

Excerpts from an upcoming paper

Calculation of Molecular Geometries, Relative Conformational Energies, Dipole Moments, and Molecular Electrostatic Potential Fitted Charges of Small Organic Molecules of Biochemical Interest by Density Functional Theory

Alain St-Amant^{*†}, Wendy D. Cornell^{*⊥}, Thomas A. Halgren[†], and Peter A. Kollman^{*}

^{*} *Department of Pharmaceutical Chemistry, University of California, San Francisco, San Francisco, California 94143-0446*

[⊥] *Graduate Group in Biophysics*

[†] *Merck Research Laboratories, Rahway, New Jersey 07065*

[‡] *Present Address: Department of Chemistry, University of Ottawa, 10 Marie Curie, Ottawa, Ontario, Canada K1N 6N5*

Methods

Density functional theory is based on the Hohenberg-Kohn Theorem [1] which states that all ground state properties of a system are uniquely determined by the electronic density, $\rho(\mathbf{r})$, and that any other density, $\rho'(\mathbf{r})$, will give rise to a higher energy. Unfortunately, the precise nature of the energy functional was not known and a route towards performing accurate *ab initio* quantum mechanical calculations using the electronic density as the basic variable was not proposed. However, a set of working equations was proposed by Kohn and Sham [2]. Following Kohn and Sham, the

energy functional, $E[\rho(\mathbf{r})]$, is partitioned in the following manner,

$$E[\rho(\mathbf{r})] = T[\rho(\mathbf{r})] + U[\rho(\mathbf{r})] + E_{xc}[\rho(\mathbf{r})]$$

where $T[\rho(\mathbf{r})]$ is the kinetic energy of a system of non-interacting particles of the same density, $\rho(\mathbf{r})$; $U[\rho(\mathbf{r})]$ is the classical electrostatic energy; and $E_{xc}[\rho(\mathbf{r})]$ is the sum of the exchange and correlation (XC) energies. If $\rho(\mathbf{r})$ is expressed as the sum of one-electron orbital densities,

$$\rho(\mathbf{r}) = \sum_i^n |\psi_i(\mathbf{r})|^2$$

where n is the number of singly occupied Kohn-Sham spin orbitals, $\{\psi_i(\mathbf{r})\}$, then the first two terms of the energy functional may be written out as

$$T[\rho(\mathbf{r})] = \sum_i^n \int d\mathbf{r} \psi_i(\mathbf{r}) \frac{-\nabla^2}{2} \psi_i(\mathbf{r})$$

and

$$U[\rho(\mathbf{r})] = \sum_i^n \int d\mathbf{r} \psi_i(\mathbf{r}) \left(\sum_A^N \frac{-Z_A}{|\mathbf{r} - \mathbf{R}_A|} \right) \psi_i(\mathbf{r}) + \frac{1}{2} \sum_i^N \sum_j^N \int d\mathbf{r} d\mathbf{r}' \frac{\psi_i(\mathbf{r}) \psi_i(\mathbf{r}) \psi_j(\mathbf{r}') \psi_j(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + \sum_A^N \sum_{B < A}^N \frac{Z_A Z_B}{|\mathbf{R}_A - \mathbf{R}_B|} \quad (1)$$

where $\{Z_A\}$ and $\{R_A\}$ are the charges and positions of the N atomic nuclei.

The Kohn-Sham orbitals are obtained by solving the Kohn-Sham equations (which are obtained by stating that the energy is stationary with respect to any arbitrary change in the density, $\delta\rho(\mathbf{r})$),

$$\left\{ \frac{-\nabla^2}{2} + \sum_A \frac{-Z_A}{|\mathbf{r} - \mathbf{R}_A|} + \int d\mathbf{r}' \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + \frac{\delta E_{xc}[\rho(\mathbf{r})]}{\delta\rho(\mathbf{r})} \right\} \psi_i(\mathbf{r}) = \varepsilon_i \psi_i(\mathbf{r})$$

where $\{\varepsilon_i\}$ are the set of eigenvalues of the Kohn-Sham equations (they *do not* carry the same Koopman's Theorem significance as do their Hartree-Fock counterparts). In principle, the Kohn-Sham approach can furnish results of arbitrary accuracy. Unfortunately, the nature of the XC energy, $E_{xc}[\rho(\mathbf{r})]$, and potential,

$$v_{xc}(\mathbf{r}) = \frac{\delta E_{xc}[\rho(\mathbf{r})]}{\delta\rho(\mathbf{r})}$$

remains unknown. Approximations to $E_{xc}[\rho(\mathbf{r})]$ must be made if the Kohn-Sham approach is to be of any practical use.

In practice, $E_{xc}[\rho(\mathbf{r})]$ is expressed as

$$E_{xc}[\rho(\mathbf{r})] = \int d\mathbf{r} \rho(\mathbf{r}) \epsilon_{xc}(\mathbf{r})$$

where $\epsilon_{xc}(\mathbf{r})$ is the XC energy density. In the LSDA, $\epsilon_{xc}(\mathbf{r})$ is but a function of the total electronic density and the net spin density. The Vosko–Wilk–Nusair (VWN) parametrization [3] of Monte Carlo calculations by Ceperley and Alder [4] on a series of homogeneous electron gases of varying total and spin densities is considered to be the “limit” of the LSDA. We have used the VWN functional in our calculations. Though one may imagine that such an approximation is severe, results obtained from geometry and vibrational analyses [5] have shown that the LSDA compares very well to Hartree-Fock and even correlated post-HF techniques, especially for systems where Hartree-Fock is notoriously bad, i.e., transition metal systems. There are limitations to the LSDA, however. Systematic overestimation of bond strengths is undoubtedly the most disturbing [6]. This systematic overestimation can begin to provide qualitatively bad descriptions of systems with weak bonds — most importantly hydrogen-bonded systems that are critical if DFT is to be used in biomolecular modelling. It has been shown that the so-called nonlocal functionals (they remain, however, true local functionals in that they depend only on local properties of $\rho(\mathbf{r})$) patch up in large part these deficiencies of the LSDA. With nonlocal functionals, $\epsilon_{xc}(\mathbf{r})$ becomes a function of not only $\rho^\alpha(\mathbf{r})$ and $\rho^\beta(\mathbf{r})$, but also their first derivatives, $\nabla\rho^\alpha(\mathbf{r})$ and $\nabla\rho^\beta(\mathbf{r})$. For this study, we have chosen to use Becke’s nonlocal exchange functional [7] and Perdew’s nonlocal correlation functional [8].

For our particular calculations, we have chosen to use the Linear Combination of Gaussian Type Orbitals-Density Functional Theory (LCGTO-DFT) method [9, 10, 11]. In this approach, the Kohn-Sham orbitals are expressed within a gaussian basis, $\{\chi_\mu(\mathbf{r})\}$,

$$\psi_i(\mathbf{r}) = \sum_{\mu} C_{\mu i} \chi_{\mu}(\mathbf{r})$$

as are the Hartree-Fock molecular orbitals in the conventional LCGTO-HF method. In the LCGTO approach, the total density is given by

$$\rho(\mathbf{r}) = \sum_{\mu\nu} P_{\mu\nu} \chi_{\mu}(\mathbf{r}) \chi_{\nu}(\mathbf{r})$$

where the density matrix elements, $\{P_{\mu\nu}\}$, are given by

$$P_{\mu\nu} = \sum_i^n C_{\mu i} C_{\nu i} \ .$$

The computational savings of the DFT LCGTO variant are obtained by reducing the four-centered two-electron integrals of equation (1) to three-centered integrals by fitting $\rho(\mathbf{r}')$ with a second set of gaussians, $\{\chi'_k(\mathbf{r})\}$, in an analytical, anti-variational (the true total energy becomes an upper bound) least-squares fitting (LSF) procedure [10],

$$\sum_i^n |\psi_i(\mathbf{r}')|^2 \approx \sum_k c_k \chi'_k(\mathbf{r}') \ .$$

In an analogous fashion, the XC energy density and potentials are fit by a common third set of gaussians, $\{\chi_l''(\mathbf{r})\}$, in a numerical, non-variational LSF procedure over a set of atom-centered grid points [10],

$$\begin{aligned}\epsilon_{xc}(\mathbf{r}) &= \sum_l e_l \chi_l''(\mathbf{r}) \\ v_{xc}(\mathbf{r}) &= \sum_l p_l \chi_l''(\mathbf{r}) \quad (\sigma = \alpha, \beta) \quad .\end{aligned}$$

The use of these two auxiliary fitting bases reduces integral evaluation from a scaling of N^4 to N^2M , where N is the number of orbital basis functions and M is the sum total of the density and XC fitting bases (which is of the same order as N).

As first suggested by the original developers of the LCGTO-DFT methodology, the Kohn-Sham matrix elements are analytically integrated once the LSF procedures are performed and are given by

$$\begin{aligned}F_{\mu\nu}^\sigma &= \int d\mathbf{r} \chi_\mu(\mathbf{r}) \left\{ \frac{-\nabla^2}{2} + \sum_A \frac{-Z_A}{|\mathbf{r} - \mathbf{R}_A|} \right\} \chi_\nu(\mathbf{r}) \\ &+ \sum_k c_k \int d\mathbf{r} d\mathbf{r}' \frac{\chi_\mu(\mathbf{r}) \chi_\nu(\mathbf{r}) \chi_k'(\mathbf{r})}{|\mathbf{r} - \mathbf{r}'|} + \sum_l p_l^\sigma \int d\mathbf{r} \chi_\mu(\mathbf{r}) \chi_\nu(\mathbf{r}) \chi_l''(\mathbf{r}) \quad .\end{aligned}$$

For the evaluation of the total energy, we have, however, deviated from the traditional expression [10],

$$\begin{aligned}E &= \sum_\mu \sum_\nu P_{\mu\nu} \int d\mathbf{r} \chi_\mu(\mathbf{r}) \left\{ \frac{-\nabla^2}{2} + \sum_A \frac{-Z_A}{|\mathbf{r} - \mathbf{R}_A|} \right\} \chi_\nu(\mathbf{r}) \\ &+ \sum_\mu \sum_\nu P_{\mu\nu} \sum_k c_k \int d\mathbf{r} d\mathbf{r}' \frac{\chi_\mu(\mathbf{r}) \chi_\nu(\mathbf{r}) \chi_k'(\mathbf{r})}{|\mathbf{r} - \mathbf{r}'|} - \frac{1}{2} \sum_k \sum_{k'} c_k c_{k'} \int d\mathbf{r} d\mathbf{r}' \frac{\chi_k'(\mathbf{r}) \chi_{k'}'(\mathbf{r})}{|\mathbf{r} - \mathbf{r}'|} \\ &+ \sum_A \sum_{B < A} \frac{Z_A Z_B}{|\mathbf{R}_A - \mathbf{R}_B|} + \sum_\mu \sum_\nu P_{\mu\nu} \sum_l e_l \int d\mathbf{r} \chi_\mu(\mathbf{r}) \chi_\nu(\mathbf{r}) \chi_l''(\mathbf{r}) \quad ,\end{aligned}$$

and opted to integrate the final term, which involves the XC energy density, numerically

$$\sum_\mu \sum_\nu P_{\mu\nu} \sum_I W_I \chi_\mu(\mathbf{R}_I) \chi_\nu(\mathbf{R}_I) \epsilon_{xc}(\mathbf{R}_I)$$

over an augmented ensemble of grid points with weights, $\{W_I\}$, and positions, $\{\mathbf{R}_I\}$. Typically, the grids used in the LSF procedures during the SCF contain approximately 800 points per atom. Once the SCF has been achieved, over 2500 points per atom are used for the numerically integrated contribution to the total energy. With such a strategy, we maintain economy within the SCF procedure by keeping the grid size moderate while achieving greater numerical stability in the final total energy, at

very little expense. In the same vein, the energy gradient expression of Fournier *et al.* [12] has been modified so that we numerically integrate the XC term of the gradient on the same augmented grid,

$$\begin{aligned}
E^{(\lambda)} = & \sum_{\mu} \sum_{\nu} P_{\mu\nu} \left(\left\langle \chi_{\mu} \left| \frac{-\nabla^2}{2} + \sum_A \frac{-Z_A}{|\mathbf{r} - \mathbf{R}_A|} \right| \chi_{\nu} \right\rangle^{(\lambda)} + \sum_k c_k \left[\chi_{\mu} \chi_{\nu} \left| \frac{1}{r_{12}} \right| \chi_k' \right]^{(\lambda)} \right) \\
& - \sum_k \sum_{k'} c_k c_{k'} \left[\chi_k' \left| \frac{1}{r_{12}} \right| \chi_{k'}' \right]^{(\lambda)} - \sum_{\mu\nu} W_{\mu\nu} \langle \chi_{\mu} || \chi_{\nu} \rangle^{(\lambda)} + \left(\sum_A \sum_{B < A} \frac{Z_A Z_B}{|\mathbf{R}_A - \mathbf{R}_B|} \right)^{(\lambda)} \\
& + \sum_{\sigma=\alpha,\beta} \sum_{\mu} \sum_{\nu} P_{\mu\nu}^{\sigma} \sum_I W_I \left\{ \chi_{\mu}^{(\lambda)}(\mathbf{R}_I) \chi_{\nu}(\mathbf{R}_I) v_{xc}^{\sigma}(\mathbf{R}_I) + \chi_{\mu}(\mathbf{R}_I) \chi_{\nu}^{(\lambda)}(\mathbf{R}_I) v_{xc}^{\sigma}(\mathbf{R}_I) \right\}
\end{aligned}$$

where the superscript (λ) denotes differentiation with respect to a nuclear coordinate, $P_{\mu\nu}^{\alpha}$ and $P_{\mu\nu}^{\beta}$ are the density matrix elements for each electronic spin, and $W_{\mu\nu}$ is an orbital energy weighted density matrix element given by

$$W_{\mu\nu} = \sum_i^n \varepsilon_i C_{\mu i} C_{\nu i} .$$

By numerically integrating the XC terms, the approximate gradient achieves greater accuracy than previously possible with the analytical approach [13].

Computational Details

The DFT calculations were performed with a modified version of the deMon LCGTO-DFT program originally developed by St-Amant and Salahub [13].

For the DFT calculations, double- ζ plus polarization and triple- ζ plus polarization orbital basis sets were placed, respectively, on the hydrogen and heavy atoms. Using Huzinaga's notation [14], the H atoms had a (41/1*) contraction pattern, the C, N, O, and F atoms had a (7111/411/1*) contraction pattern, and the S and Cl atoms had a (73111/6111/1*) contraction pattern. Auxiliary charge density and XC basis sets consisting of five s functions and four sets of s , p , and d functions constrained to the same exponent were placed on the S and Cl centers (such an auxiliary basis is denoted as (5,4;5,4)) [5]. Similarly, (4,4;4,4) auxiliary basis sets were used for the C, N, O, and F atoms while a (3,1;3,1) auxiliary basis was used for H atoms. These orbital and auxiliary basis sets were optimized specifically for LSDA calculations by Godbout *et al.* [15].

All geometry optimizations were carried out with nonlocal corrections added perturbatively to the energy and the gradient. In this approach, SCF calculations are carried out within the LSDA and the first and second partial derivatives of the resultant LSDA SCF density are used to evaluate the nonlocal XC contributions to the energy and gradients. This method is thus only marginally

more expensive than LSDA calculations. The use of such perturbative approaches is common in DFT applications [16]. It is not clear, however, that such an approach will work in the LCGTO formalism since the approximate analytical expression for the gradient was derived assuming that the XC potentials which appear in it were obtained in a self-consistent fashion [12]. As a test of this perturbative approach to calculating forces, we performed calculations on an intramolecularly hydrogen-bonded system, malonaldehyde. Malonaldehyde was chosen, since as previously shown [17], the LSDA provides a structure that is qualitatively wrong while the NLSDA goes a long way towards providing a quantitatively accurate description of the system. As seen in Table I, the results obtained by this perturbative approach, which we will denote as (N)LSDA, essentially recoupe all the nonlocal corrections brought to the structure of malonaldehyde by more time-consuming fully self-consistent NLSDA calculations. We feel that the perturbative approach to adding nonlocal corrections upon the LSDA is a satisfactory compromise between computational efficiency and theoretical rigor.

Acknowledgements

ASA thanks the Natural Sciences and Engineering Research Council (NSERC) of Canada for a postdoctoral fellowship. We are also grateful for support from the NIH: GM-29072 (PAK) and GM-08284 (WDC).

References

- [1] P. Hohenberg and W. Kohn, *Phys. Rev. B* **136**, 864 (1964).
- [2] W. Kohn and L. J. Sham, *Phys. Rev. A* **140**, 1133 (1965).
- [3] S. H. Vosko, L. Wilk, and M. Nusair, *Can. J. Phys.* **58**, 1200 (1980).
- [4] D. M. Ceperley and B. J. Alder, *Phys. Rev. Lett.* **45**, 566 (1980).
- [5] I. Papai, A. St-Amant, J. Ushio, and D. R. Salahub, *Int. J. Quantum Chem.* **S24**, 29 (1990).
- [6] A. D. Becke in *The Challenge of d and f Electrons*, D. R. Salahub and M. C. Zerner, Eds., *ACS Symp. Ser.* **394**, 165 (1989).
- [7] A. D. Becke, *Phys. Rev. A* **38**, 3098 (1988).
- [8] J. P. Perdew, *Phys. Rev. B* **33**, 8822 (1986).
- [9] H. Sambe and R. H. Felton, *J. Chem. Phys.* **62**, 1122 (1975).
- [10] B. I. Dunlap, J. W. D. Connolly, and J. R. Sabin, *J. Chem. Phys.* **71**, 3396, 4993 (1979).
- [11] J. Andzelm and E. Wimmer, *J. Chem. Phys.* **96**, 1280 (1992).
- [12] R. Fournier, J. Andzelm, and D. R. Salahub, *J. Chem. Phys.* **90**, 6371 (1989).
- [13] A. St-Amant, Ph. D. thesis, Université de Montréal (1992).
- [14] S. Huzinaga, Ed., *Gaussian Basis Sets for Molecular Calculations* (Elsevier, New York, 1984).
- [15] N. Godbout, D. R. Salahub, J. Andzelm, and E. Wimmer, *Can. J. Chem.* **70**, 560 (1992) (for H, C, N, O, and F); N. Godbout, D. R. Salahub, J. Andzelm, and E. Wimmer, private communication (for S).
- [16] L. Fan and T. Ziegler, *J. Chem. Phys.* **94**, 6057 (1991).
- [17] F. Sim, A. St-Amant, I. Papai, and D. R. Salahub, *J. Amer. Chem. Soc.* **114**, 4391 (1992).