

FINAL EXAM

This exam is due May 11th at 3:00 PM. Please email an electronic version (preferably as a pdf file) to me at cramer@chem.umn.edu. I will be out of the country, so email is critical.

Some might say that science is all about seeing important details in an otherwise enormous sea of data. Your task is to exercise your abilities in this regard using the data collected by the class over the last two problem sets and now available at

<http://pollux.chem.umn.edu/8021/C4H6SO2/>

You may work alone or as part of a group of people. You are limited to 500 words if you are working alone (exclusive of references, tables, or figures). If you are part of a group, the group is limited to 500 words per member (e.g., a six-person group can write an exam of up to 3000 words, although you need not feel compelled to hit the maximum).

Find one or more interesting things in the data and present them in the form of a scientific paper. We've read a lot of papers this semester, and I've offered editorial comments from time to time about what constitutes a good paper. Do your best to follow those guidelines. Intro/Results/Discussion/Conclusions.

The subject(s) you choose to address are entirely up to you. You can talk about theoretical issues (how does DFT compare to HF for various things? How does solvation behave as a function of structure? How might you resolve discrepancies in the computations? etc.) You can compare data for one or several molecules to experiment or other calculations for the same compounds or analogs (Are there compounds related to known tautomeric equilibria? Can you find connected reactants and transition states for particular unimolecular reactions and would predicted rate constants from, say, transition state theory compare well to experiment? Are relative energies, enthalpies, or free energies consistent with experimental values in the NIST database, if such values are available?). You can look at trends in bond lengths for different molecules, trends in energies—you can do

whatever you want. To help you with planning, I am attaching a *J. Chem. Ed.* article that describes some of the ideas that members of a past class had.

References, where provided, should be in *JACS* format. Note that if you find obvious errors in the data, these should be reported to me for repair. They are not fitting subjects for analysis, unless there is some interesting technical reason that errors might have been expected (a doubtful prospect here).

Grading will be based on clarity (30%), depth of analysis (40%), creativity (20%), and style (10%). The exam is worth 150 points total.

A Cooperative Molecular Modeling Exercise— The Hypersurface as Classroom

Christopher J. Cramer,* Bethany L. Kormos, Paul Winget, Vanessa M. Audette, Jeremy M. Beebe, Carolyn S. Brauer, W. Russ Burdick, Eric W. Cochran, Brian M. Eklov, Timothy J. Giese, Yongseok Jun, Lakshmi S. D. Kesavan, Christopher R. Kinsinger, Mikhail E. Minyaev, Ramkumar Rajamani, Jonathon S. Salsbury, John M. Stubbs, Jack T. Surek, Jason D. Thompson, Vincent A. Voelz, Collin D. Wick, and Ling Zhang

Department of Chemistry and Supercomputer Institute, University of Minnesota, Minneapolis, MN 55455-0431;

**cramer@chem.umn.edu*

Quantum chemistry has been a staple of undergraduate curricula for several decades. For the most part, however, it has been the rare course in quantum chemistry that advances much beyond the hydrogen atom (or perhaps H_2) in terms of describing the electronic structure of real molecular systems. Driven in part by staggering advances in computer technology during the last 15 years or so, which have rendered possible on a laptop calculations that once would have taxed the largest mainframes available, this situation now seems to be changing (1–3). A brief survey of modern texts in quantum chemistry finds increasing reference to electronic structure theories and their use in the approximate solution of the Schrödinger equation for many-electron systems. For instance, the fifth edition of Levine (4) devotes extensive space to Hartree–Fock (HF) theory (5) in both its *ab initio* (6) and semiempirical (7) forms; describes post-Hartree–Fock methods such as Møller–Plesset perturbation theory (8), configuration interaction (6), and coupled-cluster theory (9); presents density functional theory (DFT) (10, 11); and offers an introductory discussion of methods to carry out population analysis (12), account for relativistic effects (13, 14), and include solvent effects (15).

At the same time, computational chemistry software has become increasingly affordable and user-friendly, making it fairly simple for beginners to take advantage of sophisticated methods for predicting molecular properties. (Of course, it also makes it simple for beginners to do phenomenally misguided calculations and misinterpret their artifactual results, but that is a subject best left for another time.) There is thus an increasing need for practical problems that can take advantage of these newly available resources to improve our pedagogy. Several problems have been described recently; these were primarily developed as complements to existing organic/

inorganic laboratory experiments (16–20) or as software-specific teaching tools (21–23).

This article describes a comprehensive assignment developed for a one-semester computational chemistry class composed of roughly 25 first-year graduate and advanced undergraduate students (Chemistry 8021 [24]; one could certainly use this exercise in a more general quantum chemistry course, however). The assignment was designed to span two computational problem sets (laboratories, if you will) and was also the basis for the final exam. Key intents of the exercise were to introduce students to Web-based information storage and data manipulation, illustrate concepts associated with the utility of various theoretical models, and stress skills involved in recognizing interesting chemical patterns in a large collection of raw data. Moreover, effort was made to foster student collaboration in various aspects of the assignment, although some competitive features were also present.

This article describes the assignment, the data generated by the students, and the various approaches students took toward completing the final exam, and offers some thoughts for improving or tailoring this concept for specific situations.

The Assignment

Students had access through the Minnesota Supercomputer Institute to personal class accounts and the quantum chemistry program suite Gaussian 98 (25), with which they were familiarized. They were asked to log in to a password-protected portion of the class Web site (24) (each student had his or her own password). The initial assignment contained the following instructions (where the instructor is speaking in first person):

Select View Surface. You will see my contribution to mapping the $C_3H_4N_2O_2$ surface (click on *cramer.html* to see it in 3D—if working from home, you will need to get the Chime Plug-in if you don't already have it) [a link to Chime, at <http://www.mdli.com/cgi/dynamic/welcome.html>, was provided]. I found a [3+2] cycloaddition transition state for acetylene and *N*-nitrosomethylidene nitrene.

Your task is to find a point of your own on the potential energy surface (PES)—either a minimum or a transition state—that is not already in the table (easy if you're the first to finish, harder and harder the longer you wait). To complete an entry, you will need a frequency calculation at the HF/6-31G* level (and thus, of course, you will first have to have optimized the structure at this level of theory). When you have the data, click on Submission Form and paste in the number of imaginary frequencies (0 or 1), the HF energy, enthalpy, and free energy, and finally the optimized Cartesian coordinates (you can cut and paste right from the output file of the frequency calculation). If you make any mistakes, you can delete an entry and start again.

Each new submission will be visible to all students, so you can see if the structure you're working on has already been taken. Full credit for this problem simply consists of entering sensible data. In future exercises, we will make comparisons of different levels of theory and different isomers to learn more about the chemistry of the surface.

As an aside for those less familiar with the jargon of molecular modeling, note that the PES is defined as the surface describing the energy for all possible arrangements of the atoms within a given molecular formula. Constitutional isomers, for instance, correspond to different wells on a PES, and the lowest energy path from one well to another must pass through some highest energy point, which defines a transition-state structure.

Returning to details, the information displayed by the View Surface link was a table (24), containing as entries (i) a link to the Chime-readable Cartesian coordinates, (ii) the number of imaginary frequencies (no imaginary frequencies implies an optimized structure to be a minimum, one implies a transition state structure (i.e., a saddle point), and more than one is not chemically meaningful), (iii) the absolute energy, (iv) enthalpy, and (v) free energy in units of E_h , (vi) the free energy of solvation in units of kcal/mol (blank at this point), and (vii) a small box for comments.

The competitive aspect of permitting only one individual to submit any particular structure was, in some sense, mitigated by the vast array of chemical possibilities available for the empirical formula in question (note moreover that different conformations of the same compound qualify as different structures). The choice of empirical formula is an important one and should be tailored to the class size to some extent. A smaller class should have a simpler formula to avoid too sparse a sampling of the relevant PES. There are other issues associated with choice of formula that we will address below.

Successful completion of this part of the problem (which was one component of a larger problem set devoted to ab initio theory) simply consisted of entering sensible data, as checked by the teaching assistant and instructor for the course.

In the next problem set, which was devoted jointly to DFT and continuum solvent models, the following assignment appeared:

Here continues a problem that will carry over to the final exam. We add to the data at [the Web site]. Your present task is to update your structure on the potential energy surface (PES) with a calculation at the B3LYP/6-31G* level (since there is no double slash, this implies reoptimizing the geometry). The most efficient way to do this will be to read in the force constants and geometry from your RHF frequency calculation. Once you have a reoptimized structure, do a new frequency calculation at the B3LYP/6-31G* level and update your entry with that data.

Finally, compute the aqueous solvation free energy for your DFT structure using the SM5.42R/B3LYP/MIDI! model in MN-GSM.

MN-GSM (26) is a solvation module for G98; MIDI! (27, 28) is a highly efficient basis set optimized for use in solvation calculations. Here again, the only requirement for credit was the entry of sensible data, both for the gas-phase DFT calculations and for the continuum solvent calculation using the SM5.42R model (29, 30).

The complete data set stored on the Web site then found use as the basis for the take-home final exam, the instructions for which were:

Some might say that science is all about seeing important details in an otherwise enormous sea of data. Your task is to exercise your abilities in this regard using the data collected by the class over the last two problem sets and now available at [the Web site].

You may work alone or as part of a group of people. You are limited to 500 words if you are working alone (exclusive of references, tables, or figures). If you are part of a group, the group is limited to 500 words per member (e.g., a six-person group can write an exam of up to 3000 words).

Find one or more interesting things in the data and present them in the form of a scientific paper. We've read a lot of papers this semester, and I've offered editorial comments from time to time about what constitutes a good paper. Do your best to follow those guidelines.

The subject(s) you choose to address are entirely up to you. You can talk about theoretical issues (how does DFT compare to HF for various things? How does solvation behave as a function of structure? Are there clear errors in the data? How might you resolve discrepancies in the computations? etc.) You can compare data for one or several molecules to experiment or other calculations for the same compounds or analogs (Are there compounds related to known tautomeric equilibria? Can you find connected reactants and transition states for particular unimolecular reactions and would predicted rate constants from, say, transition state theory compare well to experimentally measured values?). You can look at trends in bond lengths for different molecules, trends in energies—you can do *whatever you want*.

References, where provided, should be in *JACS* format.

Grading will be based on clarity, depth of analysis, creativity, and style.

In the design of this overall problem, it was hoped that students would use one another's data to build connections on the PES. For example, the presence of one tautomer on the Web site might inspire a student still working on the assignment to pick a related tautomer. Or, a student might see two structures suggesting a reaction path between them and look for the transition state (TS) structure. The instructor's choice of a cycloaddition TS structure was in part to provide an obvious opportunity for the first student to choose to find the product. Here too, the choice of empirical formula is fairly important. If the formula is more limited, students must perforce develop connected structures.

As a final point, it should be noted that students were made aware of the large body of thermochemical and other data available on the Web site of the U.S. National Institute of Standards and Technology (<http://webbook.nist.gov>). For the particular formula chosen here, the Webbook contains data for only 2 molecules—namely, 2,4-imidazolidinedione (which was submitted by a student) and 3-methylsydnone (an interesting zwitterion, which was *not* submitted). A potential consideration in problem design is choice of an empirical formula that takes good advantage of this convenient online data source.

Student Data

The complete data are provided online at <http://pollux.chem.umn.edu/8021/JCEsm.html>. Here we summarize a few salient details. Most of the entries were five-membered-ring heterocycles. Of 25 entries, 9 were TS structures and the remainder were minima. The TS structures tended to be for conformational interconversions (e.g., a planar TS between alternative puckered minima for a 5-membered-ring system), although there were more unusual cases, including a TS for a [3+2] cycloaddition, a TS for the closure of a dinitroso compound to a furoxan, and a TS for proton transfer between two heteroatoms. One particularly unusual pentacyclic compound was submitted as a local minimum. The total range of electronic energies was about 250 kcal/mol, and the total range of solvation free energies was about 20 kcal/mol.

Some care was required before the final exam was issued to ensure that the data were checked for sensibility. Beginning students are sometimes prone to assume that errors are in fact interesting chemical phenomena (it's so tempting to believe in the infallibility of something from one's computer). Access of the instructor and TAs to all student accounts made it simple to fix instances where copy-paste errors were made.

Student Performance on the Final Exam

Given the opportunity to collaborate, 11 of the 25 students chose to do so, forming groups of 2, 4, and 5. Students were told when the exam was distributed (roughly three weeks before its due date) that they were welcome to do additional computations if they so chose, using their class resources, and roughly 25% of the exams included new computations.

Two student groups chose especially creative analyses. One examined whether the computed solvation free energies correlated with other molecular properties. They compared against dipole moments computed using the CM2 charge model (31) and against the fraction of polar surface area as

computed by the molecular mechanics program PC Model (32) (which had been used in the class for other purposes). The availability of Cartesian coordinates for all structures on the Web site allowed the latter task to be accomplished in a very straightforward fashion. Following their analysis, the students identified outliers in their correlations and proposed a qualitative fragment-wise contribution model to the solvation free energy. (Solvation free energies seemed to be a favorite topic for analysis; other students attempted to rationalize them in terms of contributions from hydrogen-bond donating or accepting groups in the molecules.)

Another group elected to generate intrinsic reaction coordinates (33) (IRCs) at an economical level of theory for all of the listed TS structures. That is, they attempted to follow each TS structure (after reoptimization at the more economical level of theory) down to its reactants and products using various algorithms present in Gaussian 98. This group discovered that such calculations are technically rather challenging, and transformed that aspect of the experience into a thoughtful discussion of what TS features seemed to contribute to particular difficulties. In successful cases, they analyzed the differences predicted between HF and DFT and attempted to gauge the relative performance of those two levels by undertaking a simple bond-energy analysis.

Several other student initiatives are worthy of note. Some data are available on the NIST Web site for the student-submitted molecule 2,4-imidazolidinedione, and one student compared computed ionization potentials for this molecule with the experimental value and discussed theoretical trends. Tautomeric energy differences (of which several examples were available) were analyzed in terms of simple bond-energy arguments. Some students analyzed the data from the standpoint of developing parameters for molecular mechanics force fields (which are typically not designed to handle molecules having extended linkages of heteroatoms, as were present in some of the submitted structures). The degree of pyramidity of nitrogen atoms in various rings and functionalities was analyzed. Some general discussion of the differing ranges of applicability of HF and DFT theory to the unusual functionalities present in the data set was provided, which developed nicely from lecture topics that were part of the course material.

No students focused on differences in energies versus enthalpies versus free energies, perhaps reflecting a lack of having been exposed to the thermodynamic implications of such differences. Similarly, no students successfully identified known rate constants or equilibrium constants for reactions or equilibria related to the molecules in the data set. This latter point is not surprising, perhaps, given the vastness of the literature and the somewhat exotic chemical formula under consideration.

Targeted Improvements

The greatest potential for modifying the scope of the problem described above rests with the chosen molecular formula. The greater the available molecular diversity for a given formula, the less facile it becomes for students to profitably take advantage of synergy from one another's data. The discerning instructor might choose one formula over another because of significant quantities of available experimental data (for rates, equilibrium constants, spectra, etc.)

against which to compare. Class size also influences the choice of chemical formula.

Of course, one could enforce synergy by providing a starting structure and insisting that all submitted structures make some sort of "contact" with existing structures. That is, if the instructor's structure is a TS, for instance, the first submission will be either the reactant(s) or product(s) of that reaction, at which point a new reaction may be envisioned from the new cases, and students might be required to explain how they went from a prior case to their own.

As an alternative, an instructor who makes a choice of chemical formula based on available experimental (or other theoretical) data could arrange for those data to be included before the final analysis stage. That is, the student data would be salted with additional structures offering opportunities for greater contact with studies in the literature.

One interesting exercise would be to have the class, rather than the TA or instructor, identify erroneous data. With multiple levels of theory, it is usually simple to compare relative energies, for instance, against the standard of the instructor's structure (the data for which are presumably perfect) and identify where impossible discrepancies arise.

Another modification of the problem would involve a focus on properties other than energetic ones. For instance, one could have a formula with a single occurrence of, say, a phosphorus atom, and ask the students to compute the nuclear magnetic resonance chemical shift (δ) of ^{31}P in their molecule. Analysis could then focus on the response of the chemical shift to different molecular environments, differences in predictions for different levels of theory, etc.

We conclude by noting that student feedback on this exercise was quite positive. The prevailing sentiment was that the nature of the problem required individuals to synthesize a response drawing on mastery of a number of topics that had been separately presented in the classroom (as formal lecture material), but importantly there was considerable freedom afforded individuals to choose from among a variety of possible topics and in the process to exercise initiative and creativity.

Acknowledgments

Support was provided by the Alfred P. Sloan Foundation and the University of Minnesota Supercomputer Institute.

Literature Cited

- DeKock, R. L.; Madura, J. D.; Rioux, F.; Casanova, J. *Rev. Comput. Chem.* **1993**, *4*, 149–228.
- Gasyna, Z. L.; Rice, S. A. *J. Chem. Educ.* **1999**, *76*, 1023–1029.
- Whisnant, D. M.; Howe, J. J.; Lever, L. S. *J. Chem. Educ.* **2000**, *77*, 199–201.
- Levine, I. N. *Quantum Chemistry*, 5th ed.; Prentice Hall: Upper Saddle River, NJ, 2000.
- Roothaan, C. C. J. *Rev. Mod. Phys.* **1951**, *23*, 69–133.
- Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; Wiley: New York, 1986.
- Pople, J. A.; Beveridge, D. A. *Approximate Molecular Orbital Theory*; McGraw-Hill: New York, 1970.
- Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618–622.
- Cizek, J. *Adv. Chem. Phys.* **1969**, *14*, 35–89.
- Parr, R. G.; Yang, W. *Density-Functional Theory of Atoms and Molecules*; Oxford University Press: New York, 1989.
- Ziegler, T. *Chem. Rev.* **1991**, *91*, 651–667.
- Bachrach, S. M. *Rev. Comput. Chem.* **1993**, *5*, 171–227.
- Pykkö, P. *Chem. Rev.* **1988**, *88*, 563–614.
- Almlöf, J.; Gropen, O. *Rev. Comput. Chem.* **1996**, *8*, 203–244.
- Cramer, C. J.; Truhlar, D. G. *Chem. Rev.* **1999**, *99*, 2161–2200.
- Poon, T.; Bodolosky, S. A.; Norris, C. M. *J. Chem. Educ.* **1999**, *76*, 983–985.
- Pfennig, B. W.; Frock, R. L. *J. Chem. Educ.* **1999**, *76*, 1018–1022.
- Lipkowitz, K. B.; Robertson, D. *J. Chem. Educ.* **2000**, *77*, 206–209.
- Graham, K. J.; Skoglund, K.; Schaller, C. P.; Muldoon, W. P.; Klassen, J. B. *J. Chem. Educ.* **2000**, *77*, 396–397.
- Cook, A. G.; Kreeger, P. K. *J. Chem. Educ.* **2000**, *77*, 90–92.
- Hehre, W. J.; Burke, L. D.; Shusterman, A. J.; Pietro, W. J. *Experiments in Computational Organic Chemistry*; Wavefunction Inc.: Irvine, CA, 1993.
- Pietro, W. J. *J. Chem. Educ.* **1994**, *71*, 416–420.
- Foresman, J. B.; Frisch, M. J. *Exploring Chemistry with Electronic Structure Methods*; Gaussian, Inc.: Pittsburgh, PA, 1996.
- The complete class Web site is <http://pollux.chem.umn.edu/8021>. The particular data for this article are found at <http://pollux.chem.umn.edu/8021/JCEsm.html>. An analogous Web site of interest is <http://simons.bec.utah.edu/TheoryPage/index.html> (all accessed May 2001).
- Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T. A.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Peng, C. Y.; Ayala, P. A.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A. *Gaussian 98*; Gaussian, Inc.: Pittsburgh, PA, 1998.
- Li, J.; Hawkins, G. D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G.; Frisch, M. J. *Minnesota Gaussian Solvation Module MN-GSM version 98.3.2*; University of Minnesota: Minneapolis, 1999.
- Easton, R. E.; Giesen, D. J.; Welch, A.; Cramer, C. J.; Truhlar, D. G. *Theor. Chim. Acta* **1996**, *93*, 281–301.
- Li, J.; Cramer, C. J.; Truhlar, D. G. *Theor. Chem. Acc.* **1998**, *99*, 192–196.
- Zhu, T.; Li, J.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Phys.* **1998**, *109*, 9117–9133.
- Li, J.; Zhu, T.; Hawkins, G. D.; Winget, P.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *Theor. Chem. Acc.* **1999**, *103*, 9–63.
- Li, J.; Zhu, T.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **1998**, *102*, 1820–1831.
- PC Model, version 7.0; Serena Software: Bloomington, IN, 1999.
- Truhlar, D. G.; Kupperman, A. *J. Am. Chem. Soc.* **1971**, *93*, 1840–1851.
- Bühl, M.; Kaupp, M.; Malkina, O. L.; Malkin, V. G. *J. Comp. Chem.* **1999**, *20*, 91–105.