

**CHEMISTRY 4021/8021**  
**MIDTERM EXAM – SPRING 2012 – KEY**

Q1) What is dispersion? Describe the physics underlying dispersion in a qualitative way and discuss the means by which dispersion is accounted for (or not) in the various theories that we have discussed in class to date.

An ideal answer would address the definition of dispersion (e.g., induced-dipole-induced-dipole interactions associated with the correlated motion of electrons) and proceed to discuss it in the context of molecular mechanics, semiempirical MO theory, ab initio HF theory, and post-HF theories. In MM, it is included in non-bonded terms (e.g., 6-12 potentials, or Morse potentials) and an ideal answer would note the portion of the potential energy curve over which it is operative. In HF theory, dispersion is essentially absent, as correlated electronic motion is not addressed beyond inclusion of exchange for same spin electrons. However, some semiempirical theories (e.g., MNDO, AM1, PM3, etc.) add molecular-mechanics-like corrections to the nuclear repulsion terms that, by parameterization, mimic dispersion to some extent. At the post-HF level, the formal inclusion of excited determinants in the wave function introduces dynamical electron correlation explicitly (since, for instance, a single excitation localized to a given fragment coupled with a single excitation localized to a different fragment can introduce a local dipole-dipole interaction).

Q2) Consider the unpleasant little molecule HOF. Discuss the two-electron integrals that would enter into a calculation of HOF at the CNDO, INDO, PM3, HF/6-31G(d), and MP2/6-31G(d) levels of theory. More specifically, what integrals are required for the construction of the Fock matrix and how are their various values determined?

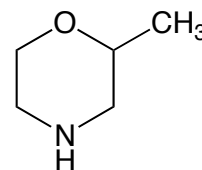
An ideal answer would start with CNDO, where the core orbitals are absorbed into a reduced nuclear charge for O and F and there are only 6 total 2-electron integrals (corresponding to  $\gamma_{HH}$ ,  $\gamma_{HO}$ ,  $\gamma_{HF}$ ,  $\gamma_{OO}$ ,  $\gamma_{OF}$ , and  $\gamma_{FF}$ ), for which the values are computed from a function involving atomic ionization potentials and electron affinities, and interatomic distance for diatomic terms. At the INDO level, the interatomic terms are unchanged, but O and F now have 5 same-center 2-electron integrals (the  $G$  and  $L$  integrals discussed in class). The new integrals are parameters of the model, typically taken from interpretation of electronic spectroscopy. At the PM3 level, the same-center integrals are done as described for INDO, but all combinations of 2-center 2-electron integrals ( $\mu\nu|\lambda\sigma$ ) are computed so long as both  $\mu$  and  $\nu$  are on the same atom, and  $\lambda$  and  $\sigma$  are both on a different atom. That will lead to 100 OF 2-electron integrals, and 10 each OH and FH integrals. The integrals are computed by replacing the atomic electron distributions with classical multipoles whose magnitudes are determined from Slater exponents for the

atomic basis functions. At the HF/6-31G(d) level, there are 32 basis functions required for HOF, so there are (ignoring symmetry)  $32^4$  integrals, all of which are computed using gaussian type orbitals, which permit analytic evaluation of all needed integrals (core electrons are now treated explicitly). At the MP2 level, the size of the basis set remains the same, but to compute the second-order correction to the energy, all of the 2-electron integrals must be computed over *molecular* orbitals, not atomic (basis-set) orbitals. Of course, the MOs are *composed* as linear combinations of AO basis functions, but the transformation of the AO 2-electron integrals to MO 2-electron integrals increases the scaling cost of this level of theory.

Q3) When sampling phase space, what are the relative advantages and disadvantages of molecular dynamics vs. Monte Carlo algorithms? What considerations might go into choosing one over the other?

An ideal answer would first note that MD sampling involves the propagation of Newton's equations of motion while Monte Carlo sampling involves a Markov chain with moves selected according to Boltzmann statistics. Some important points to note would be (i) only MD delivers a time-dependent trajectory, so MC cannot be used for time-dependent properties, like diffusion constants, for example, (ii) MC does not require the computation of energy gradients, while MD does, rendering the latter more expensive, (iii) MD is generally more prone to becoming trapped in local wells of the potential energy surface while MC, by its selection rules, can hop more readily between wells under favorable circumstances, (iv) MD is potentially difficult to parallelize on a multiprocessor system while MC is embarrassingly parallel (since every processor can run its own Markov chain), (v) MD requires a global update of the energy and gradient at every time step while MC requires only local updates associated with a moved particle when such a move is accepted, and (vi) only MC readily handles equilibrium between multiple phases. Very complete answers might also address the issue of choosing time steps for MD and choosing move ranges for MC.

Q4) Discuss how one might go about computing the 298 K heat of formation ( $\Delta H_{f,298}^{\circ}$ ) of gaseous 2-methylmorpholine, focusing, obviously, on approaches covered in class (or the reading) so far. Assign a level of confidence to the various protocols that you suggest, if not necessarily in quantitative terms, at least in a "best to worst" characterization. Note how computational constraints might play a role in limiting your range of choices, if at all.



An ideal answer would begin with MM, where strain energies may be added to heat-of-formation equivalents for strain-free atom types (determined, for

example, from Benson's equivalents or analogs), provided such atomic values are available. At the NDDO level of theory, the MNDO, AM1, PM3, etc. models equate electronic energies with enthalpies, and sum an assembly energy (the opposite of an atomization energy) with experimental atomic heats of formation to compute molecular heats of formation (thereby ignoring zero-point vibrational energy, for example, except to the extent that it is absorbed into parameterization). Other levels of theory, e.g., ab initio HF or post-HF levels, proceed by assembling the enthalpy from the calculations necessary to derive an ideal-gas, rigid-rotator, harmonic-oscillator partition function for the molecule, and again summing the assembly enthalpy with the atomic heats of formation to derive a molecular value. Post-HF levels may be used only for the electronic energy portion of the enthalpy (i.e., they may use an HF geometry and thus HF moments of inertia and vibrational frequencies for the rotational and vibrational partition functions, respectively).

A very good answer would then turn to accuracy, and note that force fields are good for what they are good for (and sometimes they are very good), while semiempirical models have substantial noise (accuracy of perhaps  $\pm 7$  kcal/mol might be about right for moderately sized organicky molecules), HF would be disastrous (as the assembly energy is very wrong owing to failure to account for dynamical electron correlation), and post-HF levels can achieve sub-kcal/mol accuracy but quickly become very expensive, thereby limiting their use to rather small molecules. By using isodesmic equations, substantial cancellation of errors in the underlying theory can be achieved. Thus, for instance, one might *compute* the enthalpy change for methylcyclohexane plus morpholine  $\rightarrow$  cyclohexane plus 2-methylmorpholine and use the *experimental* gas-phase heats of formation for methylcyclohexane, morpholine, and cyclohexane to derive a best predicted value for 2-methylmorpholine.