

CHEMISTRY 4021/8021
MIDTERM EXAM KEY – SPRING 2011

1. Molecular Mechanics and Semiempirical Molecular Orbital Theory (100 points)

We have read two papers this semester addressing the development of new parameters for a (semi)empirical model: Nicholas et al. presented zeolite force-field development and Anders et al. presented the extension of PM3 to include lithium. For this problem, compare and contrast parameter development in a molecular mechanics model vs. a semiempirical molecular orbital theory model. Thus, what similarities and differences are there in terms of numbers and kinds of parameters needed, means for determination/optimization of such parameters, kinds of validation data, etc.?

Extension of a molecular mechanics (MM) or semiempirical molecular orbital (SEMO) model fundamentally implies the addition of a new atom type to the model. One difference between MM and SEMO models is that the former may have many different atomic types for a given atomic number (e.g., different kinds of oxygen atoms typed by functional group) while SEMO models identify atoms simply by atomic number.

Generally speaking, far more parameters are needed to extend an MM model than to extend a SEMO model. This follows from the means by which energy is computed. An MM model needs to compute stretching, bending, torsional, and non-bonded energies, and these terms involve the specification of atom-type-pair specific parameters (for bond stretching and non-bonded interactions), atom-type-triple specific parameters (for angle bending), and atom-type-quadruple specific parameters (for torsions). Thus, each new atom in the force field introduces a number of new parameters in principle proportional to the 4th power of the number of total atom types. Modern SEMO models, on the other hand, are typically defined by only a handful of atom-specific parameters (on the order of 10) so that each new atom introduced is no more, nor less, work than any other. The nature of the SEMO parameters includes, inter alia, same-center one-electron integrals, different-center one-electron integrals, same-center two electron integrals, orbital exponents for the computation of overlap.

In any exercise of parameter optimization, one must define an error function (involving a quantitative weighted comparison of computed to target data over some specified data set). In the case of MM models, the data available tend to be limited to those most closely tied to the potential energy surface (PES). Thus, structural data (e.g., bond lengths, angles, etc.) and vibrational frequencies (related to the steepness of PES coordinates are useful. Energetic data are typically restricted to conformational energy changes, as complications arise when comparing molecules not composed of identical atom types. In SEMO models, on the other hand, their quantum mechanical nature permits expansion of validation data to include any expectation value of the wave function, e.g., the

dipole moment. In addition, the energetic analysis can go beyond conformational comparisons to include atomization energies (typically compared to experimental atomization enthalpies or, after addition of atomic heats of formation, to molecular heats of formation) and transition-state energies. Furthermore, the SEMO model can consider the new atom in different charge or spin states (computing ionization potentials, for instance), which is not the case for an MM model.

The sources of the validation data for both MM and SEMO parameterization efforts tend to be the same. Experimental data are ideal (when their accuracy is assured) and supplemental data derived from high-level electronic structure calculations can be useful to supplement gaps in experimental data. The time required to carry out parameterization is likely to be greater for the SEMO model than for the MM model, only because repeated quantum mechanical calculations will be required for the former while the latter requires only very fast force field calculations.

2. Ab Initio Hartree-Fock (HF) and post-Hartree-Fock Theories (50 points)

- a. Why must the HF self-consistent-field equations be solved iteratively?

HF theory assumes each electron to interact with the average field of all of the other electrons. That field is computed, electron-by-electron, from the electron density created by the occupied molecular orbitals. However, the purpose of carrying out the HF SCF equations is to *determine* the occupied MOs. Thus, one must first guess a set of MOs (or, equivalently put, a density matrix), solve the SCF equations for a new set, and proceed iteratively to convergence.

- b. What basis functions are there on a F atom and a H atom in the 6-31G(d) basis set?

The H atom has only valence 1s functions. There are two, one (tight) formed of 3 contracted primitive gaussians and one (loose) that is a single (uncontracted) gaussian. The F atom has a core 1s function formed of 6 contracted primitive gaussians and two sets of valence 2s, 2p_x, 2p_y, and 2p_z functions, one set (tight) having each basis function formed of 3 contracted primitive gaussians and one set (loose) having each basis function as a single (uncontracted) gaussian. Finally, the F atom has a single d function on it, which includes the 6 cartesian d-type functions.

c. Explain, ideally including any necessary equation(s), the following statement: An atomic orbital basis function is to a molecular orbital as a Slater determinant is to a configuration interaction (CI) wave function.

Both molecular orbitals (MOs) and CI wave functions can be expressed as linear combinations of basis functions. In the former case, the basis functions are typically atomic orbital (AO) type functions (e.g., those defining a basis set like 6-31G(d)) while in the latter case the basis functions are Slater determinants. Hartree-Fock theory provides a means to optimize the coefficients multiplying the basis functions for MOs, and CI theory provides a means to optimize the coefficients multiplying the Slater determinants for the CI wave function. In equation form, we have

$$\phi_j = \sum_i a_{ij} \varphi_i$$

and

$$\Psi_k = c_{0k} \Psi_{\text{HF}} + \sum_{a,i} c_{aik} \Psi_a^i + \sum_{ab,ij} c_{abijk} \Psi_{ab}^{ij} + \dots$$

where ϕ_j is the j th MO formed as a linear combination of the AO basis functions φ_i having coefficients a_{ij} (where i runs over the total number of basis functions in the basis set) and Ψ_k is the k th CI wave function formed as a linear combination of Slater determinants including the HF case and other Slater determinants that are themselves defined relative to the HF one by moving one or more electrons from occupied orbitals a, b , etc. to virtual orbitals i, j , etc. (so-called single, double, etc., excitations) up to a limit of including all possible excited determinants (which is dictated by the total number of electrons and the total number of available virtual orbitals).

d. What is spin contamination?

In unrestricted calculations (where orbitals are optimized separately for α and β electrons), the resulting Slater determinant need not have an expectation value for the S^2 operator that is equal to a pure spin-state eigenvalue (e.g., 0 for a singlet, 0.75 for a doublet, $S(S+1)$ in general, where S is the maximal allowed value of S_z given the number of unpaired electrons). When the optimized wave function is *not* a pure spin state, we say that it is “contaminated” by other spin states (usually higher ones). Thus, we can think of the resulting wave function as a linear combination of different spin states, with undesired spin states having coefficients whose magnitude dictates $\langle S^2 \rangle$. This affects expectation values for all other operators, too, e.g., energy, dipole moment, geometry, and in pathological instances can lead to very poor predictions with respect to the desired “pure” spin state.