

Very Large Scale Computations of the Free Energies of Eight Low-Lying Structures of Arginine in the Gas Phase

Robert J. Gdanitz,[†] Wim Cardoen, Theresa L. Windus,[‡] and Jack Simons*

University of Utah, Henry-Eyring-Center for Theoretical Chemistry, 315 South 1400 East, Room 2020, Salt Lake City, Utah 84112

Received: September 24, 2003; In Final Form: November 19, 2003

To assess the relative energies and free energies of five canonical and three zwitterionic low-lying structures of the arginine molecule, modern basis set extrapolation techniques and high-level ab initio treatments of electron correlation have been used on state-of-the-art parallel computers. The electronic energy and Gibbs free energy orderings of these eight species turn out to be consistent with previous findings [Rak, J.; Skurski, P.; Simons, J.; Gutowski M. *J. Am. Chem. Soc.* **2001**, *123*, 11695] obtained using smaller basis sets and lower level treatments of electron correlation. Nevertheless, the results presented here represent what the current state of the art can achieve for a molecule of this size and complexity and they offer the best available estimates of the relative stabilities of the eight structures.

I. Introduction

Determination of the global-minimum structure for the isolated arginine molecule is a complex problem because this species has two kinds of proton donor groups (OH and NH), six proton acceptor sites (N's and O's), and six (or seven, depending on the tautomer) bonds about which rotation may occur. Thus, its potential energy surface supports numerous tautomers and conformers stabilized by different intramolecular hydrogen bonds. The multitude of low-energy structures and the small energy and free energy differences among them complicate a definite theoretical prediction of the global minimum. Therefore, efficient minimum-energy search algorithms are required to explore the complex structural space of this molecule, and large atomic orbital basis sets and highly correlated ab initio methods should be used to determine relative energies of the various geometrically stable structures. Moreover, (thermal) entropy effects contribute differentially to the free energies of the various forms. As we now summarize, until the present effort, restrictions of computational expense have limited our treatment of some if not all of the above features in a way that leaves doubt in the predictions that have been made about the relative stabilities of the myriad of structures examined.

Recently, Maksic and Kovacevic (MK) performed an extensive computational investigation of arginine at the second-order Møller–Plesset (MP2) and density functional theory levels.¹ They concluded that the most stable structure was canonical, but the energy difference between the lowest zwitterion and canonical structures was relatively small (within 1–3 kcal/mol depending on the theoretical treatment applied). Our group extended their study and new zwitterion and canonical structures were identified with energies even lower than those previously

known.² However, the lowest canonical structure was still found to be lower in energy than the lowest zwitterion by 2.8 kcal/mol, although the internal-geometry space of this complex molecule was explored only in a limited range.

In a second study from our group,³ the relative stabilities of three zwitterionic (**Z1–Z3**) and five canonical (**C1–C5**) isomers of neutral arginine were again studied using ab initio electronic structure methods combined with a genetic algorithm for searching the multitude of structures. These structures have geometries given in ref 3 and shown in Figure 1. In that work, trial structures were first identified at the PM3 level of theory using a genetic algorithm to systematically vary geometrical parameters. Further geometry optimizations of these structures were performed at the MP2 and B3LYP levels of theory with basis sets of the 6-31++G** quality. The final energies were determined at the coupled-cluster singles and doubles (CCSD) level with the 6-31++G** basis set and were corrected for thermal effects determined at the B3LYP level. Two new nonzwitterionic structures of the neutral arginine were identified in that study, and one of them (**C5**) was found to be the lowest energy structure found to date. Nevertheless, the five lowest energy structures of neutral arginine were found to be nonzwitterionic in nature and clustered within a narrow energy range of 2.3 kcal/mol. The lowest energy zwitterion structure was suggested to be less stable than the lowest nonzwitterion structure by 4.0 kcal/mol.

There are no experimental data to suggest that the energy orderings for the canonical and zwitterionic structures offered in refs 2 and 3 are incorrect. However, because the range of energies is small and the level of theory (i.e., basis sets and treatment of correlation) was limited, these issues remain inadequately resolved. In the present effort, we therefore further refined the accuracy of our determination of the relative energies of these low-energy structures by employing enhanced computer power, larger basis sets (as well as extrapolation), and higher level treatment of electron correlation. In effect, we present in this work the best that can presently be done to evaluate the energy and free energy orderings of the eight arginine species

* Corresponding author. E-mail: simons@chemistry.utah.edu.

[†] On leave from Technical University Braunschweig, Institute for Physical and Theoretical Chemistry, Hans-Sommer-Str. 10, D-38106 Braunschweig, Germany. Present address: North Carolina A&T State University, Department of Physics, Rm. 101, Marteen Hall, Greensboro, NC 27411.

[‡] Pacific Northwest National Laboratory, Environmental Molecular Sciences Laboratory, P.O. Box 999, Richland, WA 99352.

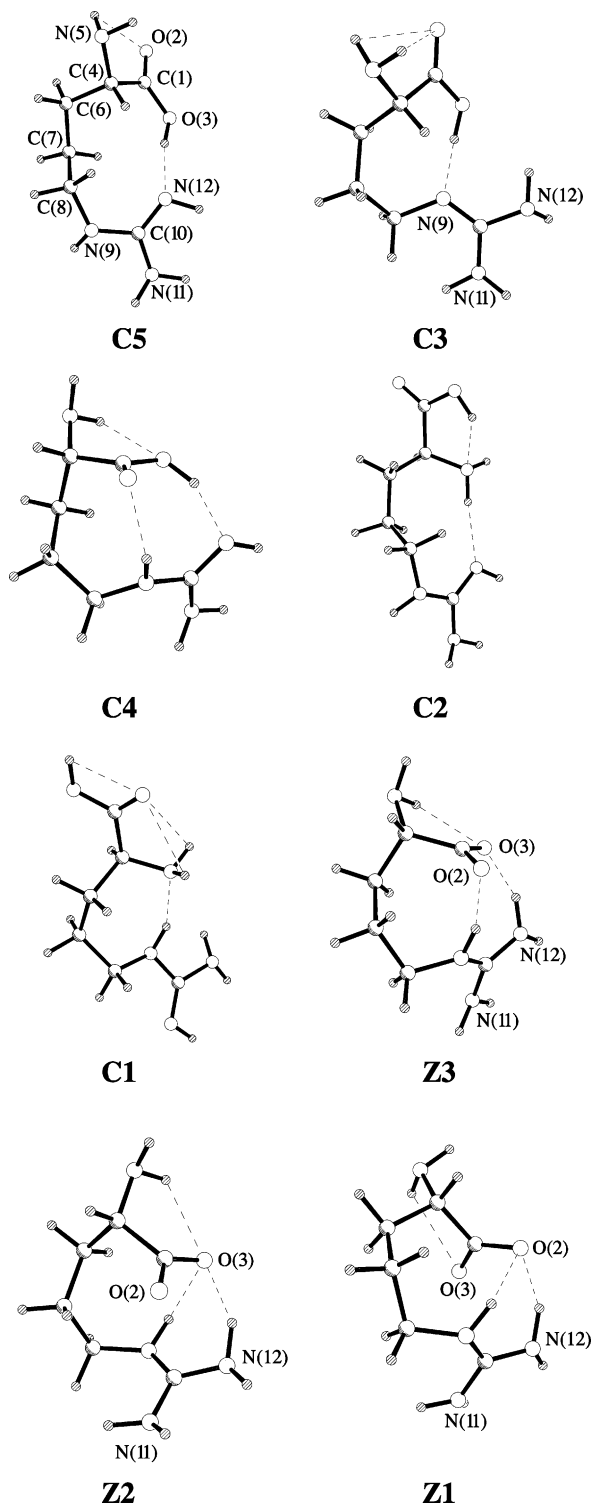


Figure 1. Five canonical (C1–C5) and three zwitterionic (Z1–Z3) locally stable geometries for arginine identified in ref 3 and used here.

studied in ref 3 and using the optimized geometries identified there.

II. Methods

Throughout this work, we employed Dunning's⁴ correlation consistent basis sets of double, triple, and quadruple- ζ quality, cc-pVxZ, (where x can be D, T, or Q). The basis sets for the heavy atoms (C, N, and O) were augmented with diffuse functions,⁵ and we denote the final such basis sets DZ, TZ, and QZ, respectively. Within these bases, there were 346, 748, and

1380 functions, respectively. The Hartree–Fock (HF) and correlation energies of each of the eight arginine isomers were computed in basis sets of consecutive sizes and separately extrapolated to the basis set limit. In evaluating the correlation energy at the highest level employed here, we used the CCSD method with approximate linked triple excitations treated at both the CCSD(T)⁶ and CCSD(T)⁷ levels. For all of the correlated calculations, the core orbitals of the C, N, and O atoms were frozen. For the extrapolation at the HF level of theory, we use a three-point formula, given below in eq 2, employing the DZ, TZ, and QZ energies, and we denote the corresponding results (DTQ)Z. For the correlation contribution, we write (DT)Z when the DZ and TZ basis results are used in the extrapolation and (TQ)Z when the TZ and QZ basis results are used.

Our final estimates for the total electronic energies are then given by a sum of various incremental contributions:

$$E_{\text{est}} = E_{\text{HF}}[(\text{DTQ})\text{Z}] + \Delta E_{\text{MP2}}[(\text{TQ})\text{Z}] + \Delta E_{\text{CCSD}}[\text{DZ}] + \Delta E_{\text{T}}[\text{DZ}] \quad (1)$$

where the basis set used is given inside the square brackets. In eq 1, E_{HF} is the Hartree–Fock energy [from the (DTQ)Z extrapolation], $\Delta E_{\text{MP2}} = E_{\text{MP2}} - E_{\text{HF}}$ is the correlation energy computed at the MP2 level of theory [using the (TQ)Z extrapolation], $\Delta E_{\text{CCSD}} = E_{\text{CCSD}} - E_{\text{MP2}}$ is the CCSD correction to the MP2 energy (computed using the DZ basis), and $\Delta E_{\text{T}} = E_{\text{CCSD(T)}} - E_{\text{CCSD}}$ is the approximate triple-excitation correction (again, using the DZ basis).

The HF limit energies were estimated using the empirical formula

$$E_{\text{HF}}[\infty] = E_{\text{HF}}[x] - B \exp(-\alpha x) \quad (2)$$

which, in a recent study of Halkier et al.,⁸ was shown to reduce the maximum absolute error of the energies of eight first-row diatomics by a factor of 3. In this extrapolation formula and those shown below, x is 2, 3, and 4, respectively, for the DZ, TZ, and QZ bases.

We estimated the ΔE_{MP2} basis set limit using the extrapolation formulas very recently proposed by Schaefer and co-workers,⁹ assuming the relations

$$\Delta E^{(1)}[\infty] = \Delta E^{(1)}[x\text{Z}] - a(x + 1/2)^{-3} \quad (3)$$

for the singlet and

$$\Delta E^{(3)}[\infty] = \Delta E^{(3)}[x\text{Z}] - b(x + 1/2)^{-5} \quad (4)$$

for the triplet contributions to ΔE_{MP2} . To employ, for example, eq 3 within the (TQ)Z approximation, we compute $\Delta E^{(1)}[\text{TZ}]$ and $\Delta E^{(1)}[\text{QZ}]$ and write eq 3 for $x = 3$ and for $x = 4$; these two equations we then solve for the parameter a and for the extrapolant $\Delta E^{(1)}[\infty]$. An analogous procedure is used to obtain $\Delta E^{(3)}[\infty]$ from eq 4.

The predictions of eqs 3 and 4 give rise to the following formula for the total extrapolation of the correlation energy on MP2 level of theory:

$$\Delta E[\infty] = \Delta E^{(1)}[\infty] + \Delta E^{(3)}[\infty] = \frac{\tilde{x}^3 E^{(1)}[x] - \tilde{y}^3 E^{(1)}[y]}{\tilde{x}^3 - \tilde{y}^3} + \frac{\tilde{x}^5 E^{(3)}[x] - \tilde{y}^5 E^{(3)}[y]}{\tilde{x}^5 - \tilde{y}^5} \quad (5)$$

where we use the notation $\tilde{x} = x + 1/2$. Equation 5 is indeed the

TABLE 1: Accumulated Relative Electronic Energies (kcal mol⁻¹) of Canonical (Nonionic) and Zwitterionic Forms of Neutral Arginine

	structure							
	C1	C2	C3	C4	C5	Z1	Z2	Z3
HF/DZ	0.442	1.733	1.992	4.017	0.000	8.533	7.827	7.845
HF/TZ	0.892	1.788	1.780	4.238	0.000	8.596	8.134	7.661
HF/QZ	0.896	1.785	1.829	4.319	0.000	8.757	8.318	7.801
HF/(DTQ)Z	0.838	1.775	1.900	4.359	0.000	8.888	8.436	7.947
HF/DZ + MP2/DZ	3.525	1.867	0.511	1.088	0.000	4.169	2.909	2.095
HF/TZ + MP2/TZ	4.036	1.686	0.218	1.260	0.000	4.431	3.419	1.829
HF/QZ + MP2/QZ	4.214	1.682	0.396	1.380	0.000	4.606	3.594	1.820
HF/TZ + MP2/(DT)Z	4.034	1.589	0.179	1.248	0.000	4.546	3.545	1.823
HF/QZ + MP2/(TQ)Z	4.337	1.681	0.493	1.422	0.000	4.633	3.610	1.718
HF/(DTQ)Z + MP2/(TQ)Z	4.278	1.671	0.564	1.462	0.000	4.764	3.728	1.864
HF/(DTQ)Z + MP2/(TQ)Z [(x + 1/2) ⁻³] ^a	4.312	1.671	0.582	1.455	0.000	4.749	3.706	1.833
HF/(DTQ)Z + MP2/(TQ)Z + CCSD/DZ	3.318	1.511	1.101	2.298	0.000	6.278	5.314	4.211
\sim^b + CCSD(T)/DZ	3.990	1.704	1.014	1.653	0.000	4.861	3.785	2.680
\sim^b + CCSD(T)/DZ	3.893	1.648	0.983	1.715	0.000	5.250	4.180	3.022

^a Both singlet and triplet contributions both extrapolated using an $(x + 1/2)^{-3}$ scaling law. ^b Results of the previous method with the following corrections added.

TABLE 2: Relative Electronic Energies (E^a) in kcal mol⁻¹ of Canonical and Zwitterionic Forms of Neutral Arginine Calculated at the B3LYP, MP2, and CCSD Levels, Relative CCSD Electronic Energies Corrected (Using the Data from Ref 3) for Zero-Point Energy ($\Delta E_{0,\text{vib}}$), Enthalpy ($\Delta H_{298,\text{corr}}$), and Free Energy ($\Delta G_{298,\text{corr}}$) Contributions

structure	E^{B3LYP}	E^{MP2}	E^{CCSD}	$E^{\text{CCSD}} + \Delta E_{0,\text{vib}}$	$E^{\text{CCSD}} + \Delta H_{298,\text{corr}}$	$E^{\text{CCSD}} + \Delta G_{298,\text{corr}}$	E^b	$E^b + \Delta G_{298,\text{corr}}$
C5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
C3	0.684	0.240	0.825	0.241	0.436	-0.528	0.983	-0.370
C4	1.959	0.823	1.618	1.318	1.128	1.962	1.715	2.059
C2	1.830	1.820	1.695	1.999	1.984	1.480	1.648	1.433
C1	2.528	3.097	2.283	1.994	2.269	0.446	3.893	2.056
Z3	1.824	1.676	3.966	3.244	3.201	3.835	3.022	2.891
Z2	4.270	2.579	4.113	3.845	3.810	4.003	4.180	4.070
Z1	4.076	3.725	5.230	4.653	4.768	4.397	5.250	4.417

^a The electronic energies of **C5** calculated at the B3LYP, MP2, and CCSD levels are -606.5998049, -604.8474817, and -604.9239685 Hartrees, respectively. All calculations were performed with 6-31++G** basis sets. ^b The frozen-core CCSD(T) energies at the estimated basis-set limit relative to that of **C5**, whose total electronic energy is -605.943 210 Hartrees.

most accurate extrapolation method available for the correlation energy of general molecules (see ref 9 for a review on the literature). In the special case of He clusters (where only the 1s orbital is occupied in the Hartree-Fock wave function), it is possible to estimate $\Delta E^{(1)}$ accurate to $O(x^{-4})$, as was shown in ref 10.

Most of the very large scale computations required in this project were carried out at the Environmental Molecular Sciences Laboratory (EMSL) in Richland, WA, on the 512-processor NWMpp1 IBM-SP system. The MP2/QZ and CCSD(T)/DZ computations both required 128 processors for ca. 2 days each and used 48 GBytes of main memory. Some of the MP2/QZ calculations were performed on 5 dual-processor nodes of the Linux-PC Beowulf cluster "Icebox" of the Center of High Performance Computing of the University of Utah where about 4 days and 10 GBytes of main memory were used. Finally, we used the new HP/Linux-Itanium system "Opus" at EMSL, which replaced NWMpp1 and which consists of 128 dual-processor Itanium 2 nodes. Each node has 12 GB of memory and 400 GB of local disk available to it. Fast communication between the nodes of 360 MB/s within NWChem is obtained using a dual rail QNet1/Elan3 interconnect from Quadrics. The MP2/QZ computations on this latter system used ca. 16 h on 100 processors and 146 GBytes of main memory. The CCSD(T)/DZ calculations used ca. 12 h. on 120 processors with 176 GBytes of main memory. All computations were performed with NWChem^{11,12} versions 4.0, 4.1, and 4.5, which have been modified¹³ to allow for the separate computation of the contributions of (spin-adapted) singlet and triplet pair functions within MP2.

III. Results

Table 1 shows the various contributions to the total energy defined in eq 1 relative to the most stable structure, **C5** (see Figure 1). Using the DZ basis set to compute the HF energy can be seen to introduce an error of up to 0.6 kcal/mol, but already in the QZ basis set, the HF energies are converged to at least 0.1 kcal/mol with respect to the estimated basis set limit.

As is well-known,¹⁴ the (DT)Z extrapolant of the correlation energy, eqs 3 and 4, is unreliable, so we discarded it in favor of the HF/(DTQ)Z + MP2/(TQ)Z results and obtained relative energies that deviate by at most 0.2 kcal/mol from the HF/QZ + MP2/QZ energies, but by up to 6.1 kcal/mol from the HF/(DTQ)Z energy. We note that in the case of arginine, extrapolating both the singlet and the triplet contributions simultaneously with the formula of eq 3 leads to a deviation of the relative energies of at most 0.03 kcal/mol, which is entirely negligible within the accuracy of the present work.

Including the CCSD correction computed in the DZ basis set changes the relative energies up to 1.6 kcal/mol. Finally, the CCSD(T) triples correction changes the relative energies up to 1.2 kcal/mol, but results of the CCSD[T] and CCSD(T) methods deviate from each other by 0.4 kcal/mol at most.

Thus, to obtain the relative energies of these eight arginine structures within an accuracy of 0.5–1.0 kcal/mol, it seems to be sufficient to compute the HF and MP2 energies in the TZ basis set and to apply a CCSD(T) correction obtained with the DZ basis. Extrapolation to the basis set limit turns out to give no significant change in the relative energies. This is especially true for the extrapolation to the HF limit and for using separate formulas for singlet and triplet contributions to the correlation

energy. However, at least CCSD(T) computations in the DZ basis set are required for the accuracy we aim to achieve. Unfortunately, due to hardware and software limitations, we had not earlier been able to compute the CCSD energy in the TZ basis set to check the basis set convergence. We note in passing that omitting the shift of $1/2$ in the formulas of eqs 3–5, as is proposed by Halkier et al.¹⁴ who used eq 3 for the whole correlation energy and later by Klopper¹⁵ separately for the singlet and triplet contributions, changed the relative final energies by 0.02 kcal/mol at most, which is entirely negligible.

In Table 2, we summarize in columns 2–7 the energy and free energy findings of ref 3 for the five canonical (**C1**–**C5**) and three zwitterion (**Z1**–**Z3**) structures identified in Figure 1 and we compare to what we find here. In particular, in columns 8 and 9 of Table 2, we give the relative total electronic energies E of these structures as well as the correction for the Gibbs free energy at room temperature, $E + \Delta G_{298, \text{corr}}$ obtained in the present work. These values are to be compared to the data given in the fourth (E^{CCSD}) and the seventh ($E^{\text{CCSD}} + \Delta G_{298, \text{corr}}$) columns, which were obtained in our earlier study.³ Recall that, in the present work, we improved the accuracy of our electronic energy calculations in two ways. We carried out systematic state-of-the-art basis set extrapolations that were not possible earlier, using as many as 1380 atomic basis functions. Second, we extended the level of treatment of electron correlation to triple excitations but in ref 3 we had to limit our treatment to the CCSD level.

As can be seen from Table 2, it turns out that the relative orderings with respect to the Gibbs free energy G at 298 K obtained in ref 3 are indeed essentially consistent with the findings of the present work where we use a much more sophisticated treatment of the electron correlation. However, there are two exceptions. The G value of **C1** is increased by 1.7 kcal/mol, rendering this structure degenerate with **C4**. Additionally, the G value of **Z3**, which was degenerate (within 0.2 kcal/mol) with **Z2** in our previous treatment, is now 1.2 kcal/mol lower than **Z2**. Finally, we observe that the difference in G between the lowest canonical (i.e., **C3**) and the lower zwitterionic (i.e., **Z3**) structure is reduced by 1 kcal/mol from 4.4 kcal/mol in our previous study to 3.3 kcal/mol in the present work.

In summary, the current estimates of the relative energies and free energies of eight structures of arginine offer the best that can be done at present. To obtain these data, we used extended atomic orbital basis sets and complete-basis extrapolations as well as coupled-cluster single and double excitation with approximate inclusion of triple excitations methods to

handle electron correlation. Moreover, we made use of large-scale parallel computer resources to achieve these results in a reasonable time frame. As a result, we believe that the results presented here describe the relative stabilities of the eight species studied to within 0.5 kcal/mol.

Acknowledgment. We thank Prof. P. Skurski for providing us with the geometries of the structures found in ref 3 and Dr. Anita Orendt, Center for High Performance Computing (CHPC) of the University of Utah, for her expert assistance. This research was performed in part using the Molecular Science Computing Facility (MSCF) in the William R. Wiley Environmental Laboratory at the Pacific Northwest National Laboratory (PNNL). The MSCF is funded by the Office of Biological and Environmental Research in the U.S. Department of Energy. PNNL is operated by Battelle for the U.S. Department of Energy under contract DE-AC06-76RLO 1830. This work has in part been supported by PNNL Laboratory Directed Research and Development, by the Office of Basic Energy Sciences in the U.S. Department of Energy, and by Grant 35708-AC6 from the Petroleum Research Fund of the American Chemical Society (to J.S.). R.J.G. acknowledges financial support from the BioGeometry project (NSF-CCR-00-86013).

References and Notes

- (1) Maksic, Z. B.; Kovacevic, B. *J. Chem. Soc., Perkin Trans.* **1999**, 2, 2623–2629.
- (2) Skurski, P.; Gutowski, M.; Barrios, R.; Simons, J. *Chem. Phys. Lett.* **2001**, 337, 143–150.
- (3) Rak, J.; Skurski, P.; Simons, J.; Gutowski M. *J. Am. Chem. Soc.* **2001**, 123, 11695–11707.
- (4) Dunning, T. H. *J. Chem. Phys.* **1989**, 90, 1007–1023.
- (5) Kendall, R. A.; Dunning, T.; Harrison, R. J. *J. Chem. Phys.* **1992**, 96, 6796–6806.
- (6) Urban, M.; Noga, J.; Cole, S. J.; Bartlett, R. J. *J. Chem. Phys.* **1985**, 83, 4041–4046.
- (7) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. *Chem. Phys. Lett.* **1989**, 157, 479–483.
- (8) Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Olsen, J. *Chem. Phys. Lett.* **1999**, 302, 437–446.
- (9) Valeev, E. F.; Allen, W. D.; Hernandez, R.; Sherrill, C. D.; Schaefer, H. F. *J. Chem. Phys.* **2003**, 118, 8594–8610.
- (10) Gdanitz, R. J. *J. Chem. Phys.* **2000**, 113, 5145–5153.
- (11) Kendall, R. A.; Apra, E.; Bernholdt, D. E.; Bylaska, E. J.; Dupuis, M.; Fann, G. I.; Harrison, R. J.; Jialin, J.; Nichols, J. A.; Nieplocha, J.; Straatsma, T. P.; Windus, T. L.; Wong, A. T. *Computer Phys. Commun.* **2000**, 128, 260–283.
- (12) High Performance Computational Chemistry Group, NWChem, a computational chemistry package for parallel computers, Pacific Northwest National Laboratory, Richland, WA 99352.
- (13) Harrison, R. J.; Gdanitz, R. J. Unpublished results.
- (14) Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. *Chem. Phys. Lett.* **1998**, 286, 243–252.
- (15) Klopper, W. *Mol. Phys.* **2001**, 99, 481–507.