

# **Chapter 1**

## **Molecular Modelation**



# Chapter 1

## Molecular Modelation

NORMA AUREA RANGEL-VÁZQUEZ

*División de Estudios de Posgrado e Investigación del Instituto Tecnológico de Aguascalientes, Ave. López Mateos # 1801 Ote. Fracc. Bona Gens CP. 20256 Aguascalientes, Aguascalientes, México*

### **Abstract**

*Model is the process of producing a model; a model is a representation of the construction and working of some system of interest. A model is similar to but simpler than the system it represents. One purpose of a model is to enable the analyst to predict the effect of changes to the system. On the one hand, a model should be a close approximation to the real system and incorporate most of its salient features. On the other hand, it should not be so complex that it is impossible to understand and experiment with it. A good model is a judicious tradeoff between realism and simplicity.*

*Simulation practitioners recommend increasing the complexity of a model iteratively. An important issue in modeling is model validity. Model validation techniques include simulating the model under known input conditions and comparing model output with system output. Generally, a model intended for a simulation study is a mathematical model developed with the help of simulation software. Mathematical model classifications include deterministic (input and output variables are fixed values) or stochastic (at least one of the input or output variables is probabilistic); static (time is not taken into account) or dynamic (time-varying interactions among variables are taken into account). Typically, simulation models are stochastic and dynamic.*

*Keywords: Modelling, PM3, AM1, Simulation*

## 1.1 Introduction

Attractive interactions between  $\pi$  systems are one of the principal noncovalent forces governing molecular recognition and play important roles in many chemical and biological systems. Attractive interaction between  $\pi$  systems is the interaction between two molecules leading to self-organization by formation of a complex structure which has lower conformation equilibrium than of the separate components and shows different geometrical arrangement with high percentage of yield. It is known that many aromatic compounds form infinite stacks with parallel molecular planes when crystallized.

Hydrogen bonding, electrostatic interactions, van der Waals interactions (van der Waals bonds are mainly constructed with a balance of the exchange repulsion and dispersion attractive interactions), donor–acceptor interactions, hydrophilic–hydrophobic interactions, and  $\pi$ – $\pi$  interactions are the main types of non-covalent interactions that are responsible for self-organization in biological systems. A lot of experimental evidence of charge transfer (CT) complexes had been reported in solid or in solution in a different field of chemistry. According to Mulliken's theory, formation of the (CT) complex involves transition of an electron from HOMO of donor to LUMO of acceptor.

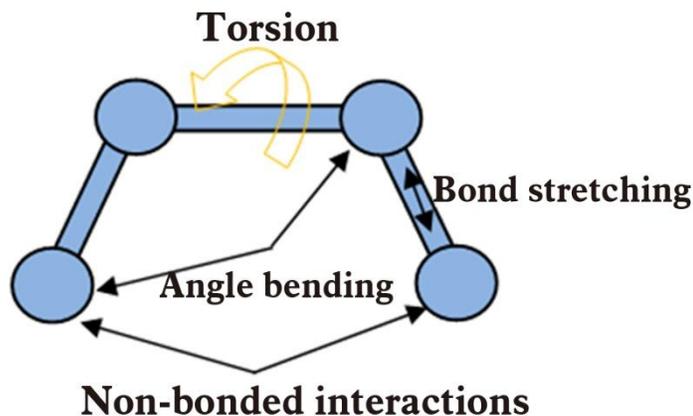
Opposing  $\pi$  systems typically adopt a parallelplaner (stacked or offset-stacked) geometry. The interaction between the donor and acceptor is characteristic of an electronic absorption band with low energy. One of these molecular complexes is  $\pi$ ,  $\pi$ -complex between neutral molecules.

Quantum chemical calculations are applied to study the (CT) complexes in order to obtain information on structures and other molecular properties like specific interaction of donor and acceptor. The interaction energy contribution comes from the effects of donor–acceptor interactions and  $\pi$ – $\pi$  interactions. The self-assembly of these molecular systems was studied using quantum mechanical semi-empirical methods in gas phase. The minimum energy configuration of the stacked molecular systems were investigated by geometry optimizations and then the other properties, such as stacking distances, heat of formation, dipole moment, and polarizability were also calculated [1].

## 1.2 Molecular Mechanics

The mechanical molecular model considers atoms as spheres and bonds as springs. The mathematics of spring deformation can be used to describe the

ability of bonds to stretch, bend, and twist (Figure 1.1).



*Figure 1.1 The mechanical molecular model.*

Non-bonded atoms (greater than two bonds apart) interact through van der Waals attraction, steric repulsion, and electrostatic attraction/repulsion. These properties are easiest to describe mathematically when atoms are considered as spheres of characteristic radii. The object of molecular mechanics is to predict the energy associated with a given conformation of a molecule. However, molecular mechanics energies have no meaning as absolute quantities. Only differences in energy between two or more conformations have meaning. A simple molecular mechanics energy equation (1) is given by:

$$\text{Energy} = \text{Stretching Energy} + \text{Bending Energy} + \text{Torsion Energy} + \text{Non - Bonded Interaction Energy} \quad (1)$$

These equations (1) together with the data (parameters) required to describe the behavior of different kinds of atoms and bonds, is called a force-field. Many different kinds of force-fields have been developed over the years. Some include additional energy terms that describe other kinds of deformations. Some force-fields account for coupling between bending and stretching in adjacent bonds in order to improve the accuracy of the mechanical model. The mathematical form of the energy terms varies from force-field to force-field. The more common forms will be described [2].

- Stretching Energy

The stretching energy equation is based on Hooke's law. The "kb" parameter controls the stiffness of the bond spring, while "ro" defines its equilibrium

length. Unique "kb" and "ro" parameters are assigned to each pair of bonded atoms based on their types (e.g. C-C, C-H, O-C, etc.). Where "r" is the distance between two atoms, this equation (2) estimates the energy associated with vibration about the equilibrium bond length. This is the equation of a parabola (Figure 1.2), as can be seen in the following plot.

$$E = \sum_{\text{bonds}} k_b (r - r_o)^2 \quad (2)$$

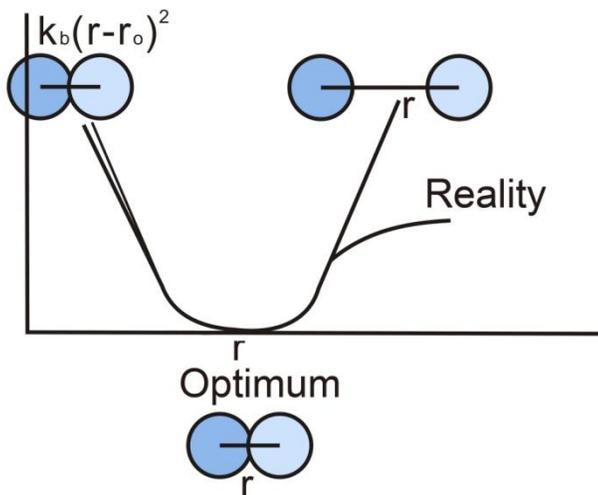


Figure 1.2 Equation of parabola.

Notice that the model tends to break down as a bond is stretched toward the point of dissociation [2].

- Bending Energy

The bending energy equation is also based on Hooke's law. The "ktheta" parameter controls the stiffness of the angle spring, while "thetao" defines its equilibrium angle (Figure 1.3). This equation (3) estimates the energy associated with vibration about the equilibrium bond angle.

$$E = \sum_{\text{angles}} k_{\theta} (\theta - \theta_o)^2 \quad (3)$$

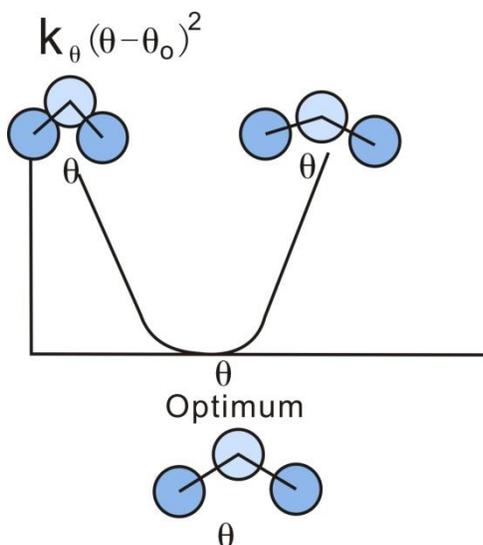


Figure 1.3 Parabola graphic.

Unique parameters for angle bending are assigned to each bonded triplet of atoms based on their types (e.g. C-C-C, C-O-C, C-C-H, etc.). The effect of the "kb" and "ktheta" parameters is to broaden or steepen the slope of the parabola. The larger the value of "k", the more energy is required to deform an angle (or bond) from its equilibrium value. Shallow potentials are achieved for "k" values between 0.0 and 1.0 [2]. The Hookeian potential is shown in the following plot for three values of "k" (Figure 1.4).

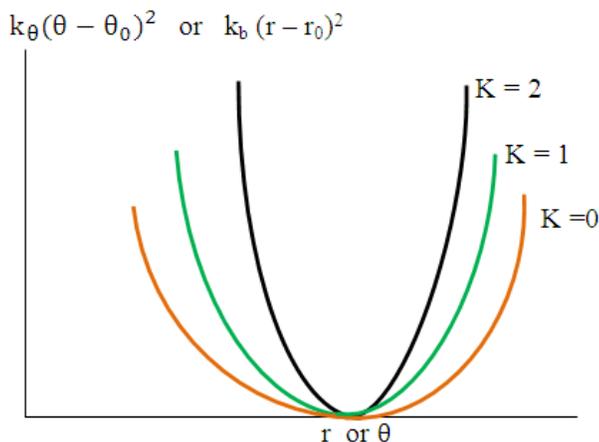
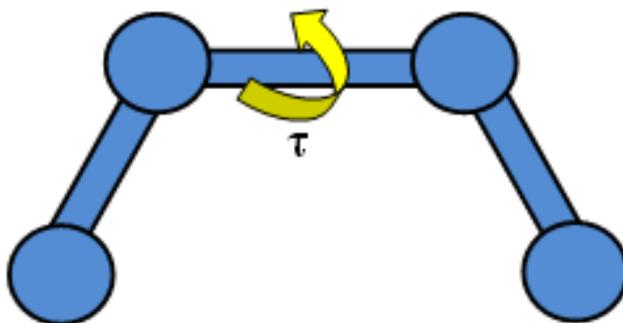


Figure 1.4 Hookein potential.

- Torsion Energy

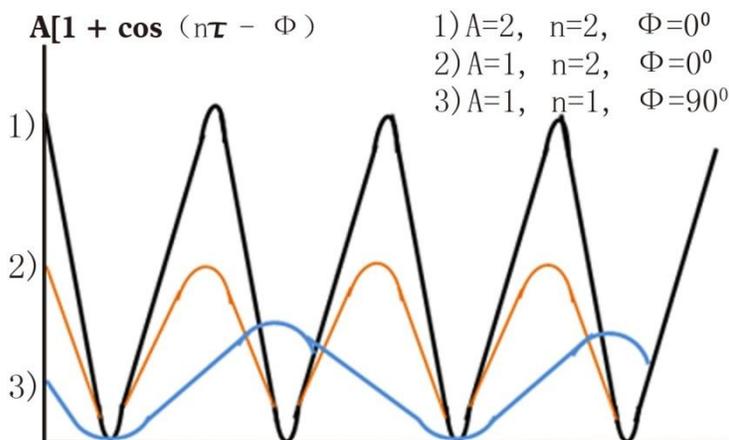
Intramolecular rotations (rotations about torsion or dihedral angles) require energy (equation 4 and Figure 1.5). Torsional energies are usually important only for single bonds because double and triple bonds are too rigid to permit rotation.

$$E = \sum_{\text{torsions}} A[1 + \cos(n\tau - \phi)] \quad (4)$$



*Figure 1.5 Torsion bond.*

The torsion energy in molecular mechanics is primarily used to correct the remaining energy terms rather than to represent a physical process. The torsional energy represents the amount of energy that must be added to or subtracted from the Stretching Energy + Bending Energy + Non-Bonded Interaction Energy terms to make the total energy agree with experiment or rigorous quantum mechanical calculation for a model dihedral angle (ethane, for example might be used a model for any H-C-C-H bond). The "A" parameter controls the amplitude of the curve, the n parameter controls its periodicity, and "phi" shifts the entire curve along the rotation angle axis (tau). The parameters are determined from curve fitting. Unique parameters for torsional rotation are assigned to each bonded quartet of atoms based on their types (e.g. C-C-C-C, C-O-C-N, H-C-C-H, etc.) (Equation 4). Torsion potentials with three combinations of "A", "n", and "phi" are shown in the following Figure 1.6.



**Figure 1.6** Torsion potentials with three combinations.

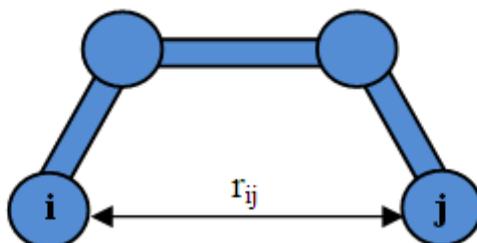
Notice that "n" reflects the type symmetry in the dihedral angle. A  $\text{CH}_3\text{-CH}_3$  bond, for example, ought to repeat its energy every 120 degrees. The *cis* conformation of a dihedral angle is assumed to be the zero torsional angles by convention. The parameter phi can be used to synchronize the torsional potential to the initial rotameric state of the molecule whose energy is being computed [2].

- Non-Bonded Energy

The non-bonded energy represents (equation 5) the pair-wise sum of the energies of all possible interacting non-bonded atoms *i* and *j* (Figure 1.7).

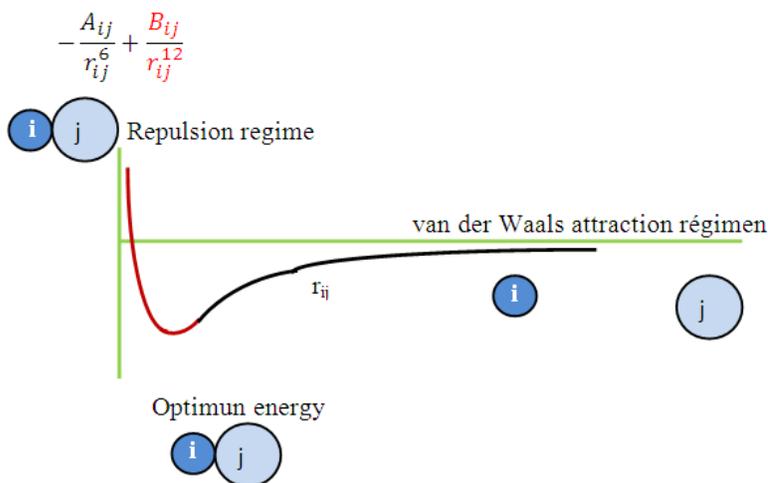
$$E = \sum_i \sum_j -\frac{A_{ij}}{r_{ij}^6} + \frac{B_{ij}}{r_{ij}^{12}} + \sum_i \sum_j \frac{q_i q_j}{r_{ij}} \quad (5)$$

van der Waals term      Electrostatic term



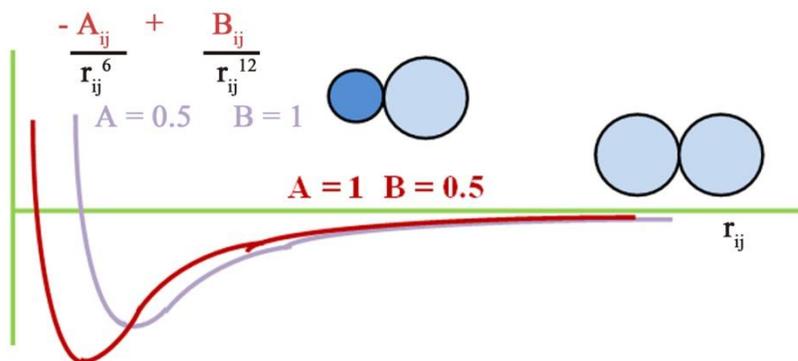
**Figure 1.7** Non-bonded atoms *i* and *j*.

The non-bonded energy accounts for repulsion, van der Waals attraction, and electrostatic interactions (Figure 1.8). Van der Waals attraction occurs at short range, and rapidly dies off as the interacting atoms move apart by a few Angstroms. Repulsion occurs when the distance between interacting atoms becomes even slightly less than the sum of their contact radii. Repulsion is modeled by an equation that is designed to rapidly blow up at close distances ( $1/r^{12}$  dependency). The energy term that describes attraction/repulsion provides for a smooth transition between these two regimes.



**Figure 1.8** Van der Waals and electrostatic attractions.

The "A" and "B" parameters control the depth and position (interatomic distance) of the potential energy (Figure 1.9) well for a given pair of non-bonded interacting atoms (e.g. C:C, O:C, O:H, etc.). In effect, "A" determines the degree of "stickiness" of the van der Waals attraction and "B" determines the degree of "hardness" of the atoms (e.g. marshmallow-like, billiard ball-like, etc.).



**Figure 1.9** *A and B parameters control of the potential energy.*

The "A" parameter can be obtained from atomic polarizability measurements, or it can be calculated quantum mechanically. The "B" parameter is typically derived from crystallographic data so as to reproduce observed average contact distances between different kinds of atoms in crystals of various molecules. The electrostatic contribution is modeled using a Coulombic potential. The electrostatic energy is a function of the charge on the non-bonded atoms, their interatomic distance, and a molecular dielectric expression that accounts for the attenuation of electrostatic interaction by the environment (e.g. solvent or the molecule itself). Often, the molecular dielectric is set to a constant value between 1.0 and 5.0. A linearly varying distance-dependent dielectric (i.e.  $1/r$ ) is sometimes used to account for the increase in environmental bulk as the separation distance between interacting atoms increases. Partial atomic charges can be calculated for small molecules using an *ab initio* or semiempirical quantum technique (usually MOPAC or AMPAC). Some programs assign charges using rules or templates, especially for macromolecules. In some force-fields, the torsional potential is calibrated to a particular charge calculation method (rarely made known to the user). Use of a different method can invalidate the force-field consistency [2].

### *AMBER Method*

The term "AMBER force field" generally refers to the functional form used by the family of AMBER force fields. This form includes a number of parameters; each member of the family of AMBER force fields provides values for these parameters and has its own name. The functional form of the AMBER force field is (equation 6).

$$\begin{aligned}
 V(\tau^N) = & \sum_{\text{bonds}} k_b (1 - l_0)^2 + \sum_{\text{angles}} k_a (\theta - \theta_0)^2 \\
 & + \sum_{\text{angles}} 0.5 V_N [1 + \cos(nw - \gamma)] \\
 & + \sum_{j=1}^{N-1} \sum_{i=j+1}^N \left\{ o_{ij} \left( \frac{ro_{ij}}{r_{ij}} \right)^{12} - 2 \left( \frac{ro_{ij}}{r_{ij}} \right)^6 + \frac{q_i q_j}{4\pi\epsilon_0 ro_{ij}} \right\}
 \end{aligned} \tag{6}$$

The meanings of right hand side terms are:

- 1) First term (summing over bonds): represents the energy between covalently bonded atoms. This harmonic (ideal spring) force is a good approximation near the equilibrium bond length, but becomes increasingly poor as atoms separate.
- 2) Second term (summing over angles): represents the energy due to the geometry of electron orbitals involved in covalent bonding.
- 3) Third term (summing over torsions): represents the energy for twisting a bond due to bond order (e.g. double bonds) and neighboring bonds or lone pairs of electrons. Note that a single bond may have more than one of these terms, such that the total torsional energy is expressed as a Fourier series.
- 4) Fourth term (double summation over  $i$  and  $j$ ): represents the non-bonded energy between all atom pairs, which can be decomposed into van der Waals (first term of summation) and electrostatic (second term of summation) energies.

The form of the van der Waals energy is calculated using the equilibrium distance ( $ro_{ij}$ ) and well depth ( $\epsilon$ ). The factor of 2 ensures that the equilibrium distance is  $ro_{ij}$ . The energy is sometimes reformulated in terms of  $o$ , where  $ro_{ij} = 2^{1/6} (o)$ , as used e.g. in the implementation of the soft core potentials. The form of the electrostatic energy used here assumes that the charges due to the protons and electrons in an atom can be represented by a single point charge (or in the case of parameter sets that employ lone pairs, a small number of point charges) [3].

### 1.3 Semi-Empirical Methods

Semi-empirical quantum chemistry methods are based on the Hartree–Fock formalism, but make many approximations and obtain some parameters from empirical data. They are very important in computational chemistry for treating

large molecules where the full Hartree–Fock method without the approximations is too expensive. The use of empirical parameters appears to allow some inclusion of electron correlation effects into the methods. Within the framework of Hartree–Fock calculations, some pieces of information (such as two-electron integrals) are sometimes approximated or completely omitted. In order to correct for this loss, semi-empirical methods are parametrized, that is their results are fitted by a set of parameters, normally in such a way as to produce results that best agree with experimental data, but sometimes to agree with *ab initio* results. Semi-empirical methods follow what are often called empirical methods where the two-electron part of the Hamiltonian is not explicitly included.

For  $\pi$ -electron systems, this was the Hückel method proposed by Erich Hückel [4-6]. For all valence electron systems, the extended Hückel method was proposed by Roald Hoffmann [7]. Semi-empirical calculations are much faster than their *ab initio* counterparts. Their results, however, can be very wrong if the molecule being computed is not similar enough to the molecules in the database used to parametrize the method. Semi-empirical calculations have been most successful in the description of organic chemistry, where only a few elements are used extensively and molecules are of moderate size. However, semi-empirical methods were also applied to solids [8] and nanostructures [9] but with different parameterization. As with empirical methods, we can distinguish methods that are: Restricted to  $\pi$ -electrons. These methods exist for the calculation of electronically excited states of polyenes, both cyclic and linear. These methods, such as the Pariser–Parr–Pople method (PPP), can provide good estimates of the  $\pi$ -electronic excited states, when parameterized well. Indeed, for many years, the PPP method outperformed *ab initio* excited state calculations [10-11].

### 1.3.1 AM1 Method

AM1 is basically a modification to and a reparameterization of the general theoretical model found in MNDO. Its major difference is the addition of Gaussian functions to the description of core repulsion function to overcome MNDO's hydrogen bond problem. Additionally, since the computer resources were limited in 1970s, in MNDO parameterization methodology, the overlap terms,  $\beta_s$  and  $\beta_p$ , and Slater orbital exponent's  $\zeta_s$  and  $\zeta_p$  for *s*- and *p*- atomic orbitals were fixed. That means they are not parameterized separately just considered as  $\beta_s = \beta_p$ , and  $\zeta_s = \zeta_p$  in MNDO. Due to the greatly increasing computer resources in 1985 comparing to 1970s, these inflexible conditions were relaxed in AM1 and then likely better parameters were obtained.

The addition of Gaussian functions significantly increased the numbers of parameters to be parameterized from 7 (in MNDO) to 13-19, but AM1 represents a very real improvement over MNDO, with no increase in the computing time needed. Dewar also concluded that the main gains of AM1 were its ability to reproduce hydrogen bonds and the promise of better estimation of activation energies for reactions. However, AM1 has some limitations. Although hypervalent molecules are improved over MNDO, they still give larger errors than the other compounds, alkyl groups are too stable, nitro compounds are too unstable, peroxide bond are too short. AM1 has been used very widely because of its performance and robustness compared to previous methods. This method has retained its popularity for modeling organic compounds and results from AM1 calculations continue to be reported in the chemical literature for many different applications [12].

### Theory

AM1 is currently one of the most commonly used of the Dewar-type methods. It was the next semiempirical method introduced by Dewar and coworkers in 1985 following MNDO. It is simply an extension, a modification to and also a reparameterization of the MNDO method. AM1 differs from MNDO by mainly two ways. The first difference is the modification of the core repulsion function. The second one is the parameterization of the overlap terms  $\beta_s$  and  $\beta_p$ , and Slater-type orbital exponents  $\zeta_s$  and  $\zeta_p$  on the same atom independently, instead of setting them equal as in MNDO. MNDO had a very strong tendency to overestimate repulsions between atoms when they are at approximately their van der Waals distance apart. To overcome this hydrogen bond problem, the net electrostatic repulsion term of MNDO,  $f(R_{AH})$  given by equation (7), was modified in MNDO/H to be

$$f(R_{AH}) = Z_A Z_B (S_A S_A | S_H S_H) \left[ e^{-\alpha \frac{R_{AH}^2}{R_{AH}}} \right] \quad (7)$$

Where  $\alpha$  was proposed to be equal to  $2.0 \text{ \AA}^{-2}$  for all A-H pairs. On the other hand, the original core repulsion function of MNDO was modified in AM1 by adding Gaussian functions to provide a weak attractive force. The core-core repulsion energy term in AM1 is given by equation 8.

$$E_{AB}^{AM1} = Z_A Z_B (S_A S_A | S_H S_H) \left[ 1 + e^{-\alpha} A^R AB + e^{-\alpha} B^R AB \right] + \frac{Z_A Z_B}{R_{AB}} [F(A) + F(B)] \quad (8)$$

The Gaussian functions  $F(A)$  and  $F(B)$  are expressed by equation 9.

$$\begin{aligned} F(A) &= \sum_i K_A i e^{-L_{A,i} (R_{AB} - M_{A,i})^2} \\ F(B) &= \sum_i K_B i e^{-L_{B,i} (R_{AB} - M_{B,i})^2} \end{aligned} \quad (9)$$

And finally AM1 core-repulsion function becomes (equation 10).

$$\begin{aligned} E_{AB}^{AM1} &= E_{AB}^{MNDO} \\ &+ \frac{Z_A Z_B}{R_{AB}} \left\{ \sum_i K_{A,i} e^{-L_{A,i} (R_{AB} - M_{A,i})^2} + \sum_j K_{B,j} e^{-L_{B,j} (R_{AB} - M_{B,j})^2} \right\} \end{aligned} \quad (10)$$

In this equation 10,  $K$ ,  $L$  and  $M$  are the Gaussian parameters. The remaining parameters have the same meaning as in the previous section.  $L$  parameters determine the widths of the Gaussians and were not found to be critical by Dewar. Therefore, a common value was used for many of the  $L$  parameters. On the other hand, all  $K$  and  $M$  parameters were optimized. Each atom has up to four of the Gaussian parameters, i.e.,  $K_1, \dots, K_4, L_1, \dots, L_4, M_1, \dots, M_4$ . Carbon has four terms in its Gaussian expansion whereas hydrogen and nitrogen have three and oxygen has two terms (only  $K_1, K_2, L_1, L_2, M_1, M_2$ ). Because in AM1 for carbon, hydrogen and nitrogen both attractive and repulsive Gaussians were used whereas for oxygen only repulsive ones considered, addition of Gaussian functions into the core-repulsion function significantly increased the number of parameters to be optimized and made the parameterization process more difficult. As for original MNDO, one-center two-electron repulsion integrals  $gss, gpp, gdd, gsp, hsp$  are assigned to atomic spectral values and not optimized. In contrast to MNDO, in which parameters were first optimized for carbon and hydrogen together and then other elements added one at a time, by increased computer resources and improved optimization procedure a larger reference parameterization dataset was used in the parameterization of AM1. All the parameters for H, C, N and O were optimized at once in a single parameterization procedure.

Optimization of the original AM1 elements was performed manually by Dewar using chemical knowledge and intuition. He also kept the size of the reference parameterization data at a minimum by very carefully selecting necessary data to be used as reference. Over the following years many of the main-group elements have been parameterized keeping the original AM1 parameters for H, C, N and O unchanged. Of course, a sequential parameterization scheme caused every new parameterization to depend on previous ones, which directly affects the quality of the results. AM1 represented

a very considerable improvement over MNDO without any increase in the computing time needed.

AM1 has been parameterized for many of the main-group elements and is very widely used, keeping its popularity in organic compounds' modeling due to its good performance and robustness. Although many of the deficiencies in MNDO were corrected in AM1, it still has some important limitations as outlined in the historical development section [12].

### 1.3.2 PM3 Method

In 1989, Stewart introduced PM3, which can be considered as a reparameterization of AM1. This method was named as parametric method 3, considering MNDO and AM1 as the methods 1 and 2, respectively, as one of the three NDDO-based methods. In both MNDO and AM1, one-center electron repulsion integrals ( $g_{ij}$ ,  $h_{ij}$ ), which are five parameters  $g_{ss}$ ,  $g_{sp}$ ,  $g_{pp}$ ,  $gp^2$ , and  $gsp$ , are assigned values determined from atomic spectra by Oleari. PM3 differs from MNDO and AM1 and these one-center electron integrals are taken as parameters to be optimized.

PM3 also differs from AM1 in the number of Gaussian terms used in the core repulsion function. PM3 uses only two Gaussian terms per atom instead of up to four used by AM1. Another difference is that PM3 uses an automated parameterization procedure, in contrast to AM1. H, C, N, O, F, Al, Si, P, S, Cl, Br, and I parameters were simultaneously parameterized, whereas AM1 parameters were adjusted manually by Dewar with the help of chemical knowledge or intuition. Since his parameter optimization algorithms permitted an efficient search of parameter space, Stewart was able to employ a significantly larger data set in evaluating his penalty function than had been true for previous efforts. Statistically, PM3 was more accurate than the other semiempirical methods available at the time, but it was found to have several deficiencies that seriously limited its usefulness. One of the most important of these is the rotational barrier of the amide bond, which is much too low and in some cases almost non-existent. The other one is that PM3 has a very strong tendency to make the environment around nitrogen pyramidal. Thus, PM3 is not suggested for use in studies where the state of hybridization of nitrogen is important. According to a search of "Current Contents" done in 1999, AM1 was the most widely used semiempirical quantum mechanical method and PM3 was second [12].

## Theory

Both MNDO and AM1 had been parameterized by hand with the help of chemical knowledge and intuition using few reference data. Stewart had a more mathematical philosophy for the parameterization procedure and thought automated search of parameter space using complex optimization algorithm might be more successful to obtain better parameters. He made an optimization process by deriving and implementing formulae for the derivative of a suitable error function with respect to the parameters (equation 11).

$$S = \sum_I (x_i^{\text{calc}} - x_i^{\text{ref}})^2 \quad (11)$$

Where  $S$  is defined as the sum of the squares of the differences between calculated or predicted ( $x_i^{\text{calc}}$ ) and reference values ( $x_i^{\text{ref}}$ ) for reference functions, the parameter set is modified to minimize the value of  $S$ , and parameters are considered as *optimized* when for a given set of parameters, the sum square of errors,  $S$ , is a minimum.

In PM3, for each of the element's parameter set consists of 18 parameters ( $U_{ss}$ ,  $U_{pp}$ ,  $\beta_s$ ,  $\beta_p$ ,  $\zeta_p$ ,  $\zeta_s$ ,  $\alpha$ ,  $g_{ss}$ ,  $g_{pp}$ ,  $g_{sp}$ ,  $gp2$ ,  $hsp$ ,  $K1$ ,  $K2$ ,  $L1$ ,  $L2$ ,  $M1$ ,  $M2$ ) except for hydrogen, which has 11 parameters only since parameters related to  $p$ -orbitals are not included.

As different from MNDO and AM1, in PM3 the one-center electron repulsion parameters are ( $g_{ij}$ ,  $h_{ij}$ ) optimized instead of assigning to atomic spectral values. PM3 also shares the same core-repulsion function with AM1 which is given as equation 12.

$$E_{AB}^{\text{PM3}} = E_{AB}^{\text{MMDO}} + \frac{Z_A Z_B}{R_{AB}} \left\{ \sum_i K_{A,i} e^{-L_{A,i}(R_{AB} - M_{A,i})^2} + \sum_j K_{B,j} e^{-L_{B,j}(R_{AB} - M_{B,j})^2} \right\} \quad (12)$$

But it uses only two Gaussian terms ( $i = 1, 2$  and  $j = 1, 2$  above) for each atom instead of four in AM1 ( $i = 1, \dots, 4$  and  $j = 1, \dots, 4$ ).

In the initial parameterization of PM3, twelve elements (H, C, N, O, F, Al, Si, P, S, Cl, Br and I) were optimized simultaneously and then following parameterizations were carried out keeping the parameters for these elements fixed. PM3 may have global minimum in comparing with MNDO and AM1, but this global minimum is obtained for a specific penalty function used and it is heavily affected by the type of compounds included in the parameterization

dataset. Thus, it does not necessarily supersede MNDO and AM1 especially for any particular type of problem [12].

The combination of quantum mechanics and molecular mechanics is a natural approach for the study of materials science. The active site or binding site is treated by the *ab initio* density functional theory or semi-empirical potentials, whereas the rest of the system is calculated by the force fields based on molecular mechanics. In the current version of *sander*, one can use the MNDO, AM1, or PM3 semi-empirical Hamiltonian for the quantum mechanical region. Interaction between the QM and MM regions includes electrostatics (based on partial charges in the MM part) and Lennard–Jones terms, designed to mimic the exchange-repulsion terms that keep QM and MM atoms from overlapping [13].

## 1.4 Gibbs Energy Free

In thermodynamics, the Gibbs free energy is a thermodynamic potential that measures the "useful" or process-initiating work obtainable from a thermodynamic system at a constant temperature and pressure (isothermal, isobaric). Just as in mechanics, where potential energy is defined as capacity to do work, similarly different potentials have different meanings. The Gibbs free energy is the *maximum* amount of non-expansion work that can be extracted from a closed system; this maximum can be attained only in a completely reversible process. When a system changes from a well-defined initial state to a well-defined final state, the Gibbs free energy  $\Delta G$  equals the work exchanged by the system with its surroundings, minus the work of the pressure forces, during a reversible transformation of the system from the same initial state to the same final state.

Gibbs energy (also referred to as  $\Delta G$ ) is also the chemical potential that is minimized when a system reaches equilibrium at constant pressure and temperature. Its derivative with respect to the reaction coordinate of the system vanishes at the equilibrium point. As such, it is a convenient criterion of spontaneity for processes with constant pressure and temperature. The Gibbs free energy is defined as (equation 13).

$$G(p, T) = U + pV - TS \quad (13)$$

Which is the same as (equation 14).

$$G(p, T) = H - TS \quad (14)$$

Where:

$U$  is the internal energy (SI unit: joule).

$p$  is pressure (SI unit: pascal).

$V$  is volume (SI unit:  $\text{m}^3$ ).

$T$  is the temperature (SI unit: kelvin).

$S$  is the entropy (SI unit: joule per kelvin).

$H$  is the enthalpy (SI unit: joule).

The expression for the infinitesimal reversible change in the Gibbs free energy as a function of its 'natural variables'  $p$  and  $T$ , for an open system, subjected to the operation of external forces (for instance electrical or magnetical)  $X_i$ , which cause the external parameters of the system  $a_i$  to change by an amount  $da_i$ , can be derived as follows from the First Law for reversible processes (equation 15).

$$\begin{aligned}
 TdS &= dU + pdV - \sum_{i=1}^K \mu_i dN_i + \sum_{i=1}^n X_i da_i + \dots \\
 d(TS) - SdT - dU + d(pV) - Vdp - \sum_{i=1}^K \mu_i dN_i + \sum_{i=1}^n X_i da_i + \dots \\
 d(U - TS + pV) - Vdp - SdT + \sum_{i=1}^K \mu_i dN_i + \sum_{i=1}^n X_i da_i + \dots \\
 dG - Vdp - SdT + \sum_{i=1}^K \mu_i dN_i + \sum_{i=1}^n X_i da_i + \dots
 \end{aligned} \tag{15}$$

Where:

- $\mu_i$  is the chemical potential of the  $i$ th chemical component. (SI unit: joules per particle or joules per mole.)
- $N_i$  is the number of particles (or number of moles) composing the  $i$ th chemical component.

This is one form of Gibbs fundamental equation. In the infinitesimal expression, the term involving the chemical potential accounts for changes in Gibbs free energy resulting from an influx or out flux of particles. In other words, it holds for an open system. For a closed system, this term may be dropped. Let the change  $\Delta G$  in Gibbs free energy be defined as equation 16.

$$\Delta G = \Delta H - T\Delta S_{\text{int}} \quad (16)$$

Notice that it is not defined in terms of any external state functions, such as  $\Delta S_{\text{ext}}$  or  $\Delta S_{\text{tot}}$ . Then the second law becomes, which also tells us about the spontaneity of the reaction.

$\Delta G < 0$  Favoured reaction (Spontaneous)

$\Delta G = 0$  Neither the forward nor the reverse reaction prevails (Equilibrium)

$\Delta G > 0$  Disfavoured reaction (Nonspontaneous)

Gibbs free energy  $G$  itself is defined as equation 17.

$$G = H - TS_{\text{int}} \quad (17)$$

But notice that to obtain equation (17) from equation (16) we must assume that  $T$  is constant. Thus, Gibbs free energy is most useful for thermochemical processes at constant temperature and pressure: both isothermal and isobaric. Such processes don't move on a  $P$ - $V$  diagram, such as phase change of a pure substance, which takes place at the saturation pressure and temperature. Chemical reactions, however, do undergo changes in chemical potential, which is a state function. Thus, thermodynamic processes are not confined to the two dimensional  $P$ - $V$  diagram. There is a third dimension for  $n$ , the quantity of gas. For the study of explosive chemicals, the processes are not necessarily isothermal and isobaric. For these studies, Helmholtz free energy is used. If an isolated system ( $Q = 0$ ) is at constant pressure ( $Q = \Delta H$ ), then

$$\Delta H = 0 \quad (18)$$

Therefore the Gibbs free energy of an isolated system is:

$$\Delta G = T\Delta S_{\text{int}} \quad (19)$$

And if  $\Delta G \leq 0$  then this implies that  $\Delta S \geq 0$ , back to where we started the derivation of  $\Delta G$  [14-15].

## 1.5 Electrostatic Potential

The electrostatic force is a conservative force. This means that the work it does on a particle depends only on the initial and final position of the particle and not on the path followed. With each conservative force, a potential energy can be associated. The introduction of the potential energy is useful since it

allows us to apply conservation of mechanical energy which simplifies the solution of a large number of problems. The potential energy  $U$  associated with a conservative force  $F$  is defined in the following manner (equation 20).

$$U(P_1) - U(P_0) = -W = \int_{P_0}^{P_1} \bar{F} \, d\bar{L} \quad (20)$$

Where  $U(P_0)$  is the potential energy at the reference position  $P_0$  (usually  $U(P_0) = 0$ ) and the path integral is along any convenient path connecting  $P_0$  and  $P_1$ . Since the force  $F$  is conservative, the integral in eq.(25.1) will not depend on the path chosen. If the work  $W$  is positive (force and displacement pointing in the same direction) the potential energy at  $P_1$  will be smaller than the potential energy at  $P_0$ . If energy is conserved, a decrease in the potential energy will result in an increase of the kinetic energy. If the work  $W$  is negative (force and displacement pointing in opposite directions) the potential energy at  $P_1$  will be larger than the potential energy at  $P_0$ . If energy is conserved, an increase in the potential energy will result in a decrease of the kinetic energy. In electrostatic problems the reference point  $P_0$  is usually chosen to correspond to an infinite distance and the potential energy at this reference point is taken to be equal to zero. Equation (25.1) can then be rewritten as:

$$U_c(P_1) = - \int_{P_0}^{P_1} \bar{F}_c \, d\bar{L} \quad (21)$$

To describe the potential energy associated with a charge distribution the concept of the electrostatic potential  $V$  is introduced. The electrostatic potential  $V$  at a given position is defined as the potential energy of a test particle divided by the charge  $q$  of this object:

$$\begin{aligned} V_c(P_1) &= \frac{U_c(P_1)}{q} = \frac{U_c(P_0)}{q} - \frac{1}{q} \int_{P_0}^{P_1} \bar{F}_c \, d\bar{L} \\ &= V_c(P_0) - \frac{1}{q} \int_{P_0}^{P_1} \bar{F}_c \, d\bar{L} \\ &= - \frac{1}{q} \int_{P_0}^{P_1} \bar{F}_c \, d\bar{L} \end{aligned} \quad (22)$$

In the last step of equation (22) we have assumed that the reference point  $P_0$  is taken at infinity, and that the electrostatic potential at that point is equal to 0. Since the force per unit charge is the electric field, equation (23) can be rewritten as:

$$V_c(P_1) = - \int_{P_0}^{P_1} \vec{E} \cdot d\vec{L} \quad (23)$$

The unit of electrostatic potential is the volt (V), and  $V = J/C = Nm/C$ . Equation (23) shows that as the unit of the electric field we can also use V/m.

A common used unit for the energy of a particle is the electron-volt (eV) which is defined as the change in kinetic energy of an electron that travels over a potential difference of 1 V. The electron-volt can be related to the Joule via eq.(23). Equation (23) shows that the change in energy of an electron when it crosses over a 1V potential difference is equal to  $1.6 \times 10^{-19}$  J and we thus conclude that  $1 \text{ eV} = 1.6 \times 10^{-19} \text{ J}$ .

In many electrostatic problems the electric field of a certain charge distribution must be evaluated. The calculation of the electric field can be carried out using two different methods:

1. The electric field can be calculated by applying Coulomb's law and vector addition of the contributions from all charges of the charge distribution.
2. The total electrostatic potential  $V$  can be obtained from the algebraic sum of the potential due to all charges that make up the charge distribution, and subsequently using equation (23) to calculate the electric field  $E$ .

In many cases method 2 is simpler since the calculation of the electrostatic potential involves an algebraic sum, while method 1 relies on the vector sum [16].

## 1.6 Molecular Orbitals

A molecular orbital (or MO) is a mathematical function describing the wave-like behavior of an electron in a molecule.

This function can be used to calculate chemical and physical properties such as the probability of finding an electron in any specific region. Molecular orbitals (MOs) represent regions in a molecule where an electron is likely to be

found. Molecular orbitals are obtained from the combination of atomic orbitals, which predict the location of an electron in an atom.

A molecular orbital can specify the electron configuration of a molecule: the spatial distribution and energy of one (or one pair of) electron(s). Most commonly an MO is represented as a linear combination of atomic orbitals (the LCAO-MO method), especially in qualitative or very approximate usage. They are invaluable in providing a simple model of bonding in molecules, understood through molecular orbital theory. Most present-day methods in computational chemistry begin by calculating the MOs of the system.

A molecular orbital describes the behavior of one electron in the electric field generated by the nuclei and some average distribution of the other electrons. In the case of two electrons occupying the same orbital, the Pauli principle demands that they have opposite spin. Necessarily this is an approximation, and highly accurate descriptions of the molecular electronic wave function do not have orbitals (see configuration interaction). Molecular orbitals arise from allowed interactions between atomic orbitals, which are allowed if the symmetries (determined from group theory) of the atomic orbitals are compatible with each other.

Efficiency of atomic orbital interactions is determined from the overlap (a measure of how well two orbitals constructively interact with one another) between two atomic orbitals, which is significant if the atomic orbitals are close in energy. Finally, the number of molecular orbitals that form must equal the number of atomic orbitals in the atoms being combined to form the molecule. The type of interaction between atomic orbitals can be further categorized by the molecular-orbital symmetry labels  $\sigma$  (sigma),  $\pi$  (pi), etc. paralleling the symmetry of the atomic orbitals s, p, etc.

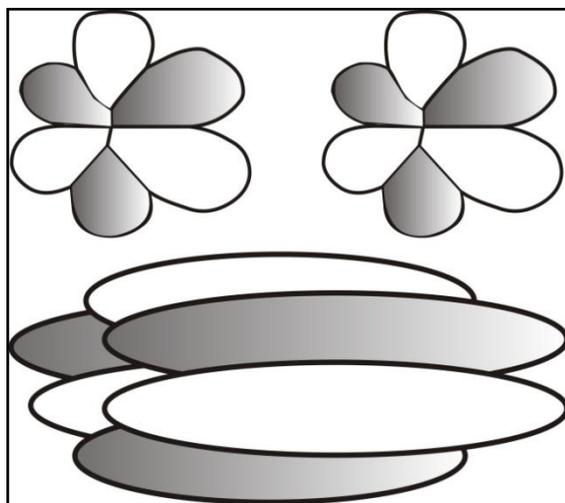
**$\sigma$  Symmetry.** A MO with  $\sigma$  symmetry results from the interaction of either two atomic s-orbitals or two atomic  $p_z$ -orbitals. A MO will have  $\sigma$ -symmetry if the orbital is symmetrical with respect to the axis joining the two nuclear centers, the internuclear axis. This means that rotation of the MO about the internuclear axis does not result in a phase change. A  $\sigma^*$  orbital, sigma antibonding orbital, also maintains the same phase when rotated about the internuclear axis. The  $\sigma^*$  orbital has a nodal plane that is between the nuclei and perpendicular to the internuclear axis.

**$\delta$  Symmetry.** A MO with  $\delta$  symmetry results from the interaction of two atomic  $d_{xy}$  or  $d_{x^2-y^2}$  orbitals. Because these molecular orbitals involve low-

energy d atomic orbitals, they are seen in transition-metal complexes.

**$\pi$  Symmetry.** A MO with  $\pi$  symmetry results from the interaction of either two atomic  $p_x$  orbitals or  $p_y$  orbitals. A MO will have  $\pi$  symmetry if the orbital is asymmetrical with respect to rotation about the internuclear axis. This means that rotation of the MO about the internuclear axis will result in a phase change. A  $\pi^*$  orbital, pi antibonding orbital, will also produce a phase change when rotated about the internuclear axis. The  $\pi^*$  orbital also has a nodal plane between the nuclei.

**$\phi$  Symmetry.** Theoretical chemists have conjectured that higher-order bonds, such as phi bonds corresponding to overlap of “f atomic orbitals”, are possible. There is as of 2005 only one known example of a molecule purported to contain a phi bond (a U–U bond, in the molecule  $U_2$ ) (Figure 1.10) [17-18].



**Figure 1.10** Suitably aligned *f* atomic orbitals overlap to form phi molecular orbital (a phi bond).

When we are dealing with interacting molecular orbitals, the two that interact are generally:

- The highest energy occupied molecular orbital (HOMO) of one molecule.
- The lowest energy unoccupied molecular orbital (LUMO) of the other molecule.
- These orbitals are the pair that lie closest in energy of any pair of orbitals in the two molecules, which allows them to interact most strongly.

- These orbitals are sometimes called the *frontier* orbitals, because they lie at the outermost boundaries of the electrons of the molecules (Figure 1.11) [19].

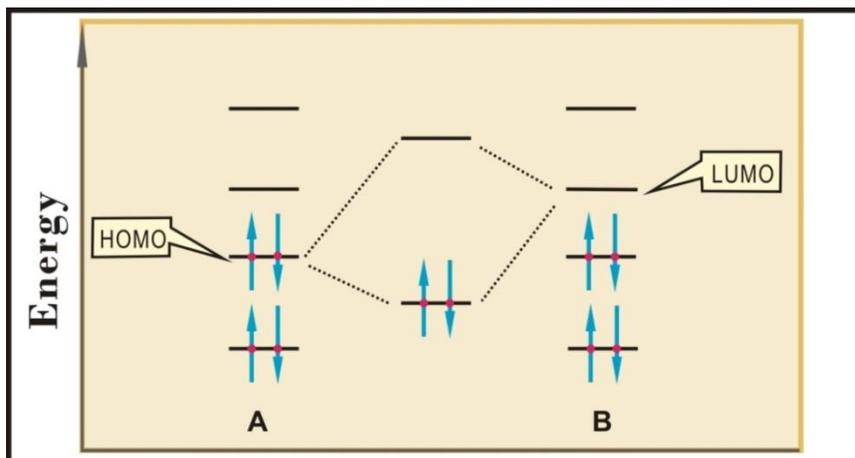


Figure 1.11 Filled-empty interactions redrawn as a HOMO-LUMO interaction.

## References

- [1] Salih S M. AM1 semi-empirical calculation to study the nature of di- and tri-molecular assembly of  $\pi$ ,  $\pi$ -aromatic interactions. *The Arabian Journal for Science and Engineering*, 35 (2A), 95-113, 2010.
- [2] [www.tau.ac.il/~ephraim/MM+MD.doc](http://www.tau.ac.il/~ephraim/MM+MD.doc).
- [3] Cornell W D, Cieplak P, Bayly C I, Gould I R, Merz K M Jr, Ferguson D M, Spellmeyer D C, Fox T, Caldwell J W, Kollman P A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* 117: 5179–5197, 1995.
- [4] Abdulsattar M A, Al-Bayati K H. Corrections and parameterization of semiempirical large unit cell method for covalent semiconductors. *Phys. Rev. B.* 75, 245201, 2007.
- [5] Abdulsattar M A. Size effects of semiempirical large unit cell method in comparison with nanoclusters properties of diamond-structured covalent semiconductors, *Physica E.* 41, 1679, 2009.

- [6] Levine I. *Quantum Chemistry*, Prentice Hall, 4th edition, 579–580, 1991.
- [7] Cramer C J. *Essentials of Computational Chemistry*. Wiley, Chichester, 126–131, 2002.
- [8] Stewart J. J. P. *Reviews in Computational Chemistry*, Volume 1, Eds. K. B. Lipkowitz and D. B. Boyd, VCH, New York, 45, 1990.
- [9] Rocha G B, Freire R O, Simas A M, Stewart J J P. RM1: A reparameterization of AM1 for H, C, N, O, P, S, F, Cl, Br, and I. *The Journal of Computational Chemistry*. 27(10), 1101–1111, 2006.
- [10] Stewart J J P. Optimization of Parameters for Semiempirical Methods V: Modification of NDDO Approximations and Application to 70 Elements. *The Journal of Molecular Modeling*. 13(12), 1173–1213, 2007.
- [11] Zerner M, *Reviews in Computational Chemistry*, Volume 2, Eds. K. B. Lipkowitz and D. B. Boyd, VCH, New York, 313, 1991.
- [12] B änsch E, Clark T, Burzlaff N. Parameterization of the am1\* semiempirical molecular orbital method for the first-row transition metals and other elements, Turqu ía, 2009.
- [13] Ryde U. Combined quantum and molecular mechanics calculations on metalloproteins. *Curr Opin Chem Biol*. 7, 136-142, 2003.
- [14] Salzman W R. *Chemical Thermodynamics*. University of Arizona, 2001. <http://web.archive.org/web/20070707224025/http://www.chem.arizona.edu/~salzmanr/480a/480ants/opensys/opensys.html>.
- [15] CRC Handbook of Chemistry and Physics, 2009, pp. 5-4 - 5-42, 90th ed., Lide Müller, Ingo (2007). *A History of Thermodynamics - the Doctrine of Energy and Entropy*. Springer. ISBN 978-3-540-46226-2.
- [16] [http://teacher.nslr.rochester.edu/phy122/Lecture\\_Notes/Chapter25/Chapter25.html](http://teacher.nslr.rochester.edu/phy122/Lecture_Notes/Chapter25/Chapter25.html).
- [17] Atkins P. Atkins J P. *Physical Chemistry*. Oxford University Press, 8th ed., 2006.
- [18] Gary L. Miessler; Donald A. Tarr. *Inorganic Chemistry*. Pearson Prentice Hall, 3rd ed., 2004.
- [19] Department of Chemistry, University of Maine, Orono, ME 04469 (<http://chemistry.umeche.maine.edu/CHY252/HOMO-LUMO.html>).