Supporting Information

THEORY

The typical approach to quantum chemistry in undergraduate textbooks is the Hartree-Fock-Roothaan(HFR) methodology, which relies on the principle of linear combination of atomic orbitals (LCAO)[1].

By using the time-independent Schrödinger equation for the ground state of an n-electron closed-shell molecule, a wave function is sought for, as the following Slater determinant:

$$\psi(q_1, q_2 \cdots, q_n) = (n!)^{-\frac{1}{2}} \det \left| \varphi_i^{\alpha}(1) \varphi_i^{\beta}(2) \varphi_j^{\alpha}(3) \cdots \varphi_{\frac{n}{2}}^{\beta}(n) \right| \tag{S1}$$

The spatial component of each molecular spin-orbital, φ_i, was developed by Roothaan as a linear combination of m atomic orbitals, χ_μ, centered in the nuclei:

$$\varphi_i = \sum_{\mu}^{m} c_{i\mu} \chi_\mu \tag{S2}$$

The correct wave function, i.e. the set of coefficients $c_{i\mu}$, is obtained variationally by minimizing the total electronic energy, which leads to the HFR equations

$$\sum_{v}^{m} c_{i\mu} \left(F_{\mu v} - E_i S_{\mu v} \right) = 0 \tag{S3}$$

where $S_{\mu v}$ is the overlap matrix, and $F_{\mu v}$ is the Fock matrix with elements:

$$F_{\mu v} = \hat{H}_{\mu v}^c + \sum_{\lambda}^{m} \sum_{\sigma}^{m} P_{\lambda \sigma} \left[(\mu v | \lambda \sigma) - \frac{1}{2} (\mu \lambda | v \sigma) \right] \tag{S4}$$

and $P_{\lambda \sigma}$ is the density matrix,

$$P_{\lambda \sigma} = 2 \sum_{j=1}^{n} c_{j\lambda}^* c_{j\sigma} \tag{S5}$$

The matrix elements $\hat{H}_{\mu v}^c$ of the core Hamiltonian correspond to the kinetic and potential energy terms of the electrons in the field of the nuclei. The last two

terms of equation (S4) are the two-electron integrals, which represent the *Coulomb* energy of repulsion and exchange energy:

$$(\mu v \mid \lambda \sigma) = \int\limits_{(2)} \int\limits_{(1)} \chi_\mu(1)\chi_v(1)\frac{1}{r_{1,2}}\chi_\lambda(2)\chi_\sigma(2)\,d\tau_1\,d\tau_2 \tag{S6}$$

Solving equations (S3) is to get the set of coefficients $c_{i\mu}$ and the corresponding energies E_i. In order to accomplish that task, one needs to compute the Fock matrix elements (S4), which depend on the set of linear combination coefficients of the density matrix (S5). However, the set of linear coefficients is precisely what is supposed to be found by solving the HFR equations. This sort of challenging mathematical problem can be solved iteratively, starting with an initial guess of coefficients and hoping the process converges to final improved values, in which case it is said that a self-consistent field (SCF) is achieved. Modern quantum chemical program packages perform the whole sequence of HFR calculations automatically, which is one of the reasons of the actual popularity of computational chemistry tools, often used as miraculous 'black boxes'. The user just has to provide the geometry of the molecular system as input, whose format is easily obtained with proper visualization programs, and a few additional parameters or keywords. Nevertheless, the initial choice of the basis set is one of the critical issues as it determines the quality of the final results. On the other hand, the computational effort increases approximately with the fourth power of the total number of basis functions, due to the calculation of the two-electron integrals (S6). Many of the integrals might be considered negligible, or might even vanish by symmetry, but those that are left to evaluate make the calculations difficult for medium size systems. Some integrals involve four different basis functions centered on different atomic nuclei. Since the decision of the type of basis set to use in a calculation is left to the user, it is important to improve our knowledge concerning the several choices we have.

The basis of functions commonly used in LCAO quantum calculations fall into either one of two classes:

Slater-type orbitals (STOs)

$$\chi_{n,l,m}^{STO} = N\, Y_{l,m}(\theta,\phi) r^{n-1} \exp(-\zeta r) \qquad\qquad (S7)$$

These functions are characterized by quantum numbers n, l and m, and the exponent ζ. N is the normalization constant and $Y_{l,m}(\theta,\phi)$ are the spherical harmonics. Slater suggested this type of functions by replacing the radial part of the exact hydrogen atom functions by nodeless and simpler functions of similar appearance[2]. Note that if $n=1$ the resulting STO is just the exact hydrogenlike *1s* orbital. Although STOs work well with small basis sets, they lead to numerical problems during integration as the number of basis increases.

Gaussian-type orbitals (GTOs)

In 1950, Boys[3] suggested the use of Gaussian-type basis functions, which represents a major advance in the evaluation of two-electron integrals. Those functions can be expressed as the following general form,

$$\chi_{n,l,m}^{GTO} = N\, Y_{l,m}(\theta,\phi) r^{2n-2-l} \exp(-\alpha r^2) \qquad\qquad (S8)$$

where the orbital exponent α is positive. The GTOs form a complete basis set, but they are not mutually orthogonal; the same happens with the STO functions[1].

Distinct features of GTOs and STOs

The dependence of the exponential on r^2 makes the GTO less accurate than the STOs in what concerns the electron description in regions close to and far away from the nucleus. In fact, the correct *cusp* (discontinuous derivative) at the nucleus, shown by the hydrogenlike wavefunctions, is well reproduced by STOs, but is dramatically absent in GTOs, where a continuous zero slope at that point is observed instead. In addition, GTOs fall off too rapidly far from the nucleus, and

those far regions of the wavefunctions are consequently poorly represented as well.

Why GTO rather than STO?

The mathematical evaluation of the two-electron integrals (S6), specially the three- and four-centered ones, is quite cumbersome if STO functions are used. However, a great simplification is obtained with GTOs, due to the fact that a product of two GTOs is still a GTO. For the sake of simplicity we can illustrate this property with one-dimension functions, but the extrapolation to three-dimensions is straightforward. Consider two one-dimension GTOs centered at x_A and x_B, respectively:

$$GTO_A(x) = \exp\left(-\alpha_A\left(x-x_A\right)^2\right)$$
$$GTO_B(x) = \exp\left(-\alpha_B\left(x-x_B\right)^2\right)$$

(S9)

The product $GTO_A(x) \times GTO_B(x)$ is another gaussian function, $GTO_P(x)$:

$$GTO_P(x) = \exp\left[-\frac{\alpha_A\alpha_B}{\alpha_A+\alpha_B}\left(x_A-x_B\right)^2\right]\exp\left[-\left(\alpha_A+\alpha_B\right)\left(x-x_P\right)^2\right]$$

(S10)

with exponent $\alpha_A+\alpha_B$ and centered at the point x_P between A and B given by:

$$x_P = \frac{\alpha_A x_A + \alpha_B x_B}{\alpha_A+\alpha_B}$$

(S11)

We can easily extend the product for more than two GTOs by repeated application of this procedure.

From the exposed above it is evident the usefulness of GTOs in the evaluation of two-electron integrals (S6), where four-center integrals can be reduced at most to two-center integrals, which are much easier to handle.

Improving with GTO

It is possible to use GTOs in quantum calculations and still get results with high accuracy. In fact, the idea of representing the STOs atomic orbitals by linear combinations of GTOs was suggested earlier by several authors[3-5], but Pople *et*

al.[6] finally proposed a systematic usage of fixed linear combinations (contracted functions) up to K=6 GTOs, obtained by least-squares methods and applicable to different molecular systems:

$$\chi_\mu^{STO} = \sum_i^K c_i \, \chi^{GTO}(\alpha_i) \tag{S12}$$

The self-consistent HFR values of atomization energies, atomic populations and electric dipole moments of some atoms and small molecules, revealed to converge rapidly to the values corresponding to pure STOs. The bigger the size of the expansion (S12), higher K, the better the accuracy of the approach.

The GTOs are called *primitives* and the resulting combination a *contracted* function. The notation used by Pople *et al.* for the *minimum basis set* (*i.e.* the smallest basis comprising one single function for each atomic orbital of the occupied shells) is STO-KG. The authors found that a good compromise between time of computation and accuracy was achieved by STO-3G, which justifies the popularity that this basis set has reached in the past. Although STO-NG basis sets are rarely used nowadays in scientific research, their historical importance, simple conception and reasonable performance in calculations makes them almost inescapable as a topic in a computational chemistry course.

Another advantage of the use of *contracted* functions is the fact that contraction coefficients c_i in equation (S12) can be chosen to have either a negative or a positive sign, thus fitting to atomic orbitals with higher principal quantum numbers which possess radial nodes.

The simpler case of STO-2G

The spreadsheet application *2-fold-secular determinant* finds the best contracted function for K=2 in terms of energy. The variation theory procedure is thus applied to the following trial function,

$$\chi_{trial} = c_1 \, \chi_{1s}^{GTO}(\alpha_1) + c_2 \, \chi_{1s}^{GTO}(\alpha_2) \tag{S13}$$

to which the normalization condition is imposed:

$$\int \chi_{trial}^2 \, d\tau = c_1^2 + c_2^2 + 2 c_1 c_2 S_{12} = 1 \tag{S14}$$

The minimization of the energy expression that results from the corresponding Rayleigh ratio leads to the following secular equations,

$$\begin{cases} c_1\left(H_{11}-E\right)+c_2\left(H_{12}-ES_{12}\right)=0 \\ c_1\left(H_{21}-ES_{21}\right)+c_2\left(H_{22}-E\right)=0 \end{cases} \tag{S15}$$

where the elements of the secular matrix are defined as:

$$H_{ii}=\frac{3}{2}\alpha_i-2\sqrt{\frac{2}{\pi}}\alpha_i \quad , \quad i=1,2 \tag{S16}$$

$$H_{12}=H_{21}=\frac{\sqrt[4]{4\alpha_1^3\alpha_2^3}}{\left(\alpha_1+\alpha_2\right)}\left[\frac{6\alpha_2}{\sqrt{\alpha_1+\alpha_2}}\left(1-\frac{\alpha_2}{\alpha_1+\alpha_2}\right)-\frac{4}{\sqrt{\pi}}\right] \tag{S17}$$

$$S_{12}=2\sqrt{2}\left[\frac{\alpha_1\alpha_2}{\left(\alpha_1+\alpha_2\right)^2}\right]^{\frac{3}{4}} \tag{S18}$$

If the set of equations (S15) is to have non-trivial solutions (i.e. $c_1 \neq 0$ and $c_2 \neq 0$) the determinant of the coefficients c_i must vanish, that is:

$$\det\left|H_{ij}-ES_{ij}\right|=0 \tag{S19}$$

A quadratic equation results from the above condition, whose lowest root corresponds to the minimum energy one can obtain with trial function (S13).

REFERENCES

(1) (a) *Molecular Quantum Mechanics*, Peter Atkins and Ronald Friedman, 5th ed., Oxford Univ. Press, Oxford, **2011**. (b) *Essentials of Computational Chemistry- Theories and Models*, Cristopher J. Cramer, 2nd ed., John Wiley & Sons, Chichester, **2004**.

(2) Slater, J.C. Atomic Shielding Constants. *Phys. Rev.* **1930**, 36, 57-64.

(3) Boys S.F. Electronic Wave Functions.1. A General Method of Calculation for the Stationary States of Any Molecular System. *Proc. R. Soc. Lond. A* **1950**, 200, 542-554.

(4) Foster M., Boys S.F. A Quantum Variational Calculation for HCHO. *Rev. Mod. Phys.* **1960**, 32, 303-304.

(5) Reeves C.M., Fletcher R. Use of Gaussian Functions in the Calculation of Wavefunctions for Small Molecules. III The Orbital Basis and Its Effect on Valence. *J. Chem. Phys.* **1965**,42, 4073-4081.

(6) Hehre W.J., Stewart R.F., Pople J.A. Self-Consistent Molecular Orbital Methods. I. Use of Gaussian Expansions of Slater-Type Atomic Orbitals. *J. Chem. Phys.* **1969**, 51, 2657-2664.