Chapter III

Geometry and Conformation of Molecules in Aqueous Solution

The first and the second chapter have been devoted to the principles of the non-empirical (ab initio) and the semi-empirical (CNDO) method which are the two main methods of calculation of this work. This chapter will try to outline some general ideas of studying solutions by quantum chemistry.

Most relevant chemical and biological systems are present in solution and hence various solvent effects will be associated with their chemical behaviour. The first step in studying such systems is the solvation. The interaction energy between the solute and solvent should be known, and the molecular conformation has to be studied under the effect of the solvent.

In the computational procedure, various computational models have been employed by theoretical chemists in the past.

The continuous model

This model does not take into account the molecular structure of the solvent, which is considered as a continuous dielectric. The general framework of the calculation is that a perturbation potential $V(r)$ will be constructed and added to the SCF Hamiltonian, which is solved iteratively.

$$H = H_{SCF} + V(r) \qquad\qquad 3.1$$

There are several approaches to set the potential $V(r)$. For example,

Klopman(15) constructed the potential V(r) by introducing the well known Born equation

$$H = H_{SCF} - z_i^2 \, e^2/2r_i 4\pi\epsilon_0 \, (1-1/\epsilon_r)$$ 3.2

Recently, Bonaccorsi, R., Palla, P., and Tomasi, E. (7) have constructed the potential V(r) as

$$V(r) = \int \frac{\sigma(s)}{\Sigma |r-s|} ds = \sum_k \sum_i \frac{q_{ki}}{|r-s|} (\Delta S_{ki})^{-1}$$ 3.3

where $\sigma(s) = \frac{\epsilon - 1 (\partial V_{tot})}{4\pi\epsilon (\partial n)}$

Where the potential V(r) is a function of the dielectric constant and also represents the cavity of the solute according to its conformation.

The super molecule approach

In this model, the molecular structure of the solvent is fully taken into consideration. Solvation of the solute is described by a formation consisting of a solute molecule and a set of solvent molecules. This formation is called the "super molecule". Rigorously speaking, the solvation is calculated in the gaseous phase due to the small amount of solvent molecules. The data from the experiment indicate that the interaction of a solvated molecule or ion with the molecules forming the first solvation layer represents the most important energy contribution to the solvation energy at infinite dilution (16). Hence the supermolecule model is quite suitable tool for the solution chemist.

The central problem is how many solvent molecules must be included in the model in order to represent the properties of the solvated molecule correctly. The more complete the solvation shell, the more information will be obtained. However, including only a part of the solvation shell has already proven to be quite useful in certain systems by Pullman, A. and Pullman, B. (1).

In this work, where we study the influence of a solvent on the internal rotation of an amino acid, we will rely on the supermolecule model and use a semi-empirical method (CNDO) and the non-empirical method (ab initio) in our calculations.

Monte-Carlo simulation

This procedure has been developed and applied by many researchers, e.g. Metropolis, N. (17), Clementi, E. etc.(18) In the first step the interaction energies of all solutes and solvent molecules with each other are evaluated by quantum chemical calculation for the whole energy surface. The interaction can be computed by ab initio or in various approximations. In the second step, these interaction potentials have to be fitted to an analytical expression that should reproduce the computed interaction energy surface. The third step makes use of these analytical potentials for the evaluation of energy for some 10^{-6} randomly generated multiparticle ensembles which allows us to evaluate structural data of the solution at a given tempearature.

However, the technique does not give intramolecular data such as the internal rotation barrier. For these effects one has to retreat to the super molecule approach again, after having determined the most probable structure elements of solution by the Monte-Carlo technique.

Determination of the interaction energy in the super-molecule approach

In quantum-chemical calculations of the interaction energy ΔE between two systems (A,B) based on the molecular orbital method, the value of ΔE is determined as the difference between the supersystem energy E_{AB} and the sum of the subsystem energies (E_A, E_B)

$$\Delta E = E_{AB} - (E_A + E_B) \qquad\qquad 3.4$$

In this work the interaction energy will always be determined by this expression, which seems to be quite simple. However, serious difficulties are encountered in the numerical treatment. The most important difficulty lies in the fact that the supersystem energy and the sum of the energies of the isolated subsystems are very large number compared to the interaction energy. Therefore, the determination of the energy of the systems must be extremly accurate $(10^{-5}$ Hartree). A further factor influencing the ΔE value is the effect of the basis set on the value of the SCF interaction energy. In general, the basis set employed should correctly express the multipole moments and the polarizability of the systems. Inclusion of polarization functions would be necessary for this.

However, in calculations of the interaction energy for more complex systems, we are forced to employ small basis sets, in most cases even minimal basis sets and a correction will be needed to compensate errors in the interaction energy.

The basis set superposition error

When insufficient basis sets are used, an artificial basis set improvement will take place in the supersystem, leading to an error

which has been known as basis set superposition error, and the interaction energies are always overestimated. To explain the error can be attempted using figure 3.1
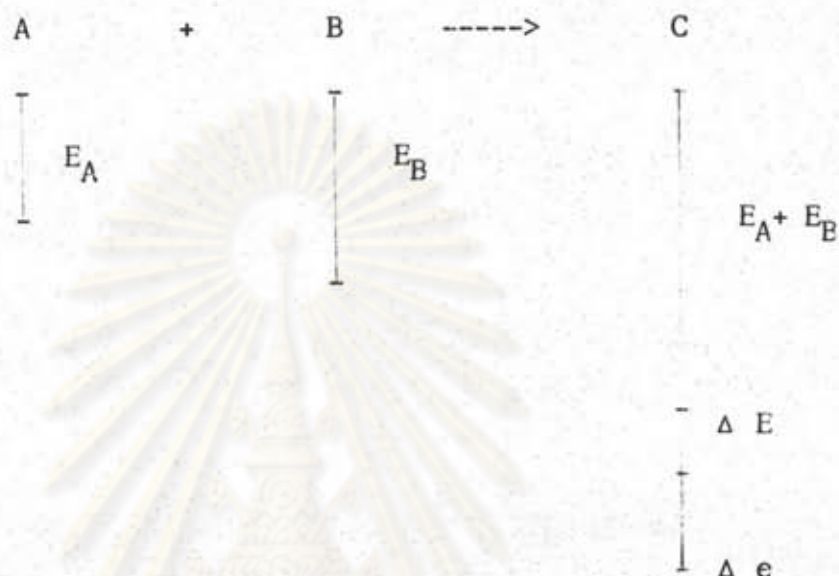
$$A \quad + \quad B \quad ----> \quad C$$

| E_A | E_B | $E_A + E_B$ |

ΔE

Δe

Figure 3.1   The origin of artificient stabilization Δe

Let there be two subsystems A and B forming supersystem (A-B). The formation of this supersystem is accompanied by a gain of energy ΔE. Subsystems A and B are described by means of equivalent basis sets. The associated energies are E_A and E_B. If the supersystem basis set quality remained identical with that of the subsystems, the supersystem energy would be the sum of the two subsystem energies plus the interaction energy. However, the supersystem energy is determined by using the basis set of both subsystems E_A and E_B and hence the supersystem is described by a larger basis set compared to the subsystem basis set. The energy of the complex (A-B) then consists of the sum of the energies of the two subsystems $(E_A + E_B)$, the interaction energy (ΔE), and the additional stabilization (Δe). However, if

extended basis sets are employed, the energy of each part of the supersystem is described sufficiently well by the basis set of each subsystem alone, so that a further extension in the basis set by function of the other subsystem has little effect.

The upper limit of this additional stabilization Δe can be estimated by the counterpoise method proposed by Boy and Bernardi (19). The method is based on the determination of the subsystem energies using the same basis set as for the determination of the supersystem energy. The procedure for the calculation of the interaction energy will be demonstrated with the interaction of two subsystems A and B. First the supersystem energy (E_{AB}) is calculated for a certain distance R with the basis set of A plus B. Second, the energies of subsystem A (E_A) and B (E_B) are computed with their own basis sets. Then a calculation of the energy of the subsystems including the basis set of the other subsystem $(E_{A(B)}$ and $E_{B(A)})$ is performed. The additional stabilization is then determined from

$$\Delta e = (E_A - E_{A(B)}) + (E_B - E_{B(A)}) \qquad 3.5$$

and the counterpoise corrected interaction energy is

$$\Delta E^{cp} = \Delta E + \Delta e \qquad 3.6$$

where ΔE is interaction energy which calculate from truncated basis set

As this procedure only gives the upper limit of additional stabilization Δe and in some case equation 3.6 yields destabilized corrected interaction energy so we have suggested another way of obtaining reasonable absolute energy values in this work. This

procedure also makes use of a scaling factor f^{HF}, which is determined by a calculation of a small but similar system at the Hartree-Fock limit and a calculation using the minimal basis set employed for the large system.

$$f^{HF} = \frac{E^{HF}}{E^{GLO}}$$
3.7

The overestimated interaction energies obtained with the small basis set are then multiplied by the factor to approximate the absolute values for stabilization energies. In our case, the water dimer was used as a model system for evaluation of E^{HF}.

Determination of interaction energies from semi-empirical method (CNDO)

In order to understand the basic errors connected with the use of the CNDO method in the calculation of the interaction energies, the theoretical principles must be briefly reconsidered. The CNDO method implies two rough physical approximations: the neglect of inner shell electrons and the neglect of most electron-electron interactions. In addition, we make use of empirical parameters. The various influences of these approximations have been the subject of many papers (16,20) and can be evaluated only roughly by comparing ab initio data with those of semi-empirical calculations in a specific system.

Population analysis.

There is no unique definition of the number of electrons to be associated with a given atom or nucleus in a molecule, but it is still sometimes useful to perform such population analyses. Since

$$N = 2 \sum_{a}^{N/2} \langle \psi_a(r) | \psi_a(r) \rangle \qquad 3.8$$

divides the total number of electrons into two electrons per molecular orbital, by substituting the basis expansion of ψ_a into (3.8), we have

$$N = \sum_{u} \sum_{v} P_{uv} S_{uv} = \sum_{u} (PS)_{uu} = tr\ PS \qquad 3.9$$

and it is possible to interpret $(PS)_{uu}$ as the number of electrons to be associated with φ_u. This is called a Mulliken population analysis. Assuming the basis functions are centered on atomic nuclei, the corresponding number of electrons to be associated with a given atom in a molecule are obtained by summing over all basis functions centered on that atom. The net charge associated with an atom is then given by

$$q_A = Z_A - \sum_{u \in A} (PS)_{uu} \qquad 3.10$$

where Z_A is the charge of atomic nucleus A; the index of summation indicated that we only sum over the basis functions centered on A. The definition (3.9) is by no means unique. Since tr AB = tr BA,

$$N = \sum_{u} (S^{\alpha} P S^{\alpha-1})_{uu} \qquad 3.11$$

for any α with $\alpha = 1/2$, we have

$$N = \sum_{u} (S^{1/2} P S^{1/2})_{uu} = \sum_{u} P'_{uu} \qquad 3.12$$

where we can show that P' is the density matrix in terms of a

symmetrically orthogonalized basis set,

$$\rho\ (r) = \underset{uv}{\Sigma\Sigma}\ P_{uv}' \left\langle \phi_u'(r) \middle| \phi_v'(r) \right\rangle \qquad\qquad 3.13$$

$$\phi_u'\ (r) = \underset{v}{\Sigma}\ (S^{-1/2})_{uv}\ \phi_v(r) \qquad\qquad 3.14$$

The diagonal elements of P' are commonly used for a Lowdin population analysis

$$q_A = Z_A - \underset{u}{\Sigma}\ \underset{v}{\Sigma} (\ S^{1/2} P S^{1/2})_{uu} \qquad\qquad 3.15$$

None of these population analysis schemes is unique, but they are often useful when comparing different molecules using the same type of basis set for each molecule.