APPENDIX

APPENDIX A: Theoretical Background

The Theory of Quantum Mechanics

There are a number of quantum theories for treating molecular systems and the one which has been most widely used is molecular orbital theory (Leach, 1996; Jensen, 1999; Cramer, 2002). The postulates and theorems of quantum mechanics based on molecular orbital theory form the foundation for the prediction of observable chemical properties. The systems are described by 'wave function, Ψ' that completely characterizes all of the physical properties of the systems.

1. Molecular Orbital Theory (MOT)

The objective of all ab initio electronic structure theories is the exact solution of the time-independent Schrödinger equation; which can be expressed in a time-independent form as shown in equation (1).

$$H\Psi = E\Psi \tag{1}$$

Where H = the Hamiltonian operator

E = energy

Ψ = the wavefunction

The typical form of Hamiltonian operator (H_{tot}) takes into account five contributions to the total energy of a system. They are the kinetic energies of the electrons (T_e) and nuclei (T_n), the attraction of the electrons to the nuclei (V_{ne}), and the interelectronic (V_{ee}) and internuclear (V_{nn}) repulsions as shown in equation (2).

$$H_{tot} = T_e + T_n + V_{ne} + V_{ee} + V_{ne} \tag{2}$$

Since the masses of nuclei are much greater than the masses of the electrons, they move more slowly. Hence, to a good approximation, one can consider the electrons in a molecule to be moving in the field of fixed nuclei. The 'Born-Oppenheimer approximation' can be used to further simplify the Schrödinger

equation. This allows the equation to be separated into electronic and nuclear terms. Within this approximation, the kinetic energy of the nuclei (T_n) can be neglected. The remaining terms in equation (2) are called the electronic hamiltonian, as written in equation (3).

$$H_{elec} = -\sum_{i}^{N} \frac{1}{2}\nabla_i^2 - \sum_{i}^{N}\sum_{a} \frac{Z_a}{|r_i - R_a|} + \sum_{i}^{N}\sum_{j>i}^{N} \frac{1}{|r_i - r_j|} + \sum_{a}\sum_{b>a} \frac{Z_a Z_b}{|R_a - R_b|} \qquad (3)$$

Where *i* and *j* represent electrons, *a* and *b* represent nuclei, *Z* are the atomic number, ∇^2 is the Laplacian operator, *r* and *R* is the distance between particles and atomic units are used in equation (3). The Laplacian has the form shown in equation (4).

$$\nabla_i^2 = \frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y_i^2} + \frac{\partial^2}{\partial z_i^2} \qquad (4)$$

Under the Born-Oppenheimer approximation, the total wave function for the molecule can be written in the following form (equation (5)).

$$\Psi_{tot}\,(\text{nuclei, electrons}) = \Psi(\text{electrons})\Psi(\text{nuclei}) \qquad (5)$$

The total energy equals the sum of the nuclear energy the electrostatic repulsion between the positively charged nuclei) and the electronic energy. The electronic energy comprises the kinetic and potential energy of the electrons moving in the electrostatic field of the nuclei, together with electron-electron repulsion (equation (6)).

$$E_{tot} = E(\text{electrons}) + E(\text{nuclei}) \qquad (6)$$

An appropriate functional form of the wave function for a polyelectronic system with noninteracting N electrons has a Hamiltonian shown in equation (7).

$$H = \sum_{i=1}^{N} h(i) \tag{7}$$

Where *h(i)* is the one-electron Hamiltonian defined by equation (8).

$$h(i) = -\frac{1}{2}\nabla_i^2 - \sum_{a=1}^{M} \frac{Z_a}{|r_i - R_a|} \tag{8}$$

Where M is the total number of nuclei.

Eigenfunctions of the one-electron Hamiltonian (equation (8)) must satisfy the corresponding one electron Schrödinger equation (a set of spin orbital, χ_j).

$$h(i)\chi_j(x_i) = \varepsilon_j\chi_j(x_i) \tag{9}$$

Because *H* is a sum of one-electron Hamiltonians, a wave function is a simple product of spin orbital wave functions for each electron, as shown in equation (10).

$$\Psi^{HP}(x_1, x_2, \ldots, x_N) = \chi_i(x_1)\,\chi_j(x_2)\ldots\,\chi_k(x_N) \tag{10}$$

A wave function of the form in equation (10) is called a 'Hartree product' and it is an eigenfunction of *H* with eigen value *E* (equation (11)).

$$H\Psi^{HP} = E\Psi^{HP} \tag{11}$$

Where *E* is the sum of the spin orbital energies of each of the spin orbitals appearing in Ψ^{HP}.

$$E = \varepsilon_i + \varepsilon_{j+\ldots+}\varepsilon_k \tag{12}$$

From equation (7), there is still a basic deficiency in the Hartree product. It takes no account of the indistinguishability of electrons, but specifically distinguishes electron-one as occupying spin orbital χ_i, electron-one as occupying spin orbital χ_j, etc. Because the Hartree product does not satisfy the antisymmetry principle, the

correct antisymmetrized wave functions can obtain as follows. For example, consider a two-electron case occupying the spin orbitals χ_i and χ_j, the electron-one and electron-two are put in χ_i and χ_j, respectively, as shown in equation (13).

$$\Psi_{12}^{HP}(x_1, x_2) = \chi_i(x_1)\, \chi_j(x_2) \tag{13}$$

On the other hand, if the electron-one and electron-two are put in χ_j and χ_i, respectively, the Hartree product is shown in equation (14).

$$\Psi_{21}^{HP}(x_1, x_2) = \chi_i(x_2)\, \chi_j(x_1) \tag{14}$$

Each of these Hartree products clearly distinguishes between electrons. The wave function which satisfies the requirement of the antisymmetry principle by taking appropriate linear combination of these two Hartree products is shown in equation (15).

$$\Psi(x_1, x_2) = 2^{-1/2}(\chi_i(x_1)\, \chi_j(x_2) - \chi_j(x_1)\, \chi_i(x_2)) \tag{15}$$

The factor $2^{-1/2}$ is a normalization factor. The minus sign insures that $\Psi(x_1, x_2)$ is antisymmetric with respect to the interchange of the coordinates of electrons one and two.

$$\Psi(x_1, x_2) = - \Psi(x_2, x_2) \tag{16}$$

From equation (15), it is evident that the wave function vanishes if both electrons occupy the same spin orbital (i.e. if $i=j$). Thus the antisymmetry requirement immediately leads to the usual statement of the Pauli exclusion principle namely, that no more than one electron can occupy a spin orbital. The antisymmetric wave function of equation (15) can be rewritten as determinant in equation (16), as called a *'Slater determinant'*.

$$\Psi(x_1, x_2) = 2^{-1/2} \begin{vmatrix} \chi_i(x_1) & \chi_j(x_1) \\ \chi_i(x_2) & \chi_j(x_2) \end{vmatrix} \quad (16)$$

This determinant is the most convenient way to write down the permitted functional forms of a polyelectronic wavefunction that satisfies the antisymmetry principle. In general, for an N electrons system with spin orbitals χ_1, χ_2,…, χ_N, an acceptable form of the wavefunction is:

$$\Psi_{SD} = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(1) & \chi_2(1) & \dots & \chi_N(1) \\ \chi_1(2) & \chi_2(2) & \dots & \chi_N(2) \\ \vdots & \vdots & & \vdots \\ \chi_1(N) & \chi_2(N) & \dots & \chi_N(N) \end{vmatrix} \quad (17)$$

Where $\chi_1(1)$ is used to indicate a function that depends on the space and spin coordinates of the electron labeled '1'. The factor $1/\sqrt{N!}$ ensures that the wavefunction is normalized.

Slater determinant implies that electron correlation is neglected, or equivalently, the electron-electron repulsion is only included as an average effect. To find a solution which simultaneously enables all the electronic motions to be taken into account, the change in the spin orbital for one electron will influence the behavior of an electron in another spin orbital due to the coupling of the electronic motions. Therefore, Fock proposed the extension of Hartree's Self-consistent field (SCF) procedure to Slater determinantal wavefunctions. The Hartree-Fock (HF) equation is shown in equation (18).

$$f(i)\chi(x_i) = \varepsilon\chi(x_i) \quad (18)$$

Where $f(i)$ is an effective one-electron operator, called the Fock operator, of the form in equation (19).

$$f(i) = -\frac{1}{2}\nabla_i^2 - \sum_{a=1}^{M} \frac{Z_a}{|r_i - R_a|} + v^{HF}(i) \quad (19)$$

Where $v^{HF}(i)$ is the average potential or Hartree-Fock potential experienced by the i^{th} electron due to the other present of the other electrons. The essence of the Hartree-

Fock approximation is to replace the complicated many-electron problem by a one-electron problem in which electron-electron repulsion is treated in an average way.

From equation (19), the first two terms represent the kinetic and potential energy of each electron moving in the field of nuclei, as called core Hamiltonian operator, $H^{core}(i)$. The last term, $v^{HF}(i)$, includes the interelectronic interactions consisting of coulomb operator ($J_j(i)$) and exchange operator ($K_j(i)$) (equation (20)). The coulomb operator contributes to the energy arises from the electrostatic repulsion between pairs of electrons. The exchange operator represents energy of the exchange interaction that it has no classical analogy and no arises because the motions of electrons with parallel spin are correlated. Electrons with the same spin thus tend to avoid each other and they show a lower coulombic interaction giving a lower (i.e. more favorable) energy.

$$v^{HF}(i) = \sum_{j}^{N} \{J_j(i) - K_j(i)\} \qquad (20$$

The coulomb operator ($J_j(1)$) and exchange operator ($K_j(1)$) can be written as in equations (21) and (22), respectively.

$$Jj(1) = \int d\tau_2 \chi_j(2) \frac{1}{r_{12}} \chi_j(2) \qquad (21)$$

$$Kj(1)\chi_i(1) = [\int d\tau_2 \chi_j(2) \frac{1}{r_{12}} \chi_i(2)]\chi_j(1) \qquad (22)$$

In equation (21), this operator corresponds to the average potential due to an electron in χ_j. The two-electron potential $(r_{12})^{-1}$ felt by electron 1 and associated with the instantaneous position of electron 2 is thus replaced by a one-electron potential, obtained by averaging the interaction $(r_{12})^{-1}$ of electron 1 and electron 2, over all space and spin coordinates τ_2 of electron 2, weighted by the probability $d\tau_2 \left|\chi_j(2)\right|^2$ that electron 2 occupies the volume element $d\tau_2$ at τ_2. By summing over all $j \neq i$, one obtains the total averaged potential acting on the electron in χ_i, arising from the *N-1* electrons in the other spin orbitals. The exchange operator (equation (22)) corresponds

to the average potential due to an electron in χ_j. It arises from the antisymmetric nature of the single determinant and has a somewhat strange form.

Equation (19) can be written as in equation (23).

$$f_i(i) = h(i) + \sum_{j=1}^{N} \{J_j(i) - K_j(i)\} \qquad (23)$$

The general strategy for solving the HF equation is called the self-consistent field (SCF) method. The way to solve this equation is as follows**.** First, a set of trial solutions χ_i to the Hartree-Fock eigenvalue equations are obtained. These are used to calculate the coulomb and exchange operators. The Hartree-Fock equations are solved, giving a second set of solutions χ_i, which are used in the next iteration. The SCF method thus gradually refines the individual electronic solutions that correspond to lower and lower total energies until the point is reached at which the results for all electrons are unchanged, when they are said to be self consistent. The solution of the Hartree-Fock equations is not a practical proposition for molecules, therefore it mostly uses linear combination of atomic orbitals (LCAO) by writing writes each spin orbital as a linear combination of single electron orbitals (equation (24)). It is introduced as a set of *K* known basis function.

$$\psi_i = \sum_{v=1}^{K} c_{vi}\phi_v \qquad i = 1, 2,\ldots,K \qquad (24)$$

The one-electron orbitals, ϕ_v, are commonly called basis functions. More sophisticated calculations use more basis than a minimal set. In accordance with the variation theorem, the set of coefficients, c_{vi}, giving the lowest energy wavefunction are required. For a given basis set and a given functional form of the wavefunction (i.e. a Slater determinant), the best set of coefficients that gives minimum energy are from the point that $\frac{\partial E}{\partial c_{vi}} = 0$. If equation (24) is substituted into the Hartree-Fock equation equation (18), the Hartree-Fock equation can be written in the form of equation (25).

$$f_i(1)\sum_{v=1}^{K} c_{vi}\phi_v(1) = \varepsilon_i \sum_{v=1}^{K} c_{vi}\phi_v(1) \qquad (25)$$

The derivation of the Hartree-Fock equations for such a system was first proposed by Roothaan and Hall. The resulting equations are known as the Roothaan equations or the Roothaan-Hall equations. Unlike the integral-differential form of the Hartree-Fock equations, Roothaan and Hall propose the equations in matrix form which can be solved using standard techniques and can be applied to systems of any geometry. By premultiplying each side of equation (25) by $\phi_\mu(1)$ (ϕ_μ is also a basis function) and integrating, the matrix equation is given in equation (26). The Fock operator, $f(1)$, is an one-electron operator, and any set of one-electron functions defines a matrix representation of this operator.

$$\sum_{v=1}^{K} c_{vi} \int d\tau_1 \phi_\mu(1) f_i(1) \phi_v(1) = \varepsilon_i \sum_{v=1}^{K} c_{vi} \int d\tau_1 \phi_\mu(1) \phi_v(1) \qquad (26)$$

The overlap matrix, $S_{\mu v}$, has been defined in equation (27), it is the overlap integral between the basis functions μ and ν.

$$S_{\mu v} = \int d\tau_1 \phi_\mu(1) \phi_v(1) \qquad (27)$$

The Fock matrix, $F_{\mu v}$, is the matrix representation of the Fock operator with the set of basis function $\{\phi_\mu\}$, as defined in equation (28).

$$F_{\mu v} = \int d\tau_1 \phi_\mu(1) f_i(1) \phi_v(1) \qquad (28)$$

With these definitions of $F_{\mu v}$ and $S_{\mu v}$, equation (26) can be written as in the following equation.

$$\sum_v F_{\mu v} c_{vi} = \varepsilon_i \sum_v S_{\mu v} c_{vi} \qquad i = 1, 2, \ldots, \mathrm{K} \qquad (29)$$

This equation is the Roothaan-Hall equation, which can be conveniently written as a matrix equation (equation (30)).

$$\mathrm{FC} = \mathrm{SCE} \qquad (30)$$

Where C is a $K \times K$ square matrix of the expansion coefficients $C_{\mu i}$.

$$C = \begin{pmatrix} C_{11} & C_{12} & \dots & C_{1K} \\ C_{21} & C_{22} & \dots & C_{2K} \\ \vdots & \vdots & & \vdots \\ C_{K1} & C_{K2} & \dots & C_{2K} \end{pmatrix} \quad (31)$$

And ε is a diagonal matrix of the orbital energies ε_i.

$$\varepsilon = \begin{pmatrix} \varepsilon_1 & 0 & \dots & 0 \\ 0 & \varepsilon_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \varepsilon_K \end{pmatrix} \quad (32)$$

From equation (24) and equation (31), it is the columns of C which describe the molecular orbitals, i.e., the coefficients describing Ψ_1 are in the first column of C, those describing Ψ_2 are in the second column of C, etc.

In molecular modeling, the ground states of molecules which have closed-shell configurations are usually concerned. In a closed-shell system containing N electrons in N/2 orbitals, there are two spin orbitals, Ψ_i: $\Psi_i\alpha$ and $\Psi_i\beta$. There is the energy of each electron moving in the field of the bare nuclei. For an electron in a molecular orbital χ_i, this contribute an energy H_{ii}^{core}. If there are two electrons in the orbital then the energy is $2H_{ii}^{core}$ and for $N/2$ orbitals the total contribution to the energy will be $\sum_{i=1}^{N/2} 2H_{ii}^{core}$. The expression of the Hartree-Fock energy for the closed-shell system is shown in equation (33).

$$E = 2\sum_{i=1}^{N/2} H_{ii}^{core} + \sum_{i=1}^{N/2}\sum_{j=1}^{N/2}(2J_{ij} - K_{ij}) \quad (33)$$

The corresponding Fock operator is shown in equation (34).

$$f_i(1) = h(1) + \sum_{j=1}^{N/2} \{2J_j(1) - K_j(1)\} \quad (34)$$

The Fock matrix, $F_{\mu\nu}$ (equation (28)), for a closed-shell system can be expanded as follows by substituting the expression for the Fock operator.

$$F_{\mu\nu} = \int d\tau_1 \phi_\mu(1) h(1) \phi_\nu(1) + \sum_{j=1}^{N/2} \int d\tau_1 \phi_\mu(1) [2J_j(1) - K_j(1)] \phi_\nu(1) \quad (35)$$

$$F_{\mu\nu} = H_{\mu\nu}^{core} + \sum_{j=1}^{N/2} \int d\tau_1 \phi_\mu(1) [2J_j(1) - K_j(1)] \phi_\nu(1) \quad (36)$$

Where a core-Hamiltonian matrix, $H_{\mu\nu}^{core}$, is defined in equation (37). This matrix is integral involving the one-electron operator $h(1)$, describing the kinetic energy and nuclear attraction of an electron, as shown in equation (8).

$$H_{\mu\nu}^{core} = \int d\tau_1 \phi_\mu(1) h(1) \phi_\nu(1) \quad (37)$$

Recall that the coulomb operator ($J_j(1)$) due to interaction with a spin orbital χ_j is shown in equation (21). The terms of appropriate linear combination of basis functions are replaced to the spin orbital terms (χ_j), as shown in equation (38).

$$Jj(1) = \int d\tau_2 \sum_{\sigma=1}^{K} c_{\sigma j} \phi_\sigma(2) \frac{1}{r_{12}} \sum_{\lambda=1}^{K} c_{\lambda j} \phi_\lambda(2) \quad (38)$$

Where σ and λ are the basis functions.

Similarly, the terms of appropriate linear combination of basis functions are replaced to the spin orbital terms (χ_j) of the exchange operator in equation (22), as shown in equation (39).

$$Kj(1)\chi_i(1) = [\int d\tau_2 \sum_{\sigma=1}^{K} c_{\sigma j} \phi_\sigma(2) \frac{1}{r_{12}} \chi_i(2)] \sum_{\lambda=1}^{K} c_{\lambda j} \phi_\lambda(2) \quad (39)$$

When the coulomb and exchange operators are expressed in terms of the basis functions and the orbital expansion is substituted for χ_i, then their contribution to the Fock matrix element, $F_{\mu\nu}$, are in the following form.

$$\sum_{j=1}^{N/2} \int d\tau_1 \phi_\mu(1)[2J_j(1) - K_j(1)]\phi_\nu(1)$$

$$= \sum_{j=1}^{N/2} \sum_{\lambda=1}^{K} \sum_{\sigma=1}^{K} c_{\lambda j} c_{\sigma j} [2\int d\tau_1 d\tau_2 \phi_\mu(1)\phi_\nu(1)\frac{1}{r_{12}}\phi_\lambda(2)\phi_\sigma(2) - \int d\tau_1 d\tau_2 \phi_\mu(1)\phi_\sigma(1)\frac{1}{r_{12}}\phi_\nu(2)\phi_\sigma(2)]$$

$$= \sum_{j=1}^{N/2} \sum_{\lambda=1}^{K} \sum_{\sigma=1}^{K} c_{\lambda j} c_{\sigma j} [2(\mu\nu \mid \lambda\sigma) - (\mu\lambda \mid \nu\sigma)] \qquad (40)$$

In equation (40), the shorthand notations for the integrals are used. The two-electron integrals may involve up to four different basis functions (μ, ν, λ, σ), which may in turn be located at four different centers. To simplify equation (40), the charge density matrix, P, is defined in equation (41).

$$P_{\mu\nu} = 2\sum_{i=1}^{N/2} c_{\mu i} c_{\nu i} \quad \text{and} \quad P_{\mu\nu} = 2\sum_{i=1}^{N/2} c_{\mu i} c_{\nu i} \qquad (41)$$

Therefore, the Fock matrix, $F_{\mu\nu}$, for a closed-shell system of N electrons can be shown in equation (42).

$$F_{\mu\nu} = H_{\mu\nu}^{core} + \sum_{\lambda=1}^{K} \sum_{\sigma=1}^{K} P_{\lambda\sigma}[(\mu\nu \mid \lambda\sigma) - \frac{1}{2}(\mu\lambda \mid \nu\sigma)] \qquad (42)$$

The electronic energy in terms of the density matrix can be calculated by using equation (43).

$$E = \frac{1}{2}\sum_{\mu=1}^{K} \sum_{\nu=1}^{K} P_{\mu\nu}(H_{\mu\nu}^{core} + F_{\mu\nu}) \qquad (43)$$

The electron density in terms of the density matrix at a point r can also be shown in equation (44).

$$\rho(r) = \sum_{\mu=1}^{K} \sum_{\nu=1}^{K} P_{\mu\nu}\phi_\mu(r)\phi_\nu(r) \qquad (44)$$

Consider the elements of Fock matrix in Roothaan-Hall equation equation (30), it should be noted that the element in the left-hand side depends on the molecular orbital coefficients c_{vi}, which also appear on the right-hand side of the equation. It indicated that the solution of this equation required an iterative procedure. The $H_{\mu\nu}^{core}$ due to the electron moving in the field of the bare nuclei do not depend on the basis set coefficients and unchanged throughout the calculation. In case of coulomb and exchange contributions depend on the coefficients and vary throughout the calculation. The individual two-electron integrals (μν|λσ) are constant throughout the calculation. To solve the Roothaan-Hall equation by using standard eigenvalue methods, Roothaan-Hall equation is required in the following form.

$$FC = CE \tag{45}$$

This equation will be in this form if S is the unit matrix (I). The basis sets, ϕ, that are used in molecular calculations are not orthonormal set. They are usually normalized but they are not orthogonal to each other. This gives rise to the overlap matrix in Roothaan-Hall equation, so they must be transformed (matrix X) (equation (46)).

$$X^{T}SX = I \tag{46}$$

Where X^{T} is the transpose of X.

To obtain two different transformation matrices X, S can be diagonalized by a unitary matrix, U, as shown in equation (47).

$$U^{T}SU = s \tag{47}$$

Where s is the diagonal matrix of the eigenvalues of S.

This expression is often written in equation (48) since for real basis function $U^{-1} = U^{T}$.

$$U^{-1}SU = s \quad (48)$$

Then the matrix X is given by equation (49).

$$X = Us^{-1/2}U^{T} \quad (49)$$

It can be considered X as $S^{-1/2}$, as shown in equation (50).

$$S^{-1/2}SS^{-1/2} = I \quad (50)$$

The Roothaan-Hall equations can now be manipulated as follows. Both sides of Roothaan-Hall equation (equation (30)) are pre-multiplied by the matrix $S^{-1/2}$ (equation (51)).

$$S^{-1/2}FC = S^{-1/2}SCE = S^{1/2}CE \quad (51)$$

Inserting the $S^{-1/2}$ $S^{1/2}$ into the left-hand side, it is shown in equation (52).

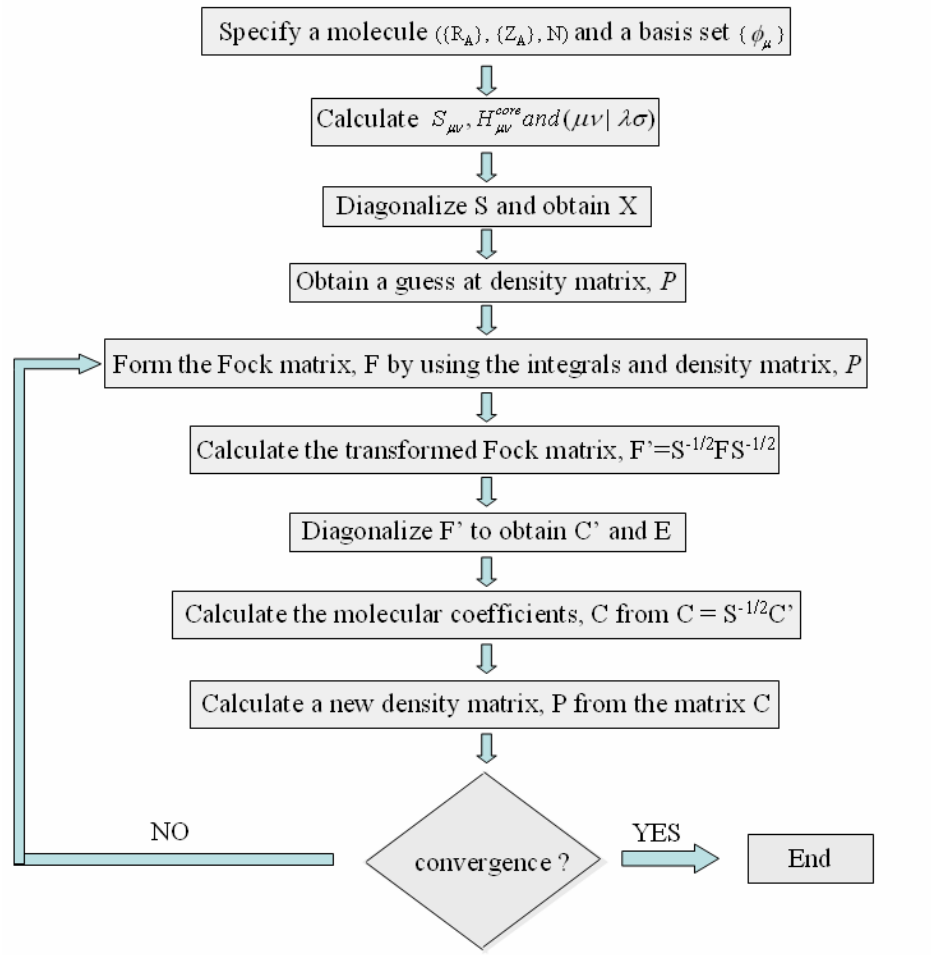
$$S^{-1/2}F(S^{-1/2}\ S^{1/2})C = S^{1/2}CE \quad (52)$$

Equation (52) can be written as in equation (53) that can be solved using standard method.

$$F'C' = C'E \quad (53)$$

Where $F' = S^{-1/2}F\ S^{-1/2}$ and $C' = S^{1/2}C$

A commom scheme for solving the Roothaan-Hall equations by using SCF procedure is shown in Appendix Figure A1.



Appendix Figure A1 A common scheme for solving the Roothaan-Hall equations by using SCF procedure.

In quantum mechanical molecular orbital calculations, it can be separated into two major categories. They are the *ab initio* and the semi-empirical methods. *Ab initio* usually refers to a calculation which uses the full Hartree-Fock/Roothaan-Hall equations, without ignoring or approximating any of the integrals or any of the terms in the Hamiltonian. By contrast, semi-empirical method reduced the CPU time during calculation by using parameters for some of the integrals and/or ignoring some of the terms in the Hamiltonian.

2. Basis set

The basis sets most commonly used in quantum mechanical calculations are composed of atomic functions. For atoms, the Hartree-Fock equations are usually solved numerically if it is assumed that the electron distribution is spherically symmetrical. For example, in case of the hydrogen atom, the analytical approximation to its successful solution has the form, as shown in equation (54).

$$\psi = R_{nl}(r)Y_{lm}(\theta,\phi) \tag{54}$$

Where *Y* is a spherical harmonic and *R* is a radial function. The radial function obtained for hydrogen atom cannot be used directly for polyelectronic atoms due to the screening of the nuclear charge by the inner shell electron. Slater has suggested a simpler analytical form for the radial function, universally known as Slater–type orbitals (STOs) (equation (55)).

$$R_{nl}(r) = (2\zeta)^{n+1/2}[(2n)!]^{-1/2} r^{n-1} e^{-\zeta r} \tag{55}$$

Unfortunately, Slater–type orbitals for many electron atoms are not suitable in implementation in molecular orbital calculation, because some of integrals are difficult to evaluate. It is common in *ab initio* calculations to replace the Slater orbitals by Gaussian function. The Gaussian function is shown in equation (56).

$$g(\alpha,r) = x^a y^b z^c \exp(-\alpha r^2) \tag{56}$$

It is found that replacing a Slater type orbital by a single Gaussian function leads to an unacceptable errors. However, this problem can be overcome by representing each atomic orbital as a linear combination of Gaussian function, as in equation (57). By using a basis set of contracted Gaussian functions, each basis

function is a fixed linear combination (contraction) of Gaussian functions (primitives).

$$\psi_i = \sum_{\mu=1}^{K} c_{\mu i} \phi_\mu \tag{57}$$

Some commonly used of basis sets are described as follows.

2.1 Minimal basis set

A minimal basis set is a relatively inexpensive one which contains just the number of functions that are required to accommodate all the filled orbitals in each atom. For instance, hydrogen and helium has a single s-type function. STO-*n*G are all minimal basis sets using a contraction of *n* primitive Gaussian functions to represent each orbital.

2.2 Split valence basis set

Because a minimal basis set uses only a contraction for each orbital, to improve this problem, more than one function (i.e. double, triple, etc.) is used for each orbital. A basis set which uses two functions for each of minimal basis set of valence electron is described as a double zeta basis set, for example 3-21G, 4-31G and 6-31G.

2.3 Polarized basis set

This type of basis set was performed by adding polarization function, ex. p-type or d-type function. The use of polarization basis function is indicated by an asterisk (*). For example, 6-31G* refers to a 6-31G basis set with polarization functions on the heavy (i.e. non-hydrogen) atoms.

2.4 Diffused basis set

To include a significant amount of electron density away from the nuclei center such as anions and molecule containing lone pairs, diffused function, denoted by +, was added to increase the amplitude or size of Gaussian basis function. For example, 3-21G+ basis set contains an additional single set of diffuse s- and p-type Gaussian functions.

3. Møller-Plesset perturbation theory (MPPT)

The difference between the exact energy (determined by the Hamiltonian) and the HF energy is known as the correlation energy, as shown in equation (58).

$$E_{correlation} = E_{exact} - E_{HF} < 0 \quad (58)$$

This difference is referred to as the electron correlation energy and is attributable to the neglect in HF theory of the instantaneous interactions (correlations) between electrons. Several approaches are known that try to calculate the correlation energy, i.e. Configuration Interaction (CI), Møller-Plesset Perturbation Theory (MPPT) and Multi-Configuration SCF (MCSCF or CASSCF).

Møller-Plesset perturbation theory treats electron correlation as a perturbation, V, on the Hartree–Fock problem and both the wavefunction and the energy are expanded in a power series of the perturbation, as shown in equation (59)-(61).

$$H = H_{HF} + \lambda V \quad (59)$$

Where λ is a parameter that can vary between 0 and 1

$$\psi_i = \psi_{HF} + \lambda\psi_i^{(1)} + \lambda^2\psi_i^{(2)} + \ldots = \sum_{n=0} \lambda^n \psi_i^{(n)} \quad (60)$$

$$E_i = E_{HF} + \lambda E_i^{(1)} + \lambda^2 E_i^{(2)} + \ldots = \sum_{n=0} \lambda^n E_i^{(n)} \quad (61)$$

The Møller-Plesset calculation is specified using the level of theory used (i.e. MP2, MP3) together with the basis set. For example, MP2/6-31G* indicated a second-order Møller-Plesset calculation with 6-31G* basis set.

4. <u>Density functional theory (DFT)</u>

Density functional theory (DFT) is presented by Hohenberg and Kohn. The basis for DFT is the Hohenberg–Kohn theorem that states that the exact ground state energy of a molecular system is a functional only of the electron density (ρ) and the fixed positions of the nuclei. The energy functional E(ρ) can be written in terms of its composite parts as follows.

$$E(\rho) = E_{KE}(\rho) + E_C(\rho) + E_H(\rho) + E_{xc}(\rho) \tag{62}$$

Where $E_{KE}(\rho)$ is the kinetic energy, $E_C(\rho)$ is the electron-nuclear interaction term, $E_H(\rho)$ is the electron-electron coulombic energy, and $E_{xc}(\rho)$ contains the exchange and correlation contributions.

All of the electron-electron interactions are contained within E_H and E_{xc} terms. Hohenberg and Kohn also proved that the exact electron density is the one which minimizes the energy, thereby providing a variational approach for finding the density.

<u>Molecular docking</u>

An important method widely used in drug discovery is molecular docking (Halperin *et al.*, 2002). Molecular docking is a computational method attempting to predict the three-dimensional structures of receptor-ligand complexes and to evaluate the relative affinity of these bound ligands. Consequently, the aims of this method are to identify correct poses of ligands in the binding pocket of a protein and to predict

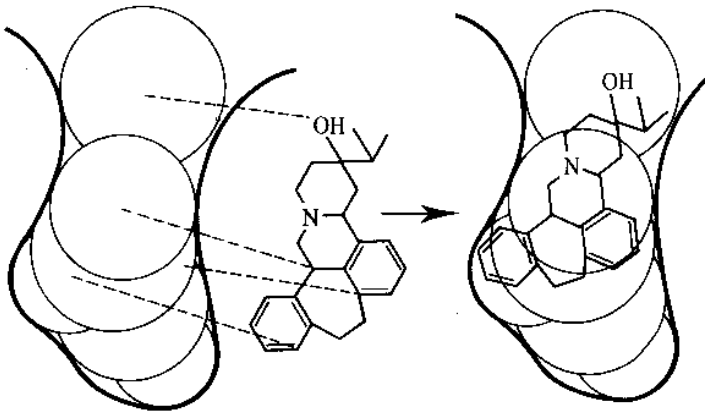
the affinity of their correct poses (Krovat *et al.*, 2005). In molecular docking, two proposes are needed in its procedure. They are docking algorithm used to generate a large number of possible structures and scoring function used to identify which are the most interest (Leach, 1996; Böhm, 2002; Brooijmans and Kuntz, 2003).

1. Docking algorithm

The first step in molecular docking method is to search the configurational and conformational degrees of freedom by using docking algorithm. The docking algorithm can be categorized by using degree of freedom. In rigid docking, the search algorithm finds the various position of ligand in the active site of receptor using translational and rotational degrees of freedom. To perform conformationally flexible docking, torsional degree of freedom of ligand is added to explore it in the active site of receptor.

1.1 Rigid docking

Rigid docking fits a rigid ligand into a rigid active site. The usual approach of this algorithm has been to generate a discrete model of the binding site consisting a set of points that define its spatial extensions with special emphasis on surface properties. Atoms of ligands are then matched onto these receptor points. The first docking program using this technique is DOCK program (Kuntz *et al.*,1982). DOCK program represents a binding site as overlapping spheres. Ligands are then matched to sphere centers by using clique technique, as shown in Appendix Figure A2.



Appendix Figure A2 DOCK algorithm. Atoms are matched to a binding site represented as a collection of overlapping spheres.

Source: Leach (1996)

1.2 Conformationally flexible docking

Docking by using rigid docking is fast but highly inaccurate for ligands with rotatable bonds, therefore the flexibility of ligands have been considered in docking algorithm, as called conformationally flexible docking. It can be divided in three types, i. e. systematic, stochastic and deterministic searches.

1.2.1 Systematic search

This algorithm is based on a grid of values for each formal degree of freedom. Each of these grid values is explored in a combinatorial fashion during a search. An example of a systematic search is the anchor-and-grow or incremental construction algorithm. In this algorithm, the ligand conformers are not an external pre-processing step. Their conformers are generated within the boundaries of the binding site. Ligands are generally divided into rigid (anchor) and flexible parts. Conformers are generated by growing the ligand from an anchor fragment. The procedure is described as follows. Firstly, one or more anchors with flexible parts are defined by perception of rotatable bonds. Secondly, the anchor fragment is docked into the active site. Finally, the flexible parts are added sequentially with systematic

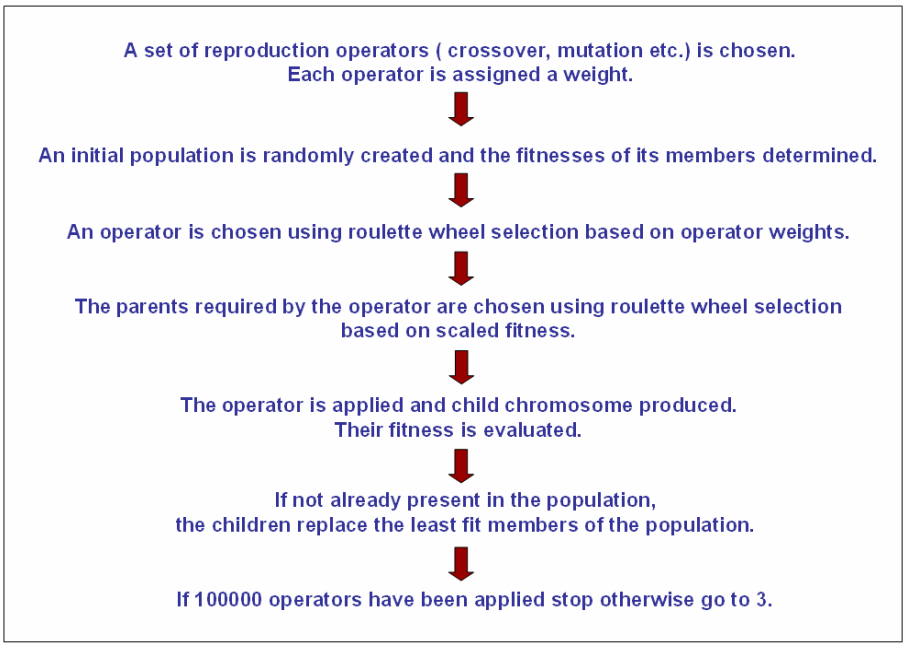
scanning of torsional angles. The anchor-and-grow or incremental construction algorithm is implemented in several docking programs, e. g. DOCK, FlexX, and Hammerhead.

1.2.2 Stochastic search

Because of the combinatorial fashion in incremental construction algorithm, it is appealing idea to search the whole orientational and conformational space in one process, e.g. stochastic search. Stochastic search algorithms make random changes, usually changing one degree of freedom of the system at a time. Advantage of the stochastic search algorithms is the less complicated data structure requirement. Examples of stochastic searches are Monte Carlo (MC) methods and evolutionary algorithms.

In Monte Carlo (MC) algorithms, a whole ligand is usually represented by a string of real value variables describing translation, rotation and variable torsional angles. Random changes in these variables form the basis of these algorithms. After each move, the structure is minimized, and the energy of the new structure is determined. Several docking programs using the Monte Carlo (MC) algorithms are AutoDock, ICM, MCDOCK, ProDock, and PRO_LEAD.

Evolutionary algorithms are stochastic methods using to find the energy minimum. Evolutionary algorithms involve generation of an initial population of ligand conformations. They did not include crossover process, a process that swaps large regions of the “parents” during evolution. When these algorithms include crossover process, it is known as genetic algorithms (GA). The common feature of evolutionary and genetic algorithms is a cyclic variation-selection process. “Parents” breed “offspring”, and the best of each generation according to a fitness measure becomes the parent of the next optimization cycle. The general procedure of GA is shown in as shown in Appendix Figure A3. The examples of programs using these algorithms are GOLD, GA-DOCK, and AutoDock.



Appendix Figure A3 Genetic algorithm (GA) procedure.

1.2.3 Deterministic search

In deterministic searches, the initial state determines the move that can be made to generate the next state, which generally has to be equal to or lower in energy than the initial state. Examples of deterministic searches are energy minimization methods and molecular dynamics (MD) simulations.

2 Scoring function

If it is assumed that docking algorithm have performed very well in searching the conformational space in the binding site, accurate scoring function is considered to be the most important and challenging step (Seifert *et. al.*, 2003). Scoring function is a function giving an estimate of binding of a molecule or molecular fragment in a given orientation and conformation (called a pose) in the binding site. The scoring function can be categorized into three main classes, i.e. force field-based methods, empirical scoring functions and knowledge-based method.

2.1 Force field-based methods

The force field-based scoring functions use the molecular mechanics force fields for the estimation of binding affinity. The interaction energies (van der Waals and electrostatic) between the receptor and ligand are computed. The overestimation of complex stability can be partially ascribed to the neglect of solute entropic terms. This scoring function can also added the empirical terms to take into account the entropy and solution changes. The AMBER and CHARMM force field are widely used as a scoring function in several docking program. Examples of the programs implemented the force field-based scoring functions are DOCK and AutoDock programs.

2.2 Empirical scoring functions

The empirical scoring functions are the most widely utilized in current drug design/discovery software. The empirical scoring functions are based on the assumption that the binding free energy ($\Delta G_{binding}$) is a sum of interactions multiplied by weighting coefficients (ΔG_i), as shown in equation (63). The energy contributions in scoring function usually contain individual terms for hydrogen bonds, ionic interactions, hydrophobic interactions, and binding entropy.

$$\Delta G_{binding} \approx \sum \Delta G_i f_i(r_l, r_{l'}) \quad (63)$$

Where each f_i is a simple geometrical function of the ligand coordinate r_l and the receptor coordinate $r_{l'}$. Most empirical scoring functions are derived by evaluating the functions f_i on a set of protein-ligand complexes and fitting the coefficients (ΔG_i) to experiment binding affinities of these complexes. The relative weight of each energy contribution depends on the training set. General problems of empirical scoring functions are that they are difficult to know what each term exactly accounts for and to access where errors come from. Binding free energy estimations can only be successful if the molecules make similar interactions to the ones in the training set complexes. Examples of the programs implemented the empirical scoring functions are FlexX, GOLD, AutoDock and LUDI.

2.3 Knowledge-based methods

The knowledge-based scoring functions use the structural data of protein-ligand complexes to observe frequencies of atom-atom interactions. Force and potentials are collected to get a score for their binding affinities, as called potential of mean force (PMF). The major different with the empirical scoring function is that no binding data are needed. This is an advantage of knowledge-based scoring functions that they can be derived from large training set. Such knowledge-based approaches have their foundation in the inverse formulation of the Boltzmann law, as shown in equation (64).

$$E_{ijk} = -kT\ln(p_{ijk}) + kT\ln(Z) \quad (64)$$

Where the energy function (E_{ijk}) is called PMF for a state defined by the variables i, j, and k, p_{ijk} is the corresponding probability density, and Z is the partition function. The second term of the sum is constant at constant temperature T and does not have to be regarded, since $Z = 1$ can be chosen by definition of a suitable reference state leading to normalized probability densities, p_{ijk}. Variables i, j, and k can be chosen to be protein atom-types, ligand atom-types, and their inter-atom distance. The reference state is defined as the state where protein and ligand do not interact each other. PMF and DrugScore are the examples of knowledge-based scoring functions.

Up to now, there is no single scoring function that can correctly rank every protein-ligand complex because the relative contribution of different protein-ligand interactions may vary between structural families (Krovat *et al.*, 2005). The scoring functions still need the improvements to enhance the reliability of discrimination correctly docked from misdocked conformation or consensus scoring that is a combination of two or three scoring function has been proposed (Kontoyianni *et al.*, 2005).

APPENDIX B: Ultrasensitive Reverse Transcriptase Assay

To measure the activities of the 32 selected compounds for inhibiting the reverse transcriptase activity of wild-type, K103N and Y181C HIV-1 RT, an ultrasensitive reverse transcriptase assay is used by Assistant Professor Dr. Arunee Thitithanyanont and coworkers at Pr-615 virus laboratory, Department of microbiology, Faculty of Science, Mahidol University.

From transfection plasmid, molecular clone of wild-type, K103N and Y181C HIV-1 RT in RT sequence are generated. These molecular clones are sensitive to the drugs. The methods are separated into three parts, i.e. reverse transcription (RT), real-time polymerase chain reaction (PCR), and the measurement of %inhibition of RT and IC_{50}.

1. Reverse transcription

The preparation of master mix contains 5x Buffer, DTT, dNTP, EMCV primer, RNase-inhibitor, EMCV RNA template, RNase-free water and NP40. The master mix is added to each tube of standard enzyme (AMV reverse transcriptase (10 units/µl), QIAGEN), wild-type HIV-1 RT, K103N HIV-1 RT, and Y181C HIV-1 RT. Standard enzyme, wild-type HIV-1 RT, K103N HIV-1 RT, and Y181C HIV-1 RT are prepared as following. Standard enzyme (AMV reverse transcriptase) is 10 fold diluted and used at 10^6, 10^7, and 10^8 nanounits/µl for making standard curve. For wild-type, K103N, and Y181C HIV-1 RT, they are 10 fold diluted and used at 10^7 nanounits/µl. They are added with the compounds for testing HIV-1 inhibition. The mixture of the master mix, AMV-RT of HIV-1 RT and the compounds are incubated in the water bath at temperature 42 ^{o}c for an hour. In this step, RT activity of AMV-RT and HIV-1 RT will convert EMCV RNA template to EMCV cDNA. If the compound has inhibitory affinity for RT, EMCV cDNA in the tube will be decreased. After an hour, RT activity is degraded by the incubation in heat block at temperature 95 ^{o}c for 5 minutes.

2. Real-time polymerase chain reaction (PCR)

The master mix consisting of 10x Buffer, $MgCl_2$, dNTP, Probe, EMCV forward and backward primers, hot tag DNA polymerase and distilled water are prepared. The products from reverse transcription are added to the master mix. They are performed real-time PCR in rotor gene for two hours for increasing the number of EMCV cDNA by DNA polymerase. Finally, the numbers of EMCV DNA product are detected by probe.

3. The measurement of %inhibition of RT and IC_{50}

RT activities of wild-type, K103N and Y181C are detected by comparing with RT activity of AMV RT. The %inhibition of RT of the selected compounds are calculated by the following formula.

$$\%inhibition\ of\ RT\ = \frac{(RT\ activity\ of\ blank\ HIV-1RT^{*} - RT\ activity\ of\ HIV-1RT^{*}\ added\ compound)}{RT\ activity\ of\ blank\ HIV-1RT^{*}}\ x\ 100$$

Where HIV-1 RT* = wild-type or K103N or Y181C HIV-1 RT

%inhibition of RT and the concentration of compound (M) are plotted the non-linear regression by using GraphPad Prism. The IC_{50} is then calculated by the following relationship.

$$Y\ =\ 1/(1+10^{(LogIC_{50}-M)Slope})$$

Where Y = %inhibition x 0.01

APPENDIX C: Oral Presentations and Poster Contributions to Conferences

Oral Presentation and Proceedings

1. Patchreenart Saparpakorn, Pornpan Pungpo and Supa Hannongbua. Interaction Energy and Bound Conformational Prediction of HIV-1 Reverse Transcriptase Inhibitor: Nevirapine Derivatives by Using Molecular Docking. The 7th Annual National Symposium, On Computational Science and Engineering (ANSCSE 2003), Mar 23-25, 2003, Department of Chemistry, Faculty of Science, Chulalongkorn University, Bangkok, Thailand.

2. Patchreenart Saparpakorn, Supa Hannongbua and Didier Rognan. Virtual screening of K103N and Y181C HIV-1 reverse transcriptase based on nevirapine and some NNRTIs. RGJ-Ph.D. Congress VI, Apr 28-30, 2005, Jomtien Palm Beach Hotel, Pattaya, Thailand.

3. Patchreenart Saparpakorn, Supa Hannongbua and Didier Rognan. Design of nevirapine derivatives insensitive to the K103N and Y181C HIV-1 Reverse Transcriptase Mutants. In the proceeding of The 3rd International Symposium Computational Methods in Toxicology & Pharmacology Integrating Internet Resources, Oct 29- Nov 2, 2005, Shanghai, Republic of China, page 47.

4. Patchreenart Saparpakorn, Jae Hyoun Kim and Supa Hannongbua. Binding study of Polycyclic Aromatic Hydrocarbons (PAHs) and Alkylphenols (APs) with Fulvic acid (FA), Humic acid (HA) and Soil Organic Matter (SOM). In the proceeding of The 10th Annual National Symposium, On Computational Science and Engineering (ANSCSE 2006), Mar 22-24, 2006, Department of Chemistry, Faculty of Science, Chiang-Mai University, Chiang-Mai, Thailand page 122-127.

Poster Contribution to Conferences

1. Patchreenart Saparpakorn, Supa Hannongbua and Didier Rognan. Virtual screening of K103N and Y181C HIV-1 reverse transcriptase mutants based on nevirapine and some NNRTIs. The 15th European Symposium on Quantitive Structure - Activity Relationships & Molecular Modelling, Sep 5-10, 2004 at Istanbul, Turkey.

2. Patchreenart Saparpakorn, Supa Hannongbua and Didier Rognan. Virtual Screening of K103N HIV-1 Reverse Transcriptase Mutant Based on PNU-142721. The 2nd Asian Pacific Conference on Theoretical and Computational Chemistry, May 2-6, 2005 at Chulalongkorn University, Bangkok, Thailand.

3. Pornpan Pungpo, Patchreenart Saparpakorn and Supa Hannongbua. Elucidating inhibitory models of efavirenz analogues as potent HIV-1 reverse transcriptase inhibitors by 3D QSAR and docking studies. The 2nd Asian Pacific Conference on Theoretical and Computational Chemistry, May 2-6, 2005 at Chulalongkorn University, Bangkok, Thailand.

4. Pornpan Pungpo, Oradee Pankwang, Patchreenart Saparpakorn, Peter Wolschann and Supa Hannongbua. Theoretical investigations on potent HIV-1 reverse transcriptase inhibitors of efavirenz analogues by using conformational analysis, molecular docking and 3D-QSAR studies. The XIIth International Congress of Quantum Chemistry (XII-ICQC2006), May 21-26, 2006, Kyoto TERRSA, kyoto, JAPAN.

5. Suwipa Saen-oon, Mayuso Kuno, Thanyada Rungrotemongkuol, Peerapol Nunrium, Patchreenart Saparpakorn and Supa Hannongbua. Drug-enzyme interaction; structural information for inhibitor design. The XIIth International Congress of Quantum Chemistry (XII-ICQC2006), May 21-26, 2006, Kyoto TERRSA, kyoto, JAPAN.

Virtual screening of K103N and Y181C HIV-1 reverse transcriptase mutants based on nevirapine and some NNRTIs.

Patchreenart Saparpakorn[1,2], Supa Hannongbua[1] and Didier Rognan[2]

Department of Chemistry, Faculty of Science, Kasetsart University, Bangkok, Thailand.[1]
Bioinformatics Group, Laboratoire de Pharmacochimie de la Communication Cellulaire, CNRS UMR 7081, Illkirch, France.[2]

Introduction

The reverse transcriptase (RT) of human immunodeficiency virus type- 1 (HIV-1) is the essential enzyme converting the single-stranded viral RNA genome into double-stranded proviral DNA prior to its integration into the host genomic DNA. Non-nucleoside reverse transcriptase inhibitors (NNRTIs) are non-competitive inhibitors binding in a hydrophobic pocket that is about 10 Å away from the enzyme's active site. They were widely used for the treatment of acquired immunodeficiency syndrome (AIDS). Nevirapine (Viramune®) was the first generation of NNRTI that has been approved by the FDA for the treatment of HIV-1 infection. Its efficiency is limited by drug resistant mutations[1], such as K103N and Y181C. Therefore, the aim of this work is to find novel inhibitors insensitive to the K103N and Y181C mutations, based on nevirapine and some NNRTIs.

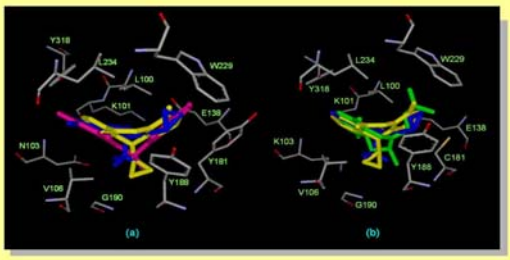


Fig1. a). The docked conformations of nevirapine (yellow), efavirenz (blue) and PNU-142721 (pink) in the K103N binding pocket.
b). The docked conformations of nevirapine (yellow), efavirenz (blue) and 8-Cl TIBO (green) in the Y181C binding pocket.

Results and Discussion

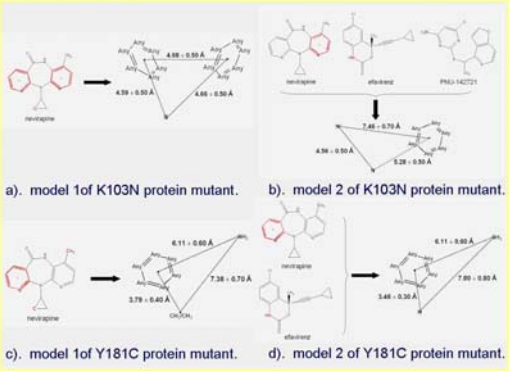
From three docking methods, GOLD was the most accurate to reproduce the X-ray bound conformation of nevirapine to both mutants. The rmsd of the docked pose from the X-ray pose of nevirapine are 0.41Å and 0.42 Å for K103N and Y181 mutants, respectively. Some NNRTIS have been docked to the K103N binding pocket and the Y181C binding pocket, (Fig1). For the K103N protein mutant, two pharmacophore models have derived from the docked conformation of nevirapine, efavirenz and PNU-142721. For the Y181C protein mutant, two pharmacophore models have derived from the docked conformation of nevirapine and efavirenz. All pharmacophore models were used as constraints for searching the 'in-house' database and were shown in fig2. From 4 pharmacophore models, there are 20,200 compounds and 19,761 compounds matching pharmacophore for the K103N and Y181C protein mutants. These compounds were selected for further molecular docking with GOLD. Finally, 6k virtual hits were selected and classified by scaffold diversity using ClassPharmer. 41 compounds were purchased and tested.

Acknowledgment

This work was supported by the Royal Golden Jubilee Ph.D. Program (3.C.KU/45/B.1) and the French Government

Methods

Three docking methods (FlexX[2], GOLD[3] and Surflex[4]) were applied to the nevirapine-RT complexes of K103N (pdb code 1fkp) and Y181C (pdb code 1jlb) mutations to validate and select the best possible docking/scoring strategy. Before performing the molecular docking with an 'in-house' database of a set of 500k commercially available drug-like compounds, these compounds were filtered by using pharmacophore searching with the UNITY4.4 program[5]. The 3D pharmacophore models were constructed based on the known important interactions between the amino acid in the binding pocket and NNRTIs. After filtering the database, the selected compounds were applied to docking. The hits from docking were selected and classified by scaffold diversity using ClassPharmer [6]. Finally, the selected compounds from the classification were tested for HIV-1 RT inhibition.



a). model 1of K103N protein mutant. b). model 2 of K103N protein mutant.
c). model 1of Y181C protein mutant. d). model 2 of Y181C protein mutant.

Fig 2. The features and constraints of all pharmacophore models.

References

1. Dyatkin AB, Brickwood JR, Proudfoot JR, Bioorg. Med. Chem. Lett. 1998, 8: 2169-2172.
2. Rarey M, Kramer B, Lengauer T, Klebe G, J. Mol. Biol. 1996, 261: 470-489.
3. Jones G, Wilett P, Glen RC, Leach AR, Taylor R, J. Mol. Biol. 1997, 267: 727-748.
4. Jain AN, J. Med. Chem. 2003, 46: 499-511.
5. UNITY 4.4; Tripos Inc., St. Louis, MO, USA.
6. ClassPharmer; Bioreason Inc., Sante Fe, NM, USA.

Virtual screening of K103N HIV-1 reverse transcriptase based on PNU-142721.

Patchreenart Saparpakorn[1,2], Supa Hannongbua[1] and Didier Rognan[2]

Department of Chemistry, Faculty of Science, Kasetsart University, Bangkok, Thailand.[1]
Bioinformatics Group, Laboratoire de Pharmacochimie de la Communication Cellulaire, CNRS UMR 7081, Illkirch, France.[2]

Introduction

The reverse transcriptase (RT) of human immunodeficiency virus type- 1 (HIV-1) is the essential enzyme converting the single-stranded viral RNA genome into double-stranded proviral DNA prior. Non-nucleoside reverse transcriptase inhibitors (NNRTIs) are non-competitive inhibitors binding in a hydrophobic pocket that is about 10 Å away from the enzyme's active site. The efficiency of these inhibitors is limited by drug resistant mutations. In the clinic, one of the important mutations frequently observed in the patients is K103N mutation. PNU-142721 is an active inhibitor against the K103N mutation[1]. Its potent activity toward the K103N mutation is explained by new interaction with the N103 (Fig 1). Therefore, the aim of this work is to find novel inhibitors insensitive to the K103N mutation, based on PNU-142721.

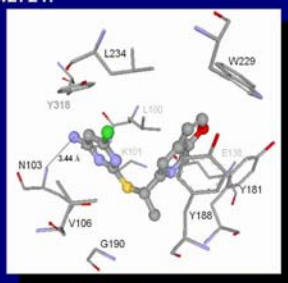


Fig 1. The orientation of PNU-142721 in the K103N binding pocket.

Methods

Three docking methods (FlexX, GOLD and Surflex) were applied to the PNU-142721-RT complex of K103N (pdb code 1ikx) mutation to validate and select the best possible docking/scoring strategy. Before performing the molecular docking with an 'in-house' database of a set of 500k commercially available drug-like compounds, these compounds were filtered by using pharmacophore searching with the UNITY4.4 program. A 3D pharmacophore model was constructed based on the known important interactions between the amino acid in the binding pocket and PNU-142721. After filtering the database, the selected compounds were applied to docking. The hits from docking were selected and classified by scaffold diversity using ClassPharmer. Finally, the selected compounds from the classification were tested for HIV-1 RT inhibition.

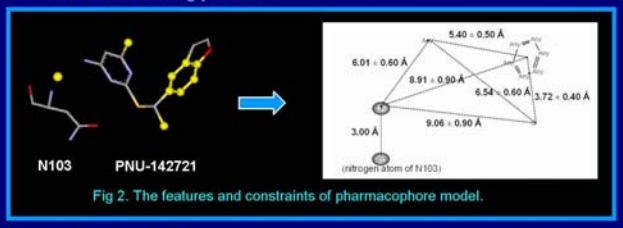
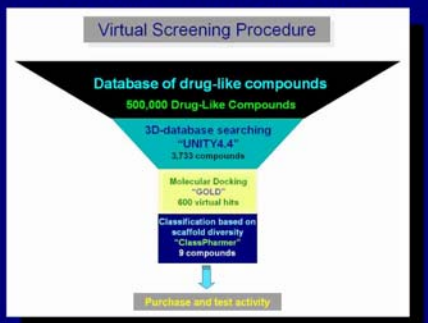
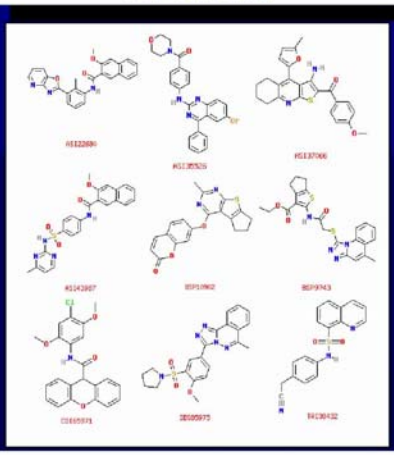


Fig 2. The features and constraints of pharmacophore model.



Results and Discussion

From three docking methods, GOLD was the most accurate to reproduce the X-ray bound conformation of PNU-142721. The rmsd of the docked pose from the X-ray pose of PNU-142721 is 0.94 Å. Pharmacophore model was derived from a centroid of six-membered aromatic ring and two atom centers of docked PNU-142721 and hydrogen bond site of N103. It was used as constraints for searching the 'in-house' database and were shown in fig2. From pharmacophore model, there is 3,733 compounds matching pharmacophore. These compounds were selected for further molecular docking with GOLD. Finally, 600 virtual hits were selected and classified by scaffold diversity using ClassPharmer. 9 compounds were purchased and tested.



References

1. Lindberg J, Sigurdsson S, Lowgren S, Andersson H O, Sahlberg C, Noreen R, Fridborg K, Zhang H, Unge T, Eur. J. Biochem. **2002**, 269: 1670-1677.

Acknowledgment

This work was supported by the Royal Golden Jubilee Ph.D. Program (3.C.KU/45/B.1), TRF grant (BRG4780007), Postgraduate on Education and Research in Petroleum and Petrochemical Technology (MUA-ADB) and French Embassy in Thailand.

APPENDIX D: Publications

Publication I

Virtual screening of K103N and Y181C HIV-1 reverse transcriptase mutants based on nevirapine and some NNRTIs.

Saparpakorn P, Hannongbua S, Rognan D.

The proceeding of the 15th European Symposium on Quantitive Structure - Activity Relationships & Molecular Modelling, Sep 5-10, 2004, at Istanbul, Turkey

EuroQSAR 2004 Proceedings, pp. 483-484
QSAR and Molecular Modelling in Rational Design of Bioactive Molecules
Eds. E. Akı (Şener) and İ. Yalçın

Virtual Screening of K103N and Y181C HIV-1 Reverse Transcriptase Mutants Based on Nevirapine and Some NNRTIs

Patchreenart Saparpakorn[1,2], Supa Hannongbua[1], Didier Rognan[2]

[1]*Department of chemistry, Faculty of science, Kasetsart University, Bangkok, Thailand*
[2]*Bioinformatics Group, Laboratoire de Pharmacochimie de la Communication Cellulaire, CNRS UMR 7081, Illkirch, France*

Introduction

The reverse transcriptase (RT) of human immunodeficiency virus type- 1 (HIV-1) is the essential enzyme converting the single-stranded viral RNA genome into double-stranded proviral DNA prior to its integration into the host genomic DNA. Non-nucleoside reverse transcriptase inhibitors (NNRTIs) are non-competitive inhibitors binding in a hydrophobic pocket that is about 10 Å away from the enzyme's active site. They were widely used for the treatment of acquired immunodeficiency syndrome (AIDS). Nevirapine (Viramune®) was the first generation of NNRTI that has been approved by the FDA for the treatment of HIV-1 infection. Its efficiency is limited by drug resistant mutations[1], such as K103N and Y181C. Therefore, the aim of this work is to find novel inhibitors insensitive to the K103N and Y181C mutations, based on nevirapine and some NNRTIs.

Methods

Three docking methods (FlexX[2], GOLD[3] and Surflex[4]) were applied to the nevirapine-RT complexes of K103N and Y181C mutations to validate and select the best possible docking/scoring strategy. Before performing the molecular docking with an 'in-house' database of a set of 500k commercially available drug-like compounds, these compounds were filtered by using pharmacophore searching with the UNITY4.4 program[5]. The 3D pharmacophore models were constructed based on the known important interactions between the amino acid in the binding pocket and NNRTIs. After filtering the database, the selected compounds were applied to docking. The hits from docking were selected and classified by scaffold diversity using ClassPharmer program[6]. Finally, the selected compounds from the classification were tested for HIV-1 RT inhibition.

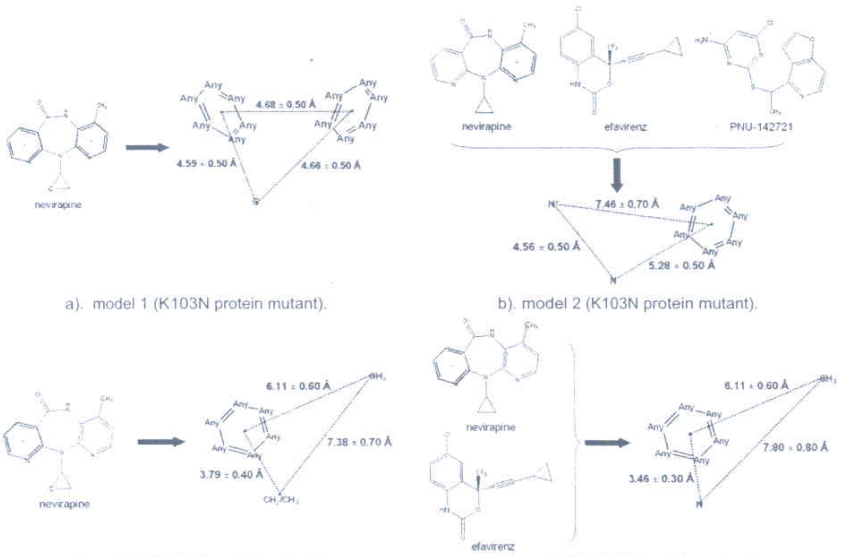
Results and discussion

Three docking methods, FlexX, GOLD and Surflex, have been used to dock nevirapine into K103N binding pocket (pdb code 1fkp) and Y181C binding pocket (pdb code 1jlb). The rmsd of the docked pose from the X-ray pose of nevirapine by using three docking methods are shown in Table 1. GOLD was the most accurate method to reproduce the X-ray bound conformation of nevirapine to both mutants, therefore the GOLD method was selected for further docking steps.

Table 1. The rmsd (Å) of the docked pose from the X-ray pose of nevirapine by using FlexX, GOLD and Surflex methods.

Method	rmsd (Å)	
	K103N (1fkp)	Y181C (1jlb)
FlexX	5.07	13.38
GOLD	0.41	0.42
Surflex	4.91	0.98

Some NNRTIS have been docked to the K103N binding pocket and the Y181C binding pocket. For the K103N protein mutant, two pharmacophore models have been derived from the docked conformation of nevirapine, efavirenz and PNU-142721. For the Y181C protein mutant, two pharmacophore models have been derived from the docked conformation of nevirapine and efavirenz. All pharmacophore models were used as constraints for searching the 'in-house' database and were shown in Fig.1.



a). model 1 (K103N protein mutant). b). model 2 (K103N protein mutant).

c). model 1 (Y181C protein mutant). d). model 2 (Y181C protein mutant).

Fig. 1. The features and constraints of all pharmacophore models.

From 4 pharmacophore models, there are 20,200 compounds and 19,761 compounds matching pharmacophore for the K103N and Y181C protein mutants. These compounds were selected for further molecular docking with GOLD. Finally, 6k virtual hits were selected and classified by scaffold diversity using ClassPharmer. 41 compounds were purchased and tested.

Acknowledgment

This work was supported by the Royal Golden Jubilee Ph.D. Program (3.C.KU/45/B.1) and the French Government.

References

[1] Dyatkin AB, Brickwood JR, Proudfoot JR, A dipyrido[2,3-b:3', 2'-f]azepine analog of the HIV-1 reversetranscriptase inhibitor nevirapine, Bioorg. Med. Chem. Lett. 1998, 8: 2169-2172.

[2] Rarey M, Kramer B, Lengauer T, Klebe G, A fast flexible docking method using an incremental construction algorithm, J. Mol. Biol. 1996, 261: 470-489.

[3] Jones G, Wilett P, Glen RC, Leach AR, Taylor R, Development and validation of a genetic algorithm for flexible docking, J. Mol. Biol. 1997, 267: 727-748.

[4] Jain AN, Surflex: Fully automatic flexible molecular docking using a molecular similarity-based search engine, J. Med. Chem. 2003, 46: 499-511.

[5] UNITY 4.4; Tripos Inc., St. Louis, MO, USA.

[6] ClassPharmer; Bioreason Inc., Sante Fe, NM, USA.

Publication II

Design of nevirapine derivatives insensitive to the K103N and Y181C HIV-1 reverse transcriptase mutants

Saparpakorn P, Hannongbua S, Rognan D.

SAR QSAR Environ Res. 2006, 17, 183-94.

SAR and QSAR in Environmental Research
Vol. 17, No. 2, April 2006, 183–194

Taylor & Francis
Taylor & Francis Group

Design of nevirapine derivatives insensitive to the K103N and Y181C HIV-1 reverse transcriptase mutants†

P. SAPARPAKORN‡, S. HANNONGBUA*‡ and D. ROGNAN§

‡Department of Chemistry, Faculty of Science, Kasetsart University, Bangkok, 10900, Thailand
§Laboratoire de Pharmacochimie de la Communication Cellulaire, UMR CNRS 7081, F-67401 Illkirch, France

(*Received 1 November 2005; in final form 10 January 2006*)

Nevirapine (Viramune®) belongs to the first generation of non-nucleoside reverse transcriptase inhibitors (NNRTIs). Its efficiency is limited by drug resistant mutations, such as K103N and Y181C, so, the aim of this work was to design novel nevirapine analogues insensitive to the K103N and Y181C HIV-1 RT. 360 Nevirapine derivatives were designed using a combinatorial library design approach and these compounds were docked into the binding pocket of mutant HIV-1 RT enzyme structures, using the GOLD program. 124 Compounds having a GoldScore higher than that of nevirapine (55.00 and 52.00 for K103N and Y181C mutants, respectively) were first retrieved and submitted to a topological analysis with the SILVER program. Consequently, 31 compounds presenting a significant percentage of the surfaces buried upon binding (>80%) and exhibiting hydrogen bonds to either N103 or C181 residues of the HIV-RT were selected. To ensure that these compounds had hydrogen bonding interaction to either N103 or C181 residues, their interaction energies were estimated by quantum chemical calculations (QCCs). Finally, QCCs represent an alternative method for performing post docking procedure.

Keywords: NNRTIs; HIV-1 RT; K103N; Y181C; Combinatorial library design; Quantum chemical calculations

1. Introduction

Acquired immunodeficiency syndrome (AIDS) is caused by the human immunodeficiency virus (HIV). The reverse transcriptase (RT) of human immunodeficiency virus type-1 (HIV-1) is an essential enzyme converting the single-stranded viral RNA genome into double-stranded proviral DNA prior to its integration into the host genomic DNA. HIV-1 RT is an asymmetric heterodimer consisting of a 560 residue chains (p66 subunit) and a 440 residue chains (p51 subunit). Nucleoside inhibitors

*Corresponding author. Email: fscisph@ku.ac.th
†Presented at CMTPI 2005: Computational Methods in Toxicology and Pharmacology Integrating Internet resources (Shanghai, China, October 29–November 1 2005).

SAR and QSAR in Environmental Research
ISSN 1062-936X print: ISSN 1029-046X online © 2006 Taylor & Francis
DOI: 10.1080/10659360600636147

(NRTIs) and non-nucleoside inhibitors (NNRTIs) are two types of HIV-1 RT inhibitors that are widely used for the treatment of AIDS. Nucleoside analog inhibitors, such as AZT, ddI and ddC, are competitive inhibitors causing termination of the growing DNA chain. Although these nucleoside inhibitors are selective for HIV-1 RT, they are not highly specific and make cellular polymerases to translate into toxic effects [1 2]. Non-nucleoside analog inhibitors, such as nevirapine, TIBO and efavirenz, are non-competitive inhibitors binding in a hydrophobic pocket that is about 10 Å from the catalytic active site in the p66 subunit. Because of the use at low concentration and the high specificity, non-competitive inhibitors are interesting for developing novel potent inhibitors. However, the efficiency of these inhibitors is limited by drug resistant mutations, such as K103N, Y181C and V106A [3]. The most important mutations of the HIV-1 RT are found to be K103N and Y181C. In the pocket of wild-type HIV-1 RT, the necessary interactions with NNRTIs show stacking interactions with the aromatic residues such as Y181, Y188, W229 and Y318, electrostatic interactions with K101, K103 and E138, van der waals interactions with L100, V106, Y181, G190, W229, L234 and Y318 and hydrogen bond interaction with the main-chain peptide bonds [4]. Recently, quantum chemical calculations (QCCs) were used to calculate the interaction energies and binding energies of NNRTIs in HIV-1 RT binding pocket [5 7]. These revealed the effects of surrounding amino acid residues to NNRTIs. The results also indicated that the mutation of HIV-1 RT reduces the binding of NNRTIs to some extent. Therefore, the mutations can affect NNRTIs binding by losing the important contacts between the binding pocket and NNRTIs, changing the size of the binding pocket and interfering the NNRTIs entry into the binding pocket. From the previous work, Hsiou *et al.* [8] indicated that the inhibitors active against K103N mutation would be expected to have favorable interactions with the mutated asparagine side chain. This suggested that design of novel NNRTIs active against protein mutants should make favorable interactions with that mutant residue.

Nevirapine (Viramune®) was the first generation of NNRTI that has been approved by the FDA for the treatment of HIV-1 infection. The binding affinity of nevirapine against wild type and mutant types RT are shown in table 1 [9]. Unfortunately, nevirapine showed a lack of affinity upon two important mutations, the K103N and Y181C mutations. From the available X-ray structures of K103N and Y181C HIV-1 RT mutants in complex with nevirapine, these revealed that the lack of affinity was caused by the lost of important interaction between nevirapine and mutant residues. The orientation of nevirapine in K103N and Y181C HIV-1 RT binding pockets are shown in figure 1. These structures were taken from pdb code 1fkp [10] and 1jlb [11], respectively. Therefore, the aim of this work was to find novel nevirapine analogues insensitive to the K103N and Y181C mutations. Nevirapine derivatives were designed using a combinatorial library design approach [12]. In this study, some docking methods have been performed with the nevirapine-RT complex to validate and select

Table 1. Binding affinities of nevirapine against wild type and mutant types RT.

	IC_{50} (μM)				
Compound	*Wild type*	*Y181C*	*P236L*	*K103N*	*L100I*
Nevirapine	0.06	3.2	0.18	1.3	0.17

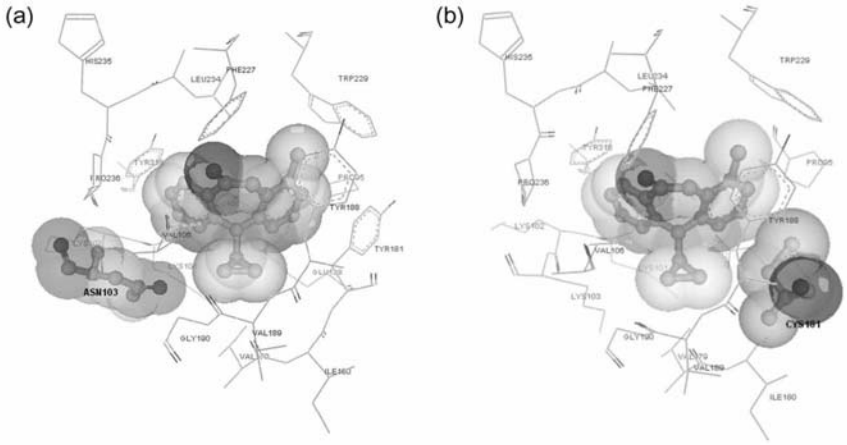


Figure 1. Orientation and molecular surfaces of nevirapine in (a) the K103N binding pocket and (b) the Y181C binding pocket.

the best possible strategy. After that, the nevirapine derivatives were docked into the binding pocket of K103N and Y181C HIV-1 RT using the selected docking program. The hits from docking were further performed for post-docking procedure by topologically analyzing with the SILVER program. The hits presented a significant percentage of their surface buried upon binding (> 80%) and exhibiting H-bonds to either N103 or C181 residues of the HIV-RT were finally selected. Furthermore, QCC was performed to calculate the interaction energies between nevirapine derivatives and mutant residues. The results can be useful to provide particular interaction information between nevirapine derivatives and N103 or C181.

2. Computational methods

2.1 *Preparation of the mutant HIV-1 RT and nevirapine coordinates*

The X-ray structures of K103N and Y181C HIV-1 RT mutants in complex with nevirapine were obtained from the Protein Data Bank (pdb code 1fkp and 1jlb). Nevirapine was removed from each structure. All-hydrogen atoms were added to the enzymes using standard SYBYL [13] geometries. For nevirapine of each complex structure, hydrogen atoms were added, and then minimized by using the default parameter of Tripos force field [14] in SYBYL 6.9 program.

2.2 *Preparation of nevirapine derivatives*

From the nevirapine-RT complexes of K103N and Y181C protein mutants, some nevirapine derivatives were designed to find the compounds having the interaction with N103 and C181. These derivatives were constructed using a combinatorial library design approach. Starting coordinates was taken from the minimized

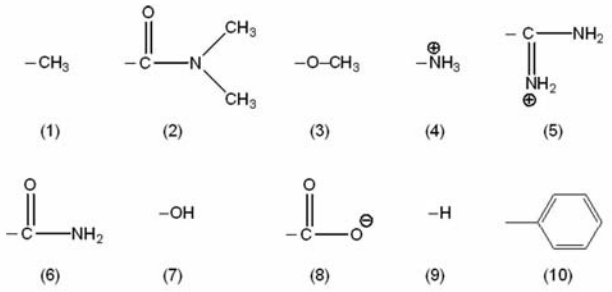


Figure 2. Molecular fragments, used in the design of nevirapine derivatives.

nevirapine structure. Substituents based on ten fragments were attached or replaced to the common structure by using SYBYL6.9 program. These fragments consist of the necessary functional groups for drug-like compound as shown in figure 2. These functional groups include diversity types of interaction, i.e. hydrophobic, hydrophilic, H-bond acceptor, and H-bond donor, aliphatic and aromatic. To design the new compounds having the interaction with the side-chain of N103, some substituent groups (R_1, R_2 and R_3) were designed by substituting or adding to nevirapine. In case of the designed compounds having the interaction with C181, R_3 and R_4 substituent groups were added or replaced to nevirapine. All designed compounds are shown in table 2.

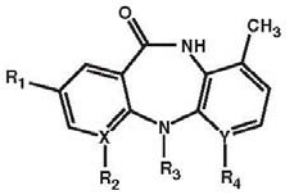
2.3 *Docking of designed nevirapine derivatives*

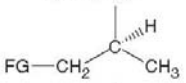
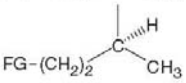
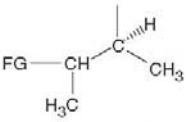
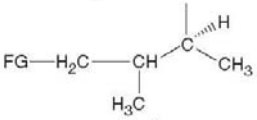
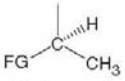
For selecting the suitable docking method of this study, three docking tools, FlexX [15], GOLD [16] and Surflex [17], were applied to the nevirapine-RT complexes of K103N and Y181C mutations with their default parameters. Nevirapine was docked back to its binding pocket of both mutants with three docking tools. The root-mean-square deviation (rmsd) of the top-scored pose from the X-ray pose was used to determine the validation of the docking method. The suitable docking method was used to investigate the orientation of designed nevirapine derivatives in K103N and Y181C binding pockets. This method was used to dock them into K103N and Y181C HIV-1 RT. The docked compounds having score higher than that of nevirapine were selected and supposed to have tighter binding than nevirapine. Then, SILVER program [18] was used to perform post-process docking results for large numbers of ligands. The hits from docking can be prioritized by topological analysis.

2.4 *Quantum chemical calculation*

To ensure that the selected compounds have attractive interactions with the mutated residues, interaction energies between selected compounds and N103 or C181 were calculated by using QCCs. These were performed at B3LYP/6-31G(d) and MP2/6-31G(d) levels of theory by using Gaussian03 program [19] and the results were compared with Nevirapine's results. The interaction energy (INT) of the selected

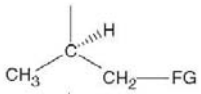
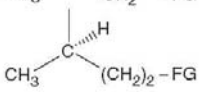
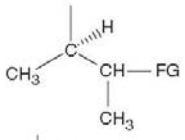
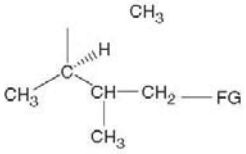
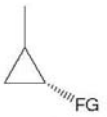
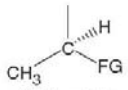
Table 2. Structures of designed nevirapine derivatives, using a combinatorial library design approach.



No.	*X*	*Y*	R_1	R_2	R_3	R_4
Str 1–10	C	N	H	O–CH_2–FG	Cyclopropyl	–
Str 11–20	C	N	H	O–$(CH_2)_2$–FG	Cyclopropyl	–
Str 21–30	C	N	H	O–$(CH_2)_3$–FG	Cyclopropyl	–
Str 31–40	C	N	H	FG	Cyclopropyl	–
Str 41–50	N	N	H	–		–
Str 51–60	N	N	H	–		–
Str 61–70	N	N	H	–		–
Str 71–80	N	N	H	–		–
Str 81–90	N	N	H	–		–
Str 91–100	N	N	H	–		–
Str 101–110	N	N	H	–	CH_2–FG	–
Str 111–120	N	N	H	–	$(CH_2)_2$–FG	–
Str 121–130	N	N	H	–		–
Str 131–140	N	C	H	–	Cyclopropyl	O–CH_2–FG
Str 141–150	N	C	H	–	Cyclopropyl	O–$(CH_2)_2$–FG
Str 151–160	N	C	H	–	Cyclopropyl	O–$(CH_2)_3$–FG

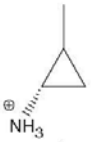
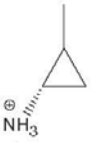
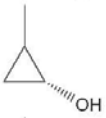
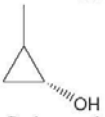
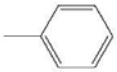
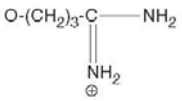
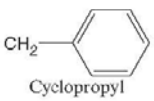
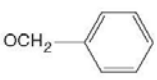
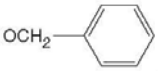
(*continued*)

Table 2. Continued.

No.	*X*	*Y*	R_1	R_2	R_3	R_4
Str 161–170	N	C	H	–	Cyclopropyl	FG
Str 171–180	N	N	H	–		–
Str 181–190	N	N	H	–		–
Str 190–200	N	N	H	–		–
Str 201–210	N	N	H	–		–
Str 211–220	N	N	H	–		–
Str 221–230	N	N	H	–		–
Str 231–240	N	N	H	–		–
Str 241–250	N	N	FG	–	Cyclopropyl	–
Str 251	N	N	CH_2–NH_3^+	–	Cyclopropyl	–
Str 252	N	N	CH_2–OH	–	Cyclopropyl	–
Str 253	N	N	CH_2–COO^-	–	Cyclopropyl	–
Str 254–263	C	N	H	O–CH_2–FG		–
Str 264–273	C	N	H	O–$(CH_2)_2$–FG		–

(*continued*)

Table 2. Continued.

No.	*X*	*Y*	R_1	R_2	R_3	R_4
Str 274–283	C	N	H	O–$(CH_2)_3$–FG		–
Str 284–293	C	N	H	FG		–
Str 294–303	C	N	H	O–CH_2–FG		–
Str 304–313	C	N	H	O–$(CH_2)_2$–FG		–
Str 314–323	C	N	H	O–$(CH_2)_3$–FG		–
Str 324–333	C	N	H	FG		–
Str 334–338	C	N	$\overset{\oplus}{N}H_3$, OH, $COO^{\ominus}$ $CH_2\overset{\oplus}{N}H_3$, CH_2OH		Cyclopropyl	–
Str 339–343	C	N	$\overset{\oplus}{N}H_3$, OH, $COO^{\ominus}$ $CH_2\overset{\oplus}{N}H_3$, CH_2OH	O–$(CH_2)_3{}^{-}COO^{-}$	Cyclopropyl	–
Str 344–348	C	N	$\overset{\oplus}{N}H_3$, OH, $COO^{\ominus}$ $CH_2\overset{\oplus}{N}H_3$, CH_2OH		Cyclopropyl	–
Str 349–353	N	N	$\overset{\oplus}{N}H_3$, OH, $COO^{\ominus}$ $CH_2\overset{\oplus}{N}H_3$, CH_2OH	–		–
Str 354–358	C	N	$\overset{\oplus}{N}H_3$, OH, $COO^{\ominus}$ $CH_2\overset{\oplus}{N}H_3$, CH_2OH		Cyclopropyl	–
Str 359–363	N	C	$\overset{\oplus}{N}H_3$, OH, $COO^{\ominus}$ $CH_2\overset{\oplus}{N}H_3$, CH_2OH	–	Cyclopropyl	

compounds with N103 or C181 is defined as follows:

$$INT = E_{(ligand+N103\ or\ C181)} - [E_{ligand} + E_{(N103\ or\ C181)}] \quad (1)$$

Where $E_{(ligand+N103\ or\ C181)}$ is the energy of the complex structure of ligand and N103 or C181. E_{ligand} and $E_{(N103\ or\ C181)}$ are the energies of ligand and N103 or C181, respectively. All energies were obtained from the single point calculation at B3LYP/6-31G(d) and MP2/6-31G(d) levels of theory. The structures of ligands were taken from their docked conformation in the binding pocket. The geometries of N103 and C181 were taken from K103N and Y181C mutants, respectively, which are starting geometries for docking.

3. Results and discussion

3.1 *Validation of the docking method*

Three docking methods, FlexX, GOLD and Surflex, were used to dock nevirapine into K103N and Y181C HIV-1 RT enzymes. The rmsd of the docked pose from the X-ray pose of nevirapine by using three docking methods are described as follows. The rmsd values of nevirapine by using FlexX are 5.07 Å and 13.38 Å for K103N and Y181C mutations, respectively. The results showed that the FlexX method failed to reproduce the X-ray bound conformation of nevirapine for both mutations. Surflex successfully reproduced the X-ray bound conformation for the Y181C mutation with a 0.98 Å rmsd, however, it failed to reproduce the X-ray pose for the K103N mutation (4.91 Å rmsd). The reason for FlexX and Surflex is due to the fact that incremental construction docking algorithm and empirical scoring function were used in both methods. Unlike FlexX and Surflex, the GOLD method uses genetic algorithm for docking and force field scoring function. Therefore, it revealed a good ability to reproduce the X-ray bound conformation with rmsd less than 1.0 Å for both mutants, producing rmsd of nevirapine for K103N and Y181C mutants as 0.41 and 0.42, respectively. Consequently, GOLD method was found to be a reliable method to reproduce the X-ray bound conformation of nevirapine to K103N and Y181C mutants with GoldScore 57.16 and 54.23, respectively. These orientations of X-ray and docked nevirapine in both mutant enzymes are shown in figure 3. Therefore, the GOLD method was used to dock the designed nevirapine derivatives into both mutant enzymes.

3.2 *Docking of nevirapine derivatives and post docking analysis*

After removing the duplicated compounds in all libraries, 363 nevirapine derivatives were docked into the K103N and the Y181C mutant enzyme structures. The orientation and GoldScore of each docked conformation were retrieved. The compounds having GoldScore higher than that of nevirapine (55.00 and 52.00 for K103N and Y181C mutants, respectively) were selected for further study. It was found that 124 compounds were retrieved to perform post-docking with SILVER program. In this study, docked compounds which have H-bonding with N103 or C181 for K103N and Y181C mutants, respectively and percentage of their surface buried upon binding higher than 80% were selected. Finally, there are 25 compounds (cpds 1 25 as shown in table 3) having H-bonding with N103 and percentage of their surface buried upon binding higher than

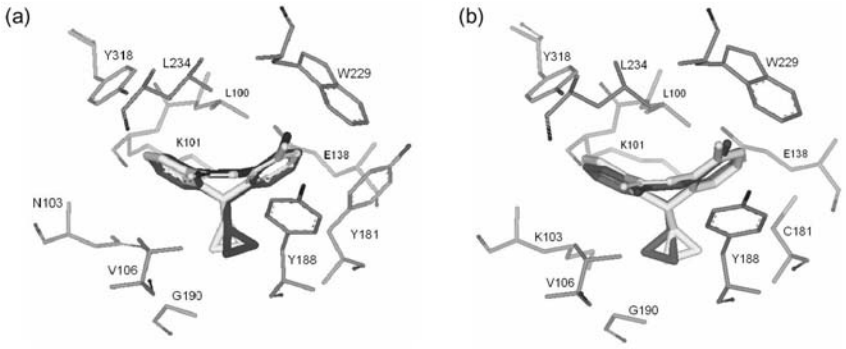


Figure 3. Orientation of nevirapine in both mutant enzymes. (a) Geometries obtained from X-ray pose (grey) and docking (dark grey) in the K103N binding pocket. (b) Geometries obtained from X-ray pose (grey) and docking (dark grey) in the Y181C binding pocket.

Table 3. Interaction energies between N103 and 25 selected compounds.

	Interaction energy (kcal mol^{-1})	
Compound	*B3LYP/6-31G(d)*	*MP2/6-31G(d)*
Nevirapine (pdb code 1fk9)	0.3	−0.7
–H-bonding with backbone N103		
Cpd 1 (Str 251)	7.3	4.5
Cpd 2 (Str 357)	−3.1	−5.6
Cpd 3 (Str 352)	6.1	2.6
–H-bonding with sidechain N103		
Cpd 4 (Str 16)	0.2	−2.4
Cpd 5 (Str 18)	−4.6	−7.6
Cpd 6 (Str 343)	−16.6	−20.6
Cpd 7 (Str 340)	−9.9	−13.5
Cpd 8 (Str 35)	−17.1	−18.8
Cpd 9 (Str 36)	8.0	6.6
Cpd 10 (Str 45)	−18.1	−19.0
Cpd 11 (Str 55)	−12.3	−13.8
Cpd 12 (Str 94)	−25.5	−25.8
Cpd 13 (Str 175)	−1.6	−3.2
Cpd 14 (Str 257)	−30.3	−31.6
Cpd 15 (Str 258)	−23.2	−25.3
Cpd 16 (Str 259)	−19.1	−21.6
Cpd 17 (Str 260)	−24.6	−26.1
Cpd 18 (Str 270)	−17.8	−20.0
Cpd 19 (Str 286)	−24.6	−26.3
Cpd 20 (Str 288)	−33.7	−35.0
Cpd 21 (Str 289)	−14.2	−14.9
Cpd 22 (Str 290)	−24.6	−25.5
Cpd 23 (Str 291)	−15.0	−16.7
Cpd 24 (Str 292)	−26.7	−27.6
Cpd 25 (Str 293)	−19.1	−21.4

Table 4. Interaction energies between C181 and 6 selected compounds.

	Interaction energy (kcal mol^{-1})	
Compound	*B3LYP/6-31G**	*MP2/6-31G**
Nevirapine (pdb code 1jlb)	1.7	−2.3
−H-bonding with backbone Cys181		
Cpd 26 (Str 227)	−1.8	−5.7
Cpd 27 (Str 235)	−3.3	−6.9
Cpd 28 (Str 297)	−0.8	−5.6
Cpd 29 (Str 322)	1.7	−4.2
Cpd 30 (Str 326)	−1.2	−5.7
Cpd 31 (Str 330)	−1.7	−4.6

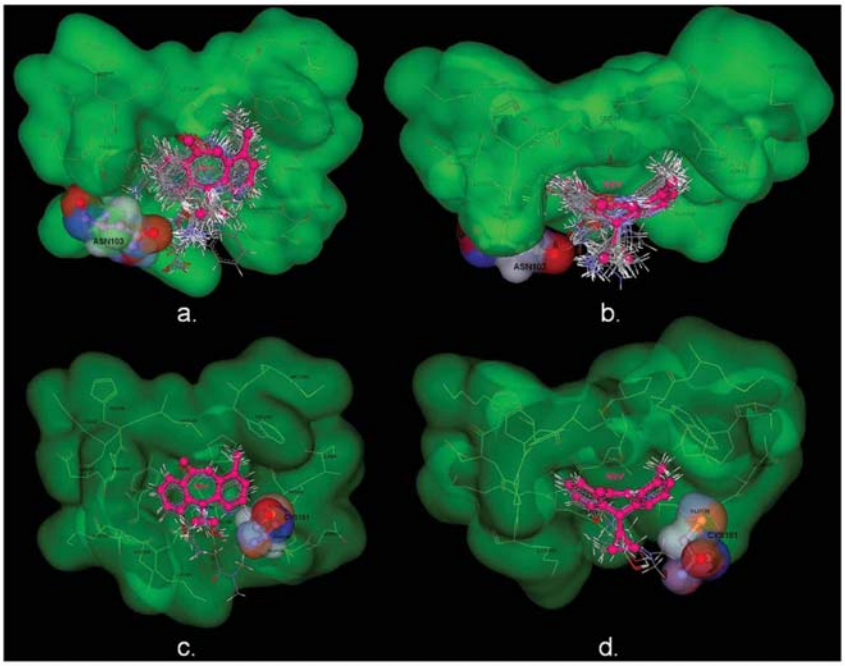


Figure 4. Orientation of selected nevirapine derivatives in the K103N binding pocket (a top view and b side view) and in the Y181C binding pocket (c top view and d side view).

80% and 3 compounds have H-bonding with the backbone atom of N103 and the others have H-bonding with sidechain atom of N103. In Y181C mutant, it was found that 6 compounds (cpds 26–31 as shown in table 4) have H-bonding with backbone atom of C181. The orientations of all 31 compounds are laid to be similarly butterfly-like shape as nevirapine in its binding pocket, as shown in figure 4. It can be noted that these compounds still kept the specific interaction of nevirapine with the surrounding amino acids, moreover, these hits can particularly interact with N103 or C181.

3.3 *Calculation of interaction energy*

For confirming the selected compounds having H-bonding interaction with N103 or C181, the calculated interaction energies were performed by using quantum chemical calculations. For K103N mutant, nevirapine (pdb code 1fk9) and each of 25 selected compounds were used to calculate interaction energies with N103 at B3LYP/6-31G(d) and MP2/6-31G(d) levels of theory and the results are shown in table 3. Interaction energies showed that nevirapine slightly interacted with N103 in both methods. The interaction energies are about 0.3 and $-0.7\,\mathrm{kcal\,mol^{-1}}$ at B3LYP/6-31G(d) and MP2/6-31G(d) levels of theory, respectively. Two of the selected compounds having H-bonding with backbone N103 (Cpd1 and Cpd3) have repulsive interactions with N103 in both methods. These repulsive interactions were caused by steric interaction of R_1 substituent (methyl ammonium group). Instead of the formation of attractive interaction with N103, the methyl ammonium group of Cpd1 and Cpd 3 has formed H-bonding interaction with P236. Attractive interaction energies of Cpd2 (-3.1 and $-5.6\,\mathrm{kcal\,mol^{-1}}$) are contributed by H-bonding interaction between nitrogen atom of backbone N103 and hydrogen atom of ammonium group in R_1 substituent with the length of 2.49 Å. In the case of other selected compounds having H-bonding with the sidechain of N103, except Cpd9, these compounds show tighter interaction energies than that of nevirapine ($0.3\,\mathrm{kcal\,mol^{-1}}$ by B3LYP/6-31G(d) calculations and $-0.7\,\mathrm{kcal\,mol^{-1}}$ by MP2/6-31G(d) calculations. The repulsive interaction of Cpd9 are due to steric hindrance between amide group at R_2 substituent and oxygen atom of N103 sidechain. In addition, the oxygen atom of amide group at R_2 substituent has attractive interaction with L100 and K101, therefore, its GoldScore is found to be higher than GoldScore of nevirapine.

In case of Y181C mutant, the interaction energy between nevirapine and C181 was calculated by both methods (table 4). The interaction energies are 1.7 and $-2.3\,\mathrm{kcal\,mol^{-1}}$ for B3LYP/6-31G(d) and MP2/6-31G(d) calculations, respectively. There are 6 selected nevirapine derivatives that interact with C181 and the results are shown in table 4. All 6 compounds showed stronger interaction energies than that of nevirapine and it was found that H-bonding with nitrogen backbone atom of C181 was formed to these nevirapine derivatives. H-bonding interactions were occurred from substituted hydroxyl group of cyclopropyl (Cpd 26, Cpds 28 30) and from amino group of R_3 substituent (Cpd27).

By using B3LYP/6-31G(d) and MP2/6-31G(d) levels of theory, the interaction energies between all hits and N103 or C181 are agreed well with the results from M. Kuno *et al.* [6] as density functional method can not handle the H-pi interaction but MP2 method has taken into account this interaction. Therefore, in this study, more attractive interaction of nevirapine derivatives were generally found by using MP2/6-31G(d) level of theory.

4. Conclusion

In this study, the information of the X-ray structures of K103N and Y181C HIV-1 RT mutants in complex with nevirapine were used to design novel nevirapine analogues insensitive to the K103N and Y181C mutations. Combinatorial library design approach was applied for designing the nevirapine derivatives supposed to have H-bonding

interaction to either N103 or C181 residues. The designed nevirapine derivatives were docked to both mutant enzymes and then the compounds having dock score higher than nevirapine's dock score were selected for further study. A program for post-processing of docking results, SILVER program, was used to select the compounds having H-bonding either N103 or C181 and percentage of their surface buried upon binding higher than 80%. Finally, quantum chemical calculations were applied to calculate the interaction energies between the hits and N103 or C181. There are 3 of 31 compounds having repulsive interaction with N103 because of the steric hindrance with N103. These indicated that QCC can be useful as an alternative method for post processing of docking results. Not only the topological analysis of the hits, but also the interaction energy calculation of the hits is a challenge approach to prioritize the hits before performing the experimental testing.

Acknowledgements

This investigation was supported by the Thailand Research Fund (BRG4780007). P.S. is grateful to the Royal Golden Jubilee Ph.D. program (3.C.KU/45/B.1) and the French Embassy in Thailand for scholarships. Laboratory for Computational and Applied Chemistry (LCAC), the Postgraduate program in education and research on Petroleum and Petrochemical Technology (ADB-MUA), HPCC/NECTEC and Bioinformatics Research Group, Laboratoire de Pharmacochimie de la Communication Cellulaire are also gratefully acknowledged for research facilities and computing resources.

References

[1] C. Tantillo, J. Ding, A. Jacobo-Molina, R.G. Nanni, P.L. Boyer, S.H. Hughes, R. Pauwels, K. Andries, P.A. Jannsen, E. Arnold. *J. Mol. Biol.*, **243**, 369 (1994).
[2] E. De Clercq. *Clin. Microbiol. Rev.*, **8**, 200 (1995).
[3] L. Menendez-Arias. *Trends Pharmacol. Sci.*, **23**, 381 (2002).
[4] E. De Clercq. *Nat. Rev. Drug. Discov.*, **1**, 13 (2002).
[5] S. Saen-oon, M. Kuno, S. Hannongbua. *PROTEINS*, **61**, 859 (2005).
[6] M. Kuno, S. Hannongbua, K. Morokuma. *Chem. Phys. Lett.*, **380**, 456 (2003).
[7] P. Nunrium, M. Kuno, S. Saen-oon, S. Hannongbua. *Chem. Phys. Lett.*, **405**, 198 (2005).
[8] Y. Hsiou, J. Ding, K. Das, A.D. Clark Jr, P.L. Boyer, P. Lewi, P.A.J. Janssen, J. Kleim, M. Rosner, S.H. Hughes, E. Arnold. *J. Mol. Biol.*, **309**, 437 (2001).
[9] A.B. Dyatkin, J.R. Brickwood, J.R. Proudfoot. *Bioorg. Med. Chem. Lett.*, **8**, 2169 (1998).
[10] J. Ren, J. Milton, K.L. Weaver, S.A. Short, D.I. Stuart, D.K. Stammers. *Structure*, **8**, 1089 (2000).
[11] J. Ren, C. Nichols, L. Bird, P. Chamberlain, K. Weaver, S. Short, D.I. Stuart, D.K. Stammers. *J. Mol. Biol.*, **312**, 795 (2001).
[12] M. Krier, J.X. de Araujo-Junior, M. Schmitt, J. Duranton, H. Justiano-Barasan, C. Lugnier, J.J. Bourguignon, D. Rognan. *J. Med. Chem.*, **48**, 3816 (2005).
[13] M. Clark, R.D. Cramer III, N. Van Opdenbosch. *J. Comp. Chem.*, **10**, 982 (1989).
[14] SYBYL 6.9; TRIPOS, Assoc., Inc.: St. Louis, MO.
[15] M. Rarey, B. Kramer, T. Lengauer, G. Klebe. *J. Mol. Biol.*, **261**, 470 (1996).
[16] G. Jones, P. Wilett, R.C. Glen, A.R. Leach, R. Taylor. *J. Mol. Biol.*, **267**, 727 (1997).
[17] A.N. Jain. *J. Med. Chem.*, **46**, 499 (2003).
[18] Available online at: www.ccdc.cam.ac.uk/products/life_sciences/gold/index.php#silver
[19] M.J. Frisch *et al.*, Gaussian03, Gaussian, Inc, Pittsburgh, PA (2003).

CURRICULUM VITAE

PERSONAL DATA

Name:	Miss Patchreenart SAPARPAKORN
Date of Birth:	January 15, 1980
Place of Birth:	Bangkok, THAILAND
Nationality	Thai
Religion	Buddhism

EDUCATION

:YEAR	INSTITUTION	DEGREE
1997-2000	Kasetsart University	B.Sc. (Chemistry)
2001-2006	Kasetsart University	Ph.D. Physical Chemistry student

SCHOLARSHIP

:YEAR	GRANT SOURCE
2000-2001	The Development and Promotion of Science and Technology talents project of Thailand (DPST)
2002-Present	The Royal Golden Jubilee (RGJ) Ph.D. program from The Thailand Research Fund (TRF), THAILAND
2003-2004	The French embassy in Thailand
2006	The financial support for attending the XIIth International Congress of Quantum Chemistry (XII-ICQC2006) Kyoto TERRSA, kyoto, JAPAN

RESEARCH VISITING

:YEAR	INSTITUTION
2003-2004	Laboratoire de Pharmacochimie de la Communication Cellulaire, UMR CNRS 7081, Illkirch, FRANCE

LIST OF PUBLICATIONS

1. Lee VS, Wittayanarakul K, Remsungnen T, Parasuk V, Sompornpisut P, Chantratita W, Sangma C, Vannarat S, Srichaikul P, Hannongbua S, Saparpakorn P, Treesuwan W, Aruksakulwong O, Promsri S, Chuakheaw D, Hannongbua S. Structure and Dynamics of SARS Coronavirus Proteinase: The Primary Key to the Designing and Screening for Anti-SARS Drugs, *ScienceAsia*, **2003**, *29*, 181-188.

2. Chanprapaph S, Saparpakorn P, Sangma C, Niyomrattanakit P, Hannongbua S, Angsuthanasombat C, Katzenmeier G. Competitive Inhibition of the Dengue Virus NS3 Serine Protease by Synthetic Peptides Representing Polyprotein Cleavage Sites. *Biochem. Biophys. Res. Commun.*, **2005**, *330*, 1237-1246.

3. Saparpakorn P, Hannongbua S, Rognan D. Design of nevirapine derivatives insensitive to the K103N and Y181C HIV-1 reverse transcriptase mutants. *SAR QSAR Environ Res.*, **2006**, *17*, 183-94.

4. Pungpo P, Saparpakorn P, Wolschann P, Hannongbua S. Computer-aided molecular design of highly potent HIV-1 RT inhibitors: 3D QSAR and molecular docking studies of efavirenz derivatives. *SAR QSAR Environ Res.* **2006**, *17*, 353-70.