METHODS OF CALCULATIONS

1. Theoretical background

1.1 ONIOM method

The computational chemistry plays a very important role in studies of chemical problems. The challenge of the computational chemistry is accurately treating the large molecular systems such as enzymes. No theoretical method is able to provide both the accuracy and acceptable computational cost. Accurate *ab initio* quantum mechanics (QM) method scale non-linearly with the size of the system while MM methods are wildely used and scales linearly but it has the weakness of the poor description of bond breaking and forming processes. Meanwhile, the actual reaction in many biological systems occurs in a relative small region. Therefore, it is not necessary to use a single computational method to treat the whole system. This result leads to development and application of variety of hybrid methods. ONIOM (*Our N-layered Integrated molecular Orbital + molecular Mechanics*) method which was developed by Morokuma and colleagues is a popular hybrid method for the theoretical treatment of large molecular systems (Dapprich *et al.*, 1999; Morokuma *et al.*, 2002; Morokuma *et al.*, 2003; Morokuma *et al*.,2006).

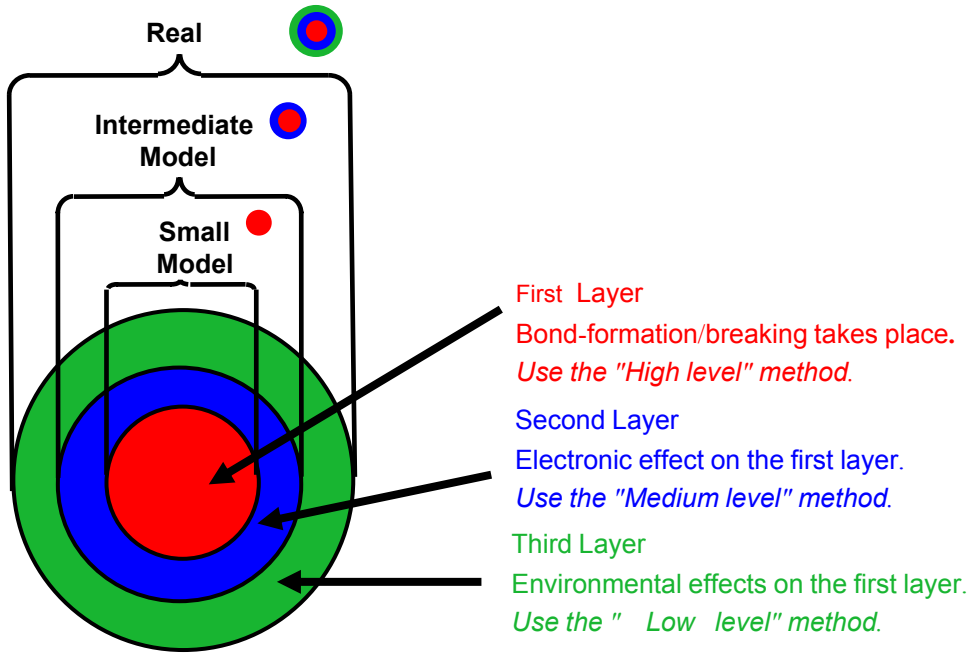

Figure 4 Schematic concept of ONIOM method.

Source Morokuma (2006)

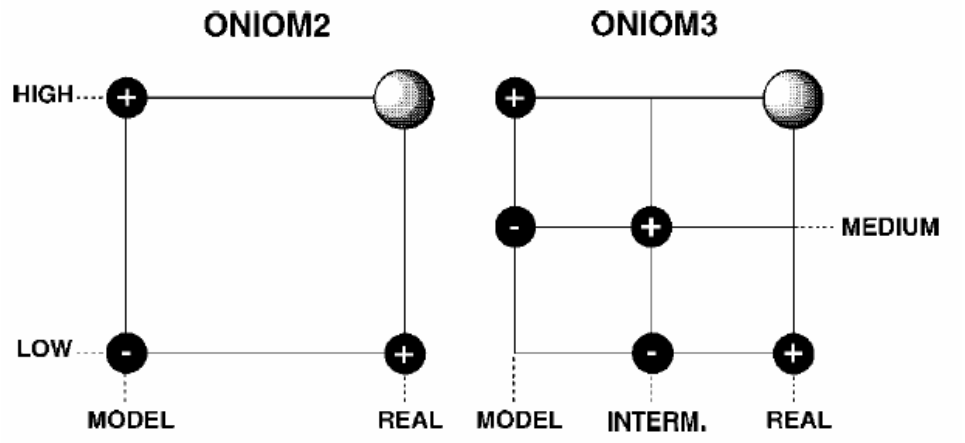

Figure 5 Schematic representation of the two- and three-layer ONIOM extrapolation scheme.

Source Morokuma (2006)

ONIOM divides the system into several onion-like layers, treating the active center with the highest level *ab initio* QM method, while outer layer can be treated with less expensive methods, such as low-level *ab initio* QM, semiempirical QM or MM methods. Figure 4 and 5 illustrate the basic concept of the multilayered ONIOM method. The method can be regarded as an extrapolation scheme.

In the two-layered ONIOM method, the total energy of the system is obtained from three independent calculations:

$$E^{ONIOM2} = E^{high}_{model} + E^{low}_{real} - E^{low}_{model} \quad (1)$$

where *real* denotes the full system, which is treated at the *low* level, while *model* denotes the part of the system of which the energy is calculated at both the *high* and *low* level. For a three-layered ONIOM (high: medium: low), energy is approximated to the energy at the high level for the real system, E(high,real), referred to as the target, and can be written as

$$E^{ONIOM3} = E^{high}_{\text{model}} + E^{medium}_{\text{intermediate}} + E^{low}_{\text{intermediate}} + E^{low}_{real} - E^{medium}_{\text{model}} \quad (2)$$

Where *high*, *medium*, and *low* signify the high-, medium- and low-level theoretical methods, respectively. *Real* signifies the entire system being studied and is only treated with a low level of theory, while the *middle* system is evaluated at low- and medium-level of theory and is only used in ONIOM3 calculations. The *small* inner layer is treated with both the medium- and high-level of theory.

1.1.2 Treatment of link atoms

An important and critical feature of the partitioning of the system into two or more parts or layers is the treatment of the boundary region. In the construction of the ONIOM model system, atoms that belong to the high level layer have the same coordinates as the corresponding atoms in the real system. Even during the geometry optimizations, these coordinates remain identical to each other. When no

bond exists between two layers, the first derivative of the energy with respect to the geometry (q) is easy to obtain as:

$$\frac{\partial E^{ONIOM}}{\partial q} = \frac{\partial E^{high}_{model}}{\partial q} + \frac{\partial E^{low}_{real}}{\partial q} - \frac{\partial E^{low}_{model}}{\partial q} \quad (3)$$

Mostly, if one is interested in the accurate description of a particular region of a large molecule, covalent bonds have to be cut in order to generate the inner model system. This process leaves dangling bonds at the border of the inner model layer, which have to be saturated in order to avoid a chemically unrealistic model. Therefore, so called *link atoms* -usually hydrogen atoms- are introduced. The link atoms are only present in the model system connected to the high level layer with the same angular and dihedral values as the link atom host (LAHs, the atoms replaced by the link atoms in the model system). The bond distances between the high level layer and the link atoms are obtained by scaling the corresponding distance the high level layer and the LAH atoms:

$$R_{link} = R_{high\ level\ atom} + g\,(R_{LAH} - R_{high\ level\ atom}) \quad (4)$$

Where $R_{high\ level\ atom}$ denotes the atom in the high level layer to which the link atom is connected. The scaling factor g is chosen so that reasonable bond lengths between the LAH atoms and high level layer atoms yield also reasonable bond lengths between the link atoms and high level layer atoms.

1.1.3 Geometry Optimization

ONIOM method that in which combines two or more computational methods in one calculation allows the accurate exploration of the chemistry of very large systems. It has been successful applied for study enzyme system. In two layer ONIOM calculation, the total energy of the system is obtained from three independent calculations as shown in equation (1). The real systems contained all the atoms calculate at low level of theory. The model system contains the part of the system that

is treated at the high level of theory, along with the link atoms that are used to cap dangling bonds resulting from cutting covalent bonds between the region and outer regions. To evaluate the ONIOM energy both regions calculations need to be carried out for model system. Because the position of the link atoms are defined in terms of the atoms in the real system, the potential energy surface (PES), and geometry optimization, is well defined (Vreven *et al*., 2003). The ONIOM gradient is obtained from

$$\frac{\partial E^{ONIOM}}{\partial q} = \frac{\partial E^{high}_{model}}{\partial q} \cdot J + \frac{\partial E^{low}_{real}}{\partial q} - \frac{\partial E^{low}_{model}}{\partial q} \cdot J \,, \tag{5}$$

where ***J*** is Jacobian. Jacobian was used to convert the coordinate system for the model system to the coordinate system for the real system.

1.1.4 The S-Value Test

The accuracy of the ONIOM method depend on the error of ONIOM (Err^{ONIOM}_{real}) (equation 6). $Err_{(ONIOM,real)}$ depends critically on the partitioning of the system in to middle and model systems and the reliability of medium- and low-level methods used in ONIOM method. The S-values of three-layer ONIOM method (Morokuma *et al*., 2006) which are defined as

$$Err^{ONIOM}_{real} = E^{ONIOM}_{real} - E^{high}_{real} \tag{6}$$

$$S^{low}_{real-middle} = E^{low}_{real} - E^{low}_{\mathrm{middle}} \tag{7}$$

$$S^{medium}_{middle-\mathrm{model}} = E^{medium}_{middle} - E^{medium}_{\mathrm{model}} \tag{8}$$

$$S^{high}_{real-\mathrm{model}} = E^{high}_{real} - E^{high}_{\mathrm{model}} \tag{9}$$

From these definitions of S-values, the ONIOM energy can be written as

$$E_{real}^{ONIOM} = E_{\mathrm{mod}el}^{high} + S_{middle-\mathrm{mod}el}^{medium} - S_{real-middle}^{low} \quad (10)$$

Therefore

$$Err_{real}^{ONIOM} = S_{real-\mathrm{mod}el}^{high} + S_{middle-\mathrm{mod}el}^{medium} - S_{real-middle}^{low}$$

$$= [S_{middle-\mathrm{mod}el}^{high} - S_{middle-\mathrm{mod}el}^{medium}] + [S_{real-middle}^{high} - S_{real-middle}^{low}] \quad (11)$$

$$Err_{real}^{ONIOM} = S_{midle-\mathrm{mod}el}^{high-medium} + \Delta S_{real-middle}^{high-low} \quad (12)$$

Therefore, the error of the three-layer ONIOM approximation is zero if $S_{middle-\mathrm{mod}el}^{high} = S_{middle-\mathrm{mod}el}^{medium}$ and $S_{real-middle}^{high} = S_{real-middle}^{low}$. The S-value is useful tool for the calibration of hybrid methods.

1.2 Quantum Chemical Calculations

Theoretical chemistry introduces very well defined mathematical approximations to chemistry problems. One of the challenging tasks for molecular orbital calculations is to approximate structures and dynamics of molecular system. This approach provides a great promise in calculating electronic structures and predicting properties of drug molecules.

Molecular orbital theory is an important method in quantum chemistry for the approximation of structures and dynamics of molecular system. In semiempirical methods, such as MNDO, AM1 and PM3, only the valence electrons of the system are considered, the core electrons are assumed into the nuclear core and parameterized Hamilton operator are used. For *ab initio* methods, calculations are performed directly from theoretical principles without inclusion of experimental data except for the fundamental constants. The most common type of *ab initio* calculations is the Hartree-Fock method, which treats the exchange part of the electron-electron interaction in an exact way while completely neglecting correlation effects. Recently, density functional

theory (DFT), which includes both exchange and correlation effects at relatively small computational cost, has been introduced to chemists as a practical tool. The detailed theoretical background of molecular orbital theory and DFT is presented in Appendix table A1.

In this study, quantum chemical calculations was used to calculate interaction energy of efavirenz with individual residues of HIV-1 RT binding pocket, Morover, the binding energies of the efavirenz inhibitor bound to the binding pocket (NNIBP) were calculated using two-and three-layered ONIOM methods. And then binding energies of Y181C and K103N/Y181C HIV-1 RT were compared with wild type

2. System studied

Recently, the interactions between efavirenz and wild type HIV-1 RT were studied by Nunrium and coworkers (Nunrium *et al*., 2005) and continually in this work. The starting models for calculations were obtained from the X-ray structures of efavirenz bound to HIV-1 RT for both the wild-type and Y181C enzymes, listed in the Protein Data Bank with PDB entry codes 1FK9 (Ren *et al*., 2000) and 1JKH (Ren *et al.,* 2001) respectively; whereas the K103N/Y181C enzyme was modeled from PDB entry code 1IKV (Lindberg *et al.,* 2002) by replacing Y181 with C181 to generate the K103N/Y181C enzyme. The studied binding pocket included residues surrounding the non-nucleoside inhibitor binding pocket (NNIBP) with at least one atom interacting with any of the atoms of the efavirenz inhibitor within the interatomic distance of 7.0 Å These residues of the studied system are Pro95, Leu100, Lys101, Lys102, Lys103(Asn103), Lys104, Ser105, Val106, Val179, Ile180, Tyr181(Cys181), Tyr188, Val189, Gly190, Phe227, Leu228, Trp229, Leu234, His235, Pro236 and Tyr318 from the p66 domain of RT, and Glu138(b) from the 51 domain of RT (Figure 6 and 7). All residues were assumed to be in their neutral form. The N- and C-terminal ends of cut residues were capped with an acetyl group (CH_3CO-) and a methyl amino group (-$NHCH_3$), respectively [$(H_3C–C(=O)–\{NH–CH(–R)–C(=O)\}_n–NH–CH_3)$]. The hydrogen atoms then were added to generate the complete structures and their positions,

including that of the C181 in K103N/Y181C HIV-1 RT, were optimized by the semi-empirical PM3 method as available in the GAUSSIAN 03 program running in Linux on a Pentium IV 3.2 GHz PC. The optimizations were carried out with fixed heavy atoms and the final structures produced were used as the starting geometries for all subsequent calculations.

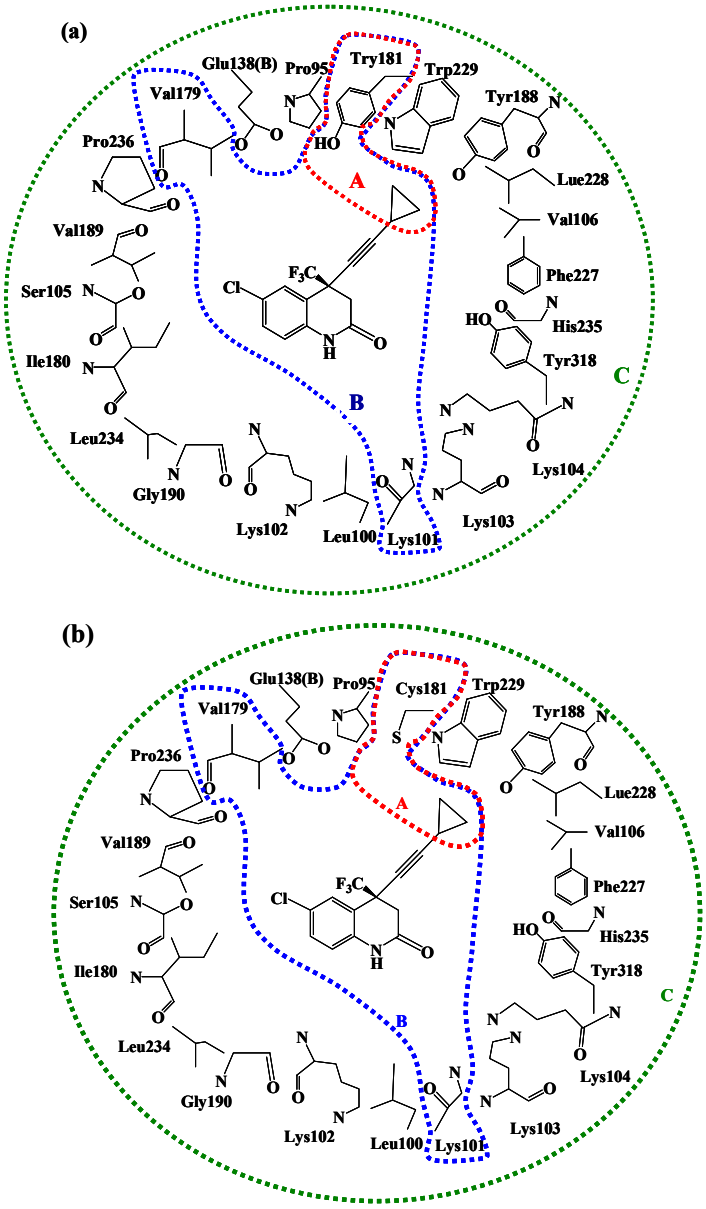

Figure 6 Model system used for efavirenz bound to allosteric site of HIV-1 RT consisting of 22 residues; (a) wild-type HIV-1 RT NNIBP, and (b) Y181C HIV-1 RT NNIBP.

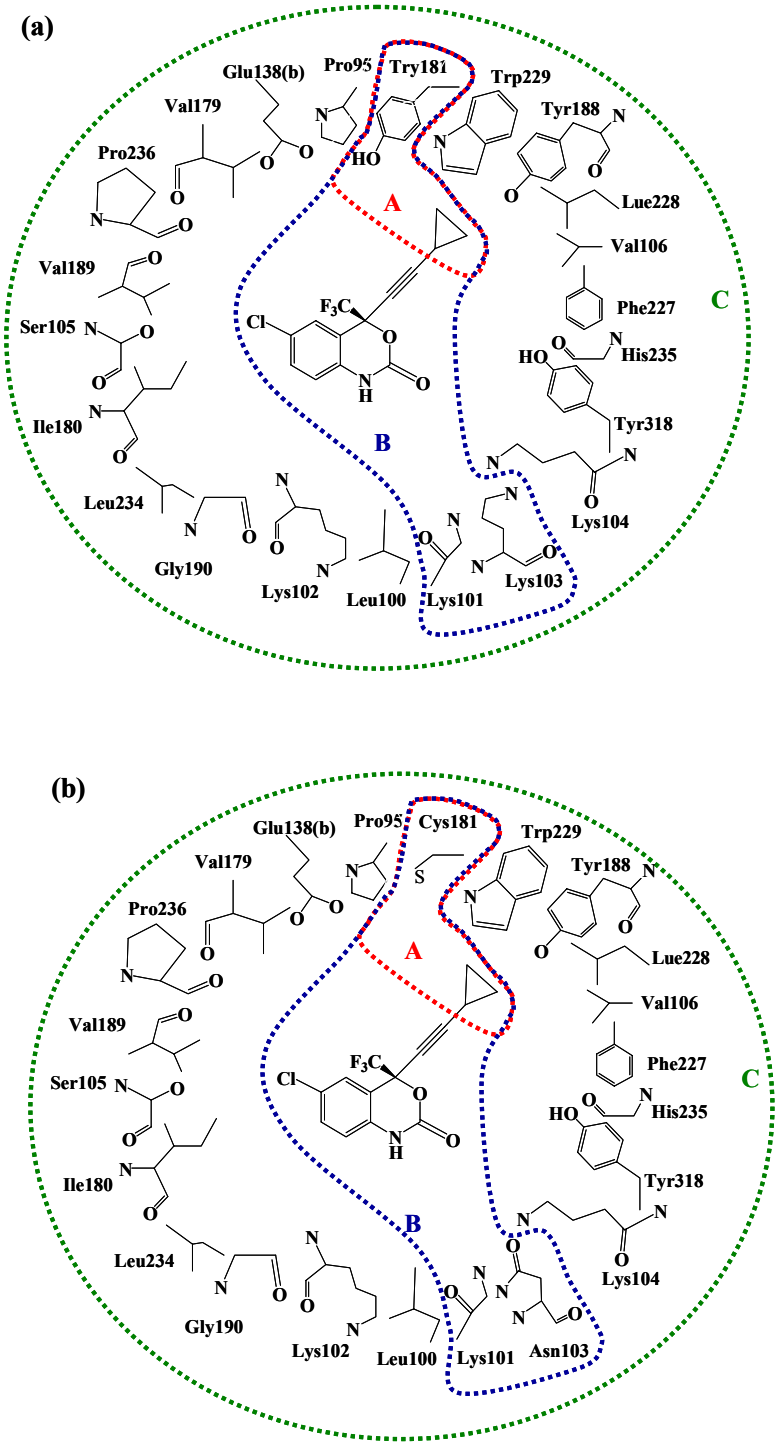

Figure 7 Model system used for efavirenz bound to allosteric site of HIV-1 RT consisting of 22 residues; (a) wild-type HIV-1 RT NNIBP, and (b) K103N/Y181C HIV-1 RT NNIBP.

3. Interaction energy calculations

The interaction energies, $E_{(EFZ+Xi)}$, between efavirenz, EFZ, and individual residues, X_i, were calculated at the B3LYP/6-31G(d,p) and MP2/6-31G(d,p) levels of theory using the geometry described above. The total interaction energy, INT, can be expressed as: (Sea-oon *et al.,* 2005)

$$INT_{(EFZ+Xi)} = E_{(EFZ+Xi)} - E_{(EFZ)} - E_{(Xi)}, \quad (13)$$

where $E_{(EFZ)}$ and $E_{(Xi)}$ are energies of efavirenz and each individual residue, respectively.

4. Binding energy calculations

Two and three layer ONIOM calculations were performed to determine the binding energy of efavirenz bound to the mutant Y181C and K103N/Y181C HIV-1 RT as compared to the wild-type. The total ONIOM energy of the entire system was obtained from three and five independent energy calculations in ONIOM2 and ONIOM3 methods, respectively, as shown in equations (1) and (2) (Morokuma, 2003). All calculations were carried out using the GAUSSIAN 03 package (Frisch *et al.,* 2003).

More precisely, the binding energy of efavirenz bound to the allosteric pocket of HIV-1 RT was determined using equations (14) and (15) for the ONIOM2 and ONIOM3 methods, respectively (Kuno *et al.,* 2003).

$$\begin{aligned}\Delta E^{ONIOM2} &= E[Cpx]_{opt} - E[P]_{opt} - E[L]_{opt} \\ &= \Delta E\,(High, A+B) + [\Delta E\,(Low, ABC) - \Delta E\,(Low, AB)] \\ &= \Delta E\,(High, A+B) + [\Delta\Delta E\,(Low, ABC-AB)] \end{aligned} \quad (14)$$

$$\begin{aligned}\Delta E^{ONIOM3} &= E\,[Cpx]_{opt} - E[P]_{opt} - E[L]_{opt} \\ &= \Delta E\,(High, A) + [\Delta E\,(Mid, AB) - \Delta E\,(Mid, A)] + [\Delta E\,(Low, ABC) - \Delta E\,(Low, AB)] \\ &= \Delta E\,(High, A) + [\Delta\Delta E(Mid, AB-A)] + [\Delta\Delta E(Low, ABC-AB)] \end{aligned} \quad (15)$$

Where $E[Cpx]_{opt}$ is the total optimized energy of the efavirenz-binding pocket complex, Cpx; $E[P]_{opt}$ is the optimized energy of binding pocket; and $E[L]_{opt}$ is the optimized energy of the efavirenz ligand. Also, ΔE (High, A) is the interaction energy in the region A which is treated at the high level of theory, $\Delta\Delta E$ (Mid, AB-A) is the interaction energy from interactions between the regions A and B and is evaluated at the medium level of theory, and $\Delta\Delta E$ (low, ABC-AB) is the interaction energy from interactions between the regions AB and C which is evaluated at the low level of theory.

Moreover, the binding energy can be decomposed into interaction energy (INT) and the deformation energy (DEF) as shown in equation (16) (Saen-oon, 2005).

$$BE = INT + DEF \tag{16}$$

$$DEF = DEF_P + DEF_L \tag{17}$$

$$DEF_p = E[P]_{sp}^{Cpx} - E[P]_{opt} \tag{18}$$

$$DEF_L = E[L]_{sp}^{Cpx} - E[L]_{opt} \tag{19}$$

Where BE, INT, DEF_p, DEF_L, are the binding energy, interaction energy, deformation energy of the pocket and deformation energy of ligand, respectively. Here, $E[P]_{opt}$ and $E[L]_{opt}$ is the total optimized energy of the pocket and ligand, respectively, while $E[P]_{sp}^{Cpx}$ and $E[P]_{sp}^{Cpx}$ is the single point energy of the pocket and ligand, respectively, by taken the configuration from the optimized complex geometry.

4.1 Comparing binding energy calculations between wild type and Y181C enzymes

The binding energy of the complex structures for the wild-type and the Y181C enzymes was determined using both ONIOM2 and ONIOM3 methods. With the ONIOM2 method, the inner layer including efavirenz, K101, Y181 or C181, and V179 (Figure 6, region A+B) was treated by the HF/6-31G(d,p) and B3LYP/6-31G(d,p) methods for HF/6-31G(d,p):PM3 and B3LYP/6-31G(d,p) calculations, respectively. The outer layer (Figure 6, region C) was treated by the PM3 method. In contrast, as a minimum MP2 method was needed to calculate the partial H-π interactions between the cyclopropylethylnyl group of efavirenz and the aromatic side-chain of Y181, therefore, the MP2/6-31G(d,p) method was used in ONIOM3 calculations to determine the dispersion interactions as the density functional theory cannot do this accurately enough (Kristya *et al.,* 1994, Tsuzuki *et al.,* 2001). In the ONIOM3 method, the system was divided into three regions as shown in Figure 6. The inner layer (Figure 6, region A) including cyclopropylethylnyl group of efavirenz (mEFZ) and Y181 or C181 was treated at the MP2/6-31G(d,p) and B3LYP/6-31G(d,p) methods. The medium layer (Figure 6, region B) including remainder of efavirenz, K101 and V179 was treated at the HF/6-31G(d,p), B3LYP/6-31G(d,p) methods. The outer layer (Figure 6, region C) was treated at the PM3 level. For this study, the following models were generated:

ONIOM2 calculations:
HF/6-31G(d,p)[EFZ+K101+V179+(Y181 or C181)]:PM3[real]
B3LYP/6-31G(d,p)[EFZ+K101+Val179+(Y181 or C181)]:PM3[real]

ONIOM3 calculations:
B3LYP/6-31G(d,p)[mEFZ+(Y181 or C181)]:HF/6-31G(d,p)[EFZ+ K101+V179]:PM3[real]
MP2/6-31G(d,p)[mEFZ+(Y181 or C181)]:HF/6-31G(d,p)[EFZ+K101+V179]: PM3[real]
MP2/6-31G(d,p)[mEFZ+(Y181 or C181)]:B3LYP/6-31G(d,p)[EFZ+K101+V179]:PM3[real]

4.2 Comparing binding energy calculations between wild-type and K103N/Y181C enzymes

With the ONIOM2 method, the inner layer (Figure 7, region (A+B)) including efavirenz, K101, K103 or N103, and Y181 or C181 was treated by the HF/6-31G(d,p) and B3LYP/6-31G(d,p) methods. The outer layer (Figure 7, region C) was treated at the PM3 method. With the ONIOM3 method, the inner layer or interaction region (Figure 7, region A) including the cyclopropylethylnyl of efavirenz and Y181 or C181 was treated by the B3LYP/6-31G(d,p) and MP2/6-31G(d,p) methods. The medium layer (Figure 7, region B) including the remainder of efavirenz, K101, and K103 or N103 were treated by the HF/6-31G(d,p) and B3LYP/6-31G(d,p) methods. The outer layer (Figure 7, region C) was treated by the PM3 method. For this study, the following models were generated:

ONIOM2 calculations:
HF/6-31G(d,p)[EFZ+K101+K103+(Y181 or C181)]: PM3[real]
B3LYP/6-31G(d,p)[EFZ+K101+K103+(Y181 or C181)]: PM3[real]

ONIOM3 calculations:
B3LYP/6-31G(d,p)[mEFZ+(Y181 or C181)]:HF/6-31G(d,p)[EFZ+K101+ (K103 or N103)]:PM3[real]
MP2/6-31G(d,p)[mEFZ+Y181 or C181]:HF/6-31G(d,p)[EFZ+K101+(K103 or N103)]:PM3[real]
MP2/6-31G(d,p)[mEFZ+(Y181 or C181)]:B3LYP/6-31G(d,p)[EFZ+K101+(K103 or N103)]:PM3[real]