THESIS APPROVAL

GRADUATE SCHOOL, KASETSART UNIVERSITY

Master of Science (Chemistry)

DEGREE

Chemistry	Chemistry
FIELD	**DEPARTMENT**

TITLE: Comparative Molecular Field Analysis and Quantum Calculation Study on Anti HIV-1 RT Diarylaniline Derivatives

NAME: Mr. Nuttapong Ithiapa

THIS THESIS HAS BEEN ACCEPTED BY

THESIS ADVISOR

(Mr. Songwut Suramitr, Ph.D.)

THESIS CO-ADVISOR

(Associate Professor Supa Hannongbua, Dr.rer.nat.)

DEPARTMENT HEAD

(Associate Professor Supa Hannongbua, Dr.rer.nat.)

APPROVED BY THE GRADUATE SCHOOL ON _____

DEAN

(Associate Professor Gunjana Theeragool, D.Agr.)

THESIS

COMPARATIVE MOLECULAR FIELD ANALYSIS AND QUANTUM CALCULATION STUDY ON ANTI HIV-1 RT DIARYLANILINE DERIVATIVE

NUTTAPONG ITHIAPA

A Thesis Submitted in Partial Fulfillment of
the Requirements for the Degree of
Master of Science (Chemistry)
Graduate School, Kasetsart University
2012

Nuttapong Ithiapa 2012: Comparative Molecular Field Analysis and Quantum Calculation Study on Anti HIV-1 RT Diarylaniline Derivatives. Master of Science (Chemistry), Major Field: Chemistry, Department of Chemistry. Thesis Advisor: Mr. Songwut Suramitr, Ph.D. 78 pages.

The goal of the research is the demonstrating the cause of the diarylaniline derivative drugs resistance using the CoMFA and quantum chemical calculations. Over all obtained results can be used to guide the new designed potent anti HIV-1 Reverse Transcriptase inhibitors for enzyme. In this study, the relationship between structural properties of 25 diarylaniline derivatives and their 50% effective concentrations (EC_{50}) to HIV-1 Reverse Transcriptase (RT) using a comparative molecular field analysis (CoMFA) were constructed. The best predictive CoMFA model gives a very good statistical result with r_{cv}^2 = 0.823, r_{nv}^2 = 0.924, S_{press} = 0.422, *SE* = 0.241, *F* = 65.055, steric contribution = 28.1% and electrostatic contribution = 71.9%. Consequently, the obtained CoMFA contour maps merging with the wild type HIV-1 RT binding site can give the informative details for understanding the structural requirements of inhibitors and can guide the new design of diarylanilline inhibitors. Deeply in molecular details, an understanding of particular interaction energy between antiHIV-1 inhibitors and surrounding residues in the binding pocket was performed by using B3LYP, M062X and MP2/6-31G(d,p) calculations. These calculations technical demonstrated the rationality of our hypothesis about main interaction between diarylaniline derivative and HIV-1 Reverse Transcriptase. The obtained results clearly demonstrate that compound 24 have more interaction and more efficiency than compound 1.

_____ / / /

Student's signature Thesis Advisor's signature

ACKNOWLEDGEMENT

I sincerely thank my advisor, Dr. Songwut Suramitr for his advice, suggestion, tremendous support and meticulous attention throughout the duration of my graduate study and research. I am particularly grateful to my advisory committee Associate Professor Dr. Supa Hannongbua, the representative of Graduate School, for their worthy suggestion and constructive criticism.

I owe my most sincere gratitude to Dr. Patchareenart Saparpakornfor her detailed, valuable advice, kind help and constructive comments, and for her important collaboration throughout this work.

I am deeply grateful to Dr. Phornphimon Maitarad, who gave me a great an opportunity to work and her suggestion and for her important collaboration throughout this work.

Financial supports were from the Thailand Research Fund (TRF), Ministry of Education through the "National Research University Project of Thailand (NRU)" and the "National Center of Excellence for Petroleum, Petrochemical and Advanced Materials (NCE-PPAM). My colleagues at Laboratory for Computational and Applied Chemistry (LCAC) are sincere thanked for their providing helpful assistance and sharing useful ideas and I would also like to thank all of staffs at Department of Chemistry, Faculty of Science, Kasetsart University for research facilities.

Finally, I am especially appreciated my parents, friends and colleagues at LCAC for their encouragement, sincere care and support throughout my entire study.

Nuttapong Ithiapa

March, 2012

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES (Continued)

LIST OF ABBREVIATION

2D	=	Two-dimension
3D-QSAR	=	Three-dimensional quantitative structure-activity relationship
Ala (A)	=	Alanine
Asn (N)	=	Asparagine
Arg (R)	=	Arginine
Asp (D)	=	Aspatic acid
B3LYP	=	Beck's three parameter hybrid functional using the LYP correlation functional
BSSE-CP	=	Basis set superposition error based on the counterpoise scheme
CoMFA	=	Comparative molecular field analysis
Comp	=	Compounds
C_{sp3} (+1)	=	Carbon sp3-hybridization with plus 1 charge probe atom
Cys (C)	=	Cysteine
Cyc	=	Cycloguanil
DNA	=	Deoxyribonucleic acid
DHF	=	Dihydrofolate
DHFR	=	Dihydrofolate reductase
dTMP	=	Deoxythymidylate
Gly (G)	=	Glycine
H (+1)	=	Hydrogen plus 1 charge probe atom
HF	=	Hartree-fock theory
HIV-1	=	Human immunodeficiency virus type 1
HQ	=	High level of quantum chemical calculations
Ile (I)	=	Isoleucine
K_i	=	Inhibition constant
Leu (L)	=	Leucine
LOO	=	Leave-one-out
LQ	=	Low level of quantum chemical calculations
Lys (K)	=	Lysine

LIST OF ABBREVIATION (Continued)

MD	=	Molecular dynamics
Met (M)	=	Methionine
MLR	=	Multiple linear regression
MM	=	Molecular mechanics
MP2	=	Second order Möller-plesset
NAD	=	Nicotinamide adenine dinucleophide
Noc	=	Number of component
ONIOM	=	Our own n-layer intergrated molecular orbital molecular mechanics
O_{sp3} (-1)	=	Oxygen sp3-hybridization with minus 1 charge probe atom
PDB	=	Protein data bank
Pf	=	*Plasmodium falciparum*
*Pf*DHFR	=	*Plasmodium falciparum* dihydrofolate reductase
Phe (F)	=	Phenylalanine
p*Ki*	=	Negative logarithm of inhibition constant
PLS	=	Partial least square
PM3	=	Modified neglect of diatomic overlap, parametric method number 3
PME	=	Particle mesh ewald
PRESS	=	Prediction error sum of squares
Pro (P)	=	Proline amino acid
Pyr	=	Pyrimethamine
QM	=	Quantum mechanics
QM/MM	=	Quantum mechanical/molecular mechanical method
QSAR	=	Quantitative structure-activity relationship
r^2_{cv}	=	Predictive ability of cross-validation
r^2_{nv}	=	Predictive ability of no-validation
RMS	=	Root mean square
RMSD	=	Root mean square deviation

LIST OF ABBREVIATION (Continued)

Ser (S)	=	Serine
SHMT	=	Serine hydroxymethyltransferase
S_{press}	=	Uncertainty of the prediction
SSY	=	Variance of the data around the mean value
THF	=	Tetrahydrofolate
Thr (T)	=	Threonine
Trp (W)	=	Tryptophan
TS	=	Thymidylate synthase
Tyr (Y)	=	Tyrosine
Val (V)	=	Valine
UFF	=	universal force field

COMPARATIVE MOLECULAR FIELD ANALYSIS AND QUANTUM CALCULATION STUDY ON ANTI HIV-1 RT DIARYLANILINE DERIVATIVES

INTRODUCTION

Human immunodeficiency virus

The human immunodeficiency virus type-1 (HIV-1) is a retrovirus that infects cells of the human immune system causing the globally disseminated disease named acquired immunodeficiency syndrome (AIDS) for over 20 years. During this time an unprecedented success has been achieved in discovering anti-HIV drugs as reflected by the fact that there are now more drugs approved for the treatment of HIV than for all other viral infections taken together. The currently Food and Drug Administration (FDA) approved anti-HIV drugs can be divided into seven groups: nucleoside reverse transcriptase inhibitors (NRTIs), nucleotide reverse transcriptase inhibitors (NtRTIs), non-nucleoside reverse transcriptase inhibitors (NNRTIs), protease inhibitors (PIs), fusion inhibitors (FIs), co-receptor inhibitors (CRIs), and integrase inhibitors (INIs). This arsenal of drugs, which is used in combinations, has moved the prognosis of HIV patients from that of high morbidity and mortality to, for many at least, a chronic, manageable but still complex disease (Kitchen *et al.*, 2001, Valenti 2001, King, 2003). However, the use of these drugs has been relatively limited by their toxicity (Carr 2003), drug resistance development (Martinez-Picado *et al.*, 2000), and more worryingly, the fact that some newly HIV-infected patients carry viruses that are already resistant to the currently approved AIDS treatments (Little *et al.*, 2002).

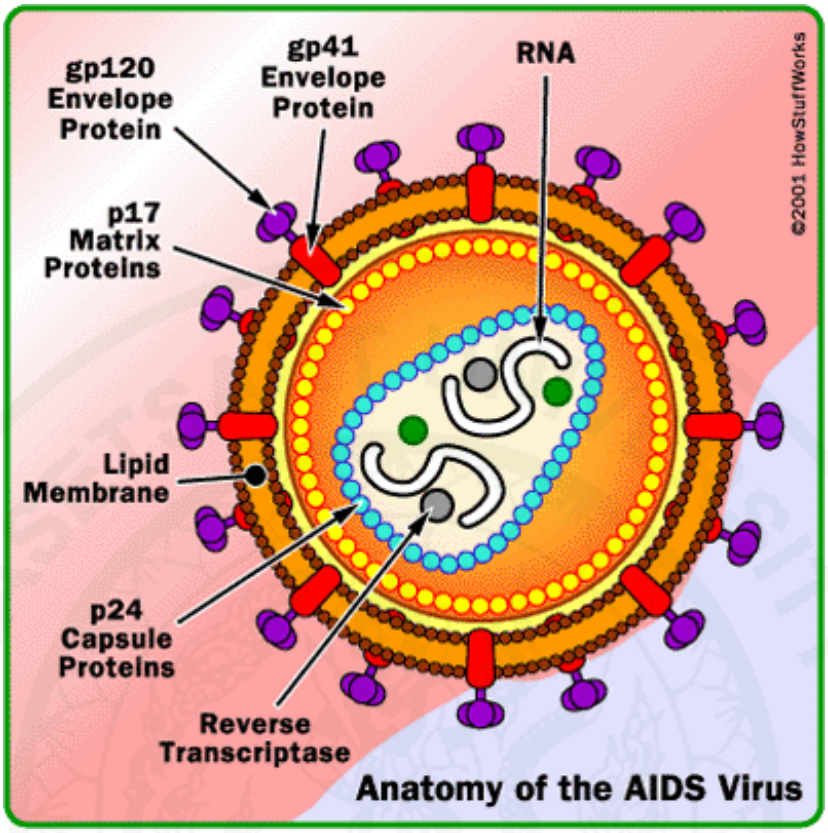


Figure 1 Anatomy of the AIDS Virus (*source: http://health.howstuffworks.com*)

HIV Life Cycle and Anti-HIV Drug Design

The HIV life cycle encompasses several crucial steps, starting from the attachment of the virus to the host cell membrane and finishing with the release of progeny virions from the cell, as summarized in Figure 2. The HIV life cycle commences by a specific interaction between the virion glycoprotein gp120 on the outer membrane and the CD4 receptor on the host cell surface. This reaction results in a conformational change allowing the interaction of gp120 with the chemokine coreceptor CXCR4 or CCR5. This is then followed by further conformational changes that expose a fusogenic peptide, which anchors into the host cell membrane. Once the viral envelope and cell membrane have fused, the virion is decapsidated releasing the viral RNA into the host cell's cytoplasm. Through the reverse transcription, the viral RNA is transcribed to viral double-stranded DNA. This process is catalyzed by a RNA-dependent DNA polymerase, also known as reverse transcriptase, which is

encoded by the viral genome. The viral DNA is then integrated into the host chromosome, and after transcription (facilitated by regulatory proteins Tat and Rev, which are themselves viral gene products) and translation into viral proteins using the cells' machinery, the assembly of the Gag and Gag-Pol poly-proteins occurs near the cell membrane (De Clercq 2002, Meadows *et al.*, 2006). During viral assembly, two copies of single-stranded viral RNA are incorporated into the virion, which then buds off from the cell, taking with it part of the host cell membrane. Soon after budding, viral protease cleaves the Gag-Pol poly-protein to generate a mature, functional virion (Meadows *et al.*, 2006).

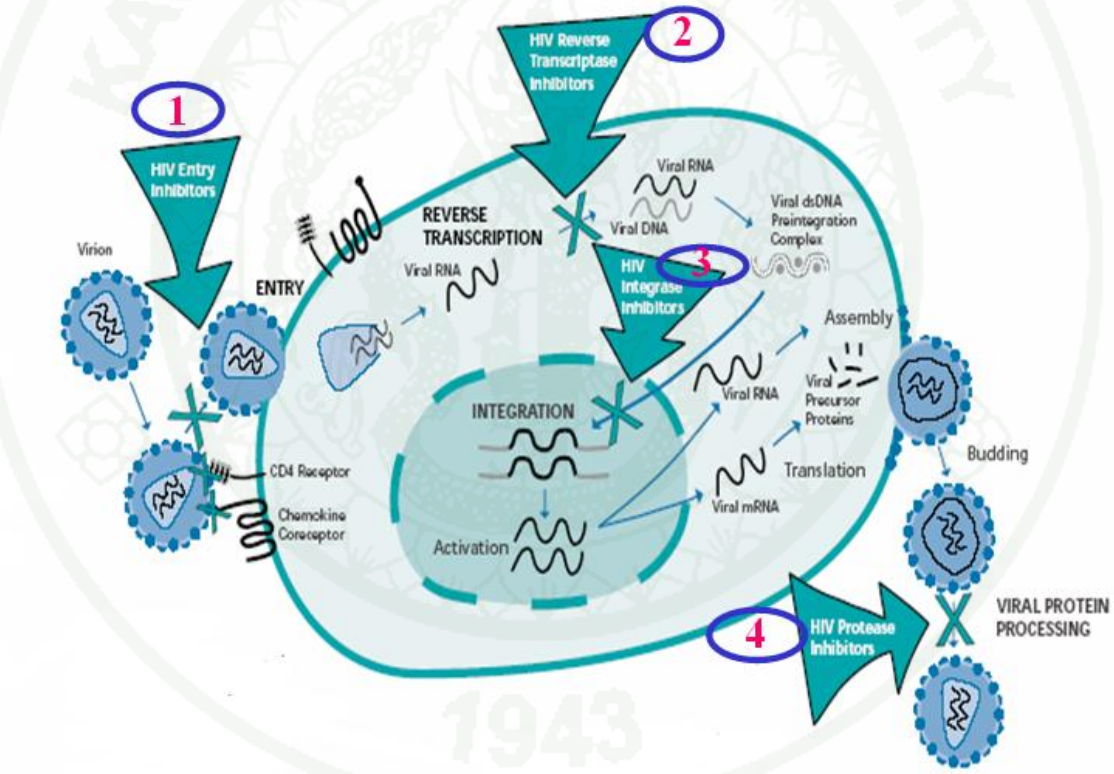


Figure 2 HIV Life Cycle (*source: http://www.biomems.co.za*)

From HIV life cycle, it can be divided into four parts for drug design: (i) nucleoside reverse transcriptase (ii) non-nucleoside reverse transcriptase (iii) integrase enzyme and (iiii) protease enzyme. In this research interests in the part of non-nucleoside reverse transcriptase.

The HIV-1 RT enzyme

Reverse transcriptase (RT) is an enzyme that can convert the single-stranded viral genomic RNA into a linear double-stranded DNA, integrated into the host chromosomes. It is an essential step in the HIV life cycle, for replication. The enzyme has two activities, (i) a DNA polymerase that can use either RNA or DNA as a template and (ii) a ribonuclease H (RNase H) that selectively degrades the RNA strand of an RNA-DNA heteroduplex. HIV-1 RT is a symmetric heterodimer composed of 66 kDa (p66) and 51 kDa (p51) subunits. Both subunits consist of four polymerase subdomains: the thumb, palm, fingers, and connection in common. Differentially, the C-terminus of p66 contains an additional 120 amino acids that form the bulk of the RNase H domain. Therefore, the p66 subunit contains two vital domains termed DNA polymerase and RNase H, related to the RT function and mechanisms (Himmel et al., 2006).

Several crystal structures of free, unliganded HIV-1 RT have been solved (Esnouf, *et al.*, 1995). The three-dimensional structure of the p66 subunit is often compared to a right hand (Figure 4), with a fingers (amino acids 1-85 and 118-155), a palm (amino acids 86-117 and 156-237) and a thumb (amino acids 238-318) domain (Kohlstaedt *et al.*, 1992). The palm domain contains the polymerase active site with its three aspartic acids (110, 185 and 186) and the YMDD characteristic motif. Co-crystals of RT with a modified oligonucleotide and a dNTP (Huang *et al.*, 1998) or double-stranded DNA (Jacobo-Molina *et al.*, 1993) have revealed that the nucleic acid passes in the cleft behind the fingers and in front of the thumb domain. The catalytic pocket is formed by the fingers folding down into the palm domain, as observed in the RT-dNTP complex (Huang *et al.*, 1998). In this structure, the nucleic acid is located in front of both the fingers and the thumb. Next to the catalytic domain, the p66 subunit also contains the RNaseH domain (amino acids 427-560), linked to the former by the connection domain (amino acids 319-426). The connection domain is also involved in interactions with the nucleic acid and the p51 subunit. Despite their sequence homology, the p66 subunit assumes a flexible and open structure, whereas

the p51 subunit is rather compact, and seems to play a structural role, devoid of catalytic activity, with the three aspartic acids buried inside (Kohlstaedt *et al.*, 1992).

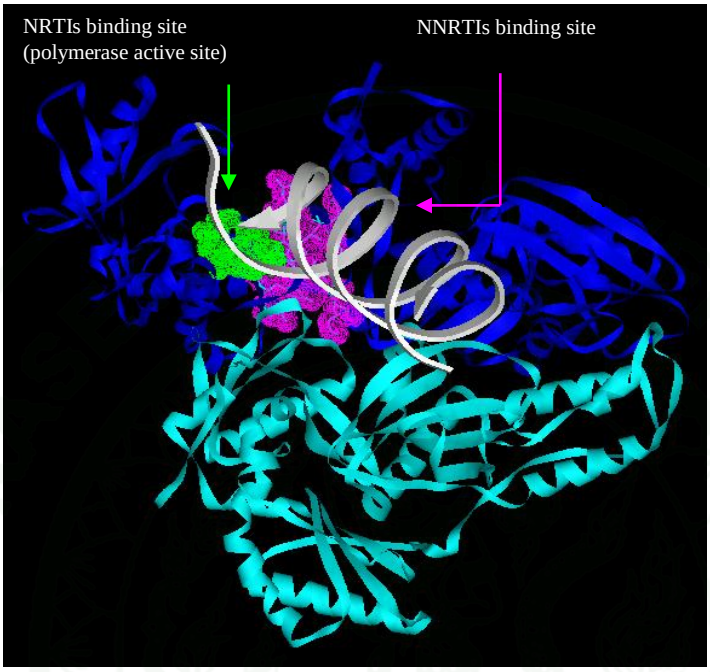


Figure 3 NRTIs and NNRTIs binding sites of HIV-1 RT structure

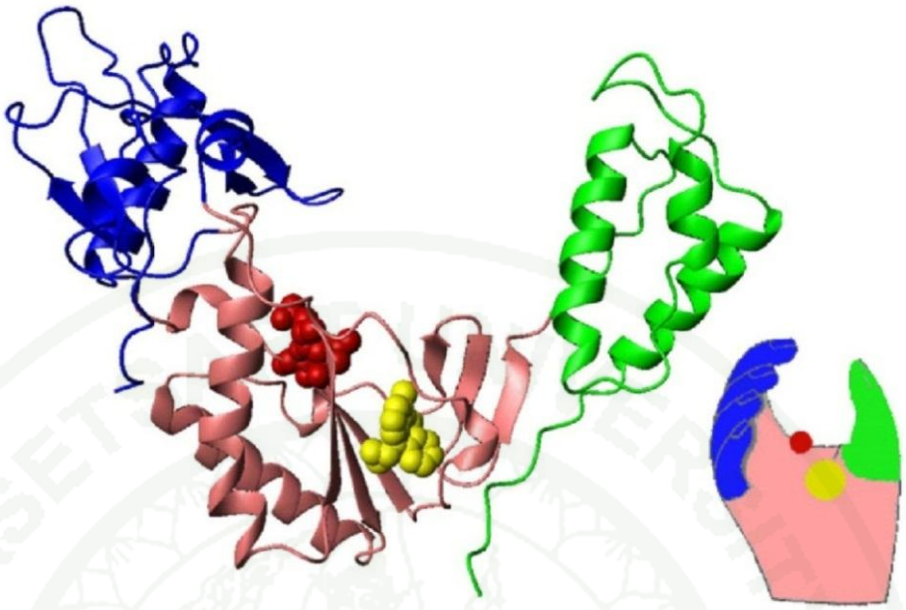


Figure 4 This ribbon representation of the RT active domain illustrates its hand-like structure, showing finger (blue), palm (pink) and thumb (green). The active site (red atom), where DNA is elongated, is in the palm region. Also shown is an RT-inhibitor drug (yellow) in the pocket where it binds. (*source: Marie-Pierre de Béthune(2010)*).

Non- Nucleoside Reverse Transcriptase Inhibitors (NNRTIs)

The second type of drug to be developed to fight HIV is non-nucleoside reverse transcriptase inhibitors, abbreviated to NNRTI, and approved in 1996. Like nucleoside RT inhibitors, NNRTIs stop the conversion of RNA to DNA previous to integration into the host cell DNA. These drugs work by changing the shape of the reverse transcriptase so they will not fit the RNA. As a result, the RNA has never converted into DNA and integrated into the host cell DNA. The first report on the ability of non-substrate analogues to inhibit the HIV reverse transcriptase appeared in 1989/1990 (Baba *et al.*, 1989, Merluzzi *et al.*, 1990, Pauwels *et al.*, 1990). These agents inhibit the HIV-reverse transcriptase by binding noncompetitively to an allosteric site located at a short distance (~15 Å) from the catalytic site (De Clercq *et al.*, 2004, Pauwels, R, 2004).

The first drug generation, there are three different NNRTIs i.e. (i) Nevirapine (**cp.1**), is a dipyridodiazepinone inhibitor of HIV-1, discovered by researchers at Boehringer Ingelheim (Merluzzi V.J., *et al.*, 1990). At the time Nevirapine was developed, the concept of HAART or Highly Active Anti-Retroviral Therapy was not yet established, and the drug was often studied in combination with a single NRTI, mostly AZT, a regimen that could not prevent the emergence of resistance (Havlir *et al.*, 1995). Nevirapine, in combination with two NRTIs is the recommended NNRTIs for first line therapy in resource limited countries. (ii) Delavirdine (cp.2) belongs to the family of bis(hetero-aryl)piperazine compounds discovered by researchers at Upjohn Laboratories (Dueweke *et al.,* 1993). It is bulkier than the other NNRTIs, and crystal structure of delavirdine with HIV- RT have shown that it protrudes outside the NNRTI binding pocket, which explains its particular resistance profile (Esnouf *et al.,* 1997). Like nevirapine, delavirdine was originally assessed in suboptimal regiments, where emergence of resistance could not be prevented (Davey Jr. *et al.,* 1996). Nowadays, delavirdine is rarely used. (iii) Efavirenz (cp.3) is a benzoxazinone discovered by the researchers at Merck (Young *et al.,* 1995), and developed jointly by DuPont and Merck. Unlike Nevirapine and delavirdine, efavirenz could be studied in Phase III trials as part of appropriate HAART regiments, and showed sustained efficiency (Staszewski *et al.,* 1999). The most frequently selected mutation efavirenz failure is K103N. Efavirenz is the most used naive patients, in combination with two NRTIs.

For the second drug generation, there are drugs in the group of etravirine (TMC125) belonging to the family of di-aryl-pyrimidine (DAPY) compounds, and are the results of a long lead optimization campaign conducted by researchers at the Janssen Research Foundation and Tibotec, aiming at identifying new NNRTIs with a better resistance profile and an increased genetic barrier to the development of resistance (Ludovici *et al.*, 2001). The screening process included the profiling of compounds against wild type and selected single and double mutant NNRTI resistant HIV-1 strains, as well as the assessment of their metabolic stability (Andries *et al.,* 2004). The resistance profile of etravirine was further confirmed by testing the compounds against thousands of NNRTI resistant HIV-1 clinical isolates,

representing the diversity of mutations patterns encountered in the clinic (De Béthune *et al.*, 2000). In vitro, etravirine shows a higher genetic barrier to the development of resistance as compared to nevirapine and efavirenz (Vingerhoets *et al.*, 2005). Co-crystals of etravirine with the K103N mutant RT helped to study the binding mode of this inhibitor to the enzyme. It is hypothesized that the inhibitor can adopt different conformations in the NNRTI binding pocket, because of its flexibility, and can thereby accommodate the mutations better than first generation NNRTIs, which are more rigid molecules (Das *et al.*, 2004, Das *et al.*, 2005).

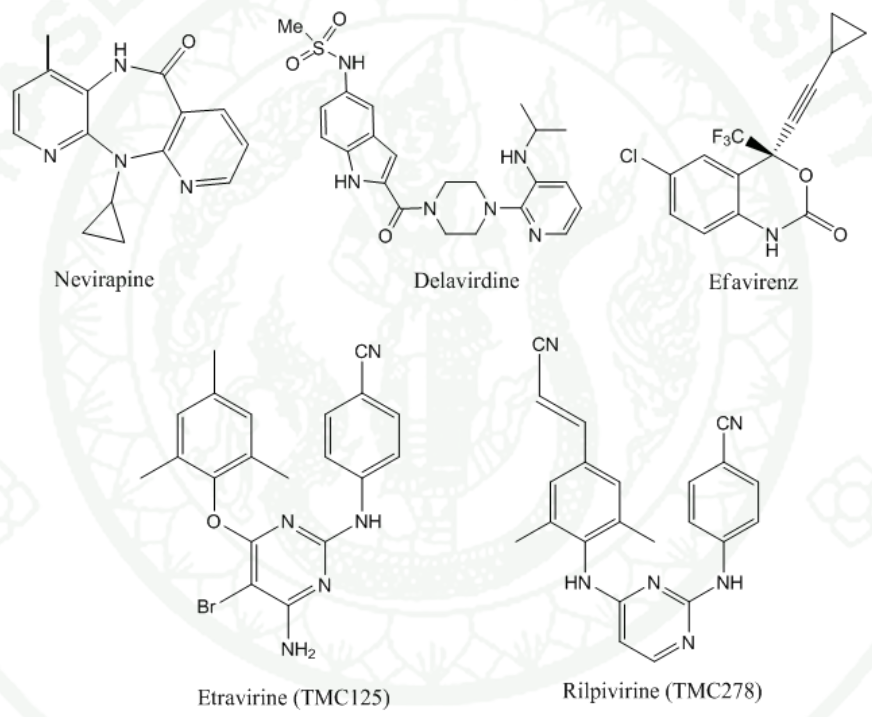


Figure 5 Chemical structures of HIV-1 NNRTIs agents

Three-dimensional Quantitative Structure-Activity Relationship and Its Applications

Classical QSAR correlates biological activities of drugs with physicochemical properties or indicator variables which encode certain structural features (Ramsden, 1990; Kubinyi, 1993; Kubinyi, 1995; Hansch and Leo, 1995; Waterbeemd, 1996). In

addition to lipophilicity, polarizability, and electronic properties, steric parameters are also frequently used to describe the different size of substituents. In some cases, indicator variables have been attributed to differentiate racemates and active enantiomers (Kubinyi, 1995). However, in general, QSAR analyses consider neither the 3D structures of drugs nor their chirality.

In 1979, Cramer and Milne made a first attempt to compare molecules by aligning them in space and by mapping their molecular fields to a 3D grid (Cramer and Milne, 1979). In the following years, this approach was further developed as the DYLOMMS (dynamic lattice-oriented molecular modelling system) method (Kubinyi, 1993) but was not very well accepted by the scientific community. Several important facts had to work together to allow a broader application of this approach.

In 1986, Svante Wold proposed the use of partial least squares (PLS) analysis, instead of principal component analysis, to correlate the field values with the biological activities.

Especially, in 1988, a key publication appeared in the Journal of the American Chemical Society (Cramer *et al*., 1988) and the method was called comparative molecular field analysis (CoMFA).

Finally, appropriate software became commercially available SYBYL/QSAR, Molecular Modelling Software, Tripos Inc., 1699 S, Hanley Road, St. Louis, MO 63944, USA.

Since 1988, many publications, several reviews and books have appeared on CoMFA subject. This analysis is a useful tool for deriving 3D-QSAR models which related between biological activity and the molecular fields of steric and electrostatic using the Lennerd-Jone and Coulomb potentials, respectively.

The three-dimensional Quantitative structure-activity relationship (3D-QSAR) models are constructed by correlating the 3D fields to the corresponding experimental activities of ligands with respect to a common target receptor. As 3D molecular

modeling becomes widely available to workers in structural biology, properties calculated from atomic coordinates are used increasingly to explain and predict biological activity (Brown and Martin, 1997). Molecular modeling can produce any desired number of explanatory descriptors for each structure-far more than the number of activity data to be explained. There are many techniques for 3D-QSAR, for example, Comparative Molecular Field Analysis (CoMFA) bases its predictions upon field values calculated at each point of a 3D grid around the molecular structures. The field values are highly correlated, having been derived from the molecular descriptors such as atomic charges and positions (Cramer et al., 1988). In the model, the linear regression by partial least squares (PLS) is used for producing a formula which fits the training data (Clark and Cramer, 1993, Bush and Nachbar, 1993). The advantage of PLS method is reducing the explanatory data to a small number of components. Within the molecular modeling package SYBYL, Partial Least Square (PLS) is the recommended regression method for analysis of CoMFA fields. The advantages of CoMFA are the ability to predict the biological activities of the molecules and to represent the relationships between steric/electrostatic properties, calculated according to Lennard-Jones and Coulomb potentials, respectively, and also the biological activity in the form of contour maps to provide the key features of both the ligand-receptor interaction and the topology of the receptor. The CoMFA results can be used for guiding the new design potent compounds, based on the same template constructions.

Quantum Chemical Calculations

As theoretical/computational chemistry has gained a major role in studies of chemical problems in the last few decades, it became a challenge for theoreticians to accurately treat large molecular systems such as ligand-enzyme in biochemistry. The quantum computational chemistry calculations can be used to investigate inhibitor-enzyme interactions at the molecular level. The obtained strong repulsive energy of some amino acids that was quantified implies that this is the cause of the HIV-1 RT resistance. On the other hand, the strong interactions between some amino acids and the inhibitor should be investigated to discover further new designs. Therefore, an understanding of the particular interaction energy of individual amino acids with the HIV-1 RT inhibitors at molecular level is required to support the identification of these new structural modifications.

OBJECTIVES

In the present work, ligand-based drug design approaches using CoMFA has been applied to the class of diarylaniline derivatives with the aims of:

1. To construct the relationship between HIV-1 biological activities and structural properties using both QSAR and 3D-QSAR.

2. To investigate the different particular interaction of poor and potent inhibitors which align into same binding pocket of HIV-1 Reverse Transcriptase.

LITERATURE REVIEW

AIDS, or acquired immunodeficiency syndrome is caused by the human immunodeficiency virus type 1 (HIV-1). HIV-1 genome encodes for three major enzymes protease, reverse transcriptase and integrase for HIV-1 replication. Reverse transcriptase is a key enzyme in the HIV replication cycle and is one of the main targets in the development of drugs for treating HIV-infection and AIDS. HIV-1 reverse transcriptase (HIV-1 RT), which is virally encoded, catalyses the conversion of viral RNA into double stranded DNA, which is then integrated in the host genome. Two types of drugs that inhibit HIV-1 polymerase activity are nucleoside and non-nucleoside inhibitors. Non-nucleoside reverse transcriptase inhibitors (NNRTIs) are important components of the first line highly active antiretroviral therapy regiments.

Specifically, we focused on NNRTIs that interact with the allosteric binding site, a highly hydrophobic cavity, in a non-compatitive manner to cause distortion of the three-dimensional structure of the enzyme and thus inhibit RT catalytic function. NNRTIs currently approved for AIDS therapy include nevirapine, delavirdine, efavirenz and etravirine (TMC125). Etravirine is the most recently approved NNRTI and is active against many drug-resistant HIV-1 strains. TMC125 (Andries, 2004), a prior clinical candidate, belong to the diarylpyrimidine (DAPY) family and is very potent against wild-type and many drug-resistant HIV-1 strains with nanomolar EC_{50} values. They have excellent pharmacological profiles, which has encouraged more research to explore next-generaion NNRTI agents (Tucker *et. al.,* 2008; Romines *et. al.*, 2006; Himmel *et. al.*, 2005).

In 2001, Donald W. Ludovici and co-worker shown synthesis and anti-HIV-1 activity of a series of DAPYs are described. Several members of this novel class of NNRTIs are extremely potent against both wild-type and a panel of clinically significant single- and double-mutant strains of HIV-1. In 2004, Yven Van Herrewege and co-worker study A new DAPY Series of diaryltriazines and diarylpyrimidines are highly potent nonnucleoside reverse transcriptase inhibitors, compared to the reference compounds UC-781 and PMPA, with possible applications as microbicides.

In 2007, Céline Mordant and co-worker,novel DAPY, which represent next generation of NNRTIs, were synthesized and their activities against HIV-1 assessed. Modulations at positions 2 and 6 of the left phenyl ring generated interesting derivatives of TMC278 displaying high potency against wild-type and mutant viruses. In 2009, Xiao-Qing Feng and co-worker, a novel series of DAPYs featuring a naphthyl moiety at the C4 position were designed, with all compounds exhibiting strong activity against wild-type HIV-1. Yong-Hong Liang and co-worker, a series of 38 2-naphthyl-substituted diarylpyrimidine (DAPY) analogues, characterized by various substitution patterns on the pyrimidine and naphthalene rings. Most of the compounds displayed strong activity against wild-type HIV-1. The most active compound, with a cyano group at position C6 on the naphthalene ring, exhibited activity against wild-type HIV-1 with an EC_{50} value of 0.002 μm and against the double mutant strain with an EC_{50} value of 0.24 μm; The structure–activity relationship (SAR) of the newly synthesized DAPYs is presented herein. In 2011, Dr. Xiao-Dong Ma and co-worker, Synthesis and Anti-HIV Activity of Aryl-2-[(4-cyanophenyl)amino]-4-pyrimidinone hydrazones as Potent Non-nucleoside Reverse Transcriptase Inhibitors. also the similarities in chemical structures lead to the emergence of crossresistance among members of the same class, where a real medical need to develop new generation of NNRTIs which do not give rise to cross-resistance and are effective against clinically relevant mutant strains (Guillemont *et. al.*, 2005).

QSAR and CoMFA method has been applied as an important tool in drug discovery and environmental risk assessment. It provides useful references for understanding the relationship between the chemical structure and biological activity of compounds, and ultimately leads to statistically robust models that can be used to make accurate and reliable predictions of the biological activity of new compounds. For many therapeutic targets of interest, structure-based approaches are not yet applicable because the structure of the target macromolecule is unknown. So, in these cases, QSAR techniques provide the best approach to rational drug design. Traditional (two-dimensional) QSAR methods attempt to correlate biological activity with local features of atoms, whole molecular properties (e.g. charge) and substituent effects (e.g. fragment hydrophobicity indices). The developments in traditional QSAR

continue to appear in the literatures. Most interest in this field focuses on three-dimensional QSAR which called 3D-QSAR.

Classical QSAR correlates biological activities of drugs with physicochemical properties or indicator variables which encode certain structural features (Ramsden, 1990; Kubinyi, 1993; Kubinyi, 1995; Hansch and Leo, 1995; Waterbeemd, 1996). In addition to lipophilicity, polarizability, and electronic properties, steric parameters are also frequently used to describe the different size of substituents. In some cases, indicator variables have been attributed to differentiate racemates and active enantiomers (Kubinyi, 1995). However, in general, QSAR analyses consider neither the 3D structures of drugs nor their chirality. Since 1988, many publications, several reviews and books have appeared on CoMFA subject. This analysis is a useful tool for deriving 3D-QSAR models which related between biological activity and the molecular fields of steric and electrostatic using the Lennerd-Jone and Coulomb potentials, respectively.

There are now a few practical applications of CoMFA for DAPY series in diarylpyrimidine derivatives. For example, recently, Joseph Rebehmed and co-worker (Rebehmed *et al.*, 2008) applied 2D and 3D QSAR methods to find the structure activity relationship of some diarylpyrimidine derivatives having good activity against resistant strain of HIV-1 reverse transcriptase (RT). 2D QSAR was performed using the heuristic method in CODESSA which had led to a linear model (R^2 = 0.928 and s^2 = 0.015) between the inhibitory activity and five descriptors. CoMFA and CoMSIA models were established using SYBYL package of programs. The better predictive ability of the CoMSIA model (q^2 = 0.730) over the CoMFA model (q^2 = 0.597) was assigned to the large contribution of hydrogen-bonding interactions to the inhibitory activity. Based on their q^2, it can be implied that the model should be re-derived to be the good model for both training set and also test set compounds. For Abhilash Thakur and co-worker (Thakur *et al.,* 2008), this work describes QSAR and SAR studies on the inhibition of reverse transcriptase by 31 novel DAPY (diarylpyrimidine) derivatives. The application of a multiple linear regression analysis

indicated that a combination of topological and physicochemical descriptors and the indicator parameters yielded a statistically significant model for the prediction of the activity, log 1/C (50% of effective concentration of DAPY derivatives for RTs). The modelling of some new potential DAPY compounds and their maximum active comformers for the inhibition of reverse transcriptase are made by quantum molecular modelling. However, DAPY series have been study are ongoing DAPY series for the QSAR techniques, such as, Hao Zhang and co-worker used revealing the drug-resistant mechanism for diarylpyrimidine analogue inhibitors of HIV-1 Reverse Transcriptase. (Zhang *et al*., 2011) In prevoius literatures, diarylaniline derivative as part of DAPY series there are no CoMFA models of diarylaniline derivatives which active against wild type of HIV-1 Reverse Transcriptase. Therefore, this research we have plan to establish the CoMFA models for diarylaniline derivatives. The obtained CoMFA results have been published as open sources for guiding to develop new and effective anti HIV-1 against in HIV-1 Reverse Transcriptase.

Generally, no single theoretical method is able to provide both the accuracy and acceptable computational cost that are required for the investigation of such chemical processes. To understand the orientation and interaction of the inhibitor in the binding pocket is important and understanding attractive interaction and repulsive interaction of weak interactions. There are a few applications of particular interaction which applied to the ligand-enzyme biological systems. Many research groups applied this method to study particular interaction on HIV1-RT enzymes which complexed with many type of ligands. The obtained results clearly demonstrated the different of binding energy between the potent and the poor ligands for wild type and mutant type of HIV1-RT (Kuno *et al*., 2003, 2006; Nunrium *et al*., 2005; Saen-oon, 2005). In addition to, the quantum chemical calculation approach was also successful to describe interactions of efavirenz with HIV-1 RT in comparison between the wild-type and mutant types. And the quantum computational chemistry calculations capable of investigating inhibitor-enzyme interactions at the molecular level are employed (Srivab, 2008).

The goal of our study is constructed the relationship between HIV-1 biological activities and structural properties using both QSAR and 3D-QSAR and investigated the particular interaction of each inhibitors in binding pocket of HIV-1 Reverse Transcriptase. Gaining insight into the particular interaction energy terms will also give us a better understanding and more detailed information on the interaction between the HIV-1 RT inhibitor and the binding pocket of the wild type HIV-1 RT. Detailed knowledge of the interactions between drug and the binding site of an enzyme can provide a structural explanation for the structure-based drug design of HIV-1 RT inhibitors and a better understanding of the action of these.

METHODS OF CALCULATIONS

CoMFA Study

Biological data

Diarylaniline

Twenty-five Diarylaniline derivatives used for the CoMFA study were selected from Bingjie Qin *et. al*. as shown in Table 1. There are four substitute positions, R_1, R_2, R_3 and R_4 as shown in Figure 6. The twenty-one compounds served as the training set. In addition, four compounds (compound number 5, 8, 14, 20) that were eliminated from the training set which removed these so-called outlier. For set of biological data, the activity, EC_{50} (μM) for inhibiting wild-type HIV-1 RT, was measured *in vitro* under the same experimental conditions. Consequently, *in vitro* HIV-1 RT inhibitor activities were converted into the corresponding pEC_{50} (-logEC_{50}) values. These values were used as dependent variables in the CoMFA study.

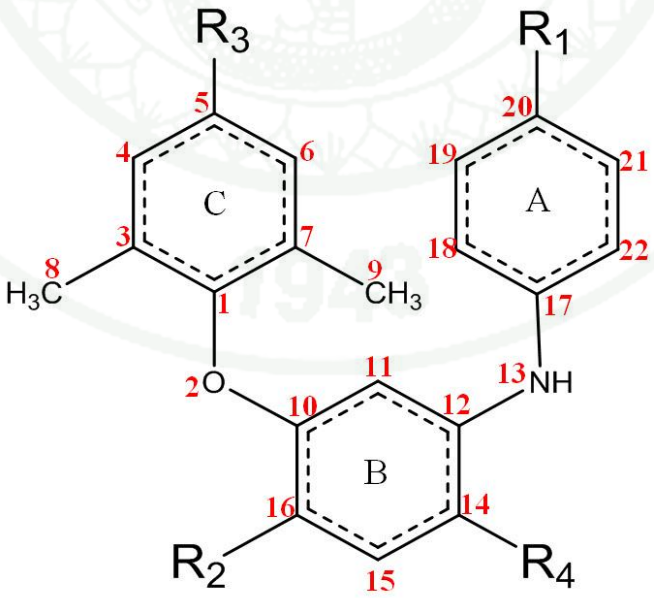


Figure 6 Template structure of Diarylaniline derivatives

Table 1 Data set used for CoMFA analysis with EC_{50} (μM) and pEC_{50} values in the Wild-type HIV-1 RT.

Name	R1	R2	R3.	R4	EC50	pEC50
Compound 1	OCH3	NO2	H	NO2	3.840	5.4157
Compound 2	OCH3	NO2	CH3	NO2	2.990	5.5243
Compound 3	OCH3	NO2	Br	NO2	3.630	5.4410
Compound 4	CH3	NO2	Br	NO2	4.310	5.3655
*Compound 5	NO2	NO2	Br	NO2	> 49.7	-
Compound 6	C≡N	NO2	Br	NO2	0.172	4.3036
Compound 7	C≡N	NO2	H	NO2	0.545	6.2636
*Compound 8	C≡N	NO2	C≡N	NO2	4.190	-
Compound 9	C≡N	NO2	CH3	NO2	0.280	6.5528
Compound 10	C≡N	NO2	CHO	NO2	1.530	5.8153
Compound 11	C≡N	H	Br	NO2	0.317	6.4989
Compound 12	C≡N	H	H	NO2	3.147	5.5021
Compound 13	C≡N	H	CN	NO2	0.208	6.6819
*Compound 14	C≡N	H	Me	NO2	0.067	-
Compound 15	C≡N	H	CHO	NO2	2.190	5.6596
Compound 16	C≡N	H	Br	NH2	0.047	7.3279
Compound 17	C≡N	H	Br	NH2	0.070	7.1549
Compound 18	C≡N	H	Br	NH2	0.073	7.1366

Table 1 Data set used for CoMFA analysis with EC_{50} (μM) and pEC_{50} values in the Wild-type HIV-1 RT.

Name	R1	R2	R3.	R4	EC50	pEC50
Compound 19	C≡N	NH2	Br	NH2	0.161	6.7932
*Compound 20	C≡N	NH2	H	NH2	3.226	-
Compound 21	C≡N	NH2	C≡N	NH2	0.030	7.5229
Compound 22	C≡N	NH2	CH3	NH2	0.070	7.1549
Compound 23	C≡N	NO2	Br	NH2	0.016	7.7959
Compound 24	C≡N	NO2	C≡N	NH2	0.003	8.5229
Compound 25	C≡N	NO2	CH3	NH2	0.062	7.2076

* outlier compounds

Structural Construction

Three-dimensional structure building was constructed using the Sybyl 8.0 program package on a Silicon Graphics Octane2 workstation at the National Electronics and Computer Technology Center of Thailand (NECTEC). The structures of diarylaniline derivatives were built using the SKETCH module in Sybyl. The skeleton and conformation of diarylaniline was extracted from the crystal structure of a TMC125 complex with wild-type HIV-1 RT with PDB code 3MEC (Eric B. Lansdon, *et al*., 2010). The other molecules were built taking compound of **TMC125** as a template and changing their substituents.

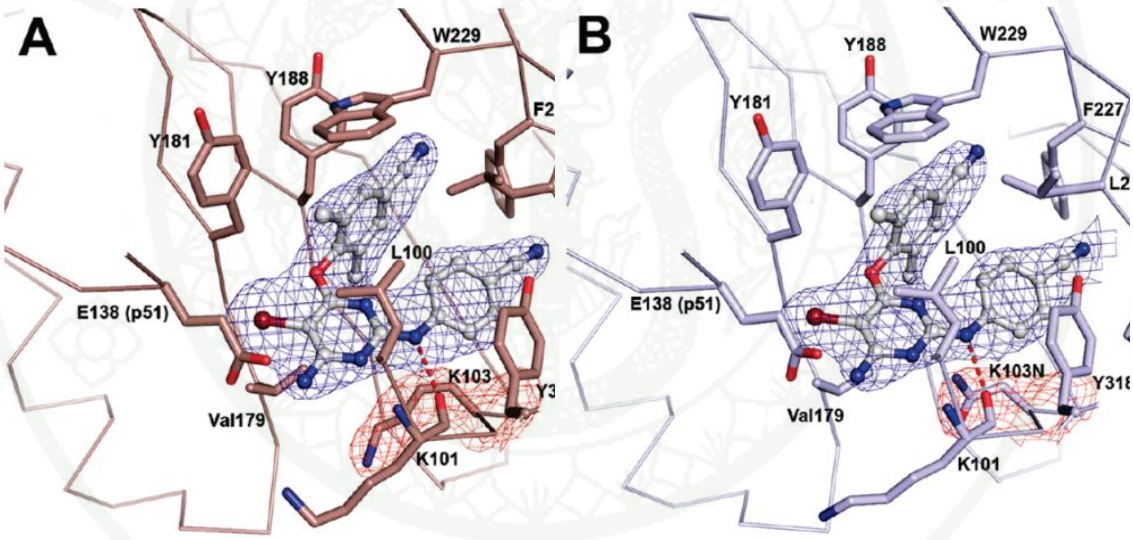


Figure 7 Cocrystal structures of (A) HIV-1 RT with WT-TMC125, (B) K103N-TMC125 Shown in blue mesh and contoured at 1.0σ is The composite omit map drawn around the inhibitor. The omit map for the K103(N) side chain is shown in red mesh.

Quatitative Structure-Activity Relationships Analysis

All derivative structural geometries were initially construction and modified by using SYBYL 7.00 (SYBYL Molecular Modelling Softwares, Version 7.00, Tripos Associates, Inc., St. Louis, MO, 63144, USA, 2003.).

Finding an accurate method for estimating the affinity of protein ligands activity is one of the most challenging tasks in computer-aided molecular design. Quantitative structure-activity relationship (QSAR) is a mathematical relationship between a biological activity of a molecule and its geometric and chemical characteristics has been proven to be the principle method used for activity prediction in drug design.

Activity should be a function of the geometric and chemical characteristics of the compounds. QSAR attempts to find consistent relationship so that can be used to evaluate activity of new compounds.

Three Dimentional Quantitative Structure-Activity Relationships Analysis (3D-QSAR)

3D-QSAR techniques are routinely used in analog-based drug design. The ability to produce quantitative correlation between three-dimensional properties of molecules and the biological activity of these compounds is of inestimable value in deciding upon the choice of further synthetic chemistry.

Comparative Molecular Field Analysis (CoMFA) is a 3D-QSAR method that search for relationship between the biological activity of a set of compounds (with specified alignment) and their three-dimensional electronic and steric properties (so called molecular fields).

In the present study, we have developed 3D-QSAR CoMFA model for the series of Diarylaniline derivatives and the contour maps derived revealed the

significance of steric and electrostatic. The structural variations in the molecular fields at particular regions in the space were studies and 3D-QSAR models generated gives an insight on the design of potent Diarylaniline inhibitor against wild type HIV-1 Reverse Transcriptase.

1. Comparative Molecular Field Analysis (CoMFA)

Comparative Molecular Field Analysis (CoMFA) was developed by Cramer *et al.* in 1988. CoMFA is a powerful 3D-QSAR technique providing further insight into the relationships between the structure and function of these Diarylaniline analogues. This methodolody is based on assumption that non-covalent forces dominate receptor-drud interactions and that these forces can be descripbed in term of steric and electrostatic fields. The changes in the biological activities of binding affinities of sample compounds correlate with changes in the steric and electrostatic fields around these molecules. For such an approach, partial least-squares statistics was used to derive the correlation between the steric and electrostatic properties and HIV-1 RT inhibitory activity.

1.1 CoMFA Set Up

1.1.1 Alignment Rule

One of the prime requirements in CoMFA study is the alignment of all compounds relative to one another, so that they have a comparable conformation and orientation in space. Since the relative interaction energies depend strongly on relative molecular position. Partial atomic charges required for calculations of electrostatic interaction were computed. The highest inhibitory activity compound in this analogues, compound 24 for wild type of HIV-1 RT was used as templates for rms-fit molecular alignments. All the structures of Diarylaniline derivatives were aligned in a 3D lattice by fitting them with the common structures that shown in Figure 8, which performed using SYBYL 8.1 molecular modeling software.

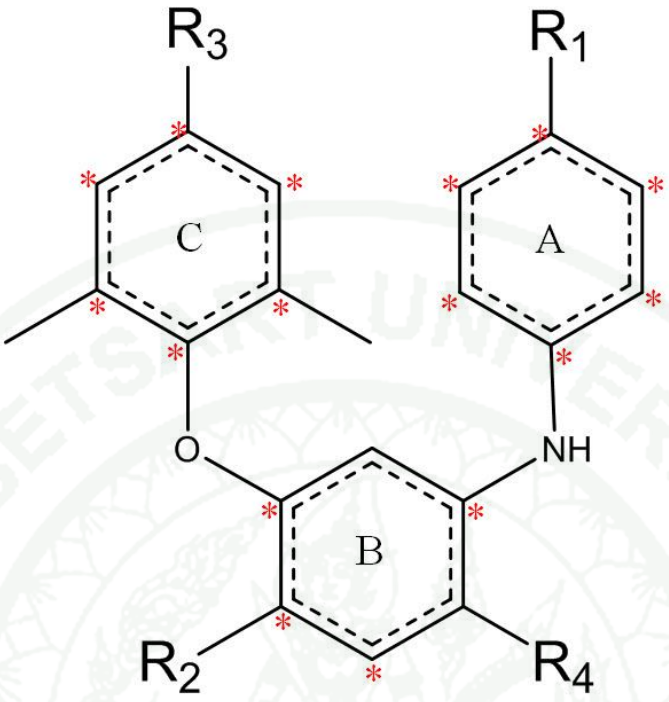


Figure 8 General structure of Diarylaniline derivatives, stars indicate the atom selected as the template for alignment rule.

1.1.2 Calculation of Interaction Energy

CoMFA cubic lattice generated around these molecules based on the molecular volume of the structure. In this investigation, three different atom, sp^3 carbon atom with +1 charge (default probe atom in SYBYL), sp^3 oxygen atom with -1 charge and sp^2 nitrogen with -1 charge, served as probe atoms. The probe atom was placed at each lattice point and their interactions of the steric and electrostatic fields with each atom molecule were all calculated with CoMFA standard scaling and then put in a CoMFA QSAR table. In order to speed up analysis and reduce the amount fo noise, the minimum sigma value was set to 2.0 kcal/mol, which omitted the analysis lattice points whose the energy variance is less than 2.0 kcal/mol, and energy cutoff values 30 kcal/mol were selected for both electrostatic and steric fields.

1.1.2.1 Steric Field

All atoms exhibit a short range interaction. This is generally referred to as the van der Waals interaction. The best known van der Waals potential function is the Lenard-Jones 12-6 potential function, which can be described in the following form:

$$E = \sum_i \sum_j \frac{A_{ij}}{r_{ij}^6} + \frac{B_{ij}}{r_{ij}^{12}} \quad (1)$$

Where A_{ij} = is the coefficient depicting repulsive heteroatomic interaction with hydrogen ($(A_iA_j)^{1/2}$)

B_{ij} = is the coefficient depicting attractive heteroatomic interaction with hydrogen ($(B_iB_j)^{1/2}$)

r_{ij} = is the distance between atom *i* of drug molecule and probe atom *j* (Å)

1.1.2.2 Electrostatic Field

Electrostatic interactions are usually calculation from Coulomb potential using a charge probe atom. Electrostatic properties of molecules are typically described by point charges at the center of atoms. In SYBYL, the electrostatic energies are usually calculated with H^+ probe atom Coulombic interaction. The general form of electrostatic interaction between two molecules is given by

$$E = \sum_i \sum_j \frac{q_i q_j}{r_{ij}} \quad (2)$$

Where q_i, q_j = are the atomic net charges of atom *i* of drug molecule and of probe atom *j*, respectively

r_{ij} = is the distance between atom *i* of drug molecule and probe atom *j* (Å)

1.2 Interpretation of CoMFA Results

The results of CoMFA are an equation showing the contribution of energy field at each lattice point. In order to facilitate their interpretation of the results, they are also displayed as coefficient (or standard deviation time coefficient or stdev*coeff) contour plot showing the regions in space where specific molecular properties increase or decrease the potency. The results of CoMFA analyses are displayed as color-coded contours around molecules, allowing visual identification of regions responsible for favorable or unfavorable interactions with the receptor.

Steric contour plots :

Green contours indicate regions where an increase in steric bulk will enhance activity.

Yellow contours indicate regions where an increase in steric bulk will reduce activity.

Electrostatic contour plots :

Blue contours correspond to region where an increase a positive charge will enhance activity.

Red contours correspond to region where an increase a negative charge will enhance activity.

2. Partial Least Squares Analysis (PLS) and Validation

Partial least squares (PLS) methodology was used for all 3D-QSAR analyses. The CoMFA descriptors were used as independent variables and log (1/EC_{50}) values were used as dependent variables in partial least squares regression analyses to derived 3D-QSAR models using the standard implementation in the SYBYL 8.10 package. PLS analysis was carried out using the leave-one-out option to obtain the optimal number of components to be used subsequently in the final analysis as show the procedure in Figure 9. Column filtering was set to 2.0 kcal/mol to omit from the analysis lattice points whose energy variance is less than 2.0 kcal/mol. This value can

speed up the analysis and reduce the noise. The cross-validated coefficient q^2 or r_{cv}^2 was calculated using equation 3:

$$q^2 = 1 - \frac{\sum(Y_{observed} - \sum Y_{predicted})^2}{\sum(Y_{observed} - Y_{mean})^2} \quad (3)$$

Where $Y_{predicted}$, $Y_{observed}$ and Y_{mean} are predicted, actual and mean values of the target property (log ($1/EC_{50}$)), respectively. $\sum(Y_{observed} - \sum Y_{predicted})^2$ is the predictive sum of squares (PRESS). To maintain the optimum number of PLS components and minimize the tendency to over fit the data, the number of components corresponding to the lowest PRESS value was used for deriving the final PLS regression models. In addition to the q^2 or r_{cv}^2 and number of components, the conventional correlation coefficient r^2 and its standard errors (SEE) were also computed.

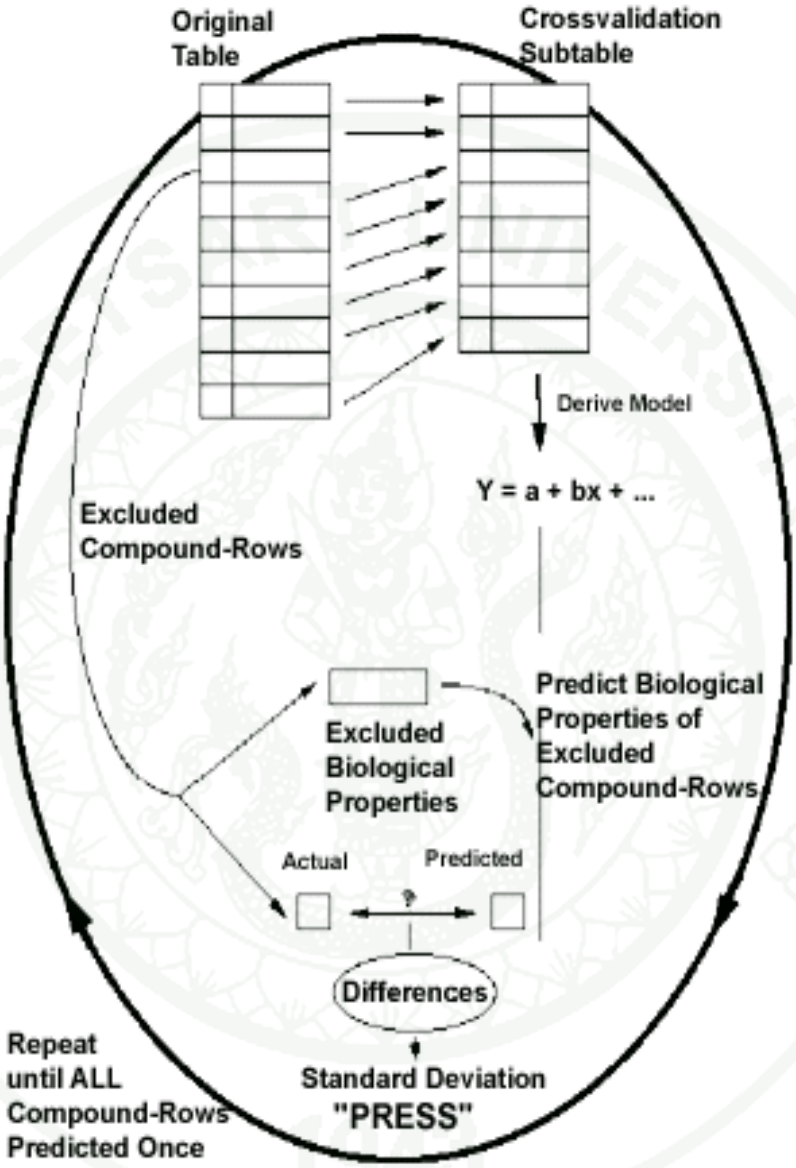


Figure 9 Cross-validation procedure Source : Kubinui (1993)

3. CoMFA Predictive Ability

The predictive ability of the model that was derived from the training set is expressed by the cross-validation predictive (r^2_{cv}) value. The r^2_{cv} value is defined as

$$r_{cv}^2 = 1.0 - \frac{PRESS}{SSY} \quad (4)$$

where, SSY is the variance of the biological activities around the mean value, and PRESS is the prediction error sum of squares derived from the LOO.

$$PRESS = \sum_y (y_{pred} - y_{actual})^2 \quad (5)$$

$$SSY = \sum_y (y_{actual} - y_{mean})^2 \quad (6)$$

The uncertainty of the prediction is defined as

$$S_{PRESS} = \sqrt{\frac{PRESS}{n-k-1}} \quad (7)$$

where *k* is the number of variables in the model and n is the number of compounds used in the study (Golbraikh, *et al*., 2002, Nilsson, *et al*., 1997, Hannongbua, *et al*., 2001).

Quantum Chemical Calculations Study

1. Model Set-up

Diarylaniline

In order to investigate specific interaction of different potency of Diarylaniline derivatives in wild-type HIV-1 RT, particular interaction was determined by quantum chemical calculations. According to comparison between good and poor inhibitor binding with wild-type complex of TMC125 (3MEC) (Eric B. Lansdon, *et al*., 2010) as shown in Figure 10.

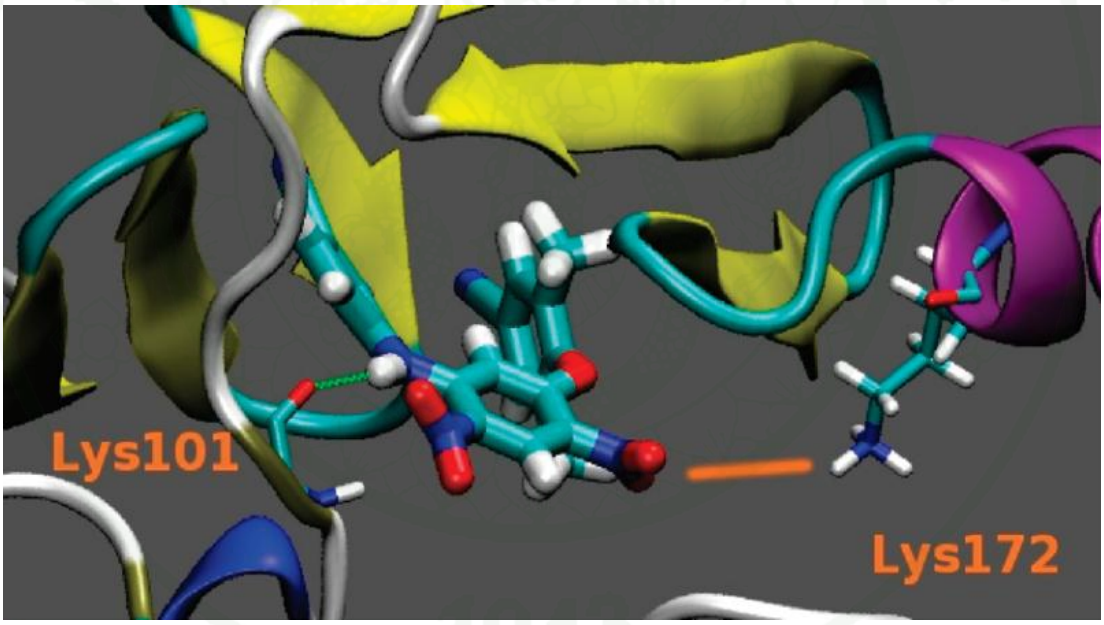


Figure 10 Compounds **24** in interaction with HIV-1 RT (source: Bingjie Qin *et. al.*, 2010)

In this study, we proposed good and poor diarylaniline derivative with wild-type complex, based on atom superposition. Considering the graphical backbone superimpostion, it can be implied that good and poor inhibitor oriented in the same binding position, therefore, inhibitor can be adapted into the K103N mutant type HIV-

1 RT to find the estimated particular interaction energy. Based on inhibitor comparison, the selected inhibitors were compounds 1 (diarylaniline drug) and 24 according to their good and poor inhibitor, which maximum and minimum pEC_{50} values (see in Table 1). Compound 1 represented a resistance to wild-type HIV-1 RT while compound 24, C≡N substituent at R_1 and R_3 NO_2 substituent at R_2 and NH_2 substituent at R_4, gave a good pEC_{50} for this enzyme. The model systems contained compounds 1 or 24 and surrounding residues in the binding pocket with at least one atom interacting with any atoms of inhibitor within the interatomic distance approximately 4 Å that covered van der Waal interactions. The 22 selected residues were Pro95, Lys100, Lys101, Lys102, Lys103, Lys104, Ser105, Val106, Val179, Ile180, Tyr181, Tyr188, Val189, Gly190, Phe227, Leu228, Trp229, Leu234, His235, Pro236, Tyr318 and Glu138(b). The one mutation Lys103Asn was also included in the system setup. The 2D scheme of the adopted model system of the inhibitor bound to the wild-type HIV-1 RT binding site is shown in Figure 11. which were retained from the backbone geometries of the nearby residues. Thus, the hydrogen atoms were added to the starting system using Sybyl8.0. Partial optimizations were performed by using the semiempirical PM3 method, implemented in the Gaussian 03 program (Frisch, *et al*., 2003) based on the 'heavy atoms fixing' approximation. Therefore, only H atoms of amino acids and all atoms of the inhibitor were optimized.

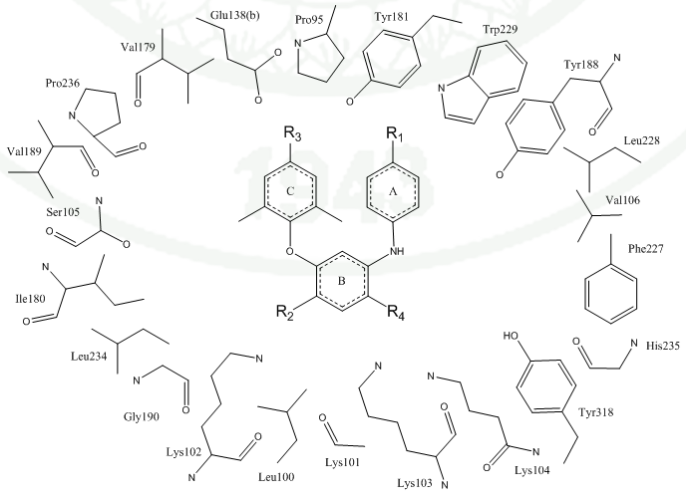


Figure 11 The 2D scheme of the adopted model system of Diarylaniline inhibitor bound to the wild type HIV-1 binding site.

2. Interaction Energy Calculations

The antiHIV-1 RT resulting geometries were used to provide different model systems of the antiHIV-1 RT inhibitor and the residues for high level of B3LYP,M06-2X and MP2 for basis set 6-31G(d,p) calculations, which then provided informations on the particular interaction energy between the inhibitor and each residue surrounding the binding site as shown in the interaction energy formula:

$$E^{INT}_{(\text{ligand–aminoacid})} = E^{AB}_{(\text{ligand–aminoacid})} - E^{A}_{(\text{ligand})} - E^{B}_{(\text{aminoacid})} \quad (8)$$

where A and B are the number of basis sets of ligands and amino acids, respectively, $E^{AB}_{(\text{ligand–aminoacid})}$ is the energy of the ligand-amino acid complex with the basis set of A plus B. $E^{A}_{(\text{ligand})}$ and $E^{B}_{(\text{aminoacid})}$ are the energies of the ligand and the amino acid with its number of basis sets.

Furthermore, the basis set superposition error based on the counterpoise scheme (BSSE-CP) of Boys-Bernardi (Boys, *et al.*, 1970) was also computed to define the interaction energy with BSSE as shown in equation 6:

$$E^{INT}_{(\text{ligand–aminoacid})} = E^{AB}_{(\text{ligand–aminoacid})} - E^{AB}_{(\text{ligand})} - E^{AB}_{(\text{aminoacid})} \quad (9)$$

where $E^{AB}_{(ligand)}$ and $E^{AB}_{(\text{aminoacid})}$ are the energies of the ligand and the amino acid, respectively, with the number of basis sets of A plus B (Saen-oon, *et al*., 2005, Kuno, *et al*., 2003 and 2006).

RESULTS AND DISCUSSION

1. CoMFA analysis

1.1 Statistical Analysis

The relationship between structural properties of twenty-five diarylaniline derivatives of HIV-1 RT is presented by using the CoMFA models. There are three models that varied the type of probe atoms, i.e. C_{sp3} (+1), O_{sp3} (-1) and N_{sp2} (-1). The statistical results are shown in Table 2. All models reveal the good prediction of pEC_{50} values for training set compounds, with the deviations lower than 0.4; these results are summarized in Table 3. Evaluation of the model prediction is assessed by abandon outlier compounds which all showed acceptable pEC_{50} prediction values, except compound 5, 8, 14 and 20.

Table 2 PLS statistical results of CoMFA models for wild type HIV-1 RT

Parameters	Probe Atoms			
	Model I C_{sp3} **(+1)**	**Model II** O_{sp3} **(-1)**	**Model III** N_{sp2} **(-1)**	**Model IV** C_{sp3} **(+1)** O_{sp3} **(-1)** N_{sp2} **(-1)**
no of molecules without outlier	21	21	21	21
r^2_{cv}	0.823	0.802	0.799	0.812
S_{press}	0.422	0.433	0.440	0.436
no of components	4	3	4	4
r^2_{nv}	0.946	0.934	0.942	0.940
s	0.241	0.250	0.241	0.245
F value	65.055	80.530	65.499	63.068
Steric field contributions	0.281	0.265	0.202	0.274
Electrostatic field contributions	0.719	0.735	0.738	0.726

Table 3 Actual (Act) and predicted (Pred) pEC_{50} values and the residuals (Δ) of the training set without outlier molecules for wild type HIV-1 RT

Compound	Act	Model I		Model II		Model III		Model IV	
	pEC_{50}	Pred pEC_{50}	Δ	Pred pEC_{50}	Δ	Pred pEC_{50}	Δ	Pred pEC_{50}	Δ
1	5.41	5.23	0.18	5.19	0.22	5.19	0.22	5.18	0.23
2	5.52	5.42	0.11	5.49	0.03	5.44	0.08	5.45	0.08
3	5.44	5.52	-0.08	5.66	-0.22	5.53	-0.09	5.55	-0.11
4	5.36	5.42	-0.06	5.45	-0.09	5.38	-0.02	5.40	-0.04
6	4.30	4.28	0.02	4.23	0.07	4.31	-0.01	4.28	0.02
7	6.26	6.35	-0.08	6.29	-0.03	6.36	-0.10	6.33	-0.07
9	6.55	6.62	-0.07	6.57	-0.02	6.58	-0.03	6.62	-0.06
10	5.81	5.97	-0.15	6.06	-0.25	5.97	-0.15	6.04	-0.22
11	6.50	6.20	0.30	6.14	0.36	6.26	0.24	6.20	0.30
12	5.50	5.87	-0.37	5.76	-0.26	5.88	-0.38	5.83	-0.33
13	6.68	6.68	-0.00	6.70	-0.02	6.65	0.03	6.65	0.03
15	5.66	5.53	0.13	5.46	0.20	5.50	0.16	5.52	0.14
16	7.33	7.08	0.25	7.07	0.25	7.15	0.18	7.10	0.23
17	7.15	7.53	-0.38	7.60	-0.44	7.53	-0.37	7.53	-0.37
18	7.14	7.04	0.09	7.01	0.12	7.10	0.04	7.07	0.07
19	6.79	6.99	-0.20	7.05	-0.25	7.02	-0.23	7.02	-0.23
21	7.52	7.44	0.08	7.56	-0.03	7.40	0.12	7.44	0.08
22	7.15	6.96	0.20	6.99	0.17	6.97	0.18	6.98	0.17
23	7.80	7.70	0.09	7.61	0.18	7.71	0.09	7.69	0.11
24	8.52	8.18	0.34	8.17	0.35	8.11	0.41	8.14	0.38
25	7.20	7.62	-0.41	7.55	-0.34	7.60	-0.39	7.62	-0.41

** compound 5, 8, 14 and 20 is outlier

By considering the statistical results in Table 2, model I-III with r^2_{cv} values higher than 0.6 (0.823, 0.802 and 0.790, respectively) can be accepted and the

conventional r^2 or no-validated r^2 (r^2_{nv}) values are found to be 0.946, 0.934 and 0.942, respectively. These mean that the four tested probe atoms (C_{sp3}, O_{sp3}, H and combine probe atom) give qualitatively very similar models. The results suggest that all four types of probe atoms form similarly important in the enzyme-ligand interactions. The best model of probe atom is model I, resulting in model I with r^2_{cv} = 0.823 and r^2_{nv} = 0.946. The carbon probe atoms in model I is superior for all model and high statistical result. Especially, the statistical error (s) of the represented model is 0.245 which is reasonably acceptable for biological activity predictions of the test set. The graphical plot between actual and predicted pEC_{50} of the set of compound is shown in Figure 12. The CoMFA field contributions of the steric interaction contribution approximate 27% and electrostatic interactions contributed approximately 73%. The results indicate that electrostatic contributions are higher effective than steric contribution affected the biological activity of wild type HIV-1 RT

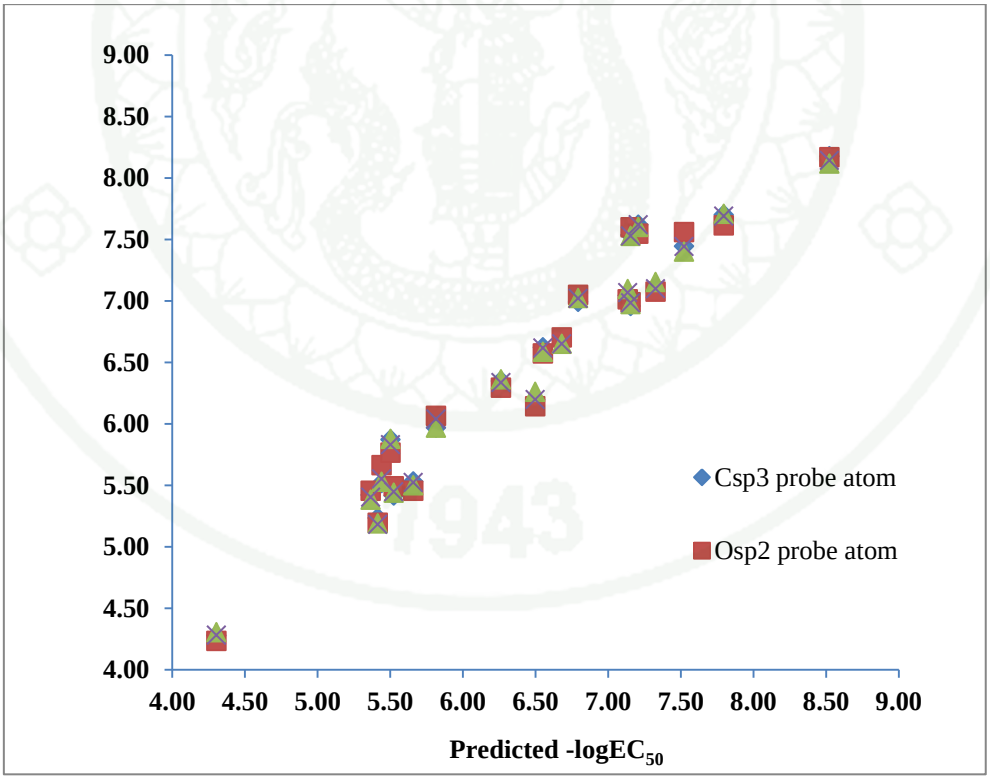


Figure 12 Plot of the predicted and actual pEC_{50} values of the test set molecules with CoMFA model I-IV

1.2 CoMFA Contour Analyses

The CoMFA analysis with hundreds or thousands terms, is usually represented as the scalar product of the associated coefficient and the standard deviation of all values in the corresponding column of the data table (STDEV*COEFF) contour plots. Moreover, the contour maps can be shown by merging with the binding pocket of a drug target. In this study, the CoMFA contour maps are merged with 4 Å of binding pocket of crystal structure of wild type HIV-1 RT which is available in the Protein Data Bank with PDB code 3MEC (Lansdon, E.B., *et al*., 2010). The template compound 37 is displayed as the inhibitor in the CoMFA contour maps.

Figure 13 shows the steric contour maps of CoMFA model I. The steric contour map indicates areas in which molecular steric bulk might have a favourable (green) or unfavourable (yellow) effect on the activity of an analogue. A sterically favoured green region is found near R_2 substituent of the aromatic ring. The location near R_2 substituent shows only a small region of favorable steric map. Therefore, based on an unclear CoMFA contour maps at R_2 and R_3, particular interactions between the partial substituent of R_2 and R_3 and the amino acids surrounding the site substitutions are needed to investigate for more understanding in the molecular interactions using the quantum chemical calculations.

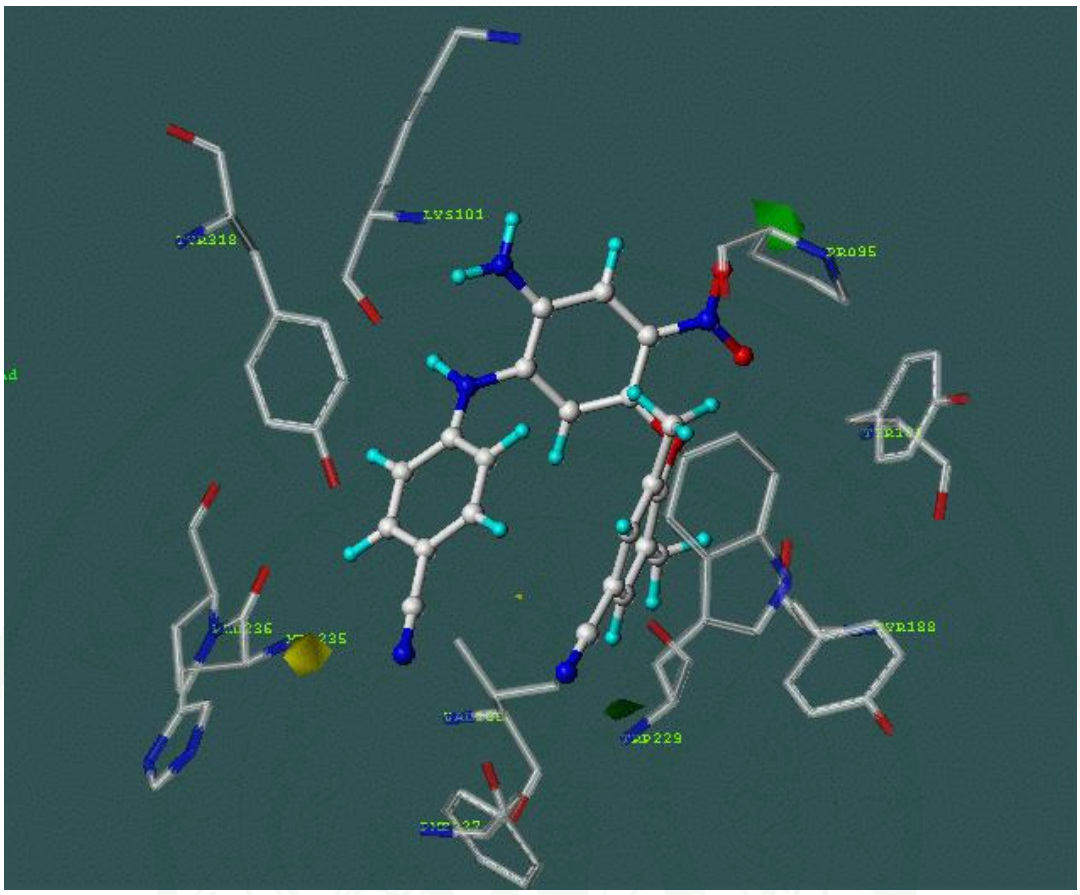


Figure 13 CoMFA (stdev.*coeff.) sterically favored areas are represented by green regions. Sterically unfavored areas are represented by yellow regions (level of steric contour contribution = 80%) and compound **37** is represented by ball and stick.

Figure 14 depicts electrostatic contour maps of CoMFA. The electrostatic contour map reveals that blue contours refer to positive charge favoured areas and red contours indicate negative charge favoured areas. The red and blue areas are found in the R_1 substitution. This is further supported by comparing R_1 substituent with CN and Me when these compounds have the same R_2, R_3 and R_4 substituents. In addition, the distribution of electrostatic contour appears around the R_1 substituent; this evidence would explain why compound 37, used as the template, is a better wild type HIV-1 RT inhibitor. An unfavorable electrostatic contour region is found at R_4 substitution which can explain the fact that compound 6, 8, 11 and 13, for all

compound R_4 substituent are electron with drawing groups, show lower pEC_{50} when compare with unfavorable electrostatic structure of compounds 16, 23 and 24 , for all compound R_4 substituent are electron with donating groups, respectively. It can be suggested R_1 that high positive charges or low electron density in this area is preferable. For R_4 substituent, donating substituents will increase the activity of the inhibitors, for example, NR_2, O^- etc.

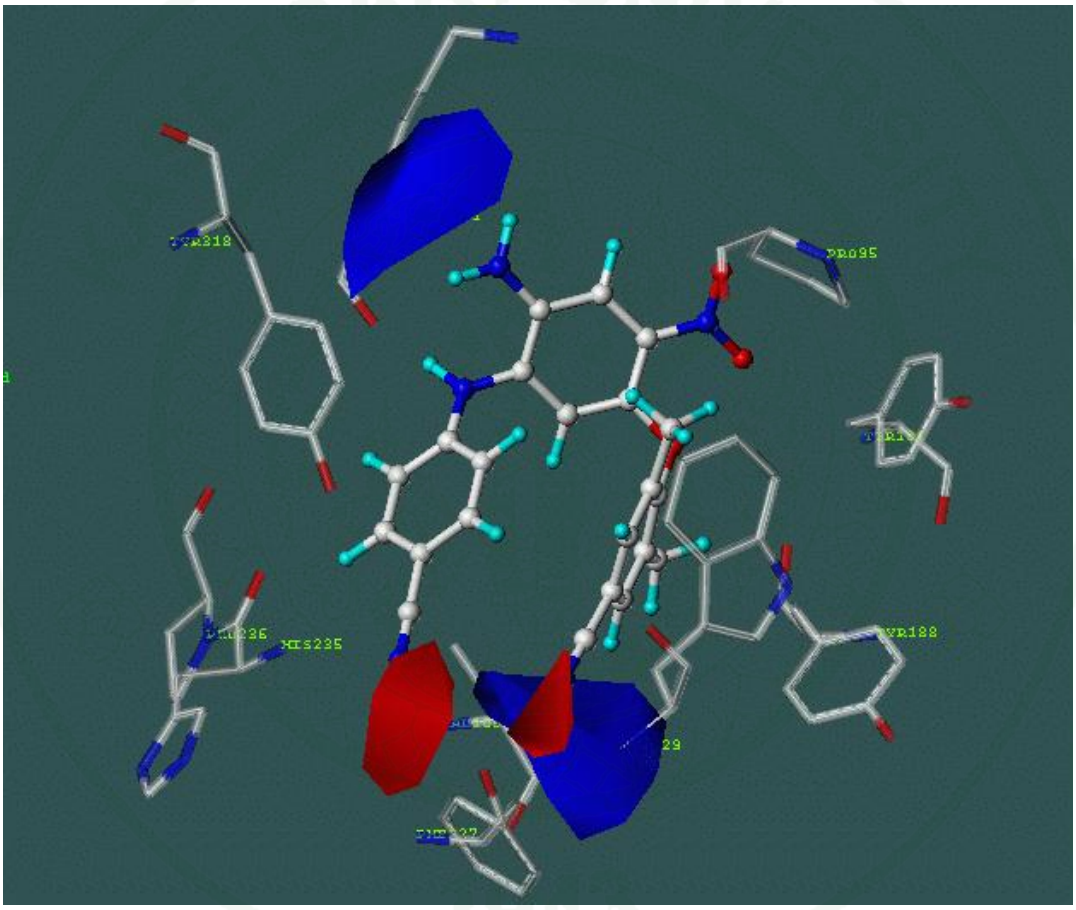


Figure 14 CoMFA (stdev.*coeff.) negative charge favored area is represented by the red region. Positive charge favored area is represented by the blue region (level of electrostatic contour contribution = 80%) and compound **24** is represented by ball-and-stick model.

2. Quantum Chemical Calculation

Particular interaction energy

In order to find the particular interaction energy between compound 1 or 24 and the amino acids surrounding the pocket of wild type were performed by B3LYP, M062X and MP2 methods with the 6-31G(d,p) basis set. In addition, the basis set superposition error (BSSE) using the counterpoise (CP) correction method was applied to calculate the interaction energy. The interaction energies are shown in Table 4. The difference between the interaction energies with no CP correction and those with CP correction are about 1-5 kcal/mol.

Figure 15 shows the interaction energies with CP correction of diarylaniline and individual amino acids surrounding the binding pocket of wild type from B3LYP, M062X, MP2 calculations with 6-31G(d,p) basis set. The main interactions are considered. His235 shows the strongest interaction energy to compound 1 and 24 with -24.37 and -16.60 kcal/mol (B3LYP), -19.70 and -13.97 kcal/mol (M062X) and -12.64 and -8.33 kcal/mol (MP2), respectively. In addition, Lys101 also reveals the strong attractive interaction to compound 1 and 24, -7.01 and -3.01 kcal/mol (B3LYP), -2.46 and -4.51 kcal/mol (M062X) and -3.42 and -5.84 kcal/mol (MP2), respectively. Lys103 shows the hydrogen bond interaction energy to compound 1 and 24 with -4.65 and -0.38 kcal/mol (B3LYP), -2.16 and -1.87 kcal/mol (M062X) and -3.57 and -1.80 kcal/mol (MP2), respectively. Glu138(b) shows the hydrogen bond interaction energy to compound 1 and 24 with -0.16 and -1.23 kcal/mol (B3LYP), -5.69 and -7.32 kcal/mol (M062X) and -3.67 and -2.94 kcal/mol (MP2), respectively. In the case of Tyr181, have a hydrogen bond and pi-pi interaction energy to compound 24 with -1.56 kcal/mol (B3LYP), -4.36 kcal/mol (M062X) and -2.64 (MP2), respectively. From the results indicated that the hybrid functional theory, B3LYP and M062X calculations, are similar with MP2 calculations. It is show that the B3LYP and M062X methods can use for this system.

Table 4 Particular interaction energy (kcal/mol) of compound 1 and 24 with individual residues, calculated by B3LYP/6-31G(d,p), M062X/6-31G(d,p) and MP2/6-31G(d,p) methods

Residue	Interaction energies non-BSSE-CP (kcal/mol)					
	Compound 1			Compound 24		
	B3LYP	M062X	MP2	B3LYP	M062X	MP2
Pro95	-0.63	-1.79	-0.16	-0.34	-0.93	0.36
Leu100	0.74	0.86	2.36	1.59	-1.34	2.54
Lys101	-11.57	-6.50	-5.76	-3.11	-5.42	-7.39
Lys102	-0.19	-0.81	-0.29	-2.91	-3.12	-1.40
Lys103	-5.84	-4.30	-3.95	-1.93	-3.39	-3.00
Lys104	-0.42	-0.53	-0.61	2.37	1.77	0.12
Ser105	0.07	0.09	0.09	-0.25	-0.59	-0.75
val106	0.82	0.65	2.06	0.50	-0.69	0.29
Val179	-0.86	-1.41	-0.23	-0.97	-1.50	-2.22
Ile180	1.32	0.75	2.23	1.09	-1.23	0.45
Tyr181	-0.12	-5.54	1.17	-0.59	-6.67	-3.82
Tyr188	3.00	-2.04	4.80	-3.06	-2.36	-0.91
Val189	-0.47	-0.58	-0.57	0.54	0.08	0.19
Gly190	0.08	-0.06	0.06	0.095	-0.07	0.04
Phe227	-0.55	-1.34	-0.20	-5.52	-2.19	-0.60
Leu228	0.12	0.04	0.12	0.02	-0.08	-0.0003
Trp229	-0.87	-4.43	-0.70	-1.53	-6.68	-1.22
Leu234	-0.82	-1.91	0.26	1.19	-2.30	0.70
His235	-26.03	-20.90	-13.83	-19.30	-15.32	-10.42
Pro236	-0.76	-1.20	-0.26	-0.63	-5.46	-4.03
Tyr318	-0.77	-1.13	-0.13	-1.17	-3.89	-0.62
Glu138(CHAIN B)	-6.77	-10.65	-4.13	-6.08	-11.39	-3.94

Table 5 Particular interaction energy (kcal/mol) of compound 1 and 24 with individual residues, calculated by B3LYP/6-31G(d,p), M062X/6-31G(d,p) and MP2/6-31G(d,p) with BSSE-CP

Residue	Interaction energies BSSE-CP (kcal/mol)					
	Compound 1			Compound 24		
	B3LYP	M062X	MP2	B3LYP	M062X	MP2
Pro95	0.10	-1.20	0.25	0.12	-0.72	0.51
Leu100	3.21	3.06	3.06	3.52	0.87	3.35
Lys101	-7.01	-2.46	-3.42	-3.01	-4.51	-5.84
Lys102	0.05	-0.56	-0.10	-2.56	-2.93	-1.13
Lys103	-4.65	-2.16	-3.57	-0.38	-1.87	-1.80
Lys104	-0.42	-0.53	-0.61	2.36	1.75	0.12
Ser105	0.06	0.09	0.09	-0.26	-0.60	-0.75
val106	1.90	1.52	2.80	1.58	-0.33	1.22
Val179	0.54	0.12	-0.23	0.24	-1.06	-1.39
Ile180	2.05	1.19	2.80	1.46	0.01	1.54
Tyr181	2.49	-3.06	1.18	-1.56	-4.36	-2.64
Tyr188	5.26	-0.97	4.87	0.74	-0.62	0.87
Val189	-0.40	-0.54	-0.49	0.60	0.13	0.23
Gly190	0.09	-0.05	0.07	0.12	-0.07	0.05
Phe227	0.68	0.22	-0.19	-2.52	-1.36	-0.16
Leu228	0.12	0.05	0.12	0.01	-0.07	0.01
Trp229	0.09	-3.06	-0.35	-0.71	-4.86	-0.70
Leu234	1.49	0.20	2.00	3.80	-1.22	1.74
His235	-24.37	-19.70	-12.64	-16.59	-13.97	-8.33
Pro236	0.56	-0.30	-0.30	1.28	-3.23	-2.11
Tyr318	0.70	0.54	1.02	0.28	-2.56	0.71
Glu138(CHAIN B)	-0.16	-5.69	-3.67	-1.23	-7.32	-2.94

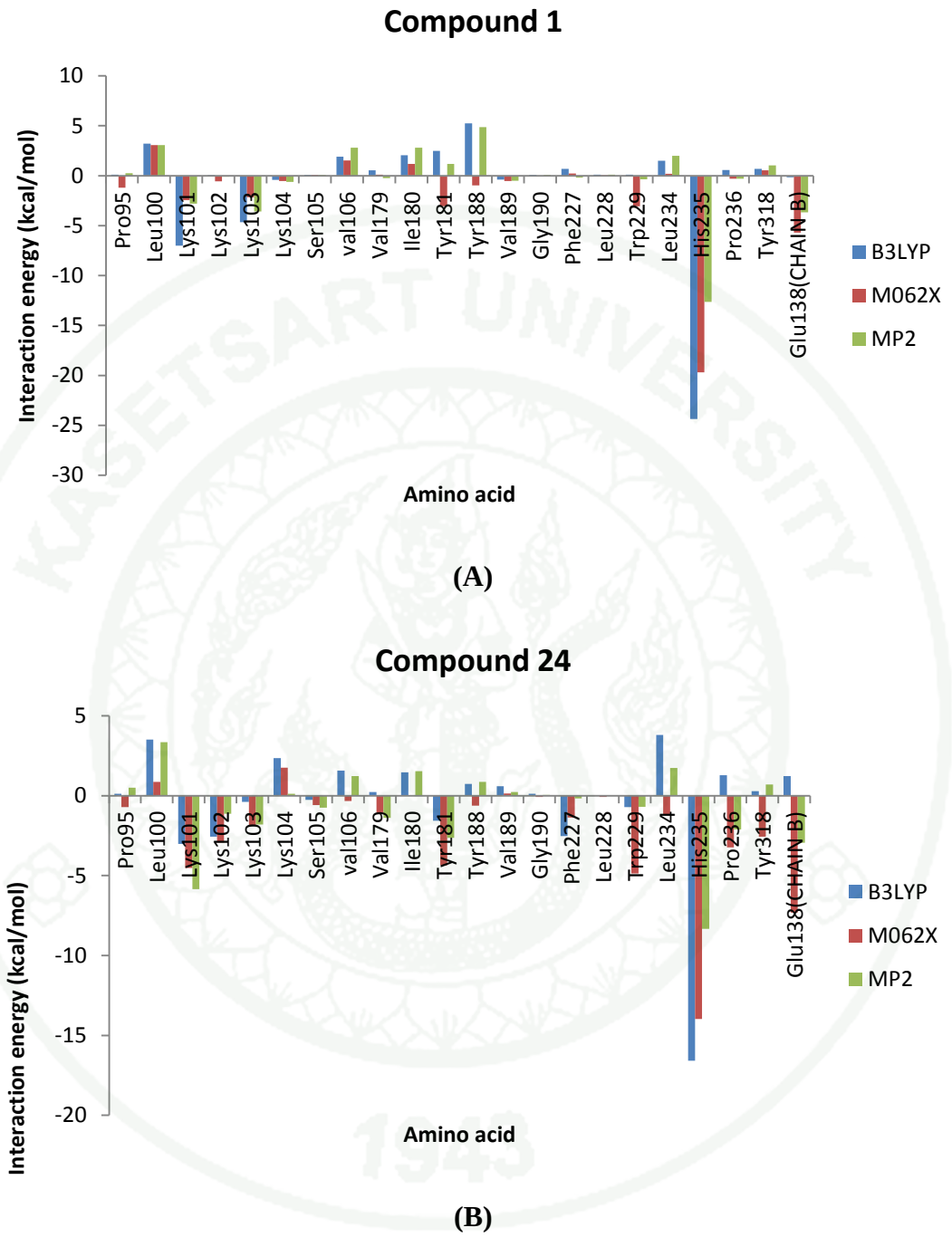
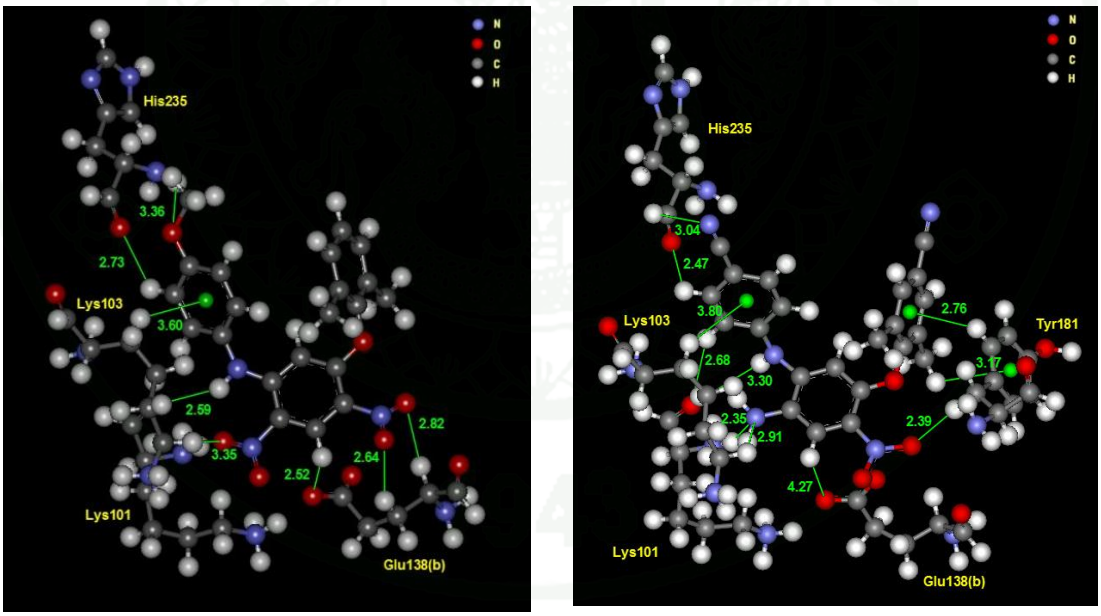


Figure 15 Interaction energies with CP correction of diarylaniline and individual amino acids surrounding the binding pocket of wild type at B3LYP, M062X, MP2 methods with 6-31G(d,p) basis set; (A) compound **1** and (B) compound **24**.

Significant H-bond interactions between the inhibitors and amino acids are displayed in Figure 16 and explained as follows. In molecular-level investigation, His235 forms hydrogen-bonding interaction with OMe of compound 1 and CN of compound 24 which are R_1 substituent in ring A. In Compound 1, Lys103 forms hydrogen-bonding interaction with NO_2 group of R_4 substitution and pi-pi interaction with ring A. Lys103 forms hydrogen-bonding interaction with NO_2 group of R_4 substitution and formed pi-pi interaction with ring A. Glu138(b) forms hydrogen-bonding interaction with NO_2 of R_2 substitution. For Compound 24, Lys103 forms hydrogen-bonding interaction with NH_2 group of R_4 substitution and pi-pi interaction with ring A. Glu138(b) forms weak hydrogen-bonding with hydrogen of C15 in ring B. Tyr181 forms hydrogen-bonding interaction with NO_2 of R_2 substitution and O2 of compound 24. Pi-pi interaction between phenyl ring of Tyr181 and ring C is also found. Trp229 forms pi-sigma with hydrogen of C4.

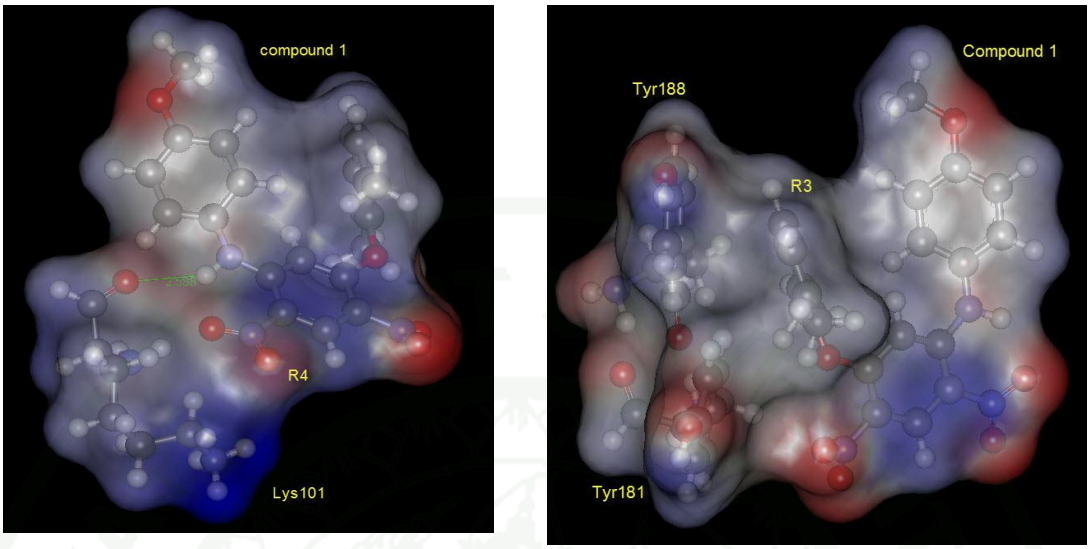
Compound 1 and compound 24 have a nitro group as the R_2 substituent on the ring B. These nitro group provides H-bond with Glu138(b) formed hydrogen-bonding interaction between hydrogen of carbon back bone with NO_2 and NO_2 of Glu138(b) with hydrogen of C15 of R_2 substitution of compound 1. Glu138(b) formed hydrogen-bonding between NO_2 of Glu138(b) with hydrogen of C15 in ring B of compound 24. For Bingjie Qin and co-worker (*Bingjie Qin et. al.(2010)*) study about R_2 substituent, the nitro group of R_2 substituent provides a small electrostatic interaction with positive charge Lys172 but R_2 replace by an amino group NH_2 , the corresponding ligand-protein electrostatic interaction decrease. For R_3 substituent, the most active compound 24 also has a CN group as the R_3 substituent on the ring. This group has a provide H-bond between NO_2 of R_2 substitution and O2 of compound 24. For pi-pi interaction, have pi pi interaction between benzene ring of Tyr181 with ring C of compound 24 and pi-sigma between ring of Trp229 with hydrogen of C4 Typ229. Compound 1 has a repulsive interaction energy with ring C of compound 1. This group has a repulsive interaction energy for Tyr181 is 1.18 kcal/mol and Tyr188 is 4.87 kcal/mol. But the M062X is in contrast to the results of MP2 as this may be due to that calculated by our system, a single point may not match the results. Because our structure does not move and may be associated with the amino acid to another. If you

need to know exactly what the system is more complicate than the calculate with ONIOM method. For R_4 substitution, compound 24 has a NH_2 group as the R_4 substituent, a hydrogen bond between the NH linker and the peptidic carbonyl oxygen of Lys101, like the DAPY compounds. However, the neighboring amino group (R_4) present in compound 24 provide two more hydrogen bonds with Lys101: one is between the peptidic carbonyl oxygen of the protein residue and hydrogen of N13 of compound 24 and the second one involves the NH atoms of Lys101 and targets the nitrogen atom of the ligand NH_2 group for R_4 substitution of compound 24. For compound 1, has a NO_2 group as the R_4 substituent, have a hydrogen bond between a carbonyl oxygen of the protein residue and the NH linker. Bingjie Qin and co-worker (*Bingjie Qin et. al.(2010)*) study about R_4 be effective electrostatic repulsion occurs between nitro group and NH group of amino acid Lys101. The electrostatic repulsive of R_3 and R_4 substitutent are displayed in Figure 17.



(A) **(B)**

Figure 16 Bond distances between inhibitor and residues in the binding pocket; (A) compound 1 and (B) compound 24 (in Å).



(A) **(B)**

Figure 17 The electrostatic potential is shown on the solvent accessible surface as red for negative and blue for positive values for compound 1 interacted with (A) Tyr181 and Tyr188 in R_3 substituent (B) and Lys101 in R_4 substituent.

CONCLUSION

Particular Interaction between Diarylaniline Derivatives and Wild Type of HIV-1 Reverse Transcriptase: CoMFA and Quantum Chemical Calculations Studies

The CoMFA analysis is a very powerful method for ligand-based drug design. Therefore, in this study, the CoMFA method was selected to build a linear equation of the quantitative structure activity relationship of the diarylaniline derivatives that are active against quadruple wild type HIV-1 RT. The C_{sp3} (+1) probe atoms model was selected to represent the CoMFA moelcular fields for accounting the different types of interactions between wild type HIV-1 RT binding site and diarylaniline derivatives. An application of the CoMFA technique was performed on diarylaniline derivatives of the wild type HIV-1 RT. Satisfactory CoMFA models of wild type was obtained with LOO cross-validation r^2_{cv} values of 0.823. According to the CoMFA contour map, the electrostatic property plays an important role around three substitutions, R_1, R_3 and R_4 which can be concluded that (i) R_1 is favored the negative charge group or electron withdrawing group. R_1 subtitution is near His235 and have a strong hydrogen bond with His235, (ii) R_3 has both negative and positive regions together which means that this site can be substituted by both withdrawing and donating groups. For quantum calculation, R_3 substitution, the most active compound 24 also has a CN group as the R_3 substituent on the ring. This group has a provides H-bond with Tyr181, pi-pi interaction with Tyr181 and Tyr188 and pi-sigma with Typ229 but compound 1 also has a hydrogen as the R_3 substituent on the ring. This group has a repulsive interaction energy for Tyr181 and Tyr188., (iii) R_4 is favored the electron donating group more than electron withdrawing group. For quantum calculation, For R_4 substitution, compound 24 has a NH_2 group as the R_4 substituent provide two more hydrogen bonds with Lys101 but compound 1 has a NO_2 group a hydrogen bond between a carbonyl oxygen of the protein residue and the NH linker. whereas, the R_2 shows only a small region of favorable steric map. Therefore, the characteristics of new design inhibitors are the bulky group on the R_2 substitution, the electron

withdrawing group on the R_1 substitution and the electron donating group on the R_4 substitution.

Moreover, we also performed MP2/6-31G(d,p) quantum chemical calculations with BSSE-CP energy correction to investigate the particular interaction energy of compounds 1 (R_1: OMe, R_2: NO_2 R_3: H and R_4: NO_2) and 24 (R_1: CN, R_2: NO_2 R_3: CN and R_4: NH_2). The obtained results clearly show that for R_1, R_2, R_3 and R_4 substitution caused result of Particular interaction show main interaction between wild type binding pocket with compound 1 and 24. The CoMFA and particular interaction energy analyses will be useful for identifying the structural features of potent diarylaniline derivatives active against wild type HIV-1 RT which is an important target of AIDs.

LITERATURE CITED

Andries, K., Azijn H., Thielemans T., Ludovici D., Kukla M., Heeres J., Janssen P., De Corte B., Vingerhoets J., Pauwels R. and de Béthune M.-P. 2004. TMC125, a novel next-generation nonnucleoside reverse transcriptase inhibitor active against nonnucleoside reverse transcriptase inhibitor-resistant human immunodeficiency virus type 1. **Antimicrob. Agents Chemother**. 48, 4680-4686.

Baba M., Tanaka H., De Clercq, E., Pauwels R., Balzarini J., Schols D., Nakashima, H., Perno C.-F., Walker R. T. and Miyasaka, T. 1989. Highly specific inhibition of human immunodeficiency virus type 1 by a novel 6-substituted acyclouridine derivative. **Biochem. Biophys. Res. Commun**. 165, 1375-1381.

Binjie Q., Xingkai J., Hong Lu., Xingtao T., Florent B., Li H., Keduo Q., Chin-Ho, C., Rong H., Shibo J., Kuo-Hsiung L. and Lan X. 2010. Diarylaniline derivative as a distinct class of HIV-1 non-nucleoside reverse transcriptase inhibitors. **J. Med. Chem.** 53, 4906-4916.

Brown, R.D. and Y.C. Martin. 1997. The information content of 2D and 3D structural descriptors relevant to ligand-receptor binding. **J. Chem. Inf. Comput. Sci.** 37: 1-9.

Bush, B. L. and R. B. Jr. Nachbar. 1993. Sample-distance Partial Least Squares: PLS optimized for many variables, with application to CoMFA. **J. Comput. Aided Mol. Des.** 7: 587-619.

Carr, A. 2003. Toxicity of antiretroviral therapy and implications for drug development. **Nat. Rev. Drug Discovery** 2003, 2, 624-634.

Case, D. A., T. A. Darden, III. T. E. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, K. M. Merz, D. A. Pearlman, M. Crowley, R. C. Walker, W. Zhang, B. Wang, S. Hayik, A. Roitberg, G. Seabra, K. F. Wong, F. Paesani, X. Wu, S. Brozell, V. Tsui, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, P. Beroza, D. H. Mathews, C. Schafmeister, W. S. Ross and P. A. Kollman. 2006. **AMBER9.0**.

Cramer, R. D. III, D. E. Patterson and J. D. Bunce. 1988. Comparative molecular field analysis CoMFA: 1. Effect of shape on binding of steroids to carrier proteins. **J. Am. Chem. Soc**. 110: 5959-5967.

Cramer, R. D. III, D. E. Patterson and J. D. Bunce. 1989. Recent advances in comparative molecular field analysis (CoMFA). **Prog. Clin. Biol. Res**. 291: 161-165.

Cho, S. J. and A. Tropsha. 1995. Cross-validated R^2-guided region selection for comparative molecular field analysis: A simple method to achieve consistent results**. J. Med. Chem**. 38: 1060-1066.

Clark, D.E. and D.R. Westhead. 1996. Evolutionary algorithms in computer-aided molecular design. **J. Comput. Aided Mol. Des**. 4: 337-358.

Clark, M. and R. D. III. Cramer. 1993. The probability of chance correlation using partial least squares (PLS). **Quant. Struct.-Act. Relat**. 12(2): 137-145.

Cornell, W. D., P. Cieplak, C. I. Bayly and P. A. Kollman. 1993. Application of RESP charges to calculate conformational energies, hydrogen-bond energies, and free-energies of solvation, **J. Am. Chem. Soc**. 115, 9620-9631.

Davey Jr., R.T., Chaitt, D.G., Reed, G.F., Freimuth, W.W., Herpin, B.R., Metcalf, J.A., Eastman, P.S., Falloon, J., Kovacs, J.A., Polis, M.A., Walker, R.E., Masur, H., Boyle, J., Coleman, S., Cox, S.R., Wathen, L., Daenzer, C.L., Lane, H.C., 1996. Randomized, controlled phase I/II trial of combination therapy with delavirdine (U-90152S) and conventional nucleosides inhumanimmunodeficiency virus type 1-infected patients. **Antimicrob. Agents Chemother**. 40, 1657-1664.

Das K., Clark J., Arthur D., Lewi P.J., Heeres J., de Jonge, M.R., Koymans, L.M.H., Vinkers, H.M., Daeyaert F., Ludovici D.W., Kukla M.J., De Corte B., Kavash, R.W., Ho, C.Y., Ye H., Lichtenstein, M.A., Andries K., Pauwels R., de Béthune M.-P., Boyer P.L., Clark P., Hughes S.H., Janssen, P.A.J. and Arnold E. 2004. Roles of conformational and positional adaptability in structure-based design of TMC125-R165335 (etravirine) and related non-nucleoside reverse transcriptase inhibitors that are highly potent and effective against wild-type and drugresistant HIV-1 variants. **J. Med. Chem**. 47, 2550-2560.

Das K., Lewi, P.J., Hughes, S.H., Arnold E. 2005. Crystallography and the design of anti-AIDS drugs: conformational flexibility and positional adaptability are important in the design of non-nucleoside HIV-1 reverse transcriptase inhibitors. **Prog. Biophys. Mol. Biol**. 88, 209-231.

De Clercq E. 2004. Non-nucleoside reverse transcriptase inhibitors (NNRTIs): past, present, and future. **Chem. Biodiversity**. 1, 44-64.

De Clercq E. 2002. Strategies in the design of antiviral drugs. **Nat. Rev. Drug Discovery**. 1, 13–25.

Donald W. Ludovici, Bart L. De Corte, Michael J. Kukla, Hong Ye, Chih Y. Ho, Mark A. Lichtenstein, Robert W. Kavash, Koen Andries, Marie-Pierre de Bethune, Hilde Azijn,c Rudi Pauwels, Paul J. Lewi, Jan Heeres, Lucien M. H. Koymans, Marc R. de Jonge, Koen J. A. Van Aken, Frederik F. D. Daeyaert,d Kalyan Das,Edward Arnolde and Paul A. J. Janssend., 2001. Evolution of Anti-HIV Drug Candidates. Part 3: Diarylpyrimidine (DAPY)Analogues. **Bioorg. Med. Chem. Lett**. 11, 2235-2239.

Dueweke T.J., Poppe S.M., Romero D.L., Swaney S.M., So A.G., Downey K.M., Althaus I.W., Reusser F., Busso M., Resnick L., Mayers D.L., Lane J., Aristoff P.A., Thomas R.C. and Tarpley W.G., 1993a. U-90152, a potent inhibitor of human immunodeficiency virus type 1 replication. **Antimicrob. Agents Chemother**. 37, 1127-1131.

Esnouf R.M., Ren J., Hopkins A.L., Ross C.K., Jones E.Y., Stammers D.K. and Stuart D.I., 1997. Unique features in the structure of the complex between HIV-1 reverse transcriptase and the bis(heteroaryl)piperazine (BHAP) U-90152 explain resistance mutations for this nonnucleoside inhibitor. **Proc. Natl. Acad. Sci**. U. S. A. 94, 3987-3989.

Frisch, M. J., G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Jr. Montgomery, T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R.

L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez and J. A. Pople. 2003. **Gaussian 03**, revision B.05;Gaussian, Inc.: Pittsburgh.

Guillemont J, Pasquier E, Palandjian P, Vernier D, Gaurrand S, Lewi PJ, Heeres J, de Jonge MR, Koymans LM, Daeyaert FF, Vinkers MH, Arnold E, Das K, Pauwels R, Andries K, de Bethune MP, Bettens E, Hertogs K, Wigerinck P, Timmerman P and Janssen PA. 2005. Synthesis of novel diarylpyrimidine analogues and their antiviral activity against human immunodeficiency Virus type 1. **J. Med. Chem**. 48:2072.

Golbraikh, A., P. Bernard and J. R. Chretien. 2000. Validation of protein-based alignment in 3D quantitative structure-activity relationship with CoMFA models. **Eur. J. Med. Chem**. 35: 123-136.

Hannongbua, S., K. Nivesanond, L. Lawtrakul, P. Pungpo and P. Wolschann. 2001. 3D-quantitative structure-activity relationships of HEPT derivatives as HIV-1 reverse transcriptase inhibitors, based on Ab initio calculations. **J. Chem. Inf. Comput. Sci.** 41: 848-855.

Hansch, C. and A. Leo. 1995. Exploring QSAR. fundamentals and applications in chemistry and biology. *American Chemical Society, Washington, DC*.

Hansch, C and T. Fujita. 1964. A method for the correlation of biological activity and chemical structure. **J. Am. Chem. Soc**. 86: 161-1626.

Havlir D., Cheeseman S.H., McLaughlin M., Murphy R., Erice A., Spector S.A., Greenough T.C., Sullivan J.L., Hall D., Myers M., Lamson M. and Richman D.D., 1995. High-dose nevirapine: safety, pharmacokinetics, and antiviral effect in patients with human immunodeficiency virus infection. **J. Infect. Dis**. 171, 537-545.

Himmel D. M., Das K., Clark A. D., Hughes S. H., Benjahad A., Oumouch S., Guillemont J.; Coupa S., Poncelet A., Csoka I., Meyer C., Andries K., Nguyen C. H., Grierson D. S., and Arnold E. 2005. Crystal structures for HIV-1 reverse transcriptase in complexes with three pyridinone derivatives: a new class of non-nucleoside inhibitors effective against a broad range of drug-resistant strains. **J. Med. Chem**. 48, 7582-7591.

Huang, H., Chopra, R., Verdine, G.L., Harrison, S.C., 1998. Structure of a covalently trapped atalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. **Science** 282, 1669-1675.

Jacobo-Molina A., Ding J., Nanni R.G., Clark Jr. A.D., Lu X., Tantillo C., Williams R.L., Kamer G., Ferris A.L., Clark P., Hizi A., Hughes S.H. and Arnold E., 1993. Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0Å shows bent DNA. **Proc. Natl. Acad. Sci.** U. S. A. 90, 6320-6324.

Kitchen C. M., Kitchen S. G., Dubin J. A. and Gottlieb M. S. 2001. Initial virological and immunologic response to highly active antiretroviral therapy predicts long-term clinical outcome. **Clin. Infect. Dis**. 33, 466-472.

Kim, K. H., G. Greco and E. Novellino. 1998. A Critical Review of Recent CoMFA Applications. **Perspect. Drug Discov. Des.** 257-315.

Kokpol, S.K., S.V. Hannongbua, N. Thongrit, S. Polman, B.M. Rode and M.G. Schwendinger. 1988. Analysis of structure-activity relation for primaquine antimatlarial drugs by a quantum pharmacological approach. **Anal. Sci**. 4: 565-568.

Kubinyi, H. 1993. QSAR: Hansch analysis and related approaches; *VCH: Weinheim*.

Kubinyi, H., G. Flokers and Y. C. Martin. 1998. 3D QSAR in drug design: ligand-protein interactions and molecular similarlity. **Perspect. Drug Discov. Des**. 3-398.

Kuno, M., R. Hongkrengkai and S. Hannongbua. 2006. ONIOM-BSSE scheme for H--¶ system and applications on HIV-1 reverse transcriptase. **Chem. Phys. Lett**. 424: 172-177.

Kuno, M., S. Hannongbua and K. Morokuma. 2003. Theoretical investigation on nevirapine and HIV-1 reverse transcriptase binding site interaction, based on ONIOM method. **Chem. Phys. Lett.** 380: 456-463.

Marshell, G. R. and R. D. III, Cramer. 1988. Three-dimensional structure-activity relationships. **Trends Pharmacol. Sci.** 9(8): 285-289.

Martinez-Picado J., De Pasquale, M. P., Kartsonis, N., Hanna, G. J., Wong J., Finzi D., Rosenberg E., Gunthard H. F., Sutton L., Savara A., Petropoulos C. J., Hellmann N., Walker B. D., Richman D. D., Siliciano R. and D'Aquila, R. T. 2000. Antiretroviral resistance during successful therapy of HIV type 1 infection. Proc. **Natl. Acad. Sci.** U.S.A. 97, 10948-10953.

Meadows D. C. and Gervay-Hague J. 2006. Current developments in HIV chemotherapy. **Chem. Bio. Chem.** 1, 16-29.

Nilsson, J., S. De Jong and A. K. Smilde. 1997. Multilinear PLS analysis application to 3D QSAR data set. **J. Chemometrics**. 11: 511-524.

Nunrium, P., M. Kuno, S. Saen-oon and S. Hannongbua. 2005. Particular interaction between efavirenz and the HIV-1 reverse transcriptase binding site as explained by the ONIOM2 method. **Chem. Phys. Lett.** 405: 198-202.

Pauwels R. 2004. New non-nucleoside reverse transcriptase inhibitors (NNRTIs) in development for the treatment of HIV infections. **Curr. Opin. Pharmacol**. 4, 437-446.

P. Maitarad, S. Kamchonwongpaisan, J. Vanichtanankul, T. Vilaivan, Y. Yuthavong, and S. Hannongbua. 2009. Interactions between cycloguanil derivatives and wild type and resistance-associated mutant *Plasmodium falciparum* dihydrofolate reductases. **J. Comput Aided Mol Des.** 23: 241-252

Redl, G., R. D. III. Cramer and C. E. Berkoff. 1974. Quantitative drug design. **Chem. Soc. Rev.** 3(3): 273-292.

Saen-oon, S., M. Kuno and S. Hannongbua. 2005. Binding energy analysis for wild-type and Y181C mutant HIV-1 RT/8-Cl TIBO complex structures: quantum chemical calculations based on the ONIOM method. **PROTEINS: Structure, Function, and Bioinformatics**. 61: 859-869.

Saen-oon, S., O. Aruksakunwong, K. Wittayanarakul, P. Sompornpisut and S. Hannongbua. 2007. Insight into analysis of interactions of saquinavir with HIV-1 protease in comparison between the wild-type and G48V and G48V/L90M mutants based on QM and QM/MM calculations. **J. Mol. Graph. Model**. 26: 720–727.

Sao, S. S. and M. Karplus. 2001. Evaluation of designed ligands by a multiple screening method: Application to glycogen phosphorylase inhibitors constructed with a variety of approaches. **J. Comput. Aided Mol. Des.** 15: 613-647.

Staszewski, S., Morales-Ramirez, J., Tashima, K.T., Rachlis, A., Skiest, D., Stanford, J., Stryker, R., Johnson, P., Labriola, D.F., Farina, D., Manion, D.J., Ruiz, N.M., 1999. Efavirenz plus zidovudine and lamivudine, efavirenz plus indinavir, and indinavir plus zidovudine and lamivudine in the treatment of HIV-1 infection in adults. **N. Engl. J. Med.** 341, 1865-1873.

Svante, W. and M. Sjoestroem. 1986. Linear free energy relationships. Local empirical rules - or fundamental laws of chemistry. A reply to Kamlet and Taft. **Acta Chemica Scandinavica, Series B: Organic Chemistry and Biochemistry.** B40(4): 270-7.

SYBYL Molecular Modelling Softwares, Version 7.0, Tripos Associates, Inc., St. Louis, MO, 63144, USA, 1998.

Valenti, W. M. 2001. HAART is cost-effective and improves outcomes. **AIDS Reader** 11, 260–262.

Vingerhoets, J., Azijn, H., Fransen, E., De Baere, I., Smeulders, L., Jochmans, D., Andries, K., Pauwels, R., de Béthune, M.-P., 2005. TMC125 displays a high genetic barrier to the development of resistance: evidence from in vitro selection experiments. **J. Virol**. 79, 12773-12782.

APPENDICES

Appendix A

Supporting information

Theoretical Background in Quantum Chemistry

Molecular Orbital Theory

Molecular orbital calculation is the important method in quantum chemistry for approximate structures and dynamics of molecular system. This approach provides a great promise in calculating electronic structures and predicting properties of drug molecules. Until now, molecular orbital investigations have been introduced into drug research to study mechanisms of action and to guide the design of more potent agents.

The quantum chemical methods are based on finding solutions to the time independent Schrödinger wave equation on molecular orbital theory

$$H\Psi = E\Psi \tag{10}$$

Where H is the Hamiltonian operator which gives the kinetic and potential energies of the system

$$H = -\frac{\bar{h}^2}{2m}\nabla^2 + V \tag{11}$$

Then, rewrite equation (11) is;

$$\left\{\frac{-\bar{h}^2}{2m}\nabla^2 + \mathrm{V}\right\}\Psi = E\Psi \tag{12}$$

Where

$$\nabla^2 = \frac{\partial^2}{\partial \mathrm{x}^2} + \frac{\partial^2}{\partial \mathrm{y}^2} + \frac{\partial^2}{\partial \mathrm{z}^2} \tag{13}$$

h is Plank's constant divided by 2π. Ψ is the wave function which characterizes the particle's proterties. E is the energy of the particle.

1. The LCAO-MO Approximation

For a molecular system, the approximate molecular orbitals Ψ_1 are customarily expanded as a linear combination of atomic orbital functions (LCAO) as

$$\Psi_{\mathrm{i}} = \sum_{\mu} C_{\mu i} \emptyset_{\mu} \tag{14}$$

Where $C_{\mu i}$ are the coefficients and $\emptyset_{\mu}$ are real atomic functions. The requirement that the orbitals are orthonormal is

$$\sum_{\mu\nu} c_{\mu i} c_{\nu i} S_{\mu\nu} = \delta_{ij} \tag{15}$$

Where δ_{ij} is the Kronecker delta and S_{μ} is overlap integral for atomic functions ϕ_{μ} and ϕ_{ν}

$$S_{\mu\nu} = \int \phi_{\mu}(1)\phi_{\nu}(1)d\tau \tag{16}$$

2. Solving for the Molecular Orbital : LCAO-MO-SCF

Introduction Eq. (5) and (6) into Eq. (1), the equation takes the final form generally known as the Roothaan equations as

$$\sum_{v}(F_{\mu v} - \varepsilon_i S_{\mu v})c_{vi} = 0 \tag{17}$$

The elements of the matrix representation of the Hartree-Fock Hamiltonian operator F are

$$F_{\mu v} = H_{\mu v}^{core} + \sum P_{\lambda\sigma}\left[\left(\frac{\mu v}{\lambda\sigma}\right)\frac{1}{2}\left(\frac{\mu\lambda}{v\sigma}\right)\right] \tag{18}$$

and density matrix defind as

$$P_{\mu v} = 2\sum_{i} c_{\mu i}^{*} c_{vi} \tag{19}$$

$$\left(\frac{\mu v}{\lambda\sigma}\right) = \int\int \phi_{\mu}(1)\phi_{v}(1)\frac{1}{r_{12}}\phi_{\lambda}(2)\phi_{\sigma}(2)d\tau_1 d\tau_2 \tag{20}$$

and one-electron orbital energy is

$$\varepsilon_i = H_i^{(1)} + \sum_{j} 2J_{ij} - K_{ij} \tag{21}$$

where

Coulomb integral,

$$J_{ij} = \sum_{\mu v\lambda\sigma} c_{\mu i}^{*} c_{\lambda j}^{*} c_{vi} c_{\sigma j}\left(\frac{\mu v}{\lambda\sigma}\right) \tag{22}$$

and Exchange integral,

$$K = \sum_{\mu v\lambda\sigma} c_{\mu i}^{*} c_{\lambda i}^{*} c_{vi} c_{\sigma i}\left(\frac{\mu\lambda}{v\sigma}\right) \tag{23}$$

the total electronic energy (ε) is

$$\varepsilon = \sum_i \left(\varepsilon_i + H_{ii}^{(1)}\right) \tag{24}$$

Therefore, Eq. (8) can be written in matrix form as

$$\mathbf{FC} = \mathbf{SCE} \tag{25}$$

where E is the diagonal matrix of the ε_i and the elements of a matrix C are the coefficients in the expansion LCAO.

Hartree-Fock or self-consistent field method introduces some elegant approximations to solve a one electron eigenvalue problem, and must be solved iteratively. Solving the Eq. (20) for the coefficient C describing the LCAO expansion of the orbital ψ_i and orbital energies ε_i which require a matrix diagonalized. Note that **F** depends on the coefficient ***C***.

They may be usefully transformed by defining new matrices

$$F^{\tau} = S^{1/2}FS^{-1/2} \tag{26}$$

$$C^{\tau} = S^{1/2}C \tag{27}$$

Then obtain

$$F^{\tau}C^{\tau} = C^{\tau}E \tag{28}$$

Matrix equation (19) can be solved using standard methods. The basis function coefficients can be obtained from C^{τ} using $C = S^{1/2}C^{\tau}$. The matrix elements of the Hartree-Fock Hamiltonian operator are dependent on the orbitals through the elements

$P_{\mu\nu}$ and the Roothaan equations are solved by first assumting an initial set of linear expansion coefficients. The whole process is then repeated until the coefficients no longer change within a given tolerance on repeated iteration. The solution is then said to be self-consistent and the method is then referred to as the SCF method. The mathematical steps required for the solution of the Roothaan-Hall equations are outlined in Figure A1.

Statistical Analysis for QSAR Analysis

Many statistical methods have been employed to generate QSAR models from descriptive variables. The most commonly used techniques are Multiple Linear Regression (MLR) and Partial Least Squares (PLS). both methods have their advantages and disadvantages. Conventional QSARs most often use MLR where the ratio of the data points to the number of descriptors should not exceed five. While PLS analyses are particularly suited to situations where the number of descriptor variables exceeds the number of observations it is often the case that the principal components extracted from the descriptor variables has unclear physical meaning. It should be noted that the CoMFA technique allows physical interpretation of PLS extracted QSAR model components in terms of 3D contour maps.

1. Multiple Linear Regressions

Multiple Linear Regression (MLR) expresses a single dependent variable (y) as a linear combination of multiple independent variable (x):

$$y = ax_1 + bx_2 \ldots + k \qquad (29)$$

where a, b are the coefficients of the regression, and k is a constant, the regression model can be built in a stepwise manner.

A number of statistical parameters are used to evaluate regression models. The overall fit of the model is given by r^2:

$$r^2 = \frac{\sum_{i=1}^{n}\left(y_{iobs}-y_{i,mean}\right)^2 - \sum_{i=1}^{n}\left(y_{i,calc}-y_{i,mean}\right)^2}{\sum_{i=1}^{n}\left(y_{i,obs}-y_{i,mean}\right)^2} = 1 - \frac{\sum_{i=1}^{n}\left(y_{i,calc}-y_{i,mean}\right)^2}{\sum_{i=1}^{n}\left(y_{i,obs}-y_{i,mean}\right)^2} \quad (30)$$

The r^2 coefficient can vary from 0 (none of the variance associated with y is explained by the model) to 1 (all the experimental variance is explained by the model). The statistical significance of the model is measured by the F value:

$$F = \frac{n-p-1}{n} \frac{\sum_{i=1}^{n}\left(y_{i,calc}-y_{i,mean}\right)^2}{\sum_{i=1}^{n}\left(y_{i,obs}-y_{i,mean}\right)^2} \quad (31)$$

The larger the F value, the greater the significance of the model. In particular F must be larger than tabulated F value with p and (n-p-1) degree of freedom at a chosen confidence level (for instance 95%)

Good statistical is a necessary condition but not sufficient for a meaningful regression model. Especially when increasing the number of variables, the number of possible models becomes larger and the risk of a chance correlation increases as well. Chance effects have been investigated on sets of random numbers and it has been shown that the higher the ratio of variables to the number of objects, the greater the risk of chance correlation. For example, given a data set of ten objects, the combination of five variables can correlate with random "activities" producing r^s superior to 0.5. For medium-size data sets (n less or equal to 30), having at least 5-6 objects for each variable has been suggested to avoid chance correlation. Finally, MLR is based on a number of assumptions about the dependent variable y (the errors on y are randomly distributed and roughly of the same size) as well as on x (predictor

variables are independent and error-free). In particular, the above conditions are generally not satisfied for data sets where the number of variables largely exceeds the number of objects, making MLR in appropriate.

2. Principal Component Regression and Partial Least Squares

Unlike MLR, Principal Component Regression (PCR) and Partial Least Squares (PLS) can be applied to data sets characterized by large numbers of descriptors and low numbers of objects. Both rely on the assumption that all the descriptors can be seen as a combination of a small number of intrinsic variables (called principal components in PCR and latent variables in PLS) plus some errors, and both are aimed at extracting this relevant information from the original descriptor matrix X and correlating it to the biological activity Y.

The PCR method accomplished this task step-wise by:

2.1 executing a Principle Component Analysis (PCA) on the X matrix and saving the scores,

2.2 selecting the optimal number p of components (based on explained variance), and

2.3 using the first p PCA scores of X to build a regression model with Y.

Because the PCA scores and the regression coefficients are calculated independently, variables important for explaining the biological response may have already been removed at the regression stage.

The two steps (PCA and regression analysis) can be effectively combined by using the PLS method. PLS is aimed at finding linear combinations of the descriptors (latent variables) that not only approximate the original matrix X, but also simultaneously correlate with the biological activity Y. latent variables (LVs) retain the same properties of PCs in the sense that they are linear combinations of the

original variables and they are an orthogonal set, but they differ because LVs are built maximizing the covariance between X and Y.

As with PCA plots of the scores and coefficients of the linear combinations can be generated, and the help for the interpretaion of the model and the identification of outliers, as well as non-linear relationships.

Regression coefficients in term of original variables can also be computed, so that the PLS solution can be still reported in the traditional form (Eq.29). unlike MLR PLS can simultaneously handle more activity.

The optimal number of LVs of a PLS model is usually estimated by cross-varidation (CV). CV means that the objects are divided in n groups, a model is derived with n-1 groups and sue to predicting the excluded group of objects. This is repeated until all groups have been excluded once at a time. The r^2_{cv} value is calculated from the predictions as follows:

$$r_{cv}^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_{i,pred} - y_{i,obs}\right)^2}{\sum_{i=1}^{n}\left(y_{i,obs} - y_{i,mean}\right)^2} \tag{32}$$

The above formula is analogous to r^2 (Eq.30) where calculated y values are replaced by predicted values. Other statistical parameters commonly calculated are the SDEP and the S_{PRESS}

$$S = \sqrt{\frac{\sum_{i=1}^{n}(y_{i,pred} - y_{i,obs})^2}{n}} \tag{33}$$

$$S_{PRESS} = \sqrt{\frac{\sum(y_{i,pred} - y_{i,obs})^2}{n - a - 1}} \tag{34}$$

Where a is the number of LVs. The optimal number of LVs corresponds to the highest q^2 or to the lowest SDEP. However, as the number of LVs increase, PLS suffers the same limitations as MLR; hence the number of latent variables should be kept as small as possible. As a rule of thumb, a new LV is added only of it leading to an increment of at least 5% in the q^2. Alternatively the S_{PRESS} can be used, because it does take into account the number of LVs and penalizes high-dimensional models.

Finally PLS, when used in predictions, provides a rapid evaluation of how test set molecules are similar to those in the training set, and hence of how reliable the predictions may be. The PLS model cam\n be easily understood by a geometrically interpretation as shown in figure A1.

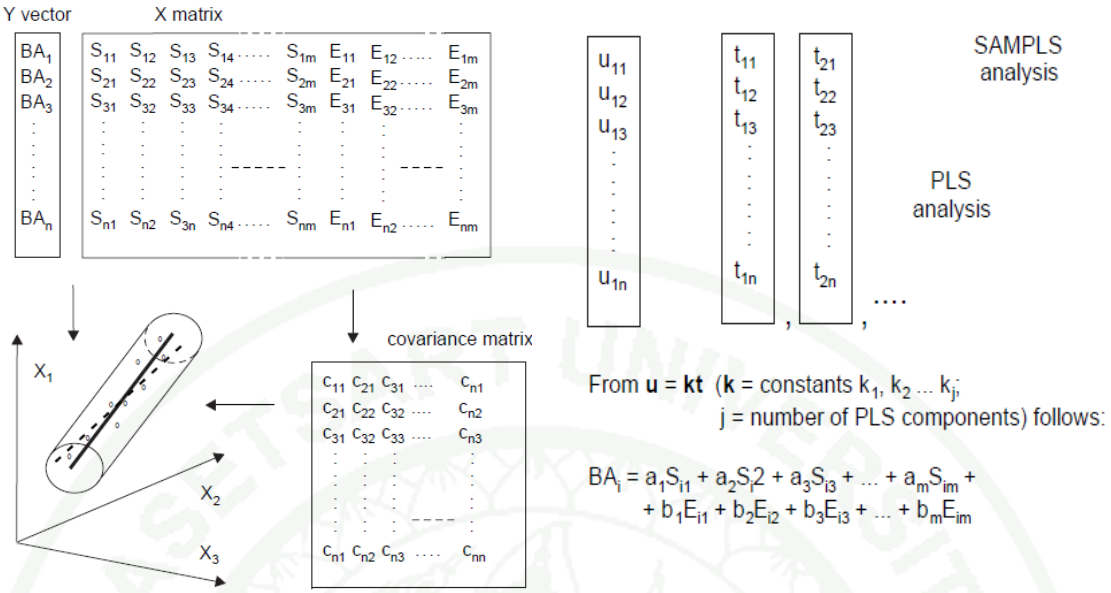


Figure A1 PLS analysis derives vectors ***u*** and ***t*** from **Y** block (or y vector; BA_i = logarithms of relative affinities or other biological activities) and the **X** block (S_{ij} = steric field variable of molecule I in the grid point j; E_{ij} = electrostatic field variable of nolecule I in the grid point j) that are related to principal components. These 'latent variable' are skewed within their confidence hyperboxes to achieve a maximum intercorrelation (diagram). SMAPLE is a PLS modification which first derives the covariance matrix of the **X** block and then the PLSresults from this covariance matrix. Especially in cross-varidation (see below), SAMPLE analysis is much faster than ordinary PLS analysis.

Appendix B

Presentation and proceeding

Proceeding

1. Nuttapong Ithiapa, , Phornphimon Maitarad, Chompoonuch Tancharoen, Songwut Suramitr, Patchreenart Saparpakorn and Supa Hannongbua. **Comparative Molecular Field Analysis Study on Anti HIV-1 RT Diarylaniline Derivatives.** 14th Asian Chemical Congress 2011 (14ACC), Queen Sirikit National Convention Center, Bangkok, Thailand, 5-8 September 2011.

OR-G7-18

COMPARATIVE MOLECULAR FIELD ANALYSIS STUDY ON ANTIHIV-1 RT DIARYLANILINE DERIVATIVES

Nuttapong Ithiapa[1,2], Phornphimon Maitarad[1,2*], Chompoonuch Tancharoen[1,2], Songwut Suramitr[1,2], Patchareenart Saparpakorn[1,2] and Supa Hannongbua[1,2*]

[1]*Department of Chemistry, Faculty of Science, Kasetsart University, Bangkok 10900, Thailand*

[2]*Center of Nanotechnology KU, Kasetsart University, Bangkok 10900, Thailand.*

**Corresponding Authors: E-mail: pmaitarad@gmail.com and fscisph@ku.ac.th, Tel: +66-2562-5555 ext. 2111*

Fax: +66-2562-5555 ext. 2176

Abstract

In the present study, we aim to construct a model of the relationship between structural properties of 25 diarylaniline derivatives and their 50% effective concentrations (EC_{50}) to HIV-1 Reverse Transcriptase (RT) using a comparative molecular field analysis (CoMFA). The best predictive CoMFA model gives a very good statistical result with r_{cv}^2 = 0.823, r_{nv}^2 = 0.924, S_{press} = 0.422, *SE* = 0.241, *F* = 65.055, steric contribution = 28.1% and electrostatic contribution = 71.9%. Consequently, the obtained CoMFA contour maps merging with the wild type HIV-1 RT binding site can give the informative details for understanding the structural requirements of inhibitors and can guide the new design of diarylanilline inhibitors.

Keywords: CoMFA, 3D-QSAR, HIV-1 Reverse Transcriptase, diarylanilline inhibitors

Introduction

AIDS, or acquired immunodeficiency syndrome is caused by the human immunodeficiency virus type 1 (HIV-1). According to UNAIDS statistics, more than 60 million people worldwide have been infected by the human immunodeficiency virus type 1 (HIV-1) and about 25 million patients have died of AIDS. HIV-1 genome encodes for three major enzymes protease, reverse transcriptase and integrase for HIV-1 replication [6-7]. Reverse transcriptase is a key enzyme in the HIV replication cycle and is one of the main targets in the development of drugs for treating HIV-infection and AIDS. HIV-1 reverse transcriptase (HIV-1 RT), which is virally encoded, catalyses the conversion of viral RNA into double stranded DNA, which is then integrated in the host genome [1-3]. Two types of drugs that inhibit HIV-1 polymerase activity are nucleoside and non-nucleoside inhibitors. Non-nucleoside reverse transcriptase inhibitors (NNRTIs) are important components of the first line highly active antiretroviral therapy regiments. NNRTIs bind to an allosteric hydrophobic pocket

and prevent the progression of DNA synthesis from the viral RNA template, located at about 10 Å from the polymerase active site. The hydrophobic pocket is formed by the key residues like Tyr181, Tyr183, Tyr188 and Trp229 [10-13]. The newly diarylaniline analogues of NNRTIs have been reported due to a real medical need to develop a new generation of NNRTIs which do not give rise to cross-validation and are effective against clinically relevant mutant strains.

Aims

Herein, we have reported the analogues based drug design of three-dimensional quantitative structure-activity relationships (3D-QSAR) studies using Comparative Molecular Field Analysis (CoMFA) technique on diarylaniline derivatives. CoMFA is one of the powerful techniques for constructing the relationship between steric and electrostatic properties and the biological activities of the inhibitor series with resulting of CoMFA contour maps leading to guide the new design of the inhibitors before synthesizing and testing activities.

Materials and methods

Data Set

A set of 25 diarylaniline derivatives [11] were reported with their EC_{50} values for inhibiting of the wild type HIV-1 RT as listed in Table 1. A dependent variable in CoMFA was defined as pEC_{50} (-logEC_{50}), where EC_{50} values were measured in vitro under the same experimental conditions. All derivatives were built based on the skeleton template of TMC125 obtaining from the X-ray structure of the PDB code 3MEC which consist of the complex structure of HIV-1 RT and TMC125 [16]. All processes of molecular buildings, Tripos force field optimizations, and Gasteiger-Hückel charge calculations were performed by the Sybyl 8.0 program package [15].

CoMFA Methodology

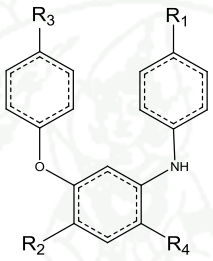
In the part of CoMFA study, the process of alignment is very important, therefore, all 25 compounds were aligned into the same template structure as shown in an insert figure of Table 1 using the “Fitting Atomic Based Alignment”. Then, CoMFA lattices with 2.0 Å grid spacing were generated around the aligned compounds based on the molecular volume of the structures. The lattices were defined automatically, and were extended by 4.0 Å in all directions. Steric and electrostatic fields were calculated by three types of probe atoms, representing for the enzyme environment, were placed at each lattice point, namely sp^3 carbon atom with +1 charge, sp^2 oxygen atom with -1 charge, and sp^3 nitrogen atom with -1 charge. An energy cut-off of 30 kcal mol^{-1} was applied to avoid infinite energy values inside the molecule.

The regression analysis was carried out using the partial least-square (PLS) to derive a CoMFA model expressing the correlation between the steric and electrostatic properties and the biological activities.

CoMFA columns with a variance of less than 2.0 kcal mol^{-1} were filtered by using column filtering to improve the signal-to-noise ratio. Leave-one-out (LOO) cross-validation method was used to check the predictive ability of the derived model and to identify the number of component (NOC) at which the difference in the r_{cv}^2 value compared with the next one was less than 0.05.

The predictive ability of the model derived from the training set is expressed as the cross-validation predictive (r_{cv}^2) value. The r_{cv}^2 value is defined as

Table 1 Structures of diarylaniline derivatives for CoMFA analysis.



Name	R_1	R_2	R_3	R_4	EC_{50}	Pred pEC_{50}
Compound 1	OCH_3	NO_2	H	NO_2	3.840	5.2347
Compound 2	OCH_3	NO_2	CH_3	NO_2	2.990	5.4173
Compound 3	OCH_3	NO_2	Br	NO_2	3.630	5.517
Compound 4	CH_3	NO_2	Br	NO_2	4.310	5.4225
***Compound 5**	NO_2	NO_2	Br	NO_2	> 49.7	-
Compound 6	C≡N	NO_2	Br	NO_2	0.172	4.2796
Compound 7	C≡N	NO_2	H	NO_2	0.545	6.3486
***Compound 8**	C≡N	NO_2	C≡N	NO_2	4.190	-
Compound 9	C≡N	NO_2	CH_3	NO_2	0.280	6.6208
Compound 10	C≡N	NO_2	CHO	NO_2	1.530	5.9673
Compound 11	C≡N	H	Br	NO_2	0.317	6.1969
Compound 12	C≡N	H	H	NO_2	3.147	5.8731
Compound 13	C≡N	H	CN	NO_2	0.208	6.6849
***Compound 14**	C≡N	H	Me	NO_2	0.067	-
Compound 15	C≡N	H	CHO	NO_2	2.190	5.5296
Compound 16	C≡N	H	Br	NH_2	0.047	7.0809
Compound 17	C≡N	H	Br	NH_2	0.070	7.5349

Compound 18	C≡N	H	Br	NH_2	0.073	7.0446
Compound 19	C≡N	NH_2	Br	NH_2	0.161	6.9912
***Compound 20**	C≡N	NH_2	H	NH_2	3.226	-
Compound 21	C≡N	NH_2	C≡N	NH_2	0.030	7.4439
Compound 22	C≡N	NH_2	CH_3	NH_2	0.070	6.9579
Compound 23	C≡N	NO_2	Br	NH_2	0.016	7.6989
Compound 24	C≡N	NO_2	C≡N	NH_2	0.003	8.1759
Compound 25	C≡N	NO_2	CH_3	NH_2	0.062	7.6196

* outlier compounds

$$r_{cv}^2 = 1.0 - \frac{PRESS}{SSY}, \quad (1)$$

Where SSY is the variance of the biological activities around the mean value, and PRESS is the prediction error sum of squares derived from LOO.

$$PRESS = \sum y\,(y_{pred} - y_{actual})^2, \quad (2)$$

$$SSY = \sum y\,(y_{actual} - y_{mean})^2, \quad (3)$$

The uncertainty of the prediction is defined as

$$S_{PRESS} = \sqrt{\frac{PRESS}{n-k-1}}, \quad (4)$$

where *k* is the number of variables in the model and *n* is the number of compounds used in the study [8,14-15].

Results

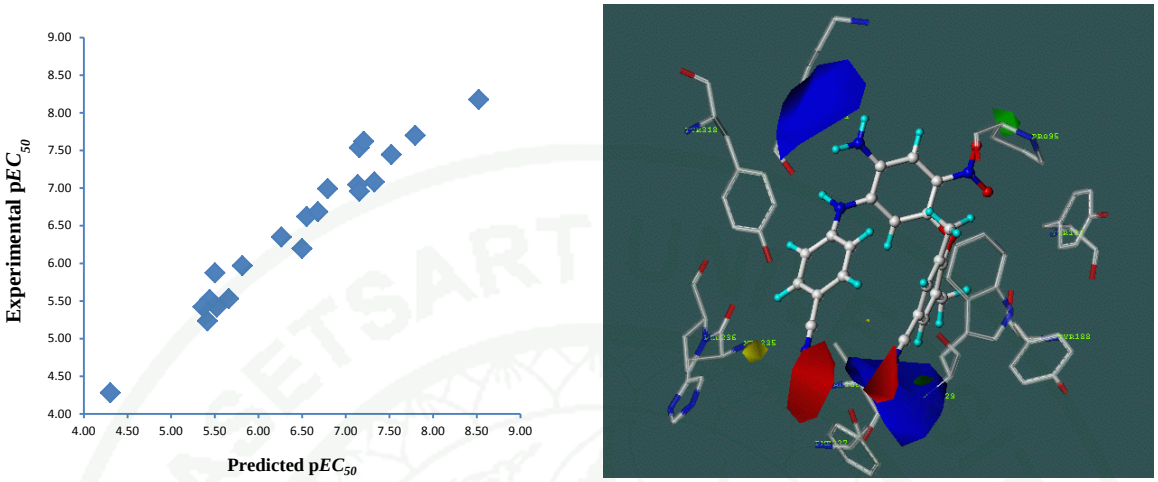
There are four CoMFA models varying the types of the probe atom at the grid spacing as representative of mainly atom types of amino acids for *in silico* receptor as summarized in Table 2. The C_{sp3}(+1) is generally represented as sterically probe atom, O_{sp2}(-1) and N_{sp3}(-1) for electrostatic interactions. The representative CoMFA model based on the best statistical results of both cross-validated r_{cv}^2 and non-validated r_{nv}^2, the CoMFA model 1 is selected for further contour map discussion and new inhibitor prediction. The graphical plots between the experimental and the predicted pEC_{50} of the diarylanilines based on CoMFA model 1 is shown in Figure 1 (a).

Discussion and Conclusion

The obtained CoMFA model 1 shows the major contribution from the electrostatic property of 71.9% and the rest of 28.1% of steric contribution. The CoMFA contour map of both 2 types of structural properties merging with the HIV-1 RT binding site is displayed in Figure 1 (b). The positive and negative steric CoMFA regions are represented in green and yellow contours, respectively, while the positive and negative electrostatic CoMFA regions are displayed in blue and red contours, respectively. According to the CoMFA contour map, the electrostatic property plays an important role around three substitutions, $\mathbf{R_1}$, $\mathbf{R_3}$ and $\mathbf{R_4}$ which can be concluded that (i) $\mathbf{R_1}$ is favored the negative charge group or electron withdrawing group, (ii) $\mathbf{R_3}$ has both negative and positive regions together which means that this site can be substituted by both withdrawing and donating groups, (iii) $\mathbf{R_4}$ is favored the electron donating group more than electron withdrawing group, whereas, the $\mathbf{R_2}$ shows only a small region of favorable steric map. Therefore, based on an unclear CoMFA contour maps at $\mathbf{R_2}$ and $\mathbf{R_3}$, particular interactions between the partial substituent of $\mathbf{R_2}$ and $\mathbf{R_3}$ and the amino acids surrounding the site substitutions are needed to investigate for more understanding in the molecular interactions using the quantum chemical calculations.

Table 2 PLS statistical results of CoMFA models of diarylaniline derivatives against wild type HIV-1 RT

Statistical terms	Model 1	Model 2	Model 3	Model 4
Probe atom	C_{sp3} (+1)	O_{sp2} (-1)	N_{sp3} (-1)	C_{sp3} (+1)O_{sp2} (-1)N_{sp3} (-1)
r^2_{cv}	0.823	0.802	0.799	0.812
NOC	4	3	4	4
S_{press}	0.422	0.433	0.440	0.436
r^2_{nv}	0.946	0.934	0.942	0.940
SE	0.241	0.250	0.241	0.245
F_{values}	65.055	80.530	65.499	63.068
%Steric	0.281	0.265	0.202	0.274
%Electrostatic	0.719	0.735	0.738	0.726



(a) (b)

Figure 1 (a) Plots of the predicted versus experimental biological activities (pEC_{50}) and (b) steric and electrostatic contour maps of CoMFA model I.

Acknowledgement

This work was supported by grants from the Thailand Research Fund (TRF) and the Commission on Higher Education (RTA5380010). The National Center of Excellence in Petroleum, Petrochemical Technology and Advanced Materials are gratefully acknowledged for research facilities. Partial supporting by National Nanotechnology Center (NANOTEC), Ministry of Science and Technology, Thailand, through its program of Center of Excellence Network is grateful. We would like to thank the Large Scale Research Laboratory of the National Electronics and Computer Technology (NECTEC), Laboratory for Computational and Applied Chemistry (LCAC) and the computing center of KU for allowing the use of their research facilities.

References

[1] Joseph Rebehmed, Florent Barbault, Cátia Teixeira & Francois Maurel. (2008) *Journal of Computer-Aided Molecular Design.* 22: 831–841.

[2] Vanangamudi Murugesan, Yenamandra S. Prabhakar & Seturam B. Katti. (2009) *Journal of Molecular Graphics and Modelling*. 27: 735–743.

[3] Zeng Li, Hao Zhang, Yixue Li, Jian Zhang & Hai-Feng Chen. (2011). *Chemical Biology & Drug Design*. 77: 63–74.

[4] Elena Cichero, Sara Cesarini, Paola Fossa, Andrea Spallarossa, & Luisa Mosti. (2009) *European Journal of Medicinal Chemistry*. 44: 2059–2070.

[5] Monique Araújo de Brito, Carlos Rangel Rodrigues, José Jair Vianna Cirino, Ricardo Bicca de Alencastro, Helena Carla Castro & Magaly Girão Albuquerque. (2008) *Journal of Chemical Information and Modeling.* 48: 1706–1715.

[6] Joseph Rebehmed, Florent Barbault, Cátia Teixeira, & Francois Maurel. (2008) *Journal of Computer-Aided Molecular Design*. 22: 831–841.

[7] Marie-Pierre de Béthune. (2010) *Antiviral Research*. 85: 75–90.

[8] Phornphimon Maitarad, Sumalee Kamchonwongpaisan, Jarunee Vanichtanankul, Tirayut Vilaivan, Yongyuth Yuthavong, & Supa Hannongbua. (2009) *Journal of Computer-Aided Molecular Design*. 23: 241–252.

[9] Hua Qin, Chang Liu, Jianfang Zhang, Ying Guo, Siwei Zhang, Zhili Zhang, Xiaowei Wang, Liangren Zhang, & Junyi Liu. (2010) *Bioorganic & Medicinal Chemistry Letters.* 20: 3003–3005.

[10] Kalyan Das, Joseph D. Bauman, Arthur D. Clark, Jr., Yulia V. Frenkel, Paul J. Lewi, Aaron J. Shatkin, Stephen H. Hughes, & Eddy Arnold. (2008) *Proceeding of the National Academy of Sciences of the United States of America.*

[11] Bingjie Qin, Xingkai Jiang, Hong Lu, Xingtao Tian, Florent Barbault, Li Huang, Keduo Qian, Chin-Ho Chen, Rong Huang, Shibo Jiang, Kuo-Hsiung Lee, & Lan Xie. (2010) *Journal of Medicinal Chemistry.* 53: 4906–4916.

[12] Elena Cichero, Sara Cesarini, Andrea Spallarossa, Luisa Mosti, & Paola Fossa. (2009) *Journal of Molecular Modeling*. 15: 871–884.

[13] Vanangamudi Murugesan, Yenamandra S. Prabhakar, & Seturam B. Katti. (2009) *Journal of Molecular Graphics and Modelling.* 27: 735–743.

[14] Rongjing Hu, Florent Barbault, Michel Delamar, & Ruisheng Zhang. (2009) *Bioorganic & Medicinal Chemistry.* 17: 2400–2409.

[15] SYBYL 8.0, Tripos Associates Inc., 1699 South Hanley Road, Suite 303, St Louis, Missouri 63144, USA.

[16] Eric B. Lansdon, Katherine M. Brendza, Magdeleine Hung, Ruth Wang, Susmith Mukund, Debi Jin, Gabriel Birkus, Nilima Kutty & Xiaohong Liu.(2010) *Journal of Medicinal Chemistry*. 53(10): 4295-4299.

CURRICULUM VITAE

NAME : Mr. Nuttapong Ithiapa

BIRTH DATE : May 7, 1986

BIRTH PLACE : Phrae, Thailand

EDUCATION :

YEAR	INSTITUTE	DEGREE/DIPLOMA
2005-2009	Kasetsart University	B.Sc. (Chemistry)
2009-2012	Kasetsart University	M.S. (Chemistry)

POSITION/TITLE : Postgraduetes student

WORKPLACE : Faculty of Science, Kasetsart University

SCHOLASHIP : The Kasetsart University Research and Development Institute (KURDI), and National Nanotechnology Center (NANOTEC)

Postgraduate Education and Research Programs in Petroleum and Petrochemicals, and Advanced Materials (NCE-PPAM)

Kasetsart University Research and Development Institute (KURDI)