relative stabilities of nucleic acid bases and model compounds with the contributions from the zero-point vibrational energies shows that the electron correlation effects are greater or of the same order as the zero-point vibrational effects. In a related case of tautomeric equilibrium ($H_3PO \rightleftarrows H_2POH$), the electron correlation contribution has been estimated to be higher by three orders of magnitude than the contribution from the zero-point vibrational energies.[38]

The current results indicate that in the case of lactim–lactam tautomeric pairs of nucleic acid bases, the electron correlation contribution to the stability is greater for the lactim tautomer, while the energies calculated at the SCF level generally favor the lactam. The 6-31G** result for P(h) is an exception. Although the order of the relative electron correlation energies depends upon the systems considered, on the geometry of the tautomers, as well as on the method and basis set used for the calculation, this result seems to be fairly general. The same tendency of electron correlation to stabilize the lactim tautomer is observed also for other systems (formamide–formamidic acid, and the lactim–lactam tautomers of 2-oxopyridine). This result is also consistent with the tesults for the Cyt–Cyt(h) pair, with our recent studies on isocytosine tautomers,[39] as well as with an ab initio study by Rodwell, Bouma, and Radom[8] for the 1,3-sigmatropic shift: vinyl alcohol → acetaldehyde, Rodwell et al. also calculated relative electron correlation contributions for this pair[40] with a (DZ + polarization) basis set to be 6.1, 1.7, and 2.3 kJ mol^{-1}, respectively, in all cases in favor of the hydroxy tautomer. Also, in the case of the isomerization reaction of the methoxy radical, $CH_3O \rightleftarrows CH_2OH$, SDQ-MBPT(4) calculations[41] predict the electron correlation energy for the hydroxymethylene radical (CH_2OH) to be much greater (by 33.6 kJ mol^{-1}) than that for the methoxy radical (CH_3O).

In the case of amino–imino forms, the electron correlation contributions favor stability of the amino forms, as does the SCF. Hence, as would be predicted, the amino–lactim form Cyt(h) shows the largest effect of correlation $\Delta E^{corr}_{Cyt,Cyt(h)}$.

Acknowledgment. This work was supported by the National Institutes of Health under Research Grant GM-32988, and by the U.S. Army Research Office under Contract No. DAAG29-84-K-0025. J.S.K. wishes also to acknowledge travel support from the ministry of Science and Higher Education (Poland) within the Project CPBP 01.06. J.S.K. wishes also to thank Dr. Samuel J. Cole (QTP, Gainesville) for his help in explaining several computational aspects of the CC/MBPT computer codes.

Registry No. Ura, 66-22-8; Gua, 73-40-5; Cyt, 71-30-7; Ade, 73-24-5; F, 75-12-7; FA, 60100-09-6; Fl, 463-52-5; P(o), 142-08-5.

(36) Laidig, W.; Purvis, G. D.; Barlett, R. J. *Int. J. Quantum Chem., Quantum Chem. Symp.* **1982**, *16*, 561; *J. Phys. Chem.* **1985**, *89*, 2161.
(37) Fôrner, W.; Cizek, J.; Otto, P.; Ladik, J.; Steinborn, E. O. *Chem. Phys.* **1985**, *97*, 235. Forner, W.; Ladik, J.; Otto, P.; Cizek, J. *Ibid.* **1985**, *97*, 251. Pipek, J.; Ladik, J. *Ibid.* **1986**, *102*, 445.
(38) Person, W. B.; Kwiatkowski, J. S.; Bartlett, R. J. *J. Mol. Struct.* **1987**, *157*, 237.
(39) Our recent results for the lactim–lactam pair of isocytosine shows that the electron correlation contribution for the lactim form is 4.1 kJ mol^{-1} greater than for the lactam.
(40) Carsky, P.; Urban, M. *Ab. Initio Calculations* ("Lecture notes in Chemistry", Vol. 16), Springer-Verlag: Berlin, 1980.
(41) Adams, G. F.; Bartlett, R. J.; Purvis, G. D. *Chem. Phys. Lett.* **1982**, *87*, 311.

Computational Evidence for an Unusual Transition State in the Fluxional Behavior of the Cyclopentadienyl Ligand in Chlorotris(cyclopentadienyl)titanium(IV)

Lillian M. Hansen and Dennis S. Marynick*

Contribution from the Department of Chemistry, University of Texas at Arlington, Arlington, Texas 76019. *Received May 18, 1987*

Abstract: The method of partial retention of diatomic differential overlap (PRDDO) was used to generate the potential surface for the η^1/η^5 rearrangement of the cyclopentadienyl ligand in chlorotris(cyclopentadienyl)titanium(IV). The calculations indicate that the ground-state structure consists of two η^5 and one η^1 Cp ligands. Prior NMR data are consistent with a rapid interconversion of these η^1 and η^5 bonding modes. At the ab inito level, the potential surface for this fluxional behavior is a symmetric curve with a maximum at 18.8 kcal/mol. The estimated transition-state structure exhibits "pseudo-η^3" coordination, with a retention of a relatively planar geometry for the Cp ring. There exists a strong σ interaction between one carbon and the titanium center with delocalization to the two adjacent carbons. A significant delocalization of the π orbitals in the Cp rings also exists, donating a total of ~0.20 e to the titanium per ring. This mechanism is unlike any other proposed for the η^1/η^5 rearrangement.

The cyclopentadienyl (Cp) ligand is one of the most common and important groups in organometallic chemistry. It exhibits a wide variety of coordination modes to metals and is known to be involved in fluxional processes which result in the apparent equivalence of all carbons on the NMR time scale. Typical fluxional behavior is observed in tetra(cyclopentadienyl)titanium,

0002-7863/88/1510-2358$01.50/0