CHAPTER 3

THEORETICAL BACKGROUND

To develop the potential function by means of quantum chemical calculations, the following steps are required: (i) Selection of representative geometries of the pairs; (ii) The *ab initio* calculations; (iii) Fitting of pair interaction energies to a functional form and (iv) Improving the quality of the function.

The program structure is given in Figure 3.1. As mentioned earlier this work focuses only on the first two steps. The other part can be found elsewhere [5].

3.1 Selection of representative geometries of the pairs

This step involves the selection for the configurations of two molecules to be used to perform *ab initio* computations. An idea is to generate as many complex configurations as possible covering the space around each other up to the distance which the interaction disappears. The appropriate choice of selection can reduce the repetition and save calculation time for the interaction energies. Consider the system in which molecule A that was fixed within the Cartesian coordinate system, and molecule B that was placed at different positions within the space around A. In this case, the distance between atoms of both molecules can not be small because of the strong repulsion between adjacency of two molecules. Therefore, this configuration will not use in the simulations, as well as in reality. In addition, the distance between the two molecules must be far enough so that the interaction between them is close to zero. There is no known algorithm to determine the number of configurations and

the distance among each molecules. Justification depends strongly on the size and symmetry of the two molecules.
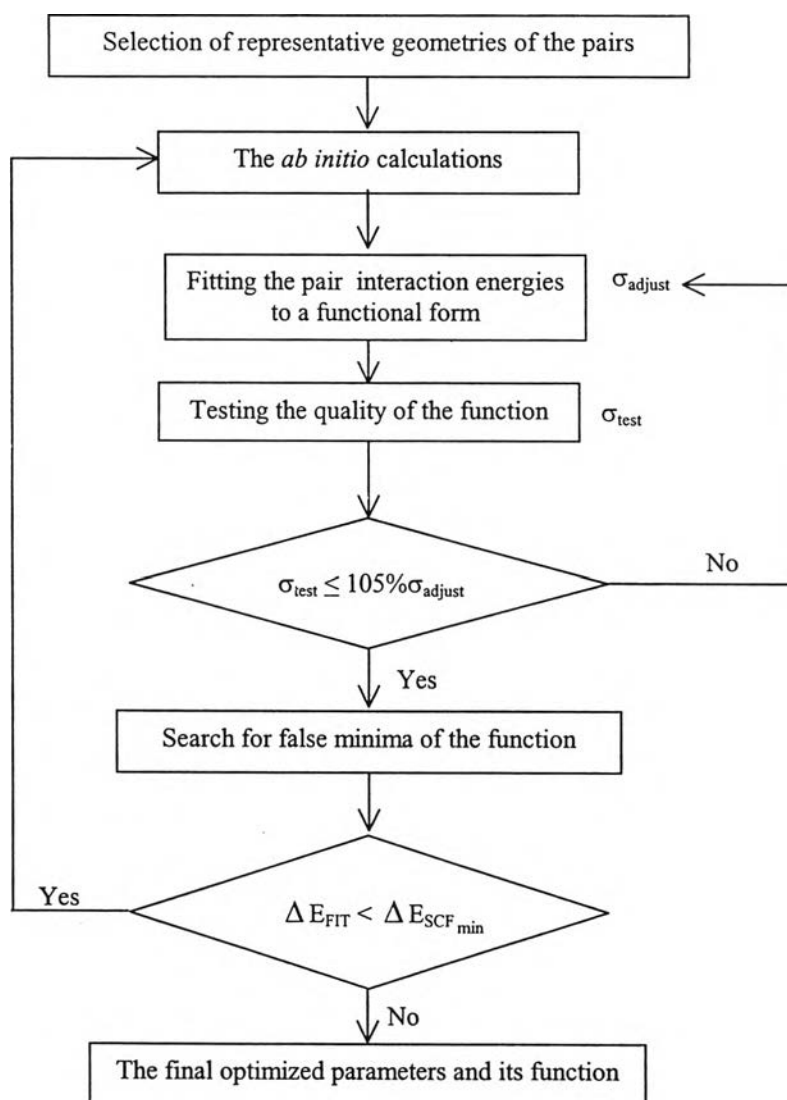

Figure 3.1 The procedure for constructing the potential function by means of quantum chemical calculations.

3.2 The *ab initio* calculations

In this step, quantum chemical, known as Self-Consistent-field (SCF), calculations were performed for the aforementioned configurations. In the SCF calculations, one must choose the appropriate method and level of accuracy. The later is normally represented by the so called "basis function" which is a function required for quantum chemical calculations. The interaction energy (ΔE_{SCF}) between the two molecules A and B based on the SCF method is determined by the difference between the super-system energy (E_{AB}) and the sum of the sub-system energies (E_A and E_B),

$$\Delta E_{SCF}(\text{kcal/mol}) = [E_{AB} - (E_A + E_B)] \times 627.5 \qquad (3.1)$$

where E_{AB}, E_A and E_B are total energies of the system given in the subscript (in atomic units). The factor of 627.5 was used to convert atomic unit to kcal/mol.

The basis set is a key factor controlling the quality of the Potential function. In addition, the case of the basis set with different size is known to lead to an error, known as "Basis Set Superposition Error (BSSE)". Therefore, we will give a treatment for this deficiency in the next section.

3.2.1 Basis Set

In quantum theory, the molecular orbitals ψ_i are expanded in the linear combination of a set of orthogonal functions

$$\psi_i = \sum_{\mu=1}^{N} c_{\mu i} \phi_\mu \qquad (3.2).$$

The functions ϕ_μ are known as one-electron basis function or simply as basis function, the set of this function is called "basis set" and $c_{\mu i}$ are the molecular orbital expansion coefficients, whose variation allows to minimize the energy.

The quality of the molecular orbitals is related to the quality of the basis set used. There are normally two types of basis functions used in calculation, Slater Type Orbital (STO), and Gaussian Type Orbital (GTO).

3.2.1.1 Slater Type Orbital (STO)

The STO basis function, which was originally introduced by Slater [17], is based on approximations of hydrogen-like atomic orbitals. Although, the advantage of using this type of orbital is that only a few functions are needed for good description of wave function since, STOs are more accurate representations of atomic orbitals, the integration of the functions is largely time consuming. The form of STO is denoted by the expression;

$$\phi^{STO} = Nr^{n-1}e^{-\zeta r}Y_l^m(\theta, \phi) \tag{3.3}$$

where r, θ, and ϕ are spherical polar coordinate. N, ζ, n, and $Y_l^m(\theta, \phi)$ are the normalization constant, the positive orbital exponent, principle quantum number, and the angular part of the wave function, respectively.

3.2.1.2 Gaussian Type Orbital (GTO)

The GTO is an alternative to STO, which represents a STO by some of Gaussian type functions of the form

$$\phi^{GTO} = Nx^a y^b z^c e^{-\alpha r^2} \qquad (3.4)$$

where x, y, and z are Cartesian coordinate and a, b, and c are nonnegative integers.

Gaussian type functions were introduced into molecular orbital computations by Boy [18]. The integration of this function is easy. Due to the different between the shape of the GTO and the shape of STO functions, a near region from nucleus, a combination of GTOs with different exponent is required to obtain equivalent results. For example, the symbol STO-3G for a basis means that each STO is approximated by a linear combination of three GTOs. The coefficients being chosen to minimize the difference between the STO and its STO-3G approximation. Currently, there are many GTO expansions in use which have been optimized for molecular calculations.

The orbital reviews of basis set can be see from reference [19].

3.2.2 Basis Set Superposition Error (BSSE)

In calculating the interaction energy for a complex system, the basis set employed should be sufficiently large and correctly express the multiple moments and the polarizability of the system. In the case that an insufficient basis set is used, an artificial basis set improvement will take place in the complex, leading to this type of error. In a complex consisting of two monomers A and B which represented by the basis set ϕ_A and ϕ_B, respectively, the error occurs when the basis set ϕ_A is contaminated by ϕ_B in computing the energy of A or conversely, when ϕ_B is contaminated by ϕ_A in computing the energy of B. In this case, the interaction energy of the complex is always overestimated [20].

The counterpoise (CP) method proposed by Boys and Bernardi [21] was carried out to estimate this error. The method is based on the determination of the monomer energies using the same basis set as for the determination of the complex energy and can be presented as follows. First, the energy of complex AB with basis sets ϕ_{AB} is computed, yielding an energy $E_{AB}(\phi_{AB})$. Second, the computation of the energy of A with ϕ_{AB}, disregarding the nuclear charge in B, is performed to obtain an energy $E_A(\phi_{AB})$. The operation is repeated for B with ϕ_{AB} as basis set, yielding the energy $E_B(\phi_{AB})$. The interaction energy of whole system, ΔE, can be defined as

$$\Delta E = E_{AB}(\phi_{AB}) - [E_A(\phi_A) + E_B(\phi_B)] \tag{3.5}$$

where

$$\phi_{AB} = \phi_A \cup \phi_B \tag{3.6}$$

$E_A(\phi_A)$ and $E_B(\phi_B)$ are total energies of A and B resulted from the computation of A with ϕ_A alone and of B with ϕ_B alone, respectively, and counterpoise correction, Δe, can be defined as

$$\Delta e = [E_A(\phi_A) - E_A(\phi_{AB})] + [E_B(\phi_B) - E_B(\phi_{AB})] \tag{3.7}.$$

The counterpoise corrected interaction energy, ΔE_{CP}, is

$$\Delta E_{CP} = \Delta E + \Delta e \tag{3.8}$$

where ΔE is the interaction energy calculated from the truncated basis set,

$$\Delta E_{CP} = E_{AB}(\phi_{AB}) - [E_A(\phi_{AB}) + E_B(\phi_{AB})] \tag{3.9}.$$

It should be mentioned, however, that this correction gives the upper limit of a possible BSSE and hence can not be regarded as an absolutely correct value for the real BSSE occurring in the simulation of a specific complex.

The ΔE_{SCF} calculated with BSSE corrections, ΔE_{BSSE}, can be defined as

$$\Delta E_{BSSE}(\text{kcal/mol}) = [E_{AB}(\phi_{AB}) - (E_A(\phi_{AB}) + E_B(\phi_{AB}))] \times 627.5 \qquad (3.10)$$

where $E_{AB}(\phi_{AB})$ is the total energy of the super-system computed with basis sets ϕ_{AB}. $E_A(\phi_{AB})$ and $E_B(\phi_{AB})$ are the total energies of the sub-system E_A and E_B with basis sets ϕ_{AB}, respectively.

3.3 Fitting of pair interaction energies to a functional form

In the energy fitting step, a suitable mathematical function is selected. In general, there are two general forms of potential functions which can be expressed as:

$$\Delta E_{MN} = \sum_i^K \sum_j^L \left(-\frac{A_{ij}}{r_{ij}^x} + \frac{B_{ij}}{r_{ij}^y} + \frac{q_i q_j}{r_{ij}} + Z_{ij} \right); x \neq y \qquad (3.11)$$

$$\Delta E_{MN} = \sum_i^K \sum_j^L \left(-\frac{A_{ij}}{r_{ij}^x} + B_{ij} \exp(-C_{ij} r_{ij}) + \frac{q_i q_j}{r_{ij}} + Z_{ij} \right) \qquad (3.12)$$

where r_{ij} is the distance between the ith atom of molecules M and the jth atom of molecules N, K and L are number of atoms on molecules M and N, respectively, A_{ij}, B_{ij} and C_{ij} are fitting parameters which represent interaction between atom i of A

and atom j of B, and q_i, q_j are the atomic net charge of each atom, atom i of A and atom j of B. These charges were obtained from Mulliken population's analysis [22] during the SCF calculations of this isolated molecule. ΔE is the interaction energy. The first two terms formally describe the short-range attractive and repulsive interactions, respectively. The third term describe the long-range Coulomb interaction, and the last terms, which has the form as in equation (3.13), will be added in order to obtain good statistical value.

$$ Z_{ij} = 0 + \frac{E_{ij}}{r_{ij}^m} + \frac{F_{ij}}{r_{ij}^n} + \frac{G_{ij}}{r_{ij}^o} + ... \qquad (3.13). $$

Once the analytical form of the potential has been chosen, the interaction energies of the pair were fitted by using a multidimensional non-linear least-squares procedure. Suppose that the fit yields the standard deviation of σ_{fit}. More details of the selection of the functional form and the numerical fitting can be found elsewhere [5].

3.4 Improving the quality of the function

3.4.1 Testing the quality of the function

This step is required in order to test the predictive capabilities of the potential functions. The well known procedure is that suggested by Beveridge et al. [23], i.e., the SCF energies (outside the original set) are calculated and compared with the values predicted by the function. These points are then included in the fitting procedure and the whole process was repeated until constancy of the fitting parameters is obtained within a range of $\pm$ 5% and a sufficiently low standard deviation is reached. The above mentioned procedure can be summarized below:

(1) Calculate the energy points using SCF method (ΔE_{SCF}) and the pair potential function (ΔE_{FIT}).

(2) Calculate the standard deviation from (1), σ_{test}.

(3) Compare all ΔE_{SCF} and ΔE_{FIT} in term of σ_{test}:

- If $\sigma_{test} > 105\%$ of σ_{adjust} then go back to step of fitting function and included tested points in the fitting procedure and the whole processes are repeated.

- If $\sigma_{test} \leq 105\%$ of σ_{adjust} then go to step of search false minima of the function.

3.4.2 Search for false minima of the function

In this step, false minima which may be available in the function will be searched. This checking procedure is also essential. The procedure is:

(1) Create the three-dimensional grid covering the whole space around molecule A. Grid size should be as small as possible, for example, 0.1 Å.

(2) Molecule B is located at the center of the grid.

(3) Calculate interactive energy using the fitted potential.

(4) Turn molecule B in all 3 axes, with a small step size, such as $5°$, and calculate energy as in (3).

(5) Compare the interaction energies yielded from (3) and (4) with these obtained from quantum chemical calculations.

- If the function contains artificial minima the data paint which lower than the SCF grobal minima go back to step of calculate energy. Then these points are included in the fitting procedure and the whole processes are repeated.

- If the artificial minima is not found, then the obtained function is quality.