Chemistry 478 — Molecular Modeling **Spring 2014**

Take-Home Final Examination

Answer each of the three questions below. Please submit your responses as either printed copy or electronic document to Jay Ponder (Louderman 453, ponder@dasher.wustl.edu) by the last day of final exams, May 7th. All of the references mentioned below can be found on the course web site. You should use these papers as a starting point for discussion, but feel free to do your own literature searches and include other research in your answers.

Question 1: Density functional theory (DFT) calculations are perhaps the most widely used of all current electronic structure computations. There are several different reasons for the popularity of DFT. List and explain these reasons. Suggest some types of problems where DFT is obviously the QM method of choice, and other situations where its use is not appropriate.

The "classical" B3LYP functional (*circa* 1992) is still very widely used, but the search of improved functionals has picked up speed in recent years. Describe the component parts of the B3LYP formalism. What are the differences and improvements in the following state-of-the-art density funcationals: XYGJ-OS, Zhang, *et al*., *PNAS,* ***108***, 19896-19900 (2011); wB97X-2, Chai and Head-Gordon, *J. Chem. Phys*., ***131***, 174105 (2009); and M06, Zhao and Truhlar, *Theor. Chem. Account*, ***120***, 215-241 (2008). For each functional, briefly describe its underlying theory, development goals, parameterization, accuracy and range of applicability.

Question 2: The Monte Carlo (MC) and Molecular Dynamics (MD) algorithms are two of the major sampling techniques in molecular modeling. Compare and contrast the advantages and disadvantages of these methods. In what respects are they similar, and what are the key differences? It is possible to combine aspects of both methods into a single sampling protocol. Find literature examples of such combinations, and briefly explain how such calculations are organized.

Two direct comparisons of the "efficiency" of MC *vs*. MD have been reported. See Jorgensen and Tirado-Rives, *J. Phys. Chem*., ***100***, 14508-14513 (1996), and Ulmschneider, *et al*., *J. Phys. Chem. B*, ***110***, 16733-16742 (2006). These papers study very different systems, but appear to reach generally similar conclusions. Compare and contrast these studies. What assumptions and simplifications are made in these "efficiency" comparisons.

Question 3: We discussed "protein design" in class. Three levels of difficulty in such design work, from easiest to hardest, are: (1) design of structure, (2) design of ligand binding, and (3) design of catalytic function. The ROSETTA package from David Baker's group has been used for several o the leading studies in this area. Design of structure is exemplified by a paper discussed in lecture, *Science*, ***302***, 1364-1368 (2003). In a more recent publication, the Baker group has attacked design of binding; see Nature, 501, 212-216 (2013). Describe the exact computational protocol used in each of these papers. Use of a flowchart may help illustrate your answer. How has the computation methodology evolved and/or improved over the decade from 2003 to 2013? How has the modeling been adapted in the *Nature* paper to specifically address the "binding" problem? Can you find literature examples of "design of function"? What are the biggest remaining issues in the protein design field?