2010

Open-Shell Formulation of the Fragment Molecular Orbital Method

Spencer Pruitt
Iowa State University

Dmitri G. Fedorov
National Institute of Advanced Industrial Science and Technology

Kazuo Kitaura
National Institute of Advanced Industrial Science and Technology

Mark S. Gordon
Iowa State University, mgordon@iastate.edu

Open-Shell Formulation of the Fragment Molecular Orbital Method

Abstract

Performing accurate calculations on large molecular systems is desirable for closed- and open-shell systems. In this work, the fragment molecular orbital method is extended to open-shell systems and implemented in the GAMESS (General Atomic and Molecular Electronic Structure System) program package. The accuracy of the method is tested, and the ability to reproduce reaction enthalpies is demonstrated. These tests also demonstrate its utility in providing an efficient means to model large open-shell systems.

Disciplines

Chemistry

Comments

JCTC Journal of Chemical Theory and Computation

Open-Shell Formulation of the Fragment Molecular Orbital Method

Spencer R. Pruitt,[†] Dmitri G. Fedorov,[‡] Kazuo Kitaura,[‡,§] and Mark S. Gordon*,[†]

Department of Chemistry, Iowa State University, Ames, Iowa 50011, RICS, National Institute of Advanced Industrial Science and Technology (AIST), 1-1-1 Umezono, Tsukuba, Ibaraki 305-8568, Japan, and Graduate School of Pharmaceutical Sciences, Kyoto University, Sakyo-ku, Kyoto 606-8501, Japan

Abstract: Performing accurate calculations on large molecular systems is desirable for closed- and open-shell systems. In this work, the fragment molecular orbital method is extended to open-shell systems and implemented in the GAMESS (General Atomic and Molecular Electronic Structure System) program package. The accuracy of the method is tested, and the ability to reproduce reaction enthalpies is demonstrated. These tests also demonstrate its utility in providing an efficient means to model large open-shell systems.

1. Introduction

Recently, a large number of fragment-based methods[1−6] have been developed, including the fragment molecular orbital (FMO) method.[7−10] The aim of these methods is to treat complex molecular species efficiently, while retaining ab initio accuracy, by dividing the system into many much smaller fragments. However, few of these methods are capable of treating open shells.[11] Several wave function types have already been interfaced with FMO;[12−17] however, most of them are for closed-shell methods. The only exception is the multiconfiguration self-consistent field (MCSCF method),[14] which can treat both closed- and open-shell species. Although the MCSCF method is very useful in some systems, it is also desirable to have a single-reference open-shell method, which could be efficiently combined with restricted open-shell (RO) second-order Møller−Plesset perturbation theory (MP2)[18] or coupled cluster theory (CC).[19]

[†] Iowa State University.
[‡] AIST.
[§] Kyoto University.
* To whom correspondence should be addressed. E-mail: mark@si.msg.chem.iastate.edu.

Open-shell systems play a very important role in many processes, for instance, in radical chemistry,[21] electron transfer,[22] and many transition metal compounds.[23] In addition, transition states in chemical reactions involve breaking chemical bonds and thus possess considerable open-shell character. Although more difficult cases such as transition metal complexes may require a more sophisticated treatment (e.g., MCSCF), in many organic, inorganic, and biochemical systems, the open-shell character can be well described by an open-shell single-reference wave function. Therefore, it is useful to be able to perform large-scale calculations of open-shell systems with single-reference methods. The FMO method is particularly appealing in this regard, since it has been shown to accurately reproduce fully ab initio calculations with high accuracy.[9,10] Due to the inherently parallel nature of the method, it also scales almost linearly with system size.

2. Methodology

The *n*-body FMO energy (FMO*n*) of *N* fragments is given by[8]

$$E^{\mathrm{FMO2}} = \sum_I^N E_I + \sum_{I>J}^N (E_{IJ} - E_I - E_J) \quad (1)$$

$$E^{\mathrm{FMO3}} = E^{\mathrm{FMO2}} + \sum_{I>J>K}^N (E_{IJK} - E_I - E_J - E_K - \Delta E_{IJ} - \Delta E_{JK} - \Delta E_{IK}) \quad (2)$$

where

$$\Delta E_{IJ} = E_{IJ} - E_I - E_J \quad (3)$$

and E_I, E_{IJ}, and E_{IJK} are the monomer (single fragment), dimer (fragment pair), and trimer (fragment triple) energies, respectively, computed in the electrostatic field of other fragments.[8]

This expression is also used for the restricted open-shell FMO method, which is an extension of the closed-shell method, described in detail elsewhere.[10] The computational scheme is methodologically similar to that of the FMO-based MCSCF or time-dependent[23] density functional theory (TDDFT), with one fragment chosen to be the open-shell fragment. There are two types of dimers and trimers in an open-shell system: (a) open-shell, if they include the open shell fragment; (b) closed-shell otherwise. The electrostatic field added to all monomer, dimer, and trimer calculations is computed from the total density of either open- or closed-shell fragments. The covalent boundaries are treated exactly as in the closed-shell FMO,[7,9,10] that is, by assigning two electrons from the detached bond to one fragment and none to the other, for the pair of fragments between which a bond is detached.
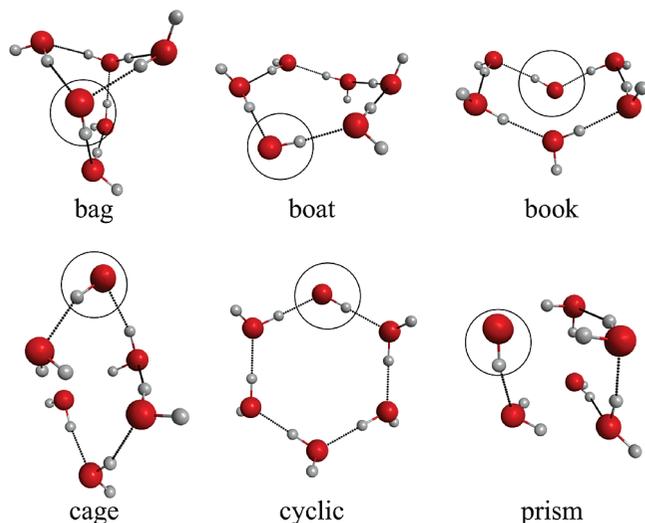
Figure 1. The six isomers of OH(H$_2$O)$_5$ clusters used for testing. Open-shell OH molecules are circled, and the naming convention is from ref 26.

Table 1. Absolute Errors between ab Initio ROMP2 and FMO-ROHF for Six OH(H$_2$O)$_5$ Clusters

	absolute errors (kcal/mol)			
	aug-cc-pVTZ			
isomer	FMO2-ROHF	FMO3-ROHF	FMO2-ROMP2	FMO3-ROMP2
prism	0.2	0.0	0.0	−0.1
cage	1.4	0.1	1.3	0.1
bag	2.8	0.0	2.7	0.1
cyclic	1.7	0.1	1.6	0.3
boat	1.7	0.1	1.6	0.3
book	1.4	0.1	1.3	0.2

The FMO energy has been implemented for the RO-based Hartree−Fock method (ROHF), MP2 (ROMP2), and CC (ROCC). In FMO-ROHF, the densities of all fragments are optimized self-consistently on the basis of the ROHF density of the open-shell fragment, whereas in FMO-TDDFT all densities are optimized for the ground state, from which excitations are evaluated following the TDDFT theory.

In addition, a multilayer scheme[25] was also implemented in which several layers with varying levels of electronic structure theory (HF, MP2, and CC) can be defined. The multilayer FMO method uses the notation of listing the wave functions in increasing order of layers, for example, FMO2-ROMP2:ROCC means that the fragment densities are obtained self-consistently at the uncorrelated level (ROHF) and used in the correlated

Table 2. Comparison of the ROMP2 Relative Energies[a] of Six OH(H$_2$O)$_5$ Clusters

	relative energies (kcal/mol)		
	aug-cc-pVTZ		
isomer	ROMP2	FMO2-ROMP2	FMO3-ROMP2
prism	0.0	0.0	0.0
cage	1.4	2.7	1.6
bag	2.5	5.2	2.7
cyclic	3.9	5.5	4.3
boat	4.7	6.3	5.0
book	8.8	10.2	9.1

[a] Zero energy for all methods is set to be the prism isomer.

calculations at the ROMP2[26] (Z-averaged perturbation theory) level for the less important fragments (substituents) and ROCC[19,20] (completely renormalized CR-CC(2,3) method) for (for example) a reaction center. Dimer calculations are performed at the lower level of the two layers to which the two fragments belong. In the case of the reaction described below, there was only one fragment in the higher level, so that all dimer calculations were done with ROMP2. The same basis set is used in both layers.

Especially for CC, which is a very steeply scaling method (N^7) with the system size N, the use of FMO is beneficial even for very small systems, such as that with only three fragments discussed below. In addition to the computational cost scaling, CC requires very large memory; the huge memory demand prevented the full ab initio CC calculations while the FMO-CC computations are feasible.

The open-shell FMO code was parallelized with the generalized distributed data interface (GDDI),[27,28] using a two-level hierarchical scheme. Since the open-shell MP2 method in GAMESS (General Atomic and Molecular Electronic Structure System)[29] is fully parallelized, it can take advantage of both levels of GDDI. This is not the case for the open-shell CC code, which can only take advantage of the inherent coarse-grained level parallelism of the FMO method using GDDI. The load balancing is dynamic (optionally static) and follows a general algorithm[28] with all fragments (closed- and open-shell) distributed over groups in GDDI. All methods discussed here have been implemented in GAMESS.

All calculations discussed here used the default values of thresholds. The exception is that, for water clusters, the Mulliken point charge representation[8] of the electrostatic potentials in the FMO method was used (i.e., fragment calculations are done in the field of point charges derived self-consistently from the fragment densities, and fragment calculations are repeated until their densities converge with respect to the field;[7,8,10] dimers are computed in the converged field only once). Spherical basis functions were used throughout, and the core electrons (e.g., 1s on C and O) were not correlated in MP2 and CC.

3. Tests

3.1. Open-Shell FMO2 and FMO3 Calculations on OH(H$_2$O)$_5$ Clusters. A preliminary test of the open-shell FMO method employed clusters of six water molecules. The ability to accurately model the solvated OH radical has implications in biological applications and atmospheric processes.[30] The large charge transfer present in solvated OH clusters adds an additional degree of difficulty, providing an excellent test case for the open-shell FMO method.

The structures of the six clusters shown in Figure 1 were determined by optimizing previously determined minima[31] with MP2 and the aug-cc-pVTZ basis set.[32] To create the open-shell test systems, one hydrogen atom was arbitrarily removed from one of the water molecules in each cluster (doublet spin multiplicity). Fragments were chosen by placing the open-shell OH in a fragment with both nearest neighbor H$_2$O molecules, while placing the other H$_2$O molecules in fragments by themselves, with the exception of the prism and bag isomers. For these two isomers, there were two other H$_2$O molecules with a significant interaction between them. This required them
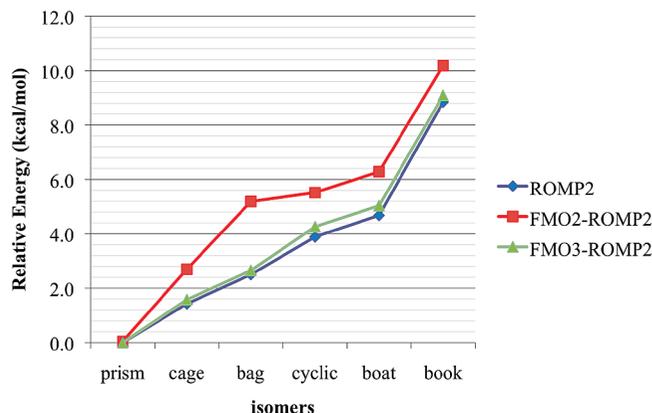
Letter

J. Chem. Theory Comput., Vol. 6, No. 1, 2010 **3**

Figure 2. Graph of the relative energies of the six OH(H$_2$O)$_5$ clusters computed using ab initio ROMP2, FMO2-ROMP2, and FMO3-ROMP2.

Table 3. Absolute Errors between ab Initio ROMP2 Calculations and FMO2-ROMP2 for the RAFT Reaction

	absolute errors (a.u.)		
	6-311G(d,p)		
	ROMP2	FMO2-ROMP2	error (kcal/mol)
reactant 1	−518.94377299	−518.94481925	−0.7
reactant 2	−1235.33260039	−1235.33407266	−0.9
product	−1754.31552043	−1754.31616825	−0.4

to be placed into one fragment together, while the open-shell OH was placed in a fragment with only one nearest neighbor. The final fragmentation scheme created four FMO monomers for all isomers.

Errors for FMO2-ROHF calculations (Table 1) relative to ab initio calculations are between 0.2 and 2.8 kcal/mol, while the errors are between 0.0 and 2.7 kcal/mol for FMO2-ROMP2. The addition of ab initio three-body interactions with FMO3-ROHF significantly reduces the error to 0.0−0.1 kcal/mol, while the error for FMO3-ROMP2 falls to 0.1−0.3 kcal/mol. The improvement in accuracy for FMO3 is not a surprising result, as the importance of three-body effects in water has been shown previously.[12,33]

Relative energetics (Table 2) are of similar accuracy, with the ordering of isomers being captured correctly with both FMO2 and FMO3, illustrated in Figure 2. The choice of fragments is important for the accurate reproduction of the relative energies, especially for FMO2 and in systems in which

fragments may have very strong interactions such as charge transfer. These strongly interacting fragments should be grouped together to improve accuracy.

For the small water clusters computed, we note that FMO2-ROMP2 takes about the same time as ab initio ROMP2, whereas FMO3-ROMP2 takes about three times more, and the crossover size (when FMO owing to its nearly linear scaling becomes faster) may be just a few more molecules. Considering the distributed memory requirements of the ab initio ROMP2 calculations (∼2 GB of RAM) versus that of the FMO2-ROMP2 calculation (∼512 MB of RAM) or the FMO3-ROMP2 calculation (∼1 GB of RAM), the open-shell FMO method is capable of providing accurate energies at a much lower cost. One can imagine that, as the size of the cluster N increases, the memory requirement of the *ab initio* calculation will also increase substantially (∼N^4); however, for FMO it will remain the same, no matter how large the cluster is.

3.2. Multilayer FMO2 Calculation of the Reversible Addition−Fragmentation Chain Transfer (RAFT) Reaction Enthalpy (Figure 3). As a further test of the open-shell FMO method, the initiation step in the RAFT reaction[34,35] was chosen as a small test case. The initiation step consists of two reactants, one an open-shell radical (doublet spin multiplicity), that combine to form an open-shell radical product. Initial structures were optimized using DFT with the B3LYP functional and the
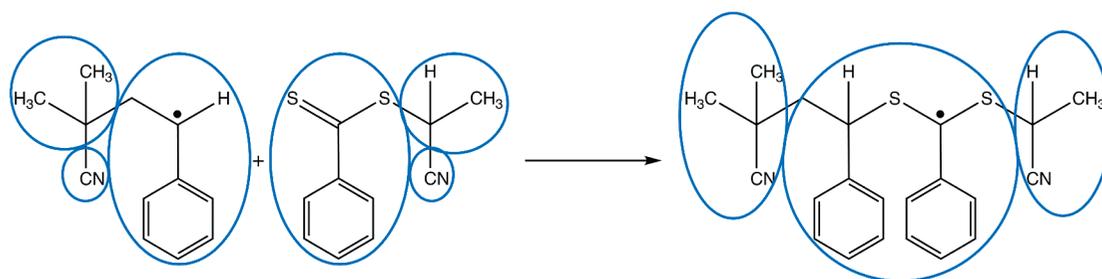

Figure 3. Reaction scheme for the RAFT reaction with the choice of FMO fragments shown in blue.
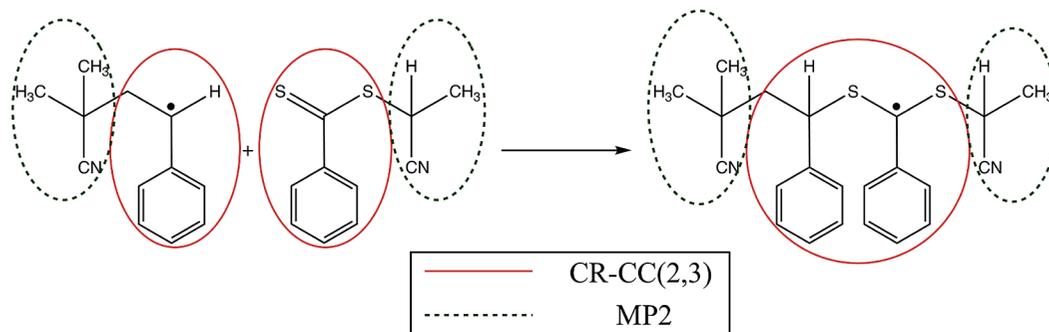

Figure 4. Reaction scheme for the RAFT reaction with the multilayer FMO details: higher layer fragments (CC) are circled in red, and lower levels fragments (MP2) are in green dashed circles.

Table 4. Calculated Enthalpy of the RAFT Reaction Using ab Initio ROMP2, FMO2-ROMP2, and FMO2-MP2:CR-CC(2,3)

basis set	method	energy (a.u.)			enthalpy (kcal/mol)
		reactant 1	reactant 2	product	
3-21G	FMO2-ROMP2	−515.27782025	−1228.15256231	−1743.46127536	−19.4
3-21G	FMO2-MP2:CR-CC(2,3)	−515.37480438	−1228.23664205	−1743.63878524	−17.2
6-311G(d,p)	FMO2-ROMP2	−518.94481925	−1235.33407266	−1754.31616825	−23.4
3-21G	ROMP2	−515.27782024	−1228.15256232	−1743.46133847	−19.4
6-311G(d,p)	ROMP2	−518.94377300	−1235.33260039	−1754.31552043	−24.6

6-31G(d) basis set.[36] FMO2-ROMP2 single point energy calculations using the 6-311G(d,p) basis set were performed with the fragmentation scheme shown in Figure 3 (the unpaired electron is shown schematically; in practice, it is delocalized). Higher-level calculations were also performed, using the completely renormalized coupled cluster single and double excitations using left eigenstates for perturbative triple excitation corrections (CR-CC(2,3))[19,20] method with the 3-21G basis set to model the open-shell fragment, and MP2 with the 3-21G basis set to model the closed-shell fragments (FMO2-MP2:CR-CC(2,3)), shown in Figure 4. The reason for using a smaller basis set is the huge memory requirements for the CC code, which even for 3-21G was 8 GB, while for the 6-31G(d) basis set, the requirements are more than 32 GB.

Table 3 gives the absolute energy differences between ab initio ROMP2 and FMO2-ROMP2 for both reactants and the product of the RAFT reaction. The FMO2-ROMP2 method gives accurate energies, with errors between 0.4 and 0.9 kcal/mol. This accuracy in absolute energies translates to equivalent accuracy when calculating the reaction enthalpy (Table 4). Comparing the enthalpy from FMO2-ROMP2 using the 6-311G(d,p) basis set with the ab initio ROMP2 enthalpy, the error is only 0.9 kcal/mol. Calculations performed with the 3-21G basis set also show very good agreement between ab initio ROMP2 and FMO2-ROMP2 calculations. With the addition of the CC correction in the FMO2-MP2:CR-CC(2,3) calculation, the enthalpy changes by +2.2 kcal/mol. This suggests that the use of ROMP2 to calculate enthalpies is adequate in this case; however, a higher level of theory may be required in other instances to properly describe open-shell systems.

3.3. Excitation Energy Calculation of a Small Protein (1L2Y). As a model application to a larger system, consider the FMO2-ROMP2 triplet excitation energy using the 6-31+G(d) basis set of the Trp-cage miniprotein construct (1L2Y), with diffuse functions added to the carboxyls. The geometry has been taken from previous studies.[9] The protein consists of 20 amino acid residues and a total of 304 atoms. When a triplet excitation of one of the Trp-6 residue in the protein is specified and also the ground-state closed-shell calculation is performed, the calculations took 160 min each on four nodes containing two 2.66 GHz quad core Intel Xeon processors (32 CPUs total). In contrast, the full ab initio ROMP2 calculation containing 2610 basis functions, if the calculation was feasible, would require 1024 CPUs with 16 GB of RAM per CPU. The FMO-ROMP2 excitation energy was found to be 93.35 kcal/mol (4.05 eV).

4. Conclusions

The open-shell FMO method has been implemented in the GAMESS program package and parallelized using GDDI for the HF, MP2, and CC levels of electronic structure theory. The accuracy of the method was tested by calculating the absolute and relative energetics of open-shell molecular clusters. The ability of the method to reproduce reaction enthalpies was also tested using the RAFT reaction. It was demonstrated that in both cases the open-shell FMO method provides energies and properties within 0.0−2.0 kcal/mol of ab initio calculations.

The need for a single reference open-shell FMO method was fulfilled through this work, providing a scalable method for use on large chemical systems such as the RAFT reaction. The combination of accuracy and reduction in computational expense provides a means for accurate calculations on much larger open-shell radical chemical systems than was previously available.

References

(1) Söderhjelm, P.; Aquilante, F.; Ryde, U. *J. Phys. Chem. B* **2009**, *113*, 11085.

(2) Pomogaev, V.; Pomogaeva, A.; Aoki, Y. *J. Phys. Chem. A* **2009**, *113*, 1429.

(3) Xie, W.; Orozco, M.; Truhlar, D. G.; Gao, J. *J. Chem. Theory Comput.* **2009**, *5*, 459.

(4) Leverentz, H. R.; Truhlar, D. G. *J. Chem. Theory Comput.* **2009**, *5*, 1573.

(5) Suárez, E.; Díaz, N.; Suárez, D. *J. Chem. Theory Comput.* **2009**, *5*, 1667.

(6) Deshmukh, M. M.; Gadre, S. R. *J. Phys. Chem. A* **2009**, *113*, 7927.

(7) Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. *Chem. Phys. Lett.* **1999**, *313*, 701.

(8) Nakano, T.; Kaminuma, T.; Sato, T.; Fukusawa, K.; Akiyama, Y.; Uebayasi, M.; Kitaura, K. *Chem. Phys. Lett.* **2002**, *351*, 475.

(9) Fedorov, D. G.; Kitaura, K. *J. Phys. Chem. A* **2007**, *111*, 6904.

(10) Fedorov, D. G.; Kitaura, K. *The Fragment Molecular Orbital Method: Practical Applications to Large Molecular Systems*; CRC Press: Boca Raton, FL, 2009.

(11) Korchowiec, J.; Gu, F. L.; Aoki, Y. *Int. J. Quantum Chem.* **2005**, *105*, 875.

(12) Fedorov, D. G.; Kitaura, K. *J. Chem. Phys.* **2004**, *120*, 6832.

(13) Fedorov, D. G.; Kitaura, K. *J. Chem. Phys.* **2004**, *121*, 2483.

(14) Fedorov, D. G.; Kitaura, K. *J. Chem. Phys.* **2005**, *122*, 054108.

(15) Mochizuki, Y.; Koikegami, S.; Amari, S.; Segawa, K.; Kitaura, K.; Nakano, T. *Chem. Phys. Lett.* **2005**, *406*, 283.

(16) Fedorov, D. G.; Kitaura, K. *J. Chem. Phys.* **2005**, *123*, 134103.

(17) Chiba, M.; Fedorov, D. G.; Nagata, T.; Kitaura, K. *Chem. Phys. Lett.* **2009**, *474*, 227.

(18) Lee, T. J.; Rendell, A. P.; Dyall, K. G.; Jayatilaka, D. *J. Chem. Phys.* **1994**, *100*, 7400.

(19) Włoch, M.; Gour, J. R.; Piecuch, P. *J. Phys. Chem. A,* **2007**, *111*, 11359.

(20) Piecuch, P.; Włoch, M. *J. Chem. Phys.* **2005**, *123*, 224105.

(21) Pryor, W. A. *Free Radicals*; McGraw-Hill: New York, 1966.

(22) Griese, B.; Wang, M.; Gao, J.; Stoltz, M.; Müller, P.; Graber, M. *J. Org. Chem.* **2009**, *74*, 3621.

(23) Astruc, D. *Electron Transfer and Radical Processes in Transition-Metal Chemistry*; Wiley-VCH: New York, 1995.

(24) Chiba, M.; Fedorov, D. G.; Kitaura, K. *J. Chem. Phys.* **2007**, *127*, 104108.

(25) Fedorov, D. G.; Ishida, T.; Kitaura, K. *J. Phys. Chem. A.* **2005**, *109*, 2638.

(26) Aikens, C. M.; Fletcher, G. D.; Schmidt, M. W.; Gordon, M. S. *J. Chem. Phys.* **2006**, *124*, 014107.

(27) Fletcher, G. D.; Schmidt, M. W.; Bode, B. M.; Gordon, M. S. *Comput. Phys. Commun.* **2000**, *128*, 190.

(28) Fedorov, D. G.; Olson, R. M.; Kitaura, K.; Gordon, M. S.; Koseki, S. *J. Comput. Chem.* **2004**, *25*, 872.

(29) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A., Jr. *J. Comput. Chem.* **1993**, *14*, 1347.

(30) Seinfeld, J. H.; Pandis, S. N. *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*; Wiley: New York, 2006.

(31) Day, P. N.; Pachter, R.; Gordon, M. S.; Merrill, G. N. *J. Chem. Phys.* **2000**, *112*, 2063.

(32) (a) Dunning, T. H., Jr. *J. Chem. Phys.* **1989**, *90*, 1007. (b) Woon, D. E.; Dunning, T. H., Jr. *J. Chem. Phys.* **1995**, *103*, 4572.

(33) Xantheas, S. *Struct. Bonding (Berlin)* **2005**, *116*, 119.

(34) Coote, M. L.; Izgorodina, E. I.; Krenske, E. H.; Busch, M.; Barner-Kowollik, C. *Macromol. Rapid Commun.* **2006**, *27*, 1015.

(35) Chiefari, J.; Chong, Y. K.; Ercole, F.; Krstina, J.; Jeffery, J.; Le, T. P. T.; Mayadunne, R. T. A.; Meijs, G. F.; Moad, C. L.; Moad, G.; Rizzardo, E.; Thang, S. H. *Macromolecules* **1998**, *31*, 5559.

(36) Structures used for the RAFT reaction calculations were obtained from Professor Michelle Coote at the Austrailian National University.

CT900442B