CHAPTER III
THEORY

3.1. Quantitative Structure Activity Relationship (QSAR)

Syntheses and screening of hundreds and thousands of potential drugs candidate to find one active drug is extremely expensive and laborious procedure. Although often guided by rational concept, drug research generally has been, over decades, a mere trial- and –error search for new leads and active analogues. More effective strategies are desirable and a rational alternative is the derivation of structure-activity hypotheses and their qualitative evaluation.

Quantitative structure activity relationship (QSAR) is a technique to quantify the relationship between biological activities and structural properties of compounds. The objectives of QSAR may be described by (i) to correlate and summarize the relationships between biological activity and physicochemical/structural property for better understanding of the mode of action of compounds, (ii) to optimize the structure for the best possible biological activity, and (iii) to predict the biological activity of other compounds. QSAR is one of the important tools for delivering new lead candidate more quickly with a low cost. The application of QSAR is used as guidance how to further modify the structure of compound to enhance its activity. Moreover, QSAR is also used to predict the potency and physicochemical properties of clinical drug candidates. Based on QSAR results, only compounds showing high inhibitory potency will be further synthesized and tested for the biological activity. This reduces time and saves a lot of money.

The binding of a molecule to a macromolecule has to occur in such a manner that both molecules are stabilized in a 3D orientation to promote the observed biological activities. In other words, the biological activity of a molecule is a consequence of its 3D shape, size, and geometry. As both drug and receptor have 3D structures, it is important to understand the nature of 3D shape and conformation of biological molecules in relation to the observed biological activities. Therefore, QSAR methods based on the 3D structure, termed 3D-QSAR, were developed. It is expected that 3D-QSAR methods could provide better information about the drug-receptor interactions. Moreover, 3D-QSAR methodology also shows advantage over the classical QSAR in the way that

compounds with structural differences can be studied. Several 3D-QSAR approaches have been proposed by different scientists. Two most popular and widely used techniques are Comparative Molecular Field Analysis (CoMFA) and Comparative Molecular Similarity Indices Analysis (CoMSIA).

3.1.1 Comparative Molecular Field Analysis (CoMFA)

The CoMFA method was introduced by Cramer et al in 1998 [79]. This technique was based on the assumption that changes in binding affinities of ligands are related to changes in shape and strength of non-covalent interaction fields surrounding the molecules. These fields are of steric and electrostatic nature. In order to compute the fields, first all compounds must be fitted or superimposed together, then, the 3D cubic lattice points are created to cover these aligned molecules and the interaction energies between each molecule and specifically defined probe atoms representing receptor are calculated for each grid point. CoMFA then compares the computed steric and electrostatic fields around the molecules and extracts important features related to the biological activity. In doing so, CoMFA tries to identify the quantitative influence of specific chemical features of molecules on their potencies. The results can be further displayed in contour maps showing the important regions in 3D space that are highly associated with the biological activity.

Although CoMFA has been successfully used to emphasize the relationship between biological activities and 3D physicochemical features of compounds, a number of problems inherent to CoMFA are known, some of them directly connected with the field calculation method. For the calculation of steric contributions, the Lennard-Jones potential is used, and for the calculation of electrostatic contributions, the Coulombic potential is used. Figure 3.1 shows the curves of these potentials. One main difficulty is the cut-off values whose application excludes very high field contributions near the molecular surface from analysis. Due to the sharp increase of the Lennard Jones potential, and a comparatively shallow increase of the Coulombic potential near the molecular surface, the application of cut-off values can be highly critical since the important contributions might be dropped for some molecules. Moreover, small shifts within the alignment can lead to dramatically altered results.

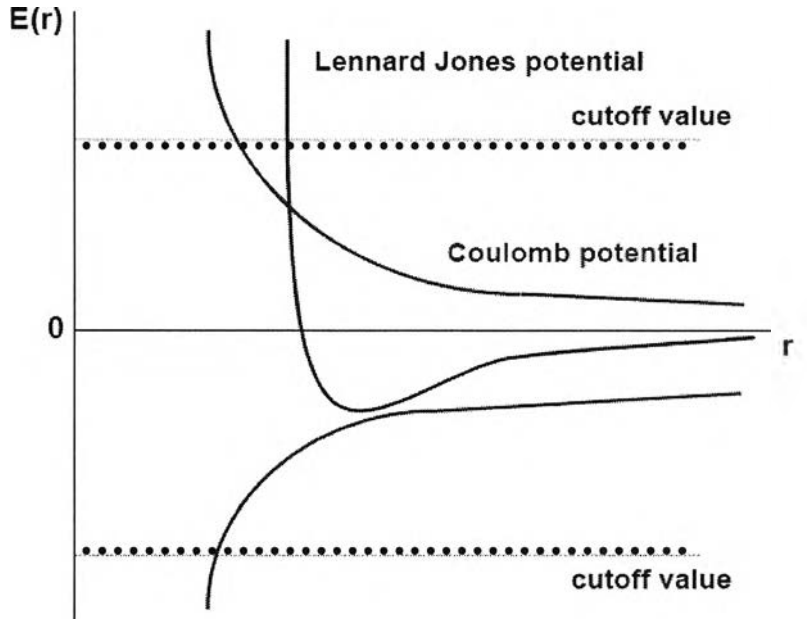


Figure 3.1 Schematic representations of the Lennard-Jones and Coulomb attractive and repulsive potentials describing steric and electrostatic contributions to CoMFA fields.

3.1.2 Comparative Molecular Similarity Indices Analysis (CoMSIA)

CoMSIA developed by Klebe et al. [80] is an extension of the CoMFA methodology. Molecular similarity is compared in terms of similarity indices (A_F). The similarity indices (A_F) between the compounds of interest and a probe atom have been determined according to [80, 81]:

$$A^{q}_{F,k}(j) = -\sum W_{probe,k} W_{ik} e^{-\alpha r^2_{iq}} \qquad (3.1)$$

where i = summation index over all atoms of the molecule j under investigation, W_{ik} = actual value of the physicochemical property k of atom, $W_{probe,k}$ = probe atom, α = attenuation factor, r_{iq} = mutual distance between probe atom at grid point q and atom i of the investigated molecules.

Its advantages over the standard CoMFA technique are reported to be a greater robustness regarding both region shifts and small shifts within the alignment, and more

intuitively interpretable contour maps [80]. Moreover, the CoMSIA is not sensitive to the type of probe atoms. This is the result of the application of similarity indices calculated by using Gaussian-type distance dependence instead of the Lennard-Jones and Coulombic potentials which make more or less arbitrary application of cut-off values unnecessary. This is shown in Figure 3.2. Additionally, not only steric and electrostatic fields are typically considered in CoMSIA technique, but also the hydrophobic, hydrogen bond donor and hydrogen bond acceptor properties.

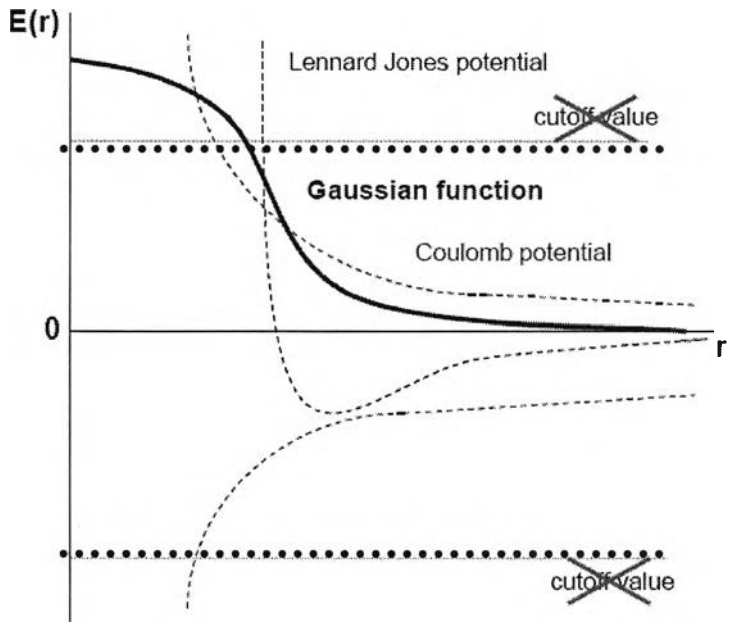


Figure 3.2 A Gaussian type function used in CoMSIA calculation.

3.1.3 3D-QSAR set up

The following steps are standard procedures of the building of 3D-QSAR model.

3.1.3.1 Compound selection

To build 3D-QSAR models, all compounds must be divided into two sets, training and test sets. The training set is used to derive 3D-QSAR models while the test set is used

for testing predictive ability of the models. Since 3D-QSAR models depend upon the structures of compounds, they should be selected carefully. The training set should contain a wide range of structurally diverse compounds and the ranges of activity should be covered as large as possible. Furthermore, the biological data must be obtained for a set of ligands using uniform protocol and ideally from a single source.

3.1.3.2 Ligand 3D structure generation

Both experimental techniques (X-ray crystallography and NMR) and molecular modeling can provide starting structures of ligands. In case of no X-ray structure available, 3D structures can be generated by molecular modeling programs. Once 3D structure is generated, it has to be refined by performing structure optimization. Examples of theoretical methods for the structure optimization are molecular mechanics, semi-empirical and *ab initio* methods. The interpretation based on molecular modeling depends on the quality of the computations. Thus, it is important to understand the concepts, the strengths and weaknesses, and the limitations of the calculations of the common computational methods associated with molecular modeling technique. Moreover, in case that various conformations of molecules might occur, conformational analyses have to be performance for the generation of multiple conformations.

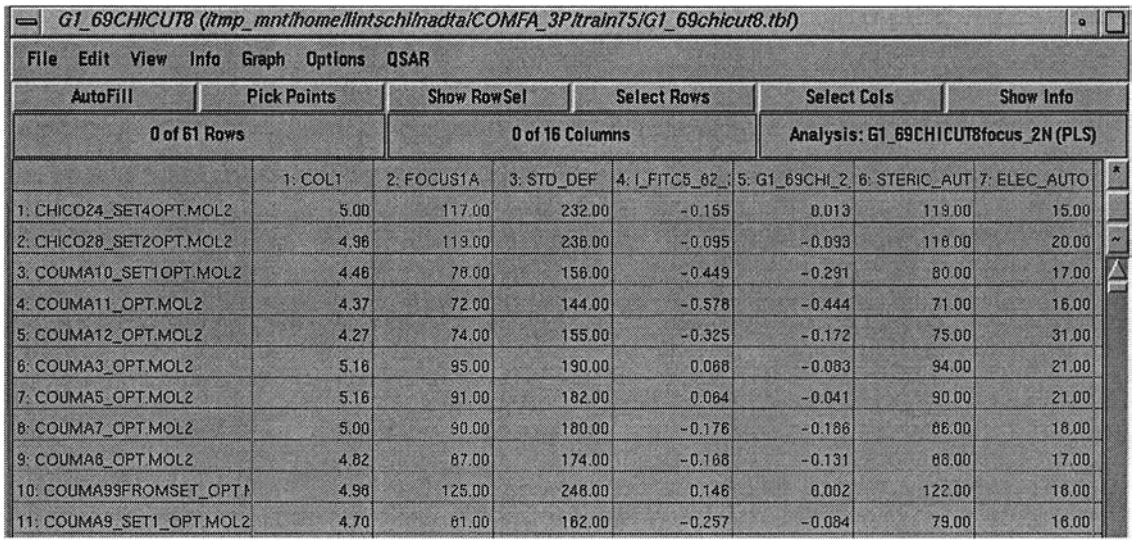
3.1.3.3 Molecule alignment

The alignment of molecules is one of the most crucial steps in 3D-QSAR, especially CoMFA. In this step, all compounds must be aligned or superimposed to the template. Typically, the template could be the most active compound, the lead or commercial compound or the compound containing the greatest number of functional groups.

3.1.3.4 Molecular field calculation

After all compounds are superimposed, 3D lattice points are created around them for calculating interaction energies with various probes at each lattice point. In order to place the lattice points around molecules, there are three aspects to concern, the size of grid spacing, size of grid box, and the location of the grid box. The typical choice for

grid spacing is 2 Å, and the size of grid box is about 3 - 4 Å larger than the union surface of the molecules. The location of grid box sometimes significantly affects the CoMFA results including the statistics and the number of components in the final CoMFA model. Since it is difficult to know a priori the best location of a grid box, the best location is often chosen after deriving the initial CoMFA models. However, possible strategy to minimize this problem is to rotate the superimposed molecules in a way that they are not parallel to any of the lattice edges. Next, the probe is placed at each lattice point and the interaction energies between probe and the molecule are calculated. For CoMFA, steric and electrostatic energies are calculated. Steric energy is computed by the Lennard Jones potential and electrostatic interaction is calculated from the Coulombic potential. For CoMSIA besides steric and electrostatic interactions, hydrophobic, hydrogen bond donor and hydrogen bond acceptor are computed as well. All of the calculated energies will be stored in QSAR table (Figure 3.3).



G1_69CHICUT8 (/tmp_mnt/home/lintschi/nadta/COMFA_3P/train75/G1_69chicut8.tbl)

File Edit View Info Graph Options QSAR

AutoFill | Pick Points | Show RowSel | Select Rows | Select Cols | Show Info

0 of 61 Rows | 0 of 16 Columns | Analysis: G1_69CHICUT8focus_2N (PLS)

	1: COL1	2: FOCUS1A	3: STD_DEF	4: I_FITC5_62_:	5: G1_69CHI_2	6: STERIC_AUT	7: ELEC_AUTO
1: CHICO24_SET4OPT.MOL2	5.00	117.00	232.00	-0.155	0.013	119.00	15.00
2: CHICO28_SET2OPT.MOL2	4.96	119.00	236.00	-0.095	-0.093	116.00	20.00
3: COUMA10_SET1OPT.MOL2	4.46	78.00	156.00	-0.449	-0.291	80.00	17.00
4: COUMA11_OPT.MOL2	4.37	72.00	144.00	-0.578	-0.444	71.00	16.00
5: COUMA12_OPT.MOL2	4.27	74.00	155.00	-0.325	-0.172	75.00	31.00
6: COUMA3_OPT.MOL2	5.16	95.00	190.00	0.066	-0.083	94.00	21.00
7: COUMA5_OPT.MOL2	5.16	91.00	182.00	0.064	-0.041	90.00	21.00
8: COUMA7_OPT.MOL2	5.00	90.00	180.00	-0.176	-0.186	86.00	18.00
9: COUMA8_OPT.MOL2	4.82	87.00	174.00	-0.166	-0.131	88.00	17.00
10: COUMA99FROMSET_OPT.M	4.96	125.00	246.00	0.146	0.002	122.00	18.00
11: COUMA9_SET1_OPT.MOL2	4.70	81.00	162.00	-0.257	-0.084	79.00	16.00

Figure 3.3 Illustration of 3D-QSAR table generated by SYBYL.

3.1.3.5 Statistical analysis

If a number of independent variables (descriptors) are very large compared to dependent variables, a multiple linear regression is not feasible to correlate such a large number of descriptors with a few activities. Partial least square (PLS) method is

developed to solve an equation having hundreds or thousands of descriptor variables with a small number of biological data. The PLS reduces the hundreds or thousands of descriptors to a few variables, called latent variable, each highly correlated with the remaining ones.

The quality of a 3D-QSAR model is mostly determined by its ability to perform predictions of compounds not included in the training set. However, it is often difficult to assemble enough compounds for sufficient large training and test sets. Therefore, the predictability is usually estimated with cross-validation by repeatedly leaving out one (or more) compound (s) at a time until each compound is excluded exactly once. During the cross-validation, the sum of squared prediction error called the predictive residual sum of squares (PRESS), the cross-validated correlation coefficient (r^2_{cv}), the standard error of estimate (SEP), and the optimum number of component (ONC) are calculated. A smaller SEP and a larger r^2_{cv} indicate the model's good predictability. Moreover, the statistical significance of the generated 3D-QSAR model is judged by some other statistical parameters like squared correlation coefficient (r^2), standard deviation (SD) and F test. The squared correlation coefficient (or coefficient of multiple determination), r^2, is a relative measure of quality of fit by the regression equation. Correspondingly, it represents the part of the variation in the observed data that is explained by the regression. The r^2 value closer to 1.0 represents the better fit of the regression. The F-test reflects the ratio of the variance explained by the model and the variance due to the error in the regression. High values of the F-test indicate that the model is statistically significant. Standard deviation is obtained from the error mean square, which expresses the variation of the residuals or the variation about the regression line. Thus standard deviation is an absolute measure of quality of fit and should have a low value for the regression to be significant.

3.1.3.6 Result interpretation

The results of CoMFA are an equation showing the contribution of energy fields at each lattice point. In order to facilitate the interpretation of the results, they are also displayed as coefficient (or standard deviation times coefficient) contour plots showing the regions in space where specific molecular properties increase or decrease the potency.

Typically, in CoMFA there are two types of contours, i.e. steric and electrostatic. The green and yellow contours mean steric favorable and steric unfavorable, respectively. The blue and red contours represent positive charge favorable and negative charge favorable, respectively. For CoMSIA, besides the steric and electrostatic fields there are hydrophobic, hydrogen donor and hydrogen acceptor contours.

3.2 Molecular dynamics simulation

Molecular dynamics (MD) simulation is one of the most powerful tools for studying the biological system as it is now widely and routinely used to investigate the structural, dynamical and thermodynamical behaviors of biological molecules and their complexes.

The starting point of MD simulation is an initial set of coordinates (obtained form X-ray or NMR experimental data). This structure is normally geometry minimized prior to the MD simulation in order to remove bad contacts and initial strain, which might disturb the subsequent MD. Once assigning velocities v_i, typically representing a low-temperature Maxwell distribution, the simulation is started by calculating the acceleration a_i for each atom i according to Newton's law.

$$-\frac{\partial V(x_i)}{\partial x_i} = F_i = m_i a_i = m_i \frac{\partial^2 x_i}{\partial t^2} \qquad (3.2)$$

F_i is the force acting on the Cartesian coordinate x_i for i = 1, …., 3N, for the N atoms in the molecule or molecular system, m_i is the atomic mass of atoms i, V is the potential energy of the system, and t is the time. The total energy (E) of the system is the sum of all kinetic $(\frac{1}{2}mv^2)$ and potential energy $V(x)$ contributions:

$$E = \frac{1}{2}m\left(\frac{\partial x}{\partial t}\right)^2 + V(x) \qquad (3.3)$$

The position $x_i(t+\Delta t)$ of an atom at the time $(t+\Delta t)$ can be calculated based on the known position $x_i(t)$ at the time t by using the integration of motion, for instance, the simple Verlet algorithm.

$$x_i(t+\Delta t) = 2x_i(t) - x_i(t-\Delta t) + \frac{F_i(t)}{m_i}\Delta t^2 \qquad (3.4)$$

Furthermore, the popular Leapfrog algorithm (a modification of Verlet algorithm) uses the position at time t and velocities at time $t-\left(\frac{\Delta t}{2}\right)$ for the update of both positions and velocities as the following equation:

$$x_i(t+\Delta t) = x_i(t) + v_i(t+\frac{\Delta t}{2})\Delta t \qquad (3.5)$$

$$v_i(t+\frac{\Delta t}{2}) = v_i(t-\frac{\Delta t}{2}) + \frac{\partial^2 x_i(t)}{\partial t^2}\Delta t \qquad (3.6)$$

The velocities, v_i, are often chosen randomly from a Maxwell-Boltzmann or Gaussian distribution at a given temperature, which gives the probability that an atom i has a velocity v_x in the x direction at a temperature T.

$$P(v_{ix}) = \left(\frac{m_i}{2\pi k_B T}\right)^{\frac{1}{2}} \exp\left[-\frac{1}{2}\frac{m_i v_{ix}^2}{k_B T}\right] \qquad (3.7)$$

where k_B is the Boltzman constant.

The temperature, T, can be calculated from the velocities using the relation:

$$T = \frac{1}{3NR}\sum_{i=1}^{N}\frac{P_i^2}{m_i} \qquad (3.8)$$

where N is the number of atoms in the system.

At present, there are several high-quality MD programs with bio-molecular focus such as CHARMM [82], AMBER [83], NAMD [84] and Gromacs [85].

3.3 Langevin dynamics [86]

The Langevin dynamics method approximates a full MD simulation of a system by eliminating unimportant or uninteresting degrees of freedom. The effects of the eliminated degrees of freedom are simulated by mean and stochastic forces. For example, instead of simulating hundreds of solvent molecules surrounding the solute molecules, the solvent can be ideally represented by a viscous fluid described in terms of dissipative and fluctuative equations. The Langevin equation incorporates two additional terms. The first term is a frictional, or damping, function intended to represent the fictional drag experienced by solute molecules in a solvent that is not explicitly simulated. The second additional term is a random force that is applied to mimic the random impulses that would be expected from both solvent and any coincident solute molecule. The Langevin equation for the motion of an atom *i* is

$$m\frac{d^2r}{dt^2} = F_i(r) - \xi_i\frac{dr}{dt} + R_i(t) \qquad (3.9)$$

where $F(r)$ is the usual term used in conventional MD, ξ_i is the friction coefficient, and $R_i(t)$ represents the random forces experienced by the atom. The temperature of the simulated system is maintained by a relationship between ξ_i and $R_i(t)$. When $\xi_i = 0$, Langevin dynamics is equivalent to conventional MD. When $\xi_i > 0$, the random impulses felt by the system can assist in propagating barrier-crossing motions and, therefore, Langevin dynamics can offer improved conformational sampling characteristics over standard MD.

3.4 Stochastic boundary

The stochastic boundary MD method [87, 88] uses a combination of both Langevin dynamics and Newtonian dynamics. The goal of the method is to eliminate atoms distant from an active site allowing detailed studies of a spatially localized portion of the reacting molecular system. With this method, the molecular system is partitioned

into a reaction region where Newtonian dynamics simulation is run, a buffer region where Langevin dynamics is run, and a reservoir region (Figure 3.4). In this way, atoms distant from the specific interactive sites in a large macromolecular system can be effectively eliminated from extensive analysis. This allows detailed studies of spatially localized portions of interacting molecular systems.

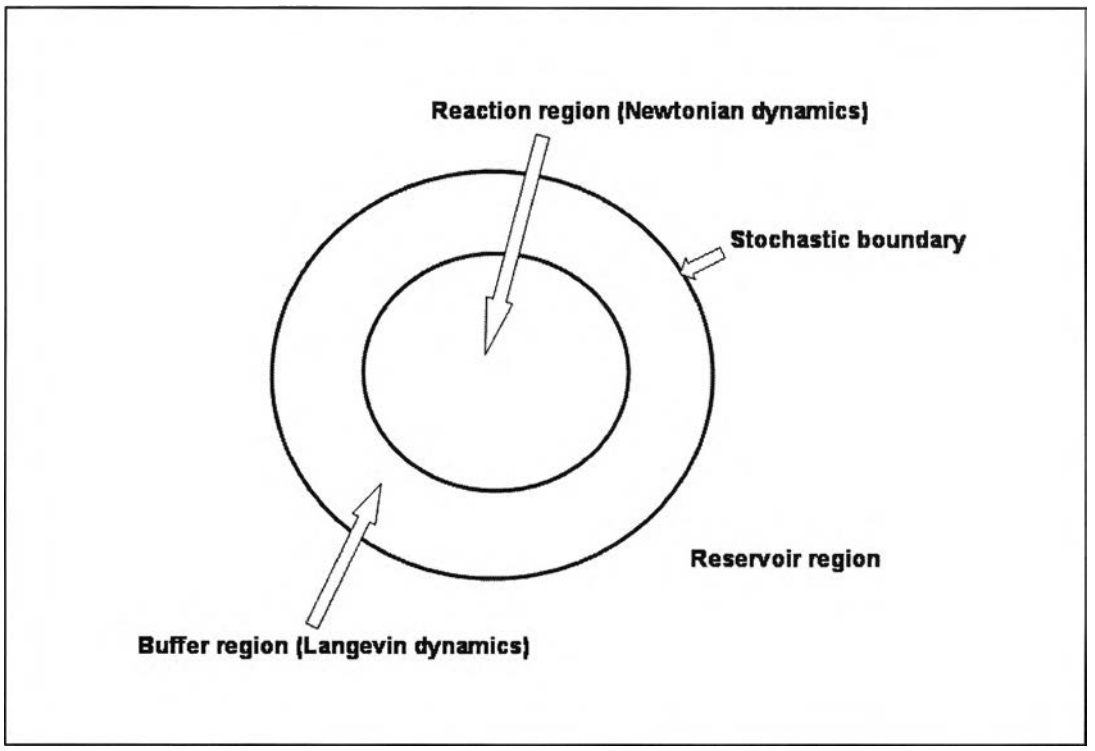


Figure 3.4 Schematic illustration of the partition of region in stochastic boundary molecular dynamics approach.

3.5 Quantum Mechanical /Molecular Mechanical (QM/MM)

3.5.1 The QM/MM approach

Although the QM approach shows high accuracy, its computational time increases with the size of the systems. Therefore, the application of QM is still limited to relatively small systems. On the other hand, MM is able to compute very large system quickly but it is unable to describe bond dissociations and formations in chemical reactions as well as

electronic interactions. Hence a combined QM and MM approach could take advantage of the applicability and accuracy of the QM methods and of computational cost for MM calculations. The fundamental concept of QM/MM approach [89-91] is to divide a large, condensed phase system into two regions, QM and MM. The reactive chemical event such as protein active site is contained within the QM area while the rest of the system is considered MM (see Figure 3.5).

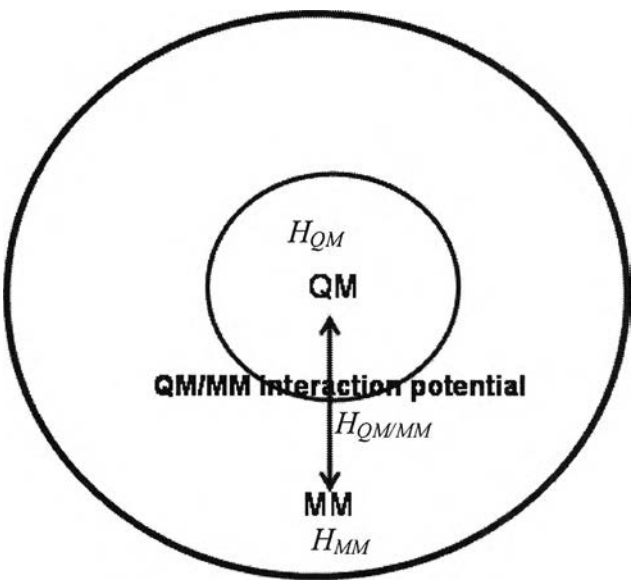


Figure 3.5 Illustration for the QM/MM method in the enzyme system. The active center is treated at the QM level and the surrounding is treated at the MM level.

The Hamiltonian describing the entire system consists of Hamiltonians for the isolated QM and MM parts and of a Hamiltonian describing the interaction between the QM and MM regions.

$$H_{total} = H_{QM} + H_{MM} + H_{QM/MM} \qquad (3.10)$$

where the H_{QM} is the Hamiltonian for the isolated QM region, H_{MM} is the Hamiltonian for MM region and $H_{QM/MM}$ is the Hamiltonian that couple QM and MM regions which takes the form:

$$H_{QM/MM} = -\sum_{i,M} \frac{eq_M}{r_{iM}} + \sum_{\alpha,M} \frac{Z_\alpha q_M}{R_{\alpha M}} + \sum_{\alpha.M} \left(\frac{A_{\alpha M}}{R_{\alpha M}^{12}} - \frac{B_{\alpha M}}{R_{\alpha M}^{6}} \right) \qquad (3.11)$$

where i = electrons, α = QM nuclei, M = MM atoms, e = electron charge, q_M = net charge on MM atoms, Z = nuclear charge, R/r = distances between particles

The first term is the electrostatic interaction between electrons in QM region and the MM nuclei, the second term is the electrostatic interaction between QM and MM nuclei and the third term is the van der Waals interaction between QM and MM atoms. The second and the third terms do not involve electronic coordinates. Based on the Eq.3.10, the total energy of the system is obtained by:

$$E = E_{QM} + E_{MM} + E_{QM/MM} \qquad (3.12)$$

QM/MM methods are useful for a wide variety of large and multi-scale applications, and they have been especially useful for molecular dynamics simulations of solvent effects, enzyme reactions, and solid-state catalysts. Nowadays various QM methods such as *ab initio*, self-consistent charge density functional tight-binding (SCC-DFTB) [92] or semi-empirical methods have been combined with force field such as CHARMM, AMBER, NWChem, Gromos and the Tripos force fields. In QM part of the system which is usually polarized by the changes in the environment, electronic properties and reaction mechanisms can be studied. The flexibility of the environment can be simulated by an appropriate force field.

3.5.2 Frontier bonds in QM/MM methods

If the QM and MM regions are separate molecules, having non-bonded interactions only, it might be sufficient. If the two regions are however parts of the same molecule, it is necessary to describe the bond connecting the two sections. A major complication in couple QM/MM methodology is the treatment of the frontier between QM part and the MM part. Various approaches have been purposed to truncate the wave function for the QM fragment at the boundary region. The most straightforward approach involves inserting a link atom, typically hydrogen atom, between the QM host atom

(QMHA, see Figure 3.6) and the MM host atom (MMHA). The link atom is treated at the QM level, and may be subjected to an angular and distance constraint to lie along the bond between QMHA and MMHA at a fixed bond distance. The link atom typically interacts with MM atoms through electrostatic terms but not through van der Waals terms. There are two types of hydrogen link atoms, QQ and HQ. HQ shows advantage over the QQ because HQ interacts with the entire MM region while QQ does not [93, 94]. Instead of regular hydrogen atoms, hydrogen like-atom or pseudohalogens have been used to terminate the QM region. The electronic nature of the link atom is modified to mimic the behavior of the MM host atom or MM host group (MMHG).

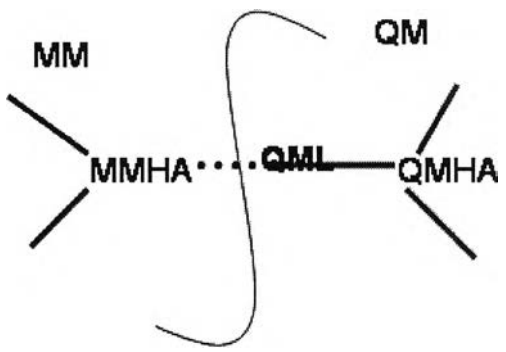


Figure 3.6 Partition between QM and MM regions in link-atom approach.

The local self-consistent field (LSCF) formalism is an alternative approach to treat the QM and MM boundary. It is based on the principle introduced by Warshel and Levitt [90] that a single hybrid sp^2 orbital with a single electron is included for each of the QM atoms at the junction. In LSCF method, the QM/MM frontier bonds are described by strictly localized bond orbitals (SLBOs) [95]. These localized bond orbitals, called frozen orbitals are excluded for the SCF procedure and defined by their hybridization coefficients and their electron population. The two parameters are usually determined by quantum chemical calculations on small model systems and are assumed to be transferable to system of interest. Since it avoids introducing extraneous hydrogen atoms in the system, the LSCF method is an attractive alternative to link atom approach. A

slightly different method, generalized hybrid orbital method, also based on the principle of hybrid frontier orbitals, was recently proposed by Gao et al [96]. Quantitative comparisons between LA and LSCF have been made and it was reported that both methods are generally of similar accuracy if care is taken in the choice of the frontier between the QM and MM regions.

3.6 Free energy MM-PB (GB) SA

In the physics-based approaches, the free energy of a given conformational state *i* relative to a chosen reference state in solution can be approximated by

$$\Delta G_i = \Delta G_{\text{int}} + \Delta G_{solv} - T\Delta S \qquad (3.13)$$

$$\Delta G_i = \Delta G_{MM} + \Delta G_{solv} - T\Delta S \qquad (3.14)$$

where ΔG_{int} are the internal energies, including bond, angle, torsional, van der Waals and electrostatics energy terms. The ΔG_{int} is often referred to ΔG_{MM} because it is usually modeled by MM force fields such as CHARMM, AMBER, OPLS and etc. ΔG_{solv} referred to the solvation free energy describes the free energy due to solvation, which includes hydrophobic packing and solvent-solute polarization. The last term (-TΔS) represents the solute configuration entropy and can be estimated by quasi-harmonic analysis of the trajectory or by normal mode analysis.

Considering the solvation free energy, ΔG_{solv}, which can be expressed by:

$$\Delta G_{solv} = \Delta G_{solv}^{polar} + \Delta G_{solv}^{non-polar} \qquad (3.15)$$

ΔG_{solv}^{polar} is generally computed using two continuum solvent approaches, the analytic generalized-Born (GB) model and numerical Poisson-Boltzmann (PB) method. The former is denoted by $\Delta G_{solv,GB}^{polar}$ where the latter is defined by $\Delta G_{solv,PB}^{polar}$.

In GB model, $\Delta G_{solv,GB}^{polar}$ is expressed as a sum of Coulombic interactions between each pair of charges (q_i, q_j) in a solvent of dielectric constant ε and the Born self solvation energy of each individual charge [97],

$$\Delta G_{solv,GB}^{polar} = -166\left(1-\frac{1}{\varepsilon}\right)\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{q_i q_j}{f_{GB}} \qquad (3.16)$$

$$f_{GB} = \sqrt{r_{ij}^2 + \alpha_i \alpha_j \exp\left\{-\frac{r_{ij}^2}{4\alpha_i \alpha_j}\right\}} \qquad (3.17)$$

f_{GB} is an effective distance function that depends on the inter-atomic distances (r_{ij}) and the effective Born radii of atoms i and j (α_i, α_j). In computing the electrostatic solvation free energies, ε is generally set to 80, the dielectric constant for bulk water.

In PB model, $\Delta G_{solv,PB}^{polar}$ is estimated according to the PB equation [98].

$$\nabla \bullet [\varepsilon(r)\nabla\phi(r)] - \varepsilon K^2 \sinh[\phi(r)] + 4\pi q \rho^f \frac{(r)}{kT} = 0 \qquad (3.18)$$

where • denote scalar product, $\phi(r)$ is the dimensionless electrostatic potential with respect to the charge density $\rho(r)$ expressed in the units of $\frac{k_B T}{q}$, k_B is the Boltzmann constant, T is the absolute temperature, q is the proton charge, ε is the dielectric constant, and ρ^f is the static (non-polarisable) charge density. K^2 is equal to $\frac{1}{l^2}$, where l is the Debye length depending on the ionic strength of the bulk solution and ε. ϕ, ε, K, *and* ρ are functions of the position vector r.

$\Delta G_{solv}^{non-polar}$ is approximated by a linear function of the solvent accessible surface area (SASA).

$$(\Delta G_{solvation}^{nonel}) = \gamma \times SASA \qquad (3.19)$$

with $\gamma = 0.00542$ kcal mol^{-1}Å^{-2}

Technically, the GB approach is considered computationally less intensive. However, solution of the PB equation has been treated as the standard method and has been used as the basis for the development of GB parameters. Thus, MM-PBSA method is considered somewhat superior in terms of accuracy although it is expected that these two methods may yield comparable results when GB is parameterized properly. Nevertheless, comparisons between these two important methods on large data sets indicate that the results of MM-PBSA approach are comparable to the computationally less expensive GB-based scoring function [99].