Introduction to Computational Chemistry

1.1 Introduction

Chemistry is traditionally an experimental science dealing with the structure, functions and reactivity of molecules and the methods for making them from convenient starting materials. With the advent of quantum mechanics, a better understanding of atomic and molecular structure was initiated but significant progress was delayed because of the difficulties in calculating all the necessary multicenter integrals and the necessity of diagonalizing large matrices. With the arrival of faster and even faster computers, these calculations are becoming possible and getting more accurate. A study of large molecular clusters and liquids was also initiated using interaction potentials and classical equations of motion. Now these methods are extended to surfaces, multiple phases and topics covering various aspects of molecular biology. In the present course, we shall investigate many aspects that touch upon the issues related to the above problems. All the above mentioned problems use the methods of numerical analysis, often applied iteratively and recursively to solve the problems of chemical interest. Hence we begin with an introduction to the methods of numerical analysis. Next, we shall study the ab initio and-semi empirical quantum chemical methods which

form an important core of theoretical/computational chemistry. Various examples will be illustrated using the Gaussian programs available in the public domain. The later part of the course deals with computer simulations involving a large number of molecules. As fully quantum mechanical calculations are not yet possible for large systems, methods of classical mechanics are used and meaningful results can be obtained for comparison with experimental data. The final section of the course is concerned with molecules of chemical as well as biological interest.

Over the years, there has been a lot of interplay between laboratory experiments and computational predictions. Often, new molecules have been predicted and later on, confirmed by experiments. When there are several pathways possible for a reaction, computations of the potential energy profiles for different pathways can indicate which of the pathways corresponds to a lower energy path and further experiments can be performed to isolate "clusters" corresponding to the lower energy paths. Classical MD/MC simulations provide very detailed pictures of molecular motions which are not accessible experimentally. In problems such as protein folding in different solvent media, molecular simulations can yield comprehensive pictures of crucial solute and solvent configurations that play an important role in the dynamics of the folding process.

1.2 Numerical Methods (Module I)

We encounter the first application of a numerical method when we fit the "best" curve or a straight line passing through a number of experimental data points. When a large number of data points are generated in an experiment, a compact way of dealing with these is through "best" fitting functions or interpolating polynomials. These functions and polynomials have the additional advantage of being amenable to differentiation and integration which are usually the requirement of a physico-chemical problem. A crucial requirement in all these problems is the ability to write/modify computer programs to suit one's own needs. While this is not a course on computer programming, some chapters are devoted to it so that a serious developer can take up other courses to enhance her skills. In applications of "theoretical" methods to solve chemical problems, good programming skills are very useful. A lot of work that has gone in developing public domain and commercial softwares. While the early programming languages were Basic or Fortran, the present languages in use are C++ or Java and these too will change with time. Often, the ability and to switch from one language to another, while a bit unpleasant, is not very difficult. The programming details are given in Chapter 2.

There are several methods for interpolation such as the Newton's forward difference form or the Lagrange interpolation scheme, Newton's method has an advantage that it is more readily adaptable to schemes for differentiation and integration. The main requirement is the availability of data points at regular intervals. Curve fitting is to be used only if there is significant error in the experimental data points (Chapter 3).

With an interpolating polynomial, we can easily do the differentiation and integration of this polynomial. For data points which are not equally spaced, separate methods need to be used. (Chapter 4)

Roots of equations are needed in different problems of science and engineering. When a large number of roots are involved, rapidly convergent methods become particularly important (Chapter 5).

Matrix methods involve finding the inverse of a matrix (whose determinant is not zero and also the eigenvalues and eigenvectors (molecular energy levels and wave functions (Chapter 6). For studying the spatial or time evolution of a problem we need to solve the appropriate differential equation that describes the physical situation correctly (Chapter 7).

Integral transforms find applications in converting data from time domain to frequency domain (e.g. NMR spectroscopy) or from coordinate space to momentum space (e.g. Xray crystallography). They are also useful in solving differential equations. These will be taken up in Chapter 8.

Public domain softwares such as Scilab have become very useful. Scilab is very easy to use. Chapter 9 gives details of its applications.

What we have seen above are just a few sample numerical methods which find frequent use in computational chemistry. There are several other elementary as well as advanced techniques which also find applications in chemical computations.

1.3 Ab initio and Semiempirical Methods (Module II)

In this module, we will be dealing with the techniques used in solving the non-relativistic Schrödinger equation for estimating molecular energy levels and wavefunctions.

In the initial lectures, the basic formulae relevant to the Hartree-Fock method and the extensions beyond the Hartree-Fork (HF) method are outlined. The form of the molecular Hamiltonian, Slater type orbitals (STOs) and the linear combinations of Gaussians (GTOs) to represent the STOs are outlined. The formulae for overlap, electron kinetic energy, nuclear attraction energy and electron repulsion integrals are given. The extensions of HF methods to include electron correlation, the many body perturbation methods and the coupled cluster methods are considered next. The methods using density functional theory have become very popular of late and are considered in a separate chapter. The public domain softwares to do the ab initio and semi empirical calculations (such as the Huckel, extended Huckel, CNDO, etc) are considered next. There are different convenient methods for providing inputs to these programs. This is the minimum

ability that one has to master to get useful results for molecules of one's interest. Computation of single point energies and the energy for an "optimized" or the "best" molecular geometry is considered next. With the voluminous results that one gets using these programs, the analysis can be quite involved.

Several public domain softwares are available to show the results pictorially. The main objects of interest are the energy levels and wave functions. Contours of wave functions, electron densities and electrostatic potentials can be plotted. Net charges on atoms and surfaces containing or enclosing a given amount of charge can be calculated from the electronic charge densities. By calculating the molecular energies at a few points in the neighbourhood of the equilibrium molecular geometry, the local potential energy surface (PES) can be estimated. Using this PES, an estimate of vibrational frequencies of the molecule can be made. Using models for molecular shapes and the three moments of inertia of the molecule, the rotational energy levels can also be estimated. Using all the data on molecular energy levels, molecular partition functions and the gas phase thermodynamic functions can be calculated at the ambient temperature. By exploring the PES in the region where the molecule is likely to fragment, the structure of transition states for dissociation (and analogously for association) can be studied. By placing the molecule in suitable cavities (dielectric medium) calculations can be made for taking into account the effect of the medium. The energy levels, wave function and molecular moments change due to the effect of solvation and the estimates of solvent effects are possible.

1.4 Classical MC/MD Simulations (Module III)

Using ab initio and semiempirical methods, one can determine intermolecular forces. This procedure is quite elaborate and for molecules containing more than 8 to 10 atoms, it is still customary to use empirical intermolecular potentials. These potentials are chosen in such a manner that the observed macroscopic properties of fluids consisting of these molecules are reproduced satisfactorily.

For liquids of chemical and biological interest, it is still expedient to use classical equations of motion to calculate the molecular trajectories of all the particles in the chosen system. Since it is not possible to consider more than a few hundred to a few thousand molecules in a system, periodic boundary conditions are used to make the system resemble a bulk fluid. From the ensemble of particle positions and velocities that are generated, different thermodynamic functions such as the internal energy, heat capacity and pressure can be calculated. Of greater interest are the spatial and time correlation functions (TCFs). The spatial (radial) interparticle distribution functions (RDFs) can be compared with the experimental X-ray and neutron diffraction data. While a time correlation function (TCF) cannot be directly compared with experimental data, the transport coefficients (such as the diffusion coefficient, conductivity and viscosity) calculated using these TCFs can be used to validate the potential models used in performing the simulations. The development of a code for performing molecular dynamics for a Lennard-Jones fluid is considered in this module. The dynamical equations have to be solved using suitable algorithms. An alternative method to obtain the

equilibrium properties of fluids is the Monte Carlo method. In this method, time does not appear explicitly during each step. Molecules are moved around their original locations by using a set of random numbers. The acceptance or rejection of such a move is decided by comparing the ratio of Boltzmann factors for the final and initial energies $\frac{e^{-\beta E_{new}/k_B T}}{e^{-\beta E_{old}/k_B T}}$ with another random number. If the ratio is greater than the random number, the new set of molecular coordinates is accepted, even if the new energy is higher than the old energy. If the new energy is lower than the old energy, the new configuration is automatically accepted. Steps in the development of the MC code will also be described in this module.

1.5 Biomolecular Simulations (Module IV)

In this module, the methods developed for fluids containing small molecules (such as water, alcohols, etc.) are extended to biomolecular systems. Different force fields (models for intermolecular interactions) have been developed for this purpose. The first requirement of the force field is that it should reproduce the correct structure of the bimolecules (protein, RNA or DNA) in different solvent media. Good models for solvents have been developed for common solvents such as water, alcohols, DMSO, DMF, acetonitrile, acetone, ether and so on. Once a good equilibrium structure is achieved, dynamical properties such as the rate of conversion of a helically coiled segment into a random configuration can be investigated. With increasing computing speeds and storage capabilities, simulations extending to realistic macroscopic times will become possible.

1.6 Summary:

In this lecture we have briefly outlined the contents covered in this course in computational chemistry. If you need to develop some programming skills, you need to go through the first module. If you have a reasonable background and are interested in liquid structure and dynamics, you may directly go to Module 3 and 4. If you want to pursue quantum chemical calculations right away, proceed to Module 2. Doing the exercises and the self assessment test will give you a good idea of the abilities that you have developed. You may contact the authors on their e-mail address if you need any special assistance or have specific queries.

References

1) Donald A. McQuarrie, Quantum Chemistry, Second Edition, Indian Edition, may be available for downloading from the internet. A good introduction to quantum chemistry.
2) Elementary Numerical Analysis: An Algorithmic Approach by S. D. Conte and C. deBoer
3) Allen, M. P. and Tildesley D. J; Computer Simulation of Liquids. Oxford University Press. This is an excellent introduction to computer simulations of liquids and includes several useful computer programmes. A version may be available on the internet for downloading.

4) McCalla, T. T; Introduction to Numerical Methods and Fortran Programming (John Wiley & Sons, Inc., New York, 1967). This is a very nice introductory book
5) Press Numerical Recipes: The Art of Scientific Computing, Third Edition in C++ (2007), Numerical Recipes in C, Second Edition (1992), Numerical Recipes in Fortran 77, Second Edition (1992), by Press, William H.; Teukolsky, Saul A.; Vetterling, William T.; Flannery, Brian P. New York: Cambridge University Press. There are Indian Reprints of this book and electronic versions are also available. This is a very good resource book.
6) Schaum's Series Textbook on Programming With Fortran 77.