Electronic Structure Properties of Solvated Biomolecules: A Quantum Approach for Macromolecular Characterization

JANA KHANDOGIN, ANCJANC HU, DARRIN M. YORK

Department of Chemistry, University of Minnesota, 207 Pleasant St. SE, Minneapolis, Minnesota 55455

Received 18 May 2000; accepted 7 July 2000

ABSTRACT: Linear-scaling electronic structure calculations of solvated biomolecules have been carried out using a semiempirical Hamiltonian and a new smooth solvation potential. These methods afford a new way of generating macromolecular properties that include quantum electronic structure. In addition to the widely used classical electrostatic potential maps based on empirically derived static point charges, now fully quantum mechanical electrostatic potentials that include electronic polarization are possible. Linear-scaling electronic structure methods provide a host of response properties of the electron density such as linear response functions, local hardness functions, and Fukui functions. It is the hope that these indices will extend insight into problems of biological macromolecular characterization. © 2000 John Wiley & Sons, Inc. J Comput Chem 21: 1562–1571, 2000

Key words: solvated biomolecules; macromolecular characterization; linear scaling; electronic structure

Introduction

The overall accuracy and predictive capability of computational models for biological systems are limited by (1) the accuracy to which intermolecular forces are described, (2) the adequacy to which relevant phase space is sampled, and (3) the degree to which the microscopic system on the computer reflects the (typically) macroscopic system in nature. (For molecular simulations, there are issues with regard to the classical vs. quantum mechanical treatment of nuclear motion, and for electronic structure methods, issues with regard to the validity of the Born–Oppenheimer approximation and the importance of relativistic effects.)

Applications targeted at the study of chemical reactions that take place in complex chemical en-

Correspondence to: D. M. York; e-mail: york@chem.umn.edu

Contract/grant sponsors: National Institutes of Health, the University of Minnesota, and the Minnesota Supercomputing Institute

vironments are of great importance in many disciplines. Often the environment has a tremendous effect on the nature of the chemistry of the interacting species relative to that which would occur in vacuum. An important and illustrative example is that of enzyme catalysis where a protein, polynucleotide, or macromolecular complex along with associated bound metals and solvent provide the catalytic environment.

"Multilevel" methods allow the use of a hierarchy of theoretical models that balance accuracy and computational efficiency. The goal, then, is to determine the weakest link in the model with respect to the particular properties of interest in the application, and design a multilevel scheme within the constraints of available computational resources to best achieve reliable, predictive answers. As an example, the diverse behavior of water molecules at the atomic level is amazingly complex, and remains an area of active research by both experimental and theoretical methods. However, to a reasonable approximation, the effect of bulk water far enough away from the active site of a biomolecule can be treated as a dielectric continuum with a set of equations that can be readily solved.[1] Similarly, hybrid quantum mechanical/molecular mechanical (QM/MM) methods are examples of a particular class of multilevel quantum models that have received a tremendous amount of attention in the past several years.[2]

In this article, linear-scaling electronic structure methods are combined with a smooth solvation potential and applied to the determination of useful properties for macromolecular characterization. These include quantum mechanical electrostatic potential surfaces, polarization response densities, solvation energies, and electronic density of states.

Scaling

A key issue in the development of models designed for very large systems is *scaling*: the rate at which the computational effort increases as the system size is increased (in some systematic way) while maintaining a fixed level of overall accuracy. For the purposes of discussion, the *formal* scaling (as described above) of computational effort is termed "linear scaling" if the computational effort scales more favorably than N^{α} where α is any real number greater than 1; i.e.,

$$\lim_{N\to\infty} \frac{CPU(N)}{N^{\alpha}} = 0 \qquad \forall\, \alpha > 1$$

where $CPU(N)$ represents the computational effort as a function of "system size" N. This definition of linear scaling thus encompasses the class of methods that scale formally as $O(N \log N)$.

Methods for Classical Electrostatics and Solvation

Classical electrostatic interactions play a key role in almost all chemical models from purely empirical force fields to *ab initio* electronic structure methods. Of prime importance to large-scale modeling applications are methods to efficiently calculate long-range Coulomb interactions and that have favorable scaling properties in the limit of large number of particles. A tremendous amount of work and progress has been made in the development of linear-scaling methods to solve the Poisson equation for the classical electrostatic potential for systems of point charges or smooth charge densities under periodic and nonperiodic boundary conditions.[3,4] The most common modern strategies can be broadly categorized into methods based on multipole expansions of the potential such as tree codes and fast-multipole methods[5–7] and methods based on plane-wave expansions of the potential such as linear-scaling Ewald methods.[3,8] These methods can be viewed as Green's function expansions of the Laplacian operator in real and reciprocal space, respectively. (Although it may be argued that methods based on spherical harmonic multipole expansions are most natural for nonperiodic systems and plane wave expansions are most natural for periodic systems, there have been generalizations of both techniques: fast-multipole methods for periodic systems, and Fourier methods for nonperiodic systems.)

For large biological applications, inclusion of the solvent environment is critical. Frequently an implicit solvation model[1,9,10] can be employed to remove the explicit solvent degrees of freedom that would otherwise dominate the calculation. It is often not clear how appropriate these models are for treatment of the first solvation layers around a biological macromolecule; however, it is generally believed that they have a useful place within the hierarchy of multilevel models (e.g., outside some domain where the effect of the bulk solvent on the region of interest can be sufficiently approximated by a continuum model). For these reasons, it is of significant importance to develop a linear-scaling solvation potential that is numerically stable and sufficiently smooth for biological applications.

The model outlined here is based on a recent general formulation of a smooth analytic solvent potential[11] based on the conductor-like screening model.[12] The method is a type of boundary-element method that uses a variational principle for the electrostatic energy appropriate for a conductor, and employs a correction factor for finite dielectric medium. A computational advantage of the method is that it utilizes a discretized 2D surface instead of a 3D grid; however, for more complicated spatially varying dielectric functions or nonlinear (e.g., ionic) terms, grid-based methods have some natural advantages. Although the COSMO method is an exact solution to the electrostatic problem only in the limit of infinite external dielectric, it has been shown to be very accurate for a high dielectric medium such as water. (Even an exact solution of the electrostatic problem for finite external dielectric is an approximation for only one component of the solvation energy.) More recently, a COSMO-RS model (conductor-like screening model for real solvents) has been developed and parameterized to improve finite dielectric screening in other organic solvents.[13, 14]

The equation for the induced solvent polarization energy is

$$E_{\mathrm{pol}} = \frac{1}{2}\boldsymbol{\sigma}_{\mathrm{pol}}^{T} \cdot \mathbf{A} \cdot \boldsymbol{\sigma}_{\mathrm{pol}} + \boldsymbol{\sigma}_{\mathrm{pol}}^{T} \cdot \mathbf{B} \cdot \boldsymbol{\rho}_0 \qquad (1)$$

where $\boldsymbol{\sigma}_{\mathrm{pol}}$ is a $N \times 1$ vector representing the polarization surface charge density, $\boldsymbol{\rho}_0$ is a $M \times 1$ vector representing the solute charge density (here treated as fixed), and $\mathbf{A}$ and $\mathbf{B}$ are $N \times N$ and $N \times M$ matrices representing the self-interaction of the surface charge vectors and the interaction of the surface charge and solute charge density, respectively. For a conducting medium (infinite dielectric), the $\mathbf{A}$ matrix is just an electrostatic interaction matrix between surface elements. In the present model, if there is an internal dielectric ϵ_1 and external dielectric ϵ_2, the electrostatic interaction matrix is scaled by a factor $1/f(\epsilon_1, \epsilon_2)$, where

$$f(\epsilon_1, \epsilon_2) = \frac{\epsilon_2 - \epsilon_1}{\epsilon_1 \epsilon_2} \qquad (2)$$

This scale factor is in accord with Gauss' law for the total surface charge density. The polarization surface charge density is obtained by a minimization procedure written in the general form

$$\delta\left\{E_{\mathrm{pol}} - \boldsymbol{\lambda}^{T} \cdot \left(\mathbf{D}^{T} \cdot \boldsymbol{\sigma}_{\mathrm{pol}} - \mathbf{y}\right)\right\} = 0 \qquad (3)$$

where $\boldsymbol{\lambda}$ is a vector of Lagrange multipliers on N_c constraint conditions

$$\mathbf{D}^{T} \cdot \boldsymbol{\sigma}_{\mathrm{pol}} = \mathbf{C} \cdot \boldsymbol{\rho}_0 = \mathbf{y} \qquad (4)$$

where $\mathbf{D}$ is a $N_c \times N$ matrix representing N_{con} linear constraint equations and the vector $\mathbf{C} \cdot \boldsymbol{\rho}_0 = \mathbf{y}$ are the constraint values. It is assumed that the constraint values are linear functions of the static density $\boldsymbol{\rho}_0$. For example, the Gauss' law constraint on the total surface charge for $\epsilon = \infty$, assuming the surface and static charge densities are expanded in a basis of L_1 normalized functions, corresponds to $N_c = 1$, $D_{i,1} = 1$, and $M_{1,i} = -1$ (which implies $y_1 = -Q_0$). Although the scale factor $f(\epsilon_1, \epsilon_2)$ is consistent with the Gauss law, it does not guarantee the exact surface charge in a numerical calculation. (There is also the issue of charge penetration that has been discussed extensively elsewhere.[15]) The use of constraints on the surface charge and the mechanisms for enforcing them are still topics of interest and discussion.

The variational solution for the surface charge vector is

$$\begin{aligned}\boldsymbol{\sigma}_{\mathrm{pol}}^{*}(\lambda) &= -\mathbf{A}^{-1}(\mathbf{B} \cdot \boldsymbol{\rho}_0 - \mathbf{D} \cdot \boldsymbol{\lambda}) \\ &= \boldsymbol{\sigma}_{\mathrm{pol}}^{*}(0) + \delta\boldsymbol{\sigma}_{\mathrm{pol}}^{*}(\lambda)\end{aligned} \qquad (5)$$

and

$$\boldsymbol{\lambda} = \mathbf{Q}^{-1} \cdot \mathbf{R} \cdot \boldsymbol{\rho}_0 \qquad (6)$$

where the matrices $\mathbf{Q} = \mathbf{D}^{T} \cdot \mathbf{A}^{-1} \cdot \mathbf{D}$ and $\mathbf{R} = (\mathbf{C} + \mathbf{D}^{T} \cdot \mathbf{A}^{-1} \cdot \mathbf{B})$ have been introduced, $\boldsymbol{\sigma}^{*}(0) = -\mathbf{A}^{-1} \cdot \mathbf{B} \cdot \boldsymbol{\rho}_0$ is the unconstrained surface charge vector and $\delta\boldsymbol{\sigma}_{\mathrm{pol}}^{*}(\lambda) = \mathbf{A}^{-1} \cdot \mathbf{D} \cdot \boldsymbol{\lambda}$ is the constraint correction. The resulting energy expression is

$$\begin{aligned}E_{\mathrm{pol}}(\lambda) &= \frac{1}{2}\boldsymbol{\sigma}_{\mathrm{pol}}^{*}(0)^{T} \cdot \mathbf{B} \cdot \boldsymbol{\rho}_0 \\ &\quad + \frac{1}{2}\delta\boldsymbol{\sigma}_{\mathrm{pol}}^{*}(\lambda)^{T} \cdot \mathbf{A} \cdot \delta\boldsymbol{\sigma}_{\mathrm{pol}}^{*}(\lambda) \\ &= \frac{1}{2}\boldsymbol{\rho}_0^{T} \cdot \left[\mathbf{G}_{\mathrm{pol}}(0) + \delta\mathbf{G}_{\mathrm{pol}}(\lambda)\right] \cdot \boldsymbol{\rho}_0 \\ &= \frac{1}{2}\boldsymbol{\rho}_0^{T} \cdot \mathbf{G}_{\mathrm{pol}}(\lambda) \cdot \boldsymbol{\rho}_0\end{aligned} \qquad (7)$$

where $\mathbf{G}_{\mathrm{pol}}(\lambda) = \mathbf{G}_{\mathrm{pol}}(0) + \delta\mathbf{G}_{\mathrm{pol}}(\lambda)$ is the Green's function of the constrained variational procedure, $\mathbf{G}_{\mathrm{pol}}(0) = -\mathbf{B}^{T} \cdot \mathbf{A}^{-1} \cdot \mathbf{B}$ is the unconstrained Green's function, and $\delta\mathbf{G}_{\mathrm{pol}}(\lambda) = \mathbf{R}^{T} \cdot \mathbf{Q}^{-1} \cdot \mathbf{R}$ is the constraint correction.

The key feature in the new model is that the solvation potential is a numerically stable and rigorously *smooth* function of the atomic positions (Fig. 1). This derives from the definition of the $\mathbf{A}$ and $\mathbf{B}$ matrix elements that can change dimension as the conformation of the molecule changes, but do so smoothly with respect to variations in the energy.[11] This is accommodated by modeling surface elements associated with the polarizatron surface charge density as three-dimensional

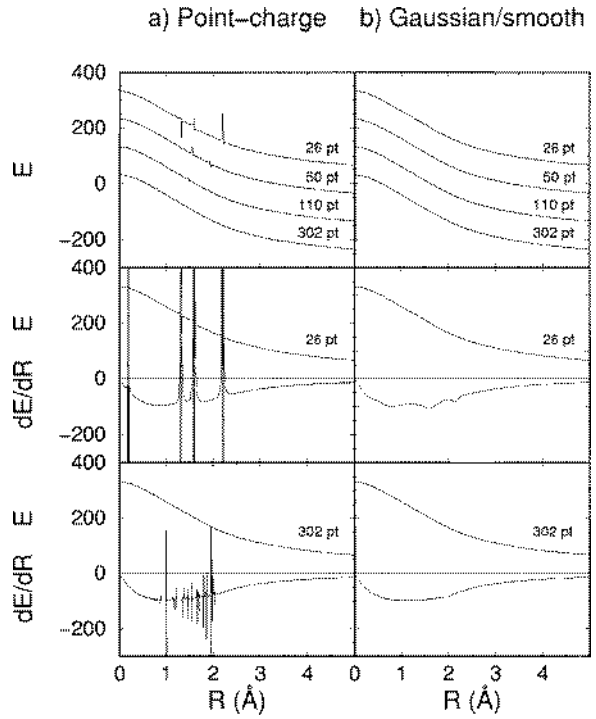

FIGURE 1. Smoothness of the solvation potential. Curves are shown at several atomic sphere discretization levels for two ions with unit radii and oppositely signed unit charges as they separate along the *z* axis. (a) *Point-charge* model: interactions between surface elements were modeled by point-charge interactions. (b) *Gaussian/smooth* model: interactions between surface elements were modeled by Gaussian Coulomb interactions that occur smoothly as they become exposed or buried. The top row shows the relative solvation energy *E* (kcal/mol, shifted for clarity) as a function of separation distance **R**, and the middle and bottom rows include the corresponding gradient curves (kcal/mol-Å) below the horizontal zero axis. Gradients were computed by finite differences ($dE/dR = \Delta E/\Delta R$ with $\Delta R = 0.01$ Å) to graphically depict the relative area associated with each singularity.

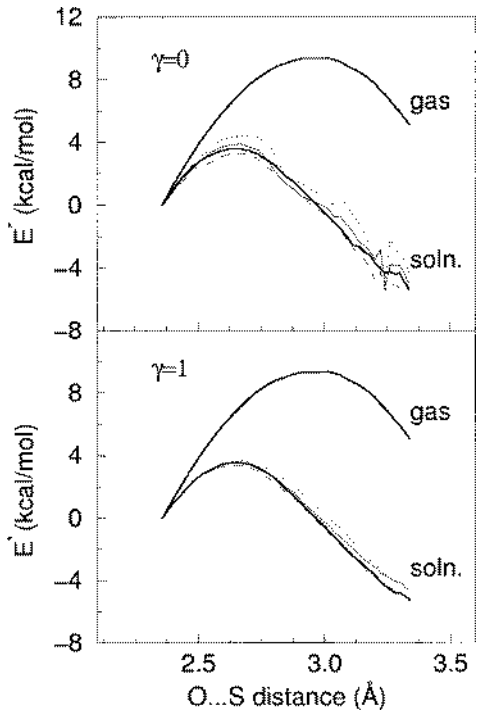

FIGURE 2. Potential energy surfaces for OH^- attack on ethylene sulfate at different levels of smoothing. Curves at several discretization levels (14, 26, 50, and 110 points/atomic sphere, corresponding to dotted, solid, dashed, and thick solid lines, respectively) are shown for smoothing levels $\gamma = 0$ (no smoothing, top), and $\gamma = 1$ (full smoothing, bottom). See ref. 11.

Gaussian functions distributed on spheres according to the rules of high-order numerical quadrature schemes for spherical harmonics. The surface elements are smoothly "switched on" when they become exposed on the solvent accessible surface and "switched off" when they become buried. The smoothness of the solvation potential is problematic for many boundary element methods,[1, 9] and is an important factor when performing gradient-based geometry optimizations, transition state searches, and molecular dynamics simulations. Figure 2 shows the effect of using the smooth solvation model without smoothing ($\gamma = 0$) and with smoothing ($\gamma = 1$) for OH^- attack on ethylene sulfate.

The conventional conductor-like screening model has been made into a linear-scaling algorithm,[10] and this technique can be extended to the Gaussian solvation model described here. The key features needed to linearize the computational effort are to (1) circumvent the $O(M^2)$ and $O(M \times N)$ computational and memory requirements to construct the **A** and **B** matrices, and (2) avoid the formal $O(M^3)$ matrix inversion procedure. These problems can be overcome by minimizing directly the energy functional of eq. (1) and recognizing that the matrix operations are essentially the evaluation of Coulomb's law. (Any deviation from Coulomb's law, such as the diagonal elements of the **A** matrix, can be included as an $O(M)$ short-ranged correction.) The minimization procedure can be greatly sped up using a preconditioned conjugate gradient method,[10] combined with an adaptive recursive bisection method for linear-scaling evaluation of the electrostatic potential and field.[16] The method is applied in the following section for the linear-scaling electronic structure calculation of biological macromolecules in solution.

Application of Linear-Scaling Electronic Structure Methods to Biomolecules

Linear-scaling electronic structure methods[15–19] (or so-called "order-N" methods) are beginning to find wide-spread application to macromolecular problems. In the near future, these methods will begin to be applied at the *ab initio* level using quality basis sets; however, it is currently a considerably strenuous computationally task to reach the size limit where the orthogonality/idempotency condition becomes a truly dominating factor superseding matrix element construction. On the other hand, for quantum methods that employ a semiempirical Hamiltonian model, linear-scaling electronic structure methods[10, 20–22] have allowed access to a host of new large-scale problems. These include the calculation of solvation and polarization energies of biomolecules,[23–25] solute to solvent charge transfer,[26, 27] electronic density of states,[28] enzyme active site structure,[29, 30] reaction energy profiles,[31] and the polarization of ligands.[32]

Semiempirical methods,[33, 34] by virtue of their efficiency and flexibility, are likely to make a major impact in biomolecular modeling over the next few years. This is largely because the complexity of biological systems requires high accuracy combined with extensive configurational sampling. Semiempirical methods are typically two orders of magnitude faster than *ab initio* methods, and offer the added flexibility of parameters that can be adjusted to obtain very high accuracy within a relatively narrow range of chemistry. Certainly *ab initio* techniques will play an increasingly important role; nonetheless, semiempirical methods have a useful place in the construction of multilevel quantum models for biological applications.

In this section, a survey of linear-scaling quantum calculations on biological macromolecules in solution is presented to illustrate some of the types of information that these calculations can provide. The field is rapidly advancing, and new insights continue to emerge through ongoing applications. The focus here is on the effect of solute polarizability in the process of solvation.

With linear-scaling electronic structure methods, one can directly access the many-body contribution of solute polarization in the process of solvation. Consider, for example, the thermodynamic cycle depicted by:

$$\begin{array}{ccc} |\rho_{gas}|_{gas} & \xrightarrow{\Delta G_{int}} & |\rho_{sol}|_{gas} \\ \Big\downarrow{\scriptstyle \Delta G_{el,gas}} & & \Big\downarrow{\scriptstyle \Delta G_{el,sol}} \\ |\rho_{gas}|_{sol} & \xrightarrow[\Delta G_{pol}]{} & |\rho_{sol}|_{sol} \end{array}$$

where the quantities in brackets, ρ_{gas} and ρ_{sol}, represent the relaxed electron density in the gas phase and in solution, respectively. The subscript on the outside of the brackets indicates the environment (*gas* phase or *solution*). The solvation energy is the energy associated with going from $|\rho_{gas}|_{gas} \rightarrow |\rho_{sol}|_{sol}$; thus $\Delta G_{sol} = \Delta G_{el,gas} + \Delta G_{pol} = \Delta G_{int} + \Delta G_{el,sol}$.

Table I shows the solvation free-energy components estimated from linear-scaling electronic structure calculations of an ensemble of structures derived from solution NMR data of ovomucoid turkey third domain.[35] The mean solvation free energy of the 12 NMR structures is −32.5 eV, and the mean polarization contribution is −4.3 eV (13.2% of ΔG_{sol}), ranging from −3.7 eV (12.2%) to −4.7 eV (14.6%). The solvation free energy of the crystallographic structure is −24.9 eV, about 23% less than the average NMR value, with a polarization contribution of −3.1 eV (12.4%). This is not surprising, because the solution structure is expected to be more flexible and have increased interaction with solvent. The contribution of ΔG_{pol} has been estimated with semiempirical methods to range typically between 5–15% for proteins and around 2% for DNA (the latter resulting from the large overall solvation effect for highly charged systems).

A useful macromolecular index in molecular biology is the electrostatic potential $\phi(\mathbf{r})$. If the electrostatic component of the solvation energy is modeled

TABLE I.
Solvation Free-Energy Components for OMTKY 3rd Domain.[a]

	ΔG_{sol}	$\Delta G_{el,gas}$	ΔG_{pol}	ΔG_{int}	$\Delta G_{el,sol}$
X-ray	−24.9	−21.8	−3.1	3.7	−28.6
⟨NMR⟩	−32.5	−28.2	−4.3	5.2	−37.3
std. dev.	1.3	1.0	0.4	0.5	1.7

[a] "X-ray" indicates that the X-ray crystallographic structure was used with hydrogens added and the structure refined with molecular mechanics minimization. ⟨NMR⟩ indicates the calculation average over the 12 NMR structures.[35] All electronic structure calculations were performed on a single protein (814 atoms) using the linear-scaling electronic structure and solvation methods described previously[10, 11, 36] with the AM1 Hamiltonian.[37] All units are in eV.

as a uniform isotropic polarizable medium, the total electrostatic potential $\phi(\mathbf{r})$ is a solution of the Poisson equation

$$\nabla \cdot \epsilon(\mathbf{r}) \nabla \phi(\mathbf{r}) = -4\pi\rho(\mathbf{r}) \tag{8}$$

where $\rho(\mathbf{r})$ is the solute charge distribution. The conductor-like screening model described previously provides an efficient means for the approximate solution of the Poisson equation. Frequently $\phi(\mathbf{r})$ provides quantitative insight into the electrostatic features at the molecular surface, and is used as a measure of molecular similarity.[38] Electrostatic properties have proven useful for the identification of functional homology of proteins, prediction of counterion binding sites in proteins and nucleic acids, and explanation of hydrophobic interactions and electron-transfer rates.[39–43] More recently, a study of electrostatics of RNA, based on numerical solutions to the nonlinear Poisson–Boltzmann equation using force field-derived point charge parameters has further revealed the significance of electrostatic complementarity in RNA recognition and stabilization.[44] Another molecular descriptor to measure electrostatic complementary that may prove useful in biological applications is the profile of induced charge density on the dielectric boundary.[14] This quantity is readily available from boundary element methods.

Typically, classical treatments of the electrostatic potential are based on a static point charge representation of the macromolecular solute constructed from an assembly of fixed charge fragments. This type of representation neglects (1) the contribution of atomic multipoles to the electrostatic potential, (2) explicit electronic relaxation of the fragments in the macromolecular environment, and (3) polarization of the solute by the induced solvent reaction field.

Linear-scaling electronic structure methods can be used as an explicit quantum model for the electronic polarization of biological macromolecules. The method is applied here to the determination of quantum mechanical electrostatic potential surfaces calculated in the gas phase and self-consistently in the reaction field of a dielectric continuum. A more detailed study of the method and application to biomolecules is forthcoming. The focus here is on the effect of solute electronic polarization in the process of solvation. Consider the solvation process:

$$(\rho_{\text{gas}})_{\text{gas}} \rightarrow (\rho_{\text{gas}})_{\text{sol}} \rightarrow (\rho_{\text{sol}})_{\text{sol}} \tag{9}$$

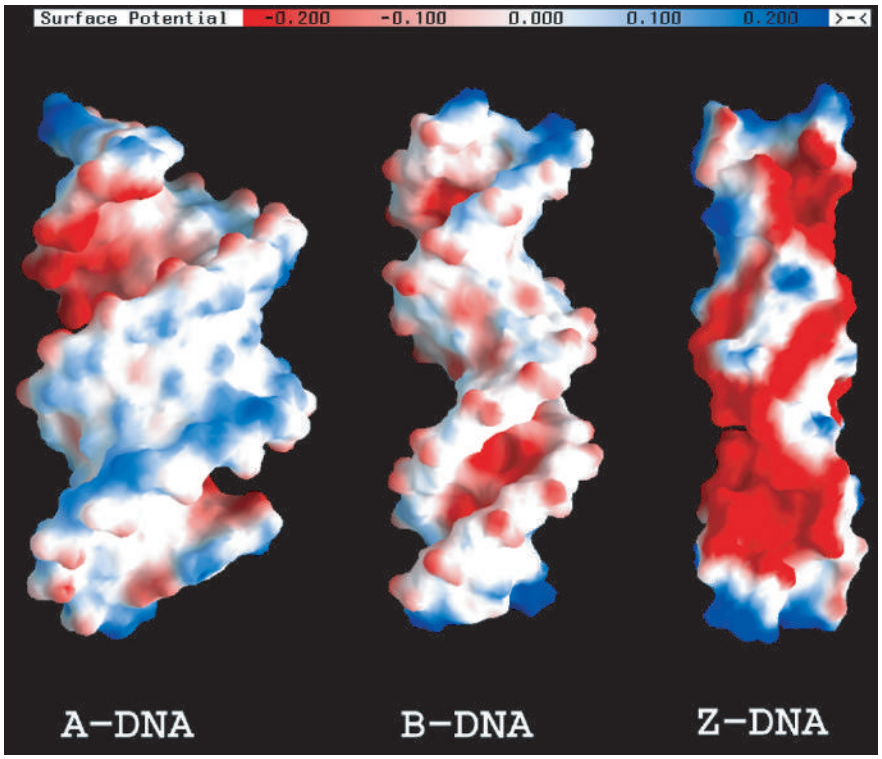

FIGURE 3. The electrostatic potential surface of the electronic response density $\delta\rho(\mathbf{r}) = \rho_{\text{sol}}(\mathbf{r}) - \rho_{\text{gas}}(\mathbf{r})$ (note: this density is neutral). Units are in kT/e ($T = 300$ K). Calculations were performed using the PM3 Hamiltonian[50] and a buffer/matrix cutoff of 8/9 Å.

The *electronic polarization response density* of this process is defined as the difference density $\rho_{pol} = \rho_{sol} - \rho_{gas}$. Figure 3 shows the electrostatic potential difference map of the solvated densities ρ_{sol} and ρ_{gas} (*i.e.*, $\Phi_{sol} - \Phi_{gas}$). The solvation potential used is a linear equation in the density; hence, the potential in Figure 3 is equivalent to the electrostatic potential of the solvated electronic polarization response density ρ_{pol}. Henceforth, this type of potential map is referred to simply as an *electronic polarization map*. Regions colored in red show areas of increased electronegative potential, and regions colored in blue show regions of decreased electronegative potential. The local features of the electronic polarization map are affected by the polarizability of residues near the region, and by the magnitude of the solvent reaction field potential. In A- and Z-form DNA, the major grooves become more negatively polarized, whereas in B-form DNA the minor grove is more negatively polarized. In all cases, the phosphates become more electronegative due to the solvent stabilization of the negative charges, and the ends of the helices are slightly electropositively polarized, most likely due to the net charge stabilization by the solvent. (One might expect that charge distribution in the "gas phase" to be as delocalized as possible because of the net −30 charge of the DNA. The solvent reaction field tends to stabilize localized net charge, leading to a slight buildup of charge in the middle of the helix relative to the gas-phase charge distribution. It should be pointed out that the "gas-phase" DNA is an artificial state, and not meant to be interpreted as a stable species, although it is useful as a reference in discussion of the effects of solvent stabilization.) Recent experiments have shown that monovalent cations selectively partition into the minor groove of B-DNA.[45] A- and Z-forms of DNA are not the forms normally found under physiological conditions; however, double helices of RNA are typically A-form. It has been suggested that the A-DNA and A-RNA are very similar in their electrostatic features.[44] The negatively polarized major groove in the A-DNA calculation is, therefore, consistent with experimental observations[46, 47] and theoretical studies[44, 48] that divalent metal ions tend to bind to the major groove of A-RNA.

Figure 4 shows the solvated electronic polarization map of the Mg^{2+}-bound hammerhead ri-

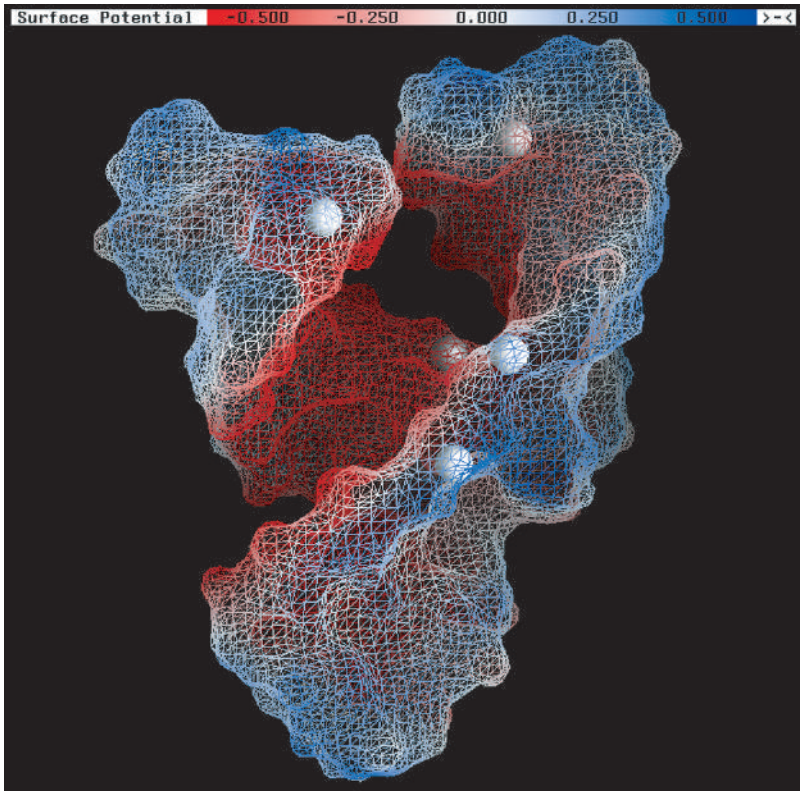

FIGURE 4. The electronic polarization response density of the hammerhead ribozyme in solution. Units are in kT/e ($T = 300$ K). Calculations were performed using the PM3 Hamiltonian[50] and a buffer/matrix cutoff of 8/9 Å.

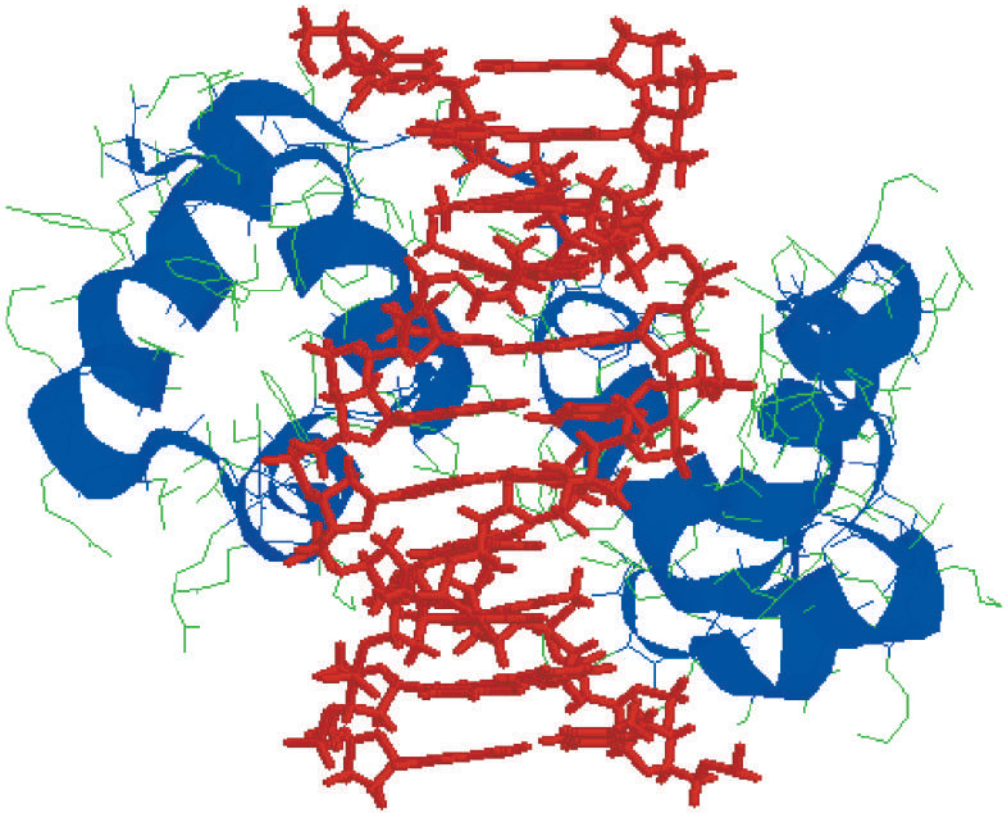

FIGURE 5. The relaxed Myb–DNA complex in solution.[51]

bozyme.[49] The groove becomes considerably more electronegative in potential. The Mg^{2+} binding sites are relatively unaffected in terms of polarization from the gas-phase calculation. Future work will be concentrated on the prediction of metal ion binding sites in RNA.

Another useful property used to characterize macromolecular electronic structure is the electronic density of states. As an example, consider the ionic binding of the Myb DNA binding domain with DNA (Fig. 5). The electronic density of states of the isolated domains and the complex are compared in Figure 6. The states at the edges of the energy gap are most perturbed upon ionic binding of DNA. The DOS immediately below the Fermi level increase, and a slight overall shift of the DOS towards the edge of the gap occurs. The spatial resolution of the molecular orbitals as a function of orbital energy can provide additional insight into the reactivity of macromolecules, and is a focus of future work.

Conclusion

A survey of linear-scaling electronic structure applications to solvated biological macromolecules has been presented. Solvation energies, electrostatic potential surfaces, and electronic density of states have been calculated to assess the role of electronic polarization that occurs in the process of solvation and substrate binding (in the case of the DNA binding domain of Myb). The methods presented here provide a new approach to the calculation of macromolecular electronic properties that will hopefully extend the insight and predictive capability of the more widely used classical methods. In addition to the electrostatic potential, other useful quantities used to characterize macromolecular electronic structure are being pursued including response properties of the density such as linear response functions, local hardness and softness functions, and Fukui functions. It is the hope that further application of these techniques to real biological problems will help to uncover new molecular indices correlated to activity and guide future directions in methodological development.

Acknowledgment

Acknowledgment is made to the Donors of the Petroleum Research Fund, administered by the

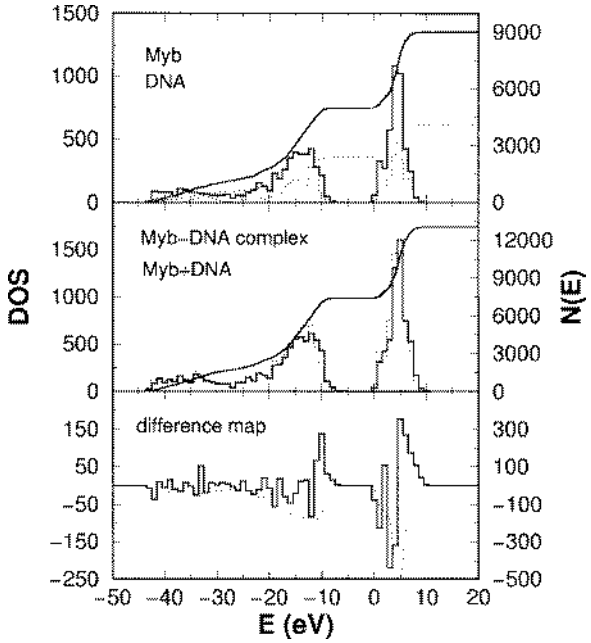

FIGURE 6. The electronic density of states for the complex of Myb with DNA: (1) the isolated molecules (top: Myb solid line, DNA dotted line); (2) complex and superposition of states of the isolated species (middle); and (3) a difference map between the complex and superposition of states (bottom) showing how they are perturbed upon complexation. Calculations were performed using the AM1 Hamiltonian[37] and a buffer/matrix cutoff of 8/9 Å.

American Chemical Society, for partial support of this research. Computational resources were provided by the Minnesota Supercomputing Institute.

References

1. Cramer, C. J.; Truhlar, D. G. Chem Rev 1999, 99, 2161.
2. Gao, J. Rev Comput Chem 1995, 7, 119.
3. Sagui, C.; Darden, T. A. Annu Rev Biophys Biomol Struct 1999, 28, 155.
4. Crengard, L. Science 1994, 265, 909.
5. Crengard, L.; Rokhlin, V. J Comput Phys 1997, 135, 280.
6. Challacombe, M.; White, C.; Head–Gordon, M. J Chem Phys 1997, 107, 10131.
7. Strain, M. C.; Scuseria, G. E.; Frisch, M. J. Science 1996, 271, 51.
8. Hockney, R. W.; Eastwood, J. W. Computer Simulation Using Particles; A. Hilger: New York, 1988.
9. Tomasi, J.; Persico, M. Chem Rev 1994, 94, 2027.
10. York, D. M.; Lee, T.-S.; Yang, W. Chem Phys Lett 1996, 263, 297.
11. York, D. M.; Karplus, M. J Phys Chem A 1999, 103, 11060.
12. Klamt, A.; Schüürmann, G. J Chem Soc Perkin Trans 1993, 2, 799.
13. Klamt, A. J Phys Chem 1995, 99, 2224.
14. Klamt, A.; Jonas, V.; Bürger, T.; Lohrenz, J. C. W. J Phys Chem A 1998, 102, 5074.
15. Chipman, D. M. J Chem Phys 1997, 106, 10194.
16. Perez–Jordá, J. M.; Yang, W. J Chem Phys 1997, 107, 1218.
17. Goedecker, S. Rev Mod Phys 1999, 71, 1085.
18. Scuseria, G. E. J Phys Chem A 1999, 103, 4782.
19. Head–Gordon, M. J Phys Chem 1996, 100, 13213.
20. Yang, W.; Lee, T.-S. J Chem Phys 1995, 103, 5674.
21. Dixon, S. L.; Merz, K. M., Jr. J Chem Phys 1996, 104, 6643.
22. Stewart, J. J. P. Int J Quant Chem 1996, 58, 133.
23. York, D.; Lee, T.-S.; Yang, W. J Am Chem Soc 1996, 118, 10940.
24. York, D. M. In Combined Quantum Mechanical and Molecular Mechanical Methods; Gao, J.; Thompson, M., Eds.; ACS Symposium Series 712; Oxford University Press: New York, 1998, p. 255, chapt. 18.
25. Gogonea, V.; Merz, K. M., Jr. J Phys Chem A 1999, 103, 5171.
26. Nadig, G.; Van Vleet, L. C. W.; Dixon, S. L.; Merz, K. M., Jr. J Am Chem Soc 1998, 120, 5593.
27. Gogonea, V.; Merz, K. M., Jr. J Chem Phys 2000, 112, 3225.
28. York, D. M.; Lee, T.-S.; Yang, W. Phys Rev Lett 1998, 80, 5011.
29. Lewis, J. P.; Carter, C. W., Jr.; Hermans, J.; Pan, W.; Lee, T.-S.; Yang, W. J Am Chem Soc 1998, 120, 5405.
30. Lewis, J.; Liu, S.; Lee, T.-S.; Yang, W. J Comput Phys 1999, 151, 242.
31. Titmuss, S. J.; Cummins, P. L.; Bliznyuk, A. A.; Rendell, A. P.; Gready, J. E. Chem Phys Lett 2000, 320, 169.
32. Greatbanks, S. P.; Gready, J. E.; Limaye, A. C.; Rendell, A. P. J Comput Chem 2000, 21, 788.
33. Stewart, J. J. P. Rev Comput Chem 1990, 1, 45.
34. Thiel, W. In Advances in Chemistry Physics; Prigogine, I.; Rice, S. A., Eds.; John Wiley and Sons: New York, 1996, p. 703, vol. 93.
35. Krezel, A. M.; Darba, P.; Robertson, A. J Mol Biol 1994, 242, 203.
36. Lee, T.-S.; York, D. M.; Yang, W. J Chem Phys 1996, 105, 2744.
37. Dewar, M. J. S.; Zoebisch, E.; Healy, E. F.; Stewart, J. J. P. J Am Chem Soc 1985, 107, 3902.
38. Bagdassarian, C. K.; Schramm, V. L.; Schwartz, S. D. J Am Chem Soc 1996, 118, 8825.
39. Sharp, K. A.; Honig, B. Annu Rev Biophys Chem 1990, 19, 301.
40. Honig, B.; Nicholls, A. Science 1995, 268, 1144.
41. Gunner, M. R.; Nicholls, A.; Honig, B. J Phys Chem 1996, 100, 4277.
42. Ullmann, G. M.; Hauswald, M.; Jensen, A.; Kostić, N. M.; Knapp, E.-W. Biochemistry 1997, 36, 16187.
43. Gilson, M. K. Curr Opin Struct Biol 1995, 5, 216.
44. Chin, K.; Sharp, K. A.; Honig, B.; Pyle, A. M. Nat Struct Biol 1999, 6, 1055.
45. McFail–Isom, L.; Sines, C. C.; Williams, L. D. Curr Opin Struct Biol 1999, 9, 298.

46. Cate, J. H.; Doudna, J. A. Structure 1996, 4, 1221.
47. Kieft, J. S.; Tinoco, I., Jr. Structure 1997, 5, 713.
48. Le, S.-Y.; Chen, J.-H.; Maizel, N. P. J., Jr. J Biomol Struct Dyn 1998, 6, 1.
49. Scott, W. G.; Murray, J. B.; Arnold, J. R. P.; Stoddard, B. L.; Klug, A. Science 1996, 274, 2065.
50. Stewart, J. J. P. J Comput Chem 1989, 10, 221.
51. Ogata, K. Cell 1994, 79, 639.