CHAPTER 2

METHODOLOGY

2.1 INTRODUCTION

In the *ab initio* calculation we solve the time independent Schrodinger equation, which is given as follows.

$$\hat{H}\Psi = E\Psi \qquad \cdots\cdots\cdots\cdots (1)$$

In contrasts to semi-empirical methods, the equation is solved without reference to experimental data.

Here the total Hamiltonian operator can be written as the sum of kinetic and potential energy operators of the nuclei and electron.

$$H_{total} = T_n + T_e + V_{ne} + V_{ee} + V_{nn} \qquad \cdots\cdots\cdots\cdots (2)$$

The electronic Hamiltonian operator is given by

$$H_{el} = T_e + V_{ne} + V_{ee} + V_{nn} \qquad \cdots\cdots\cdots\cdots (3)$$

And it depends on the potential of the nuclei.

$$H_p = -\frac{1}{2M_{tot}}(\sum_{i=0}^{N}\nabla_i)^2 \qquad \cdots\cdots\cdots\cdots 4)$$

The H_p Is called the mass polarization

The solutions are chosen for the orthogonal and normalized wave functions

$$\int \Psi_i^* \Psi_j \delta\tau = \delta_{ij}, \quad \delta_{ij} = 1 \text{ if } i = j \text{ and } \delta_{ij} = 0 \text{ if } i \neq j$$

The total wave function is written as an expansion of the complete set of electronic wave function, with the coefficient being a function of nuclear coordinates.

$$\Psi_i(R,r) = \sum \Psi_{ni}(R) \ \Psi_i(R,r)$$

Here the Born-Oppenheimer approximation was taken into account, and the potential energy surface of the system is obtained for the electronic motion. This is given by the solution of electronic Schrodinger equation. Once the Born-Oppenheimer approximation

is introduced the system reduces to the electronic Schrodinger equation and the electronic Hamiltonian is given by equation 3.

2.2 ELECTRON CORRELATION

In the HF calculation, the electron-electron interaction is replaced by an average interaction of all elections and so if the basis set taken is large enough, then almost 99% of the total energy can be obtained from the HF wave function [1-3]. But for some chemical systems the remaining small amount of energy may be important. So the electron correlation is given by the difference between the HF and lowest possible energy in a given basis set. Here the electrons are taken to correlate with each other. The basis set determines the size of the one electron basis i.e. it limits the description of one electron function. On the other hand the number of determinants included in the wave function determines the size of the many electron basis and thus limits the description of electron correlation.

So the Moller-Plesset perturbation theory can be applied for calculating perturbation energy. In this case the solution to the unperturbed problem is considered and there must be an infinite number of functions which cannot be covered in the calculation. The energy of the unperturbed problem is the HF wave function and additional higher energy solutions are excited state determinants. Depending on the finite basis set taken one can generate excited state determinants. The second order correction of the perturbation method, the matrix elements of the perturbations operate between HF and all other excited states are considered.

The total electron correlation energy can be calculated as

$$\Delta E_C = \Delta E_{(2)} + \Delta E_{(3)} + \Delta E_{(4)} \qquad \dots\dots\dots\dots\dots (5)$$

and the correction to interaction energy is given by the following equation

$$\Delta E = \Delta E_{(HF)} + \Delta E_C \qquad \dots\dots\dots\dots\dots (6)$$

Where $\Delta E_{(2)}$, $\Delta E_{(3)}$, $\Delta E_{(4)}$.... are the second, third and fourth order contribution. Generally the electron correlation is found by using the most economical MP2 method (second order). But it is feasible for calculating medium sized molecular systems. In some cases, the energies computed by MP2 methods sometimes overestimate the electron correlation, whereas it gives quite accurate values in certain cases [2-4]. Though the method is very

essential, the applicability of this method is limited as computation time and capability of machine requirment is very high. However, it is rather essential to use MP2 method for evaluating interaction energies of some particular systems such as π -π and π-σ stacking.

The formula used in the second order Moller-Plesset equation is

$$E = \sum_{i<j}^{occ} \sum_{a<b}^{virt} \frac{\left[\langle \Phi_i \Phi_j | \Phi_a \Phi_b \rangle - \langle \Phi_i \Phi_j | \Phi_b \Phi_a \rangle\right]^2}{E_i + E_j - E_a - E_b} \quad\quad \dots\dots\dots\dots\dots\dots(7)$$

MP2 level of calculation covers most of the electron correlation energy. However expensive level of calculation like MP3, MP4 and MP5 methods are difficult even for medium size molecules.

2.3 BASIS SET

The accuracy of the MOs depends on the size of the basis function. As the number of basis function increases up to the limit of complete basis set (infinite number of basis function), the result is accurate and equal to the numerical HF methods. However only the best single determinant wave function can be obtained, and in practice the incomplete basis set is normally used [1-7]. According to variation theory, as the basis set is increased, the computed energy values become better. So by increasing the basis sets the computed results are tested. In the iteration procedure the variation of HF energy is arbitrary with respect to the variation of MO, in the sense that the first derivative of the energy, with respect to the MO coefficient is zero. In order to justify the energy minimum, the second derivative can be calculated and the eigen value of the matrix should be minimum if the energy minimum correspond to local minimum.

In all approaches of calculating the energy of system the choice of basis set is important. The core electrons account large part of the total energy and optimizing basis set tend to make optimum basis set. However the valence shell electrons are very important in determining many chemical properties. So in all energy optimization the basis set should be taken for describing the other part of the wave function. In that case the basis set should be large enough to take up the energetic of the system in a more efficient way. Hence the use of diffuse function i.e basis set with small exponents is always recommended. Indeed, it is important to study weak interactions as well as loosely bound electron where the energy depends on the polarizability.

According to Pople the basis set is written as STO-NG where the Slater type orbital consisting of nPGTOs (primitive GTOs) are used. This is the minimum basis set (n=2-6) and STO-3G basis set is generally used. This basis set has been designed by Pople *[5, 6]*.

Again there are other basis set such as 3-21G, 6-31G,6-31G* and 6-31G** etc. 3-21G is the split valence basis set where the core electrons are contracted into 3PGTOs and the valence is contracted to two PGTOs and other most part is represented by one PGTO. Similar to this 6-31G is another split valence basis in which the core orbitals are contraction of six PGTOs, inner part of valence orbital is three PGTOs and one outer PGTO. This is a triple split valence basis set where the orbitals are taken as six contracted PGTOs and valence split into three functions, as three, one and one PGTOs respectively. Then the diffuse or polarization functions are added to these basis sets. In 6-31G*(or 6-31G (d)) basis set, one set of polarization function is used. 6-31G** is identical to 6-31G (d, p) where diffuse sp function, two d and one f on heavy atoms and diffuse s and two p and one d on hydrogen atom are taken. It is a triple split valence basis set.

2.4 DENSITY FUNCTIONAL THEORY (DFT)

The DFT theory is used to design functions for connecting the electron density and the energy of a system. It is conveniently applicable in medium and large molecular system. It consumes less computer time compared to the other expensive *ab initio* calculations *[5, 6]*. The method is very similar to HF methods if the exchange part of the HF procedure is determined by exchange correlation functional, and also includes some electron correlation, but there are some differences between these methods. In DFT method if the exact exchange correlation functional is known then it would provide total energy value with electron correlation part. However the evaluation of exchange correlation functional is also a complicated task. The application of DFT method is limited, particularly in systems where the dispersion part is the dominant part. In that case the calculated interaction energy values are always underestimated. The commonly used B3LYP method fails to predict dispersion energy. In such cases inclusion of dispersion energy is very difficult for large system and the dispersion energy is especially calculated empirically and added to the values of DFT calculations. The method is applied when the electron density is minimized to charge density fluctuation at the same charge density function with respect to a particular direction. There are many DFT methods, which depend on the choice of the

functional form of the exchange correlation energy. The accuracy of the results obtained from a DFT method can be known by comparing the results obtained from accurate *ab initio* method or some experimental values. However there are not much information that show clear calibration of DFT results with *ab initio* and MP2 level of calculations.

2.5 JOINMOLECULES PACKAGE

This package allows one to join two already constructed or structure-optimized molecule-models, whether through bond-breaking or through van der Waals interactions *[8]*. Though it is basically a numerical program with numeric coordinates working as inputs and obtained as the output, it can produce its joined-model output in XYZ (.xyz) or Mopac-Input (.mop) file-formats so as to be pictorially viewed with Ortep / Rasmol / Protein Explorer (PE)/ PCModel etc. As Gaussian is one of the most widely used computational software, JoinMolecules almost always produces a Gaussian input (.gjf) file as its main output. This hardly causes any loss of generality, as because the atomic-number and coordinate specification included in this output could be a part of input file of any other computational software. However, an XYZ file always named as nuclei_xyz.xyz and always residing in the working folder of JoinMolecules is also produced as an additional output, which may be viewed with Ortep-3 or Rasmol etc. JoinMoleculess can be used in three main ways to generate such outputs.

(1)   It can extract the coordinate specification from a Chem-3D or an XYZ (nuclear Cartesian coordinates) file, or from the bottom part of a Gaussian or Mopac output file to automatically place it in a Gaussian input format.

(2)   It can extract two sets of molecular coordinate specification from two such Chem-3D/ XYZ/ Mopac/ Gaussian (input/ output) file, and join them, in a manner to be specified by the user, into a combined-molecule output file in a Gaussian input format.

(3)   It can even facilitate in the generation of the partial optimization (POPT) input format for the Gaussian program.

We have used this package to perform tailor-made stacked supermolecules involving the drug and the base-pair, for all such stacking requirements.

2.6 COMPARISON OF HF/6-31G, HF/6-31G*, HF/6-31G**, DFT AND MP2 RESULTS

For typical application to computationally expensive molecules it is necessary to compare the results obtained from different theoretical methods. The *ab initio* methods with large basis set and MP2 calculation with diffused function are very much time consuming and sometimes even impossible to carry out calculations within limited time. So there must be some technical skill required for locating optimum stacked structures which are usually important for understanding sequence specificity of drugs. Hence the stacked models between tricyclic ring of azaacridine-4-carboxamide or acridine with sequences of DNA are similar to the stacking of aromatic rings like benzene and benzene, benzene and pyridine. It should be noted that the size of the molecule taken up for study are large enough for applying MP2 level of calculation. So it is proper to compare the optimum structure and interaction energies computed from different level of theory. Furthermore, the results open up an idea for the possibility of using even DFT and *ab initio* method with medium sized basis set in the calculation *[8]*.

References

1(a). Hobza P & Sponer J, *Chem Rev*, 99, **1999**, 3247.

(b). Hobza P, Sponer J, Polasek M, *J Am Chem Soc*, **1995**, 117, 792.

(c). Hobza P, Sandorfy C, *J Am Chem Soc*, **1987**, 109, 1302.

2(a). Sponer J, Kypr J, *J Biomol Struct Dyn*, **1990**, 7, 1211.

(b). Sponer J, Kypr J, *J Biomol Struct Dyn*, **1991**, 221, 761.

(c). Sponer J, Kypr J, *J Biomol Struct Dyn*, **1993**, 11, 27.

(d). Sponer J, Jursa J, Kypr J, *Nucleoside Nucleotide*, **1994**, 13, 1669.

(e). Sponer J, Kypr J, *J Biomol Struct Dyn*, **1993**, 11, 277.

(f). Sponer J, Hobza P, *J Am Chem Soc*, **1994**, 116, 709.

(g). Sponer J, Hobza P, *J Biomol Struct Dyn*, **1994**, 11, 1357.

(h). Sponer J, Hobza P, *J Biomol Struct Dyn*, **1994**, 12, 671.

(i). Sponer J, Hobza P, *J Phys Chem*, **1994**, 98, 3161.

3(a). Colson A O, Besler B, Sevilla M D, *J Phys Chem*, **1993**, 97, 13852.

(b). Colson A O, Besler B, Close D M, Sevilla M D, *J Phys Chem*, **1992**, 96, 661.

4. Hrouda V, Florian J, Hobza P, *J Phys Chem*, **1993**, 97, 1542.

5. Gould I R, Kollman P A, *J Am Chem Soc*, **1994**, 116, 2493.

6. Florian J, Leszezynski J, *J Biomol Struct Dyn*, **1995**, 12, 1055.

7(a). Sponer J, Leszezynski J, Hobza P, *J Phys Chem*, **1996**, 100, 1965.

(b). Sponer J, Leszezynski J, Hobza P, *Computational Chemistry: Reviews of Current Trends, World Scientific Publisher, Singapore*, **1996**, p 185.

8. Kalita R & Medhi C, *Proc of Curr Trends in Chem Res (Guwahati)*, **2003**, 41.

9. Frisch M J, Trucks G W, Schlegel H B, Gill P M W, Johnson B G, Robb M A, Cheeseman J R, Keith T, Petersson G A, Montgomery J A, Raghavachari K, Al-Laham, M A, Zakrzewaki V G, Ortiz J V, Foresmann J B, Ciolowski J, Stefanov B B, Namayakkara A, Challacombe M, Peng C Y, Ayala P Y, Chen W, 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 **94**; Gaussian Inc, Pittsburgh PA, **1995**.