Quantum Mechanical Study of Aqueous Polarization Effects on Biological Macromolecules

Darrin M. York,* Tai-Sung Lee, and Weitao Yang

Department of Chemistry, Duke University
Durham, North Carolina 27708-0354

Received June 10, 1996
Revised Manuscript Received August 16, 1996

Methods to determine the electronic structure of atoms and molecules are crucial for a reliable description of complex chemical processes that are inaccessible to conventional empirical models. Standard quantum mechanical techniques are limited to fairly small molecular systems due to unfavorable scaling properties inherent in the computational algorithms. We report the application of a recently developed linear-scaling quantum mechanical method to the study of aqueous polarization effects on biological macromolecules. The polarization contribution to the solvation free energy is in the range of 5−15% for proteins and 1−3% for DNA. Results suggest that polarization of proteins and DNA in the process of solvation can be well approximated by a linear response model. The developments presented here extend the realm of quantum chemical techniques to applications of macromolecular systems in solution.

The computational effort of conventional Hartree−Fock molecular orbital or Kohn−Sham density functional methods scales formally as N^3 or higher, where N is the number of electrons in the system.[1] The $O(N^3)$ scaling derives from the requirement that the occupied molecular orbitals be orthogonal. These orbitals are generally delocalized over the entire system. In the past several years, there has been much work devoted to the development of linear-scaling electronic structure methods.[2,3] The first method introduced was the divide-and-conquer (DAC) density functional approach,[2] which utilized partitioning techniques to localize the electronic degrees of freedom and obviate the need for global matrix construction and diagonalization. Although this method overcomes the formal $O(N^3)$ bottleneck, in practice extension to very large systems for *ab initio* methods is difficult to realize since a great deal of effort must be spent in the computation of effective Hamiltonian or Fock matrix elements that require multidimensional multicenter integration.

Semiempirical methods circumvent the difficulty of time-consuming matrix construction by approximating the integrals empirically with recourse to experimental data.[4] For these methods, the $O(N^3)$ computational bottleneck becomes limiting in practice. Recently, a density matrix formulation of the DAC method has been introduced[5] that extends linear-scaling capability to Hartree−Fock-based methods, including semiempirical methods.

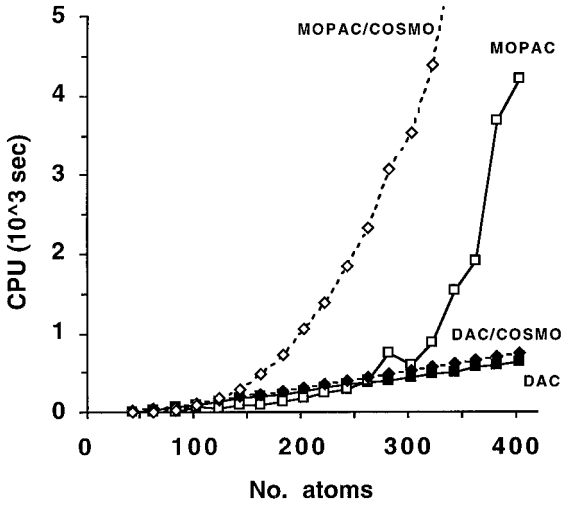

Figure 1. Comparison of the timing results for a series of polyalanine alpha helices $(Ala)_{2n}$, $n = 2, \ldots 10$, using the unmodified MOPAC program and modified DAC method:[6,9] gas phase with standard MOPAC (□), MOPAC/COSMO solvation (◇), gas phase DAC method (■), and DAC/COSMO solvation (◆) with a modified COSMO algorithm useful for large molecules.[9] Timings are for 20 SCF iterations on an IBM RISC/6000 3CT, 128 MB memory.

We have implemented the density matrix formulation of the DAC method into a semiempirical framework so that electronic structure calculations can be extended to macromolecular systems.[6] The computational advantage of the semiempirical/DAC method for large molecules are clear (Figure 1). On a modern single-processor workstation, systems up to around 10 000 atoms can be treated.[6] Here, we present results of the first application of the method to the determination of solvation free energies of protein and DNA systems in solution.

To obtain a realistic quantum mechanical description of a biological system, it is necessary to include solvation effects.[7] We employ a conductor-like screening model[8] for the reaction field of high-dielectric media such as water. We have modified the algorithm to accommodate large molecules by employing conjugate gradient methods combined with fast-multipole techniques[9] to circumvent the $O(N^3)$ matrix inversion procedure conventionally employed for small molecules.[8,10]

Efficient algorithms for solving the Poisson and Poisson−Boltzmann equations have been applied extensively to the study of macromolecular systems in solution.[11] Typically, the solute charge density is represented as a set of static point charges derived empirically from chemical fragments and assembled to form a macromolecule. This representation ignores electronic relaxation of the fragments in their macromolecular environment and polarization effects induced by the solvent. The polarization energy of a molecule is the net energetic stabilization that results from the electronic relaxation (polarization) in response to an applied field. With an explicit treatment of the electronic degrees of freedom, we can directly determine the contribution of the polarization energy to the total solvation energy (Scheme 1). Similar studies have been reported for small molecules using continuum solvation methods[12] and combined quantum me-

* Current address for correspondence: Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138

(1) Friesner, R. A. *Annu. Rev. Phys. Chem.* **1991**, *42*, 341.
(2) Yang, W. *Phys. Rev. Lett.* **1991**, *66*, 438.
(3) Cortona, P. *Phys. Rev. B* **1991**, *44*, 8454. Baroni, S.; Giannozzi, P. *Europhys. Lett.* **1992**, *17*, 547. Galli, G.; Parinello, M. *Phys. Rev. Lett.* **1992**, *69*, 3547. Li, X.-P.; Nunes, W.; Vanderbilt, D. *Phys. Rev. B.* **1993**, *47*, 10891. Mauri, F.; Galli, G.; Car, R. *Phys. Rev. B* **1993**, *47*, 9973. Ordejón, P.; Drabold, D. A.; Grumback, M. P.; Martin, R. M. *Phys. Rev. B* **1993**, *48*, 14646. Stechel, E. B.; Williams, A. P.; Feibelman, P. J. *Phys. Rev. B* **1993**, *49*, 3898. Daw, M. S. *Phys. Rev. B* **1993**, *47*, 10899. Drabold, D. A.; Sankey, O. F. *Phys. Rev. Lett.* **1993**, *70*, 3631. Kohn, W. *Chem. Phys. Lett.* **1993**, *208*, 167. Stewart, J. J. P. *Int. J. Quantum Chem.* **1996**, *58*, 133.
(4) Stewart, J. J. P. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH Publishers: New York, 1990; Vol. 1. Zerner, M. C. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH Publishers: New York, 1994; Vol. 5.
(5) Yang, W.; Lee, T.-S. *J. Chem. Phys.* **1995**, *103*, 5674.
(6) Lee, T.-S.; York, D. M.; Yang, W. *J. Chem. Phys.* **1996**, *105*, 2744.
(7) Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2027. Cramer, C. J.; Truhlar, D. G. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH Publishers: New York, 1995; Vol. 6.
(8) Klamt, A.; Schüürmann, G. *J. Chem. Soc., Perkin Trans.* **1993**, *2*, 799.
(9) York, D. M.; Lee, T.-S.; Yang, W. *Chem. Phys. Lett.* In press.
(10) Andzelm, J.; Kölmel, C.; Klamt, A. *J. Chem. Phys.* **1995**, *103*, 9312. Truong, T. N.; Stefanovich, E. V. *Chem. Phys. Lett.* **1995**, *240*, 253. Truong, T. N.; Stefanovich, E. V. *J. Chem. Phys.* **1995**, *103*, 3709.
(11) Gilson, M. K. *Curr. Opin. Struct. Biol.* **1995**, *5*, 216. Honig, B.; Nicholls, A. *Science* **1995**, *268*, 1144.

S0002-7863(96)01937-3 CCC: $12.00

Scheme 1

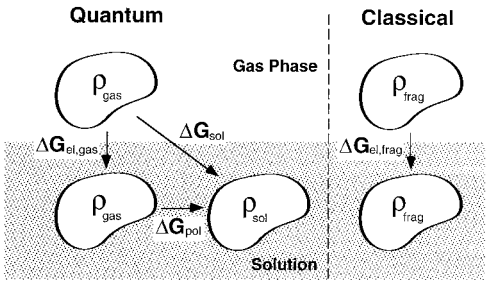

Table 1. Solvation Free Energies of Proteins and DNA[a]

molecule	atoms	ΔG_{sol}	ΔG_{pol} (%)	$\Delta G_{QM\text{-}Cl}$[b] (%)
		Proteins		
amyloid β-peptide	438	−27.3	−3.1 (11.5)	1.4 (5.2)
crambin	642	−9.6	−1.3 (13.5)	0.7 (7.8)
bpti	892	−44.3	−3.0 (6.7)	0.4 (1.0)
calbindin P43G	1195	−66.5	−5.4 (8.1)	6.1 (9.3)
ribonuclease A	1856	−53.7	−6.4 (11.9)	5.0 (9.3)
lysozyme	1960	−68.3	−5.6 (8.2)	2.3 (10.1)
HIV-1 PR dimer	3118	−69.7	−7.6 (10.9)	6.0 (8.6)
subtilisin BPN	3837	−52.4	−7.1 (13.5)	5.3 (10.2)
trypsin/bpti complex	4112	−119.8	−10.1 (8.5)	4.9 (4.1)
superoxide dismutase	4380	−113.2	−8.7 (7.7)	2.2 (1.9)
		DNA		
d(CGCGAATTCGCG)$_2$	758	−281.2	−5.3 (2.0)	−1.9 (0.7)
A-DNA, ideal (CG)$_8$	1006	−444.5	−8.9 (2.0)	−5.2 (1.2)
B-DNA, ideal (CG)$_8$	1006	−445.3	−7.6 (1.7)	−0.1 (0.0)
Z-DNA, ideal (CG)$_8$	1006	−464.5	−12.4 (2.7)	4.5 (1.0)

[a] Free energies are in electron volts. Initial structures were obtained from the Brookhaven Protein Data Bank and refined with 1000 steps of conjugate gradient energy minimization using AMBER4.1.[16] [b] $\Delta G_{QM\text{-}Cl} = \Delta G_{sol} - \Delta G_{el,frag}$ is the difference between the quantum mechanical solvation energy and the classical reaction field energy obtained from the fragment density charge distribution (Scheme 1).

chanical/molecular mechanical potentials;[13] however, to date, this has not been feasible for macromolecules.

Solvation energy calculations were performed for several protein and DNA systems (Table 1).[14] The polarization contribution to the free energy is typically in the range of 5−15% of the total solvation energy for proteins and 1−3% for DNA. The dominant contribution to the solvation energy for DNA results from the solvent effect induced by the static charge distribution. The polarization contribution, although fairly large in magnitude with respect to that of similar sized proteins, is relatively small in comparison with the total solvation energy.

An important question arises as to whether the behavior of biological macromolecules can be adequately described by a linear response model. It can be shown that if the solute is described by linear response theory, the *solute polarization energy E* is equal to $E_{LR} = {}^1/_2\int\delta\rho(\mathbf{r})\nu_{RF}(\mathbf{r})d^3r$, where ν_{RF} is the solvent reaction field and $\delta\rho = \rho_{sol} - \rho_{gas}$ is the *polarization density*. The average root-mean-square deviation (rmsd) of these two quantities for the molecules in Table 1 is 0.32 eV (E ranges from −1.59 to −12.6 eV), and regression analysis results in a straight line fit (slope = 1.05, intercept = 0.19 eV, rmsd = 0.24 eV). This result provides support for the use of linear response theory for modeling biological macromolecules in solution.

It is instructive to examine the errors involved if we employ a static charge model to calculate solvation energies, similar to conventional empirical methods routinely applied to macromolecules.[11] Static charge distributions were generated by performing calculations on isolated amino acids and nucleotides and assembling the atomic fragment density matrix elements in the macromolecular structure.[9] The quantity $\Delta G_{el,frag}$ (Scheme 1) was calculated as the reaction field energy for the fragment charge distribution. The errors in $\Delta G_{el,frag}$ are typically less than 10%. The difference between $\Delta G_{el,frag}$ and $\Delta G_{el,gas}$ reflects the degree to which the reaction field is influenced by relaxation of the electron density from its fragment representation to its gas phase distribution. Biological macromolecules frequently contain interacting charged and polar groups, such as salt bridges and hydrogen bonds, that are well separated in the fragment density representation. Electronic relaxation in the gas phase disfavors charge separation, and the resulting charge distribution has a smaller reaction field energy. The polarization component ΔG_{pol} increases the solvation energy by stabilizing charged regions. The reasonable agreement between the values of $\Delta G_{el,frag}$ and ΔG_{sol} thus derives from a cancellation of errors involving the neglect of electronic relaxation in the macromolecule and polarization in response to the solvent reaction field.

We have applied the density matrix divide-and-conquer method to the study of biological macromolecules in solution with explicit treatment of the electronic degrees of freedom. Our results indicate that in many instances it is necessary to take into account electronic relaxation in the macromolecular environment and polarization effects induced by the solvent. It is further suggested that aqueous polarization effects may be accurately described by linear response models, which lends support to the development of biomolecular force fields that incorporate many-body effects through linear response theory. We have demonstrated the feasibility of extending electronic structure calculations to the study of macromolecular systems and anticipate these methods will open doors to a host of new problems accessible to quantum mechanical investigation.

Acknowledgment. The authors gratefully acknowledge L. Bartolotti for technical assistance and financial support from a subcontract agreement with the NIH Research Resource Program at the University of North Carolina at Chapel Hill. This work was supported by the U.S. Environmental Protection Agency, the National Science Foundation, and the Exxon Education Foundation. D.Y. acknowledges support through an NSF postdoctoral fellowship jointly funded by the North Carolina Supercomputing Center. W.Y. acknowledges support from the Alfred P. Sloan Foundation.

JA961937W

(12) Cramer, C. J.; Truhlar, D. G. *Science* **1992**, *256*, 213. Cramer, C. J.; Truhlar, D. G. *Chem. Phys. Lett.* **1992**, *198*, 74. Yu, J.; Zerner, M. C. *J. Chem. Phys.* **1994**, *100*, 7487. C. J. Rashin, A. A.; Young, L.; Topol, I. A. *Biophys. Chem.* **1994**, *51*, 359. Rashin, A. A.; Bukatin, M. A.; Andzelm, J.; Hagler, A. T. *Biophys. Chem.* **1994**, *51*, 375. Tannor, D. J.; Marten, B.; Murphy, R.; Friesner, R. A.; Sitkoff, D.; Nicholls, A.; Ringnalda, M.; Goddard, W. A., III; Honig, B. *J. Am. Chem. Soc.* **1994**, *116*, 11875. Chen, J. L.; Noodleman, L.; Case, D. A.; Bashford, D. *J. Phys. Chem.* **1994**, *98*, 11059.

(13) Gao, J.; Xia, X. *Science* **1992**, *258*, 631. Gao, J. *Biophys. Chem.* **1994**, *51*, 253. Orozco, M.; Luque, F. J.; Habibollahzadeh, D.; Gao, J. *J. Chem. Phys.* **1995**, *102*, 6145. Thompson, M. A.; Schenter, G. K. *J. Phys. Chem.* **1995**, *99*, 6374.

(14) The semiempirical/DAC method[6,9] was used with the AM1 Hamiltonian[15] for all calculations. Macromolecules were divided into subsystems consisting of a single amino acid or nucleotide residue. Basis functions for projection of the Fock matrix for each subsystem were chosen to be the set of basis functions on all atoms within 6.0 Å of any subsystem atom. A 7.0 Å cutoff was used in the computation of density matrix elements. This protocol has been shown to give accuracy on the order of 10^{-4} eV/atom.[6] Atomic radii were obtained from fitting to experimental and theoretical data of relevant small molecules.[9] The convergence criterion for the SCF procedure was 10^{-5} eV/atom.

(15) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.

(16) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E., III; Ferguson, D. M.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. AMBER 4.1, University of California, San Francisco, 1995. Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S., Jr.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765. Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J. Comput. Chem.* **1986**, *7*, 230.