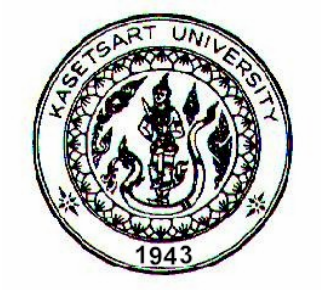


THESIS

THEORETICAL INVESTIGATION ON STRUCTURES, INTERACTION ENERGY AND 3D-QSAR STUDY OF WR99210 DERIVATIVES, DHFR INHIBITORS

WANWIMON MOKMAK

GRADUATE SCHOOL, KASETSART UNIVERSITY

2006

THESIS APPROVAL

GRADUATE SCHOOL, KASETSART UNIVERSITY

Master of Science (Chemistry)

DEGREE

Physical Chemistry — **FIELD**

Chemistry — **DEPARTMENT**

TITLE: Theoretical Investigation on Structures, Interaction Energy and 3D-QSAR Study of WR99210 Derivatives, DHFR Inhibitors

NAME: Miss Wanwimon Mokmak

THIS THESIS HAS BEEN ACCEPTED BY

Supa Hannongbua — **THESIS ADVISOR**

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

Sumalee Kamchonwongpaisan — **COMMITTEE MEMBER**

(Miss Sumalee Kamchonwongpaisan, Ph.D.)

Chak Sangma — **COMMITTEE MEMBER**

(Mr. Chak Sangma, Ph.D.)

Supanna Techasakul — **COMMITTEE MEMBER**

(Associate Professor Supanna Techasakul, Ph.D.)

Yerry Mahatumarattana — **DEPARTMENT HEAD**

(Assistant Professor Yerry Mahatumarattana, B.Sc.)

APPROVED BY THE GRADUATE SCHOOL ON 23 March 2006

Vinai Artkongharn — **DEAN**

(Associate Professor Vinai Artkongharn, M.A.)

THESIS

THEORETICAL INVESTIGATION ON STRUCTURES, INTERACTION ENERGY AND 3D-QSAR STUDY OF WR99210 DERIVATIVES, DHFR INHIBITORS

WANWIMON MOKMAK

A Thesis Submitted in Partial Fulfillment of
the Requirements for the Degree of
Master of Science (Chemistry)
Graduate School, Kasetsart University
2006

ISBN 974-16-1219-2

Wanwimon Mokmak 2006: Theoretical Investigation on Structures, Interaction Energy and 3D-QSAR Study of WR99210 Derivatives, DHFR Inhibitors. Master of Science (Chemistry), Major Field: Physical Chemistry, Department of Chemistry. Thesis Advisor: Associate Professor Supa Hannongbua, Dr.rer.nat. 118 pages.
ISBN 974-16-1219-2

Dihydrofolate reductase-thymidylate synthase of *Plasmodium falciparum* (*Pf*DHFR-TS) is an important target of malarial chemotherapy. WR99210 (2,2-dimethyl-1-[3-(2,4,5-trichlorophenoxy)propoxy]-1,3,5-triazine-2,4-diamine is the most effective inhibitor against wild-type and mutant *Pf*DHFR enzymes in vitro. A series of WR99210 analogs has been designed to have only one substituent at the position 2 of the triazine ring, so as to avoid possible steric clash with Ala16 of the enzyme. These analogs have been synthesized and tested for antimalarial activities in racemic mixture and found to have comparable binding affinity with the parent compound. To probe which enantiomeric form, (*R*)- or (*S*)-form, having higher binding affinity, molecular docking calculations were performed to investigate the orientation of ligands in the binding pocket and to estimate free energies of binding between ligands and the bound enzyme. We found that the estimated free energies of binding of the (*R*)-configuration analogs were lower than that of (*S*)-configuration analogs suggesting that the enzyme preferentially binds the inhibitors with (*R*)-configuration. The energetic results as obtained from the interaction energy calculations from quantum chemical calculations between WR99210 and the active site indicated that Ile14, Asp54, Phe58 and Ile164 are the main residue contributors. Moreover, Three-dimensional Quantitative Structure-Activity Relationship (3D-QSAR) by using Comparative Molecular Field Analysis (CoMFA) method was applied to investigate the relationships between structures and their steric and electrostatic interactions of WR99210 derivatives against wild type of *Pf*DHFR. Their geometric structures were constructed and fully optimized, based on semiempirical (PM3) method. The obtained results with CoMFA standard fields combined steric and electrostatic properties are satisfying based on statistical significance and predictive ability. Consequently, The CoMFA contour maps revealed that steric interactions play more significant role of the variance in the data corresponding to wild type inhibitions with relative contribution values equal to 70.0%. The structural requirements of the present series of compounds identified through 3D-QSAR contour plots will be useful in the design of new WR99210 inhibitors with better activity.

Wanwimon Mokmak — Student's signature

Supa Hannongbua — Thesis Advisor's signature — 20 / 03 / 2006

ACKNOWLEDGEMENT

I wish to express my deep gratitude to a number of people who giving me the guidance, help and support to reach my goal of this thesis. First of all, most of credits in this thesis should justifiably go to my advisor, Associate Professor Dr. Supa Hannongbua, for her valuable guidance, continuous support, kindness and encouragement throughout the course of my graduate. I am also thankful to Dr. Sumalee Kamchonwongpaisan, Dr. Chak Sangma and Dr. Pornpan Pungpo for their kind comments and criticism to make thesis complete.

Furthermore, I would like to thank my committee, Associate Professor Dr. Supanna Techasakul and Associate Professor Dr. Supreya Trivijitkasem the representative of Graduate School, for their helpful criticism and suggestion.

Full financial support by the Thai Graduate Institute of Science and Technology scholarship, National Science and Technology Development Agency and supporting from the Postgraduate Education and Research Program in Petroleum and Petrochemical Technology, Department of Chemistry, Kasetsart University, were gratefully acknowledged.

Special thanks are due to the High Performance Computing Center of the National Electric and Computer Technology (NECTEC), for providing SYBYL program. The Laboratory for Computational and Applied Chemistry (LCAC) at Kasetsart University are also gratefully acknowledged for computational resource and software facilities.

I appreciate Miss Phornphimon maitarad and my colleagues at LCAC at Kasetsart University for their supporting, helpful assistance and sharing useful ideas during the past three years.

Finally, I would like to thank my family for their love, sincere care and great encouragement during this work.

Wanwimon Mokmak

February, 2006

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES (cont'd)

LIST OF FIGURES (cont'd)

LIST OF FIGURES (cont'd)

LIST OF FIGURES (cont'd)

LIST OF FIGURES (cont'd)

LIST OF ABBREVIATIONS

Ala	=	Alanine amino acid
Asp	=	Aspatic amino acid
BIOTEC	=	National Center for Genetic Engineering and Biotechnology, National Science and Technology Development Agency, Thailand
CoMFA	=	Comparative molecular field analysis
cv	=	Cross-validation
CYC	=	Cycloguanil
Cys	=	Cysteine amino acid
DHF	=	7,8-dihydrofolate
DHFR	=	Dihydrofolate Reductase
DHFR-TS	=	Dihydrofolate Reductase Thymidylate Synthase
F	=	F-test value
Gly	=	Glycine amino acid
HF	=	Hartree-Fock theory
I	=	Isoleucine amino acid
Ile	=	Isoleucine amino acid
K_i	=	Inhibition Binding Constant
L	=	Leucine amino acid
LCAO-MO	=	Linear combination of atomic orbitals to molecular Orbitals
Leu	=	Leucine amino acid
log $(1/K_i)$	=	Negative logarithms of the dissociation constant
log P	=	Log of the octanol-water partition coefficient
Met	=	Methionine amino acid
MLR	=	Multiple linear regression
MNDO	=	Modified neglect of diatomic overlap
MP2	=	Second order M∅ller-Plesset
MR	=	Molar refractivity
n	=	Number of cases/observations

LIST OF ABBREVIATIONS (Cont'd)

N	=	Asparagine amino acid
NADPH	=	Nicotinamide-adenine dinucleotide phosphate
noc	=	Number of Component
NDO	=	Neglect of diatomic differential overlap
*Pf*DHFR-TS	=	*Plasmodium falciparum* Dihydrofolate Reductase-Thymidylate Synthase
Phe	=	Phenylalanine amino acid
PLS	=	Partial least-squares
PM3	=	Modified neglect of diatomic overlap, parametric method number3
PRESS	=	Predictive sum of square
Pro	=	Proline amino acid
PYR	=	Pyrimethamine
q^2	=	Predictive ability
QSAR	=	Quantitative structure-activity relationships
3D-QSAR	=	Three dimensional quantitative structure-activity relationships
R	=	Arginine amino acid
r^2	=	Squared correlation coefficient
r^2_{cv}	=	Predictive ability from cross-validation method
rms	=	Root mean square
s	=	Standard deviation
S	=	Serine amino acid
SCF	=	Self-consistent field
SEE	=	Standard error
Ser	=	Serine amino acid
SHMT	=	serine hydroxymethyltransferase
S_{PRESS}	=	The standard of error of prediction
SSY	=	Sum of squares of deviation between the affinities of the fitted set and their mean affinity
THF	=	5,6,7,8-tetrahydrofolate

LIST OF ABBREVIATIONS (Cont'd)

Thr	=	Threonine amino acid
Trp	=	Tryptophan amino acid
TS	=	Thymidylate synthase
Tyr	=	Tyrosine amino acid
Val	=	Valine amino acid
WR99210	=	2,2-dimethyl-1-[3-(2,4,5-trichlorophenoxy)propoxy]-1,3,5-triazine-2,4-diamine

THEORETICAL INVESTIGATION ON STRUCTURES, INTERACTION ENERGY AND 3D-QSAR STUDY OF WR99210 DERIVATIVES, DHFR INHIBITORS

INTRODUCTION

Significant and background of the problem

Malaria (Gilles and Warrell, 1993) is the most important disease of human over the world. It is heavily endemic in most tropical or subtropical countries but also endemic in some temperate countries. Every year, 500-800 million people are newly infected and 2-3 million of them died due to the ion caused by four species of the genus *Plasmodium*: *Plasmodium falciparum, Plasmodium vivax, Plasmodium ovale and Plasmodium malariae.* Malaria is a common and serious infect disease (Miller *et al*., 1994; Breman, 2001). *P. falciparum* is the most pathogenic species of malaria parasites (Rathod *et al*., 1997), which leads to various categories of malaria, including febrile, cerebral, gastrointestinal and pulmonary type malaria, and intravascular haemolysis. Some severe cases may die due to cerebral infarct, renal failure or other compications (Gardner *et al*., 1998; Bowman, 1999). The life cycle of *Plasmodium* spp. is complex and somewhat specific to the parasite species. *P. falciparum* infection in humans begins when an infected female Anopheles sp. mosquito takes a blood meal and injects infective sporozoites into the peripheral circulation. Within minutes, these sporozoites invade hepatocytes in the liver and, over approximately one week, undergo asexual multiplication, producing tens of thousands of merozoite forms of the parasite. When the infected hepatocyte ruptures, merozoites are released into the peripheral circulation. The merozoites invade red blood cells (rbcs) and complete another round of multiplication within 48-72 hours, with the production of additional merozoites, which devour the rbc haemoglobin in the process. The released merozoites invade additional rbcs and carry on the cycle. It is the synchronous release of merozoites that is though to be responsible for the periodic fevers associated with malaria. Some invading merozoites do not divide, but differentiate into male

(microgametocyte) and female (macrogametocyte) sexual forms. These sexual forms are taken from the bloodstream by a feeding Anopheles sp. mosquito and fertilise in the mosquito midgut to form zygotes. These zygotes further differentiate into motile forms, called ookinetes, migrate through the mosquito gut wall and divide within oocysts on the external gut wall to form thousands of sporozoites. The infective sporozoites are released into the mosquito haemocoele and move to the salivary gland, where they await injection into another human host, thus completing the life cycle (Warhurst, 2002), as shown in Figure 1.

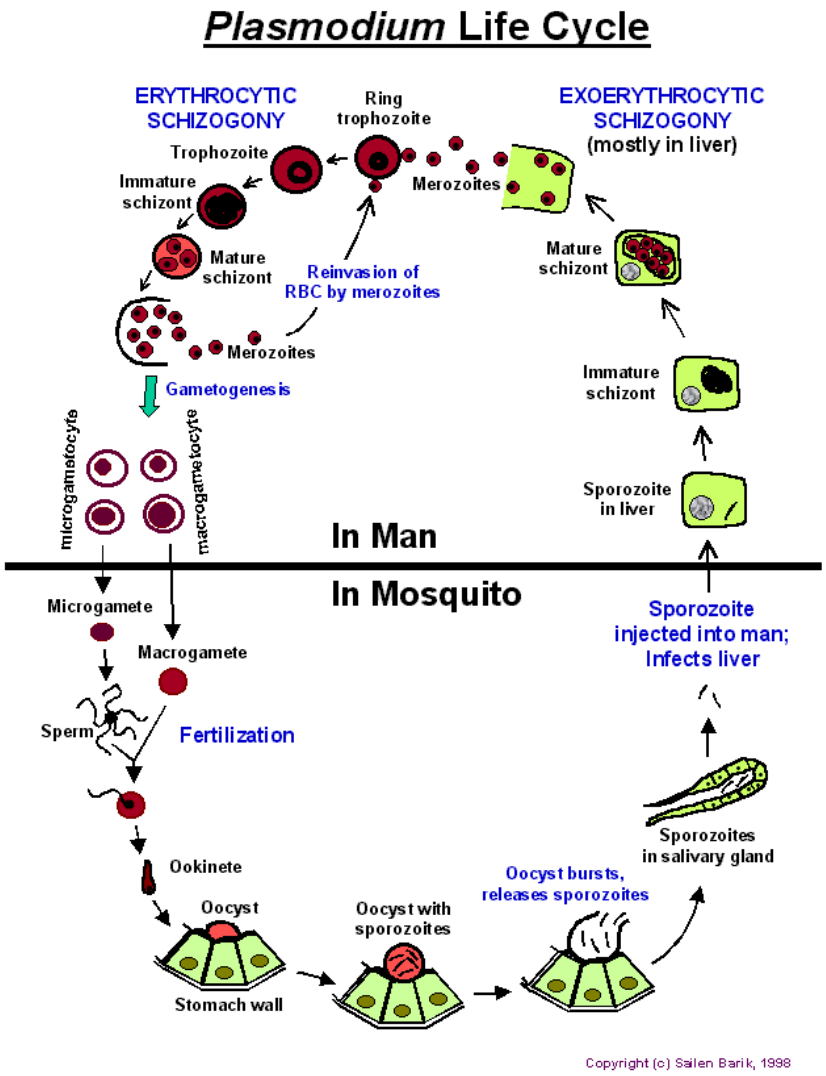


Figure 1 The life cycle of *Plasmodium falciparum* malaria parasite. Source: Barik (1998)

DHFR as a Molecular Target for drug therapy

Chemotherapy is one of the main strategies employed in malaria control. Such an approach has been used against the disease since the 17th century, and extensive reviews on antimalarial drugs have been published (Santos-Filho *et al*., 2001; Winstanley *et al*., 2000; Olliaro *et al*., 1999; Macreadie *et al*., 2000). One of the main targets for antimalarial chemotherapy has been the dihydrofolate reductase (DHFR) an enzyme found in nearly all living cells. Its function is to catalyze the reduction of 7,8-dihydrofolate (DHF) to 5,6,7,8-tetradihydrofolate (THF), in a biochemical reaction whose coenzyme is the reduced form of nicotinamide adenine phosphate (NADPH).

$$DHF + NADPH + H^{+} \xrightarrow[DHFR]{} THF + NADP^{+}$$

Structural formulas for folate, DHF, THF and NADPH are shown in Figure 2. DHF is transformed in THF by reduction of the N5 – C6 double bond. Although the mechanism of this reaction is not fully understood, it is known that NADPH acts as an electron-donor or, more precisely, as a hydride donor, in reducing biosynthetic processes (Olliaro., 2001). In this way, it is believed that the hydride is transferred from NADPH to the substrate C6 atom, while a proton is given to the N5 atom by a water molecule. Although there is not yet conclusive experimental evidence that proves that the reaction occurs in this way, this mechanism is largely accepted (Zuccotto *et al*., 1998).

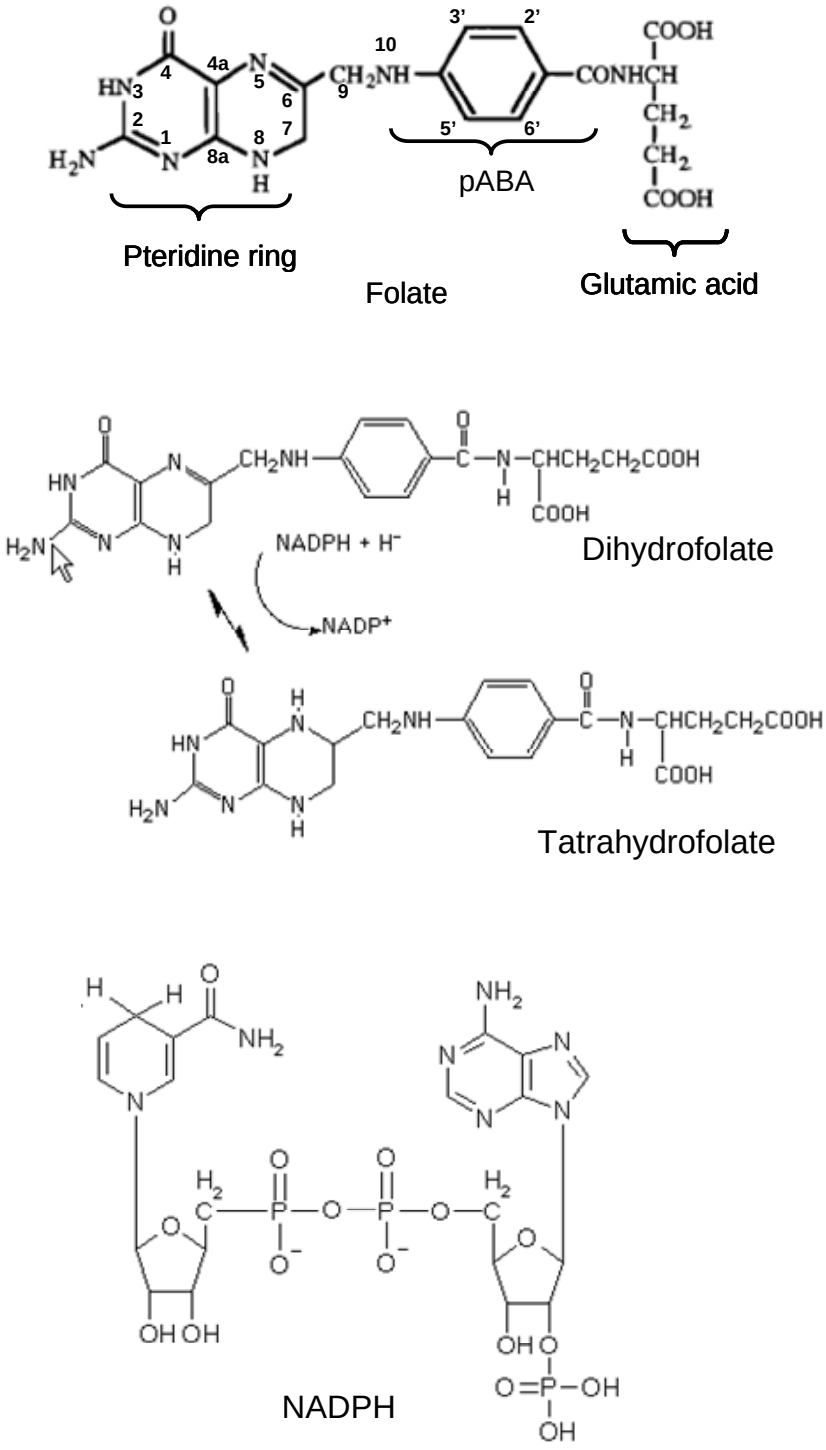


Figure 2 Structures of folate, its reduced derivatives and NADPH. The structure of folate is usually discussed in terms of three components shown above (pABA is the p-aminobenzoyl group).

Source: Delfino (2002)

DHFR plays an important role in the cellular metabolism of the vast majority of living creatures, since it is essential for maintenance of adequate levels of THF, which is fundamental in the metabolic cycle of biosynthesis of deoxythymidylate (dTMP), represented in Figure 3, In this cycle, thymidylate synthase (TS) catalyzes the conversion of deoxyuridylate (dUMP) and 5,10-methylenetetrahydrofolate to dTMP and DHF, respectively. DHFR catalyzes the subsequent reduction of DHF to THF, a compound that is indispensable for the biochemical transference of single-carbon units. Finally, serine hydroxymethyltransferase (SHMT) catalyzes the regeneration of 5,10-methylenetetrahydrofolate, which guarantees the continuation of dTMP biosynthesis. During the cycle, for each THF molecule oxidized to DHF, one dTMP molecule is formed. It is worthy of note that THF, in its derived forms, act as a coenzyme (transferring single-carbon unit) not only for dTMP synthesis, but also for the syntheses of purine nucleotides, methionine and other essential metabolites. Inhibition of DHFR or of any other enzyme in the cycle prevents the formation of new dTMP molecules, which interrupts DNA synthesis. Since there is no alternative route for dTMP biosynthesis, the consequence is cellular death in malarial (Sirawaraporn., 1998).

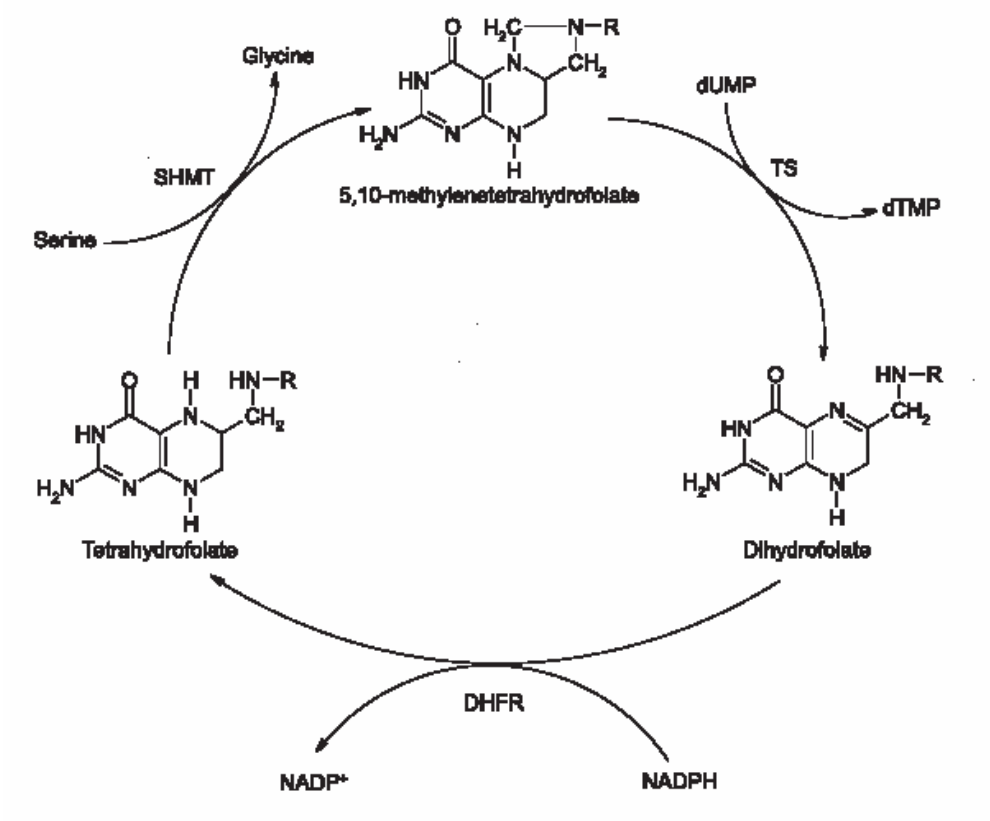


Figure 3 Biosynthetic cycle of dTMP (R = pABA + L-glutamate).
Source: Delfino (2002)

DHFR is an enzyme that has been studied for many years ago. In the past, only the primary structure of *Pf*DHFR can been determined but the tertiary structure of the plasmodial enzyme was not available. Homology modelling (Lemcke *et al*., 1996; Santos-Filho *et al*., 2001) can explain the structure-activity relationships and design of new antimalarial agents. Recently, the x-ray crystal structures of the *P. falciparum* enzyme dihydrofolate reductase-thymidylate synthase (*Pf*DHFR-TS) were crystallized by Yuvaniyama *et al*., 2003. These structures are available in the Protein Data Bank (PDB), a database with thousands of protein structures (http://www.rcsb.org/pdb/). The overall structure of the wild-type is two *Pf*DHFR-TS molecules per asymmetric unit related by a pseudo two-fold symmetry. The intersubunit contact of this dimer mainly involves interactions between the TS domains. The two DHFR domains are not in contact, as shown in Figure 4.

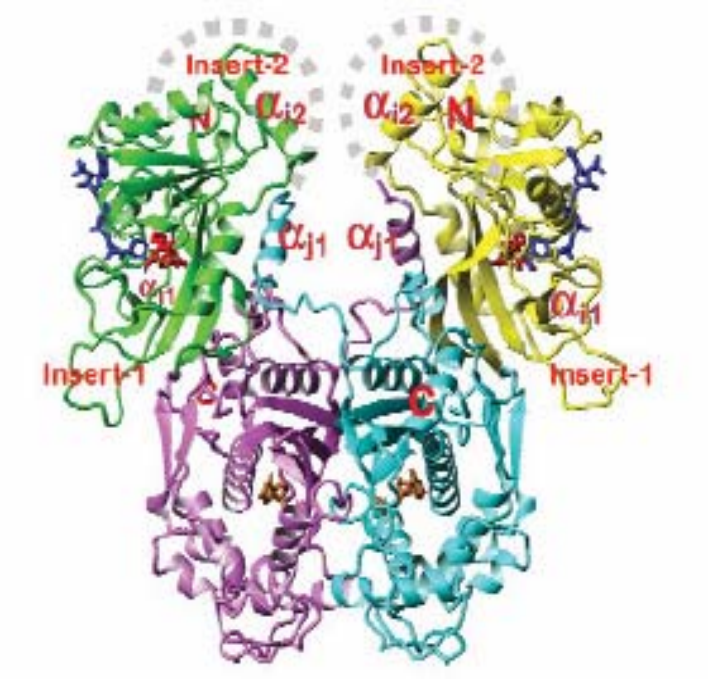


Figure 4 Overview of the wild type *Pf*DHFR-TS structure (PDB code 1J3I). Ribbon diagram of overall structure with bound WR99210, NADPH and dUMP drawn in red, blue and orange, respectively. N-terminal DHFR domains are in green and yellow; C-terminal junction regions and domains are in magenta and cyan. N and C terminal and the inserts unique to plasmodial DHFR-TS are indicated. A short helix in insert1 and long helix in insert2 are labeled as α_{i1} and α_{i2}, respectively. Terminal and α_{i1} helix on the lift part of the structure are on the back of the molecules in this orientation. The putative links between DHFR and TS domains shown as dashed gray α_{i1} curves are based on intermolecular spaces in crystal packing around the regions of unobserved residues. The helices α_{j1} in the junction region are involved in domain attractment, thus orienting the TS domains for dimerization into a functional unit.

Source: Yuvaniyama (2003)

As mentioned before, antifolates, also known folate antagonists, constitute a class of antimalarials belonging to the group of the nucleic acid biosynthesis inhibitors (Olliaro., 2001). All pharmaceuticals that interfere in the folate cycle by inhibiting any of its enzymes are called antifolates, their action results in a decrease on the synthesis rates of dTMP, purine nucleotides and metionine, among other consequences. The antifolates are divided into two groups, denominated Type 1 and Type 2 (Olliaro., 2001).

Type 1 of antifolates is sulfones and sulfonamides whose structures are similar to pABA, with which they compete for the DHPS active site, preventing or slowing down the formation of 7, 8-dihydropteroate, which is a precursor of DHF. It is worthy of note that, in *P. falciparum*, DHPS and PPPK constitute a bifunctional enzyme, in the same way as DHFR and TS. The main Type 1 antifolates are the sulfone dapsone and the sulfonamides sulfadoxine, sulfadiazine and sulfalene, whose structures are shown in Figure 5.

The chemotherapy with Type 1 antifolates has some inconveniences. Dapsone is a very toxic drug, while the cited sulfonamides are antimalarials of easy absorption, but difficult excretion by the human organism. Moreover, there strong evidences that the parasite can use efficiently exogenous folates, either as folic acid or folinic acid; this salvage pathway can bypass the blockage of the endogenous folate biosynthetic pathway induced by Type 1 antifolates. Finally, mutations in DHPS primary structure have been associated to an increases on the resistance of *P. falciparum* to these antifolates, mainly to sulfadoxine.

Type 2 of antifolates inhibits the parasite DHFR, preventing the DHF reduction to THF described before. These compounds are structurally similar to the pteridine ring of DHF, with which they compete for the active site of DHFR. The common structures of Type 2 antifolates are illustrated in Figure 6. The most extensively used antifolates are pyrimethamine (PYR) and cycloguanil (CYC). PYR is a potent and selective inhibitor of *Pf*DHFR, used alone or in combination with other drugs, such as sulfadoxine (combination known as SP). SP is cheap, practicable and

highly effective in some African countries (Winstanley., 2002), however, it is prone to the rapid emergence of resistance. CYC is the metabolic product of proguanil, a drug developed in Britain during World War II. It is also a potent drug, and can be used alone or in combination with chloroquine or atovaquone. PYR and CYC have a low level of toxicity and, if used in the correct dose, are completely free of side-effects. But its use is also affected by increasing parasite resistance. WR99210 (2,2-dimethyl-1-[2,4,5-trichlorophenoxy)propoxy]-1,3,5-triazine2,4-diamine) is an experimental drug, which is highly effective in vitro against *Pf*DHFR. However, WR99210, which differs from the compromised inhibitor in having a flexible side-chain, still binds tightly with the mutant enzyme and retrain its antimalarial efficacy. Understanding the structural basis of interaction between drugs and the *Pf*DHFR is an important for the potential development of novel antimalarial drug.

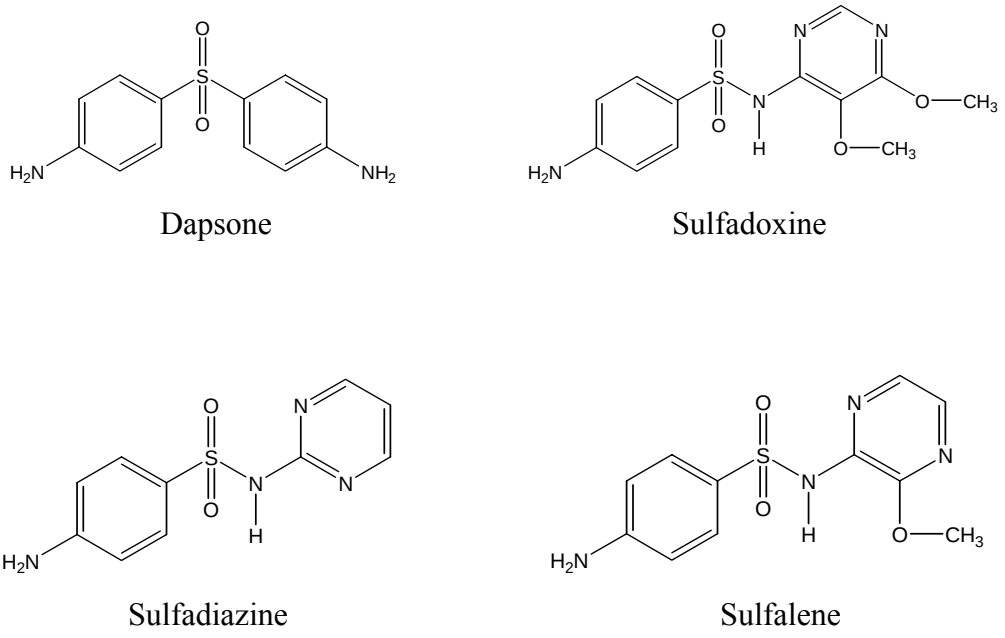
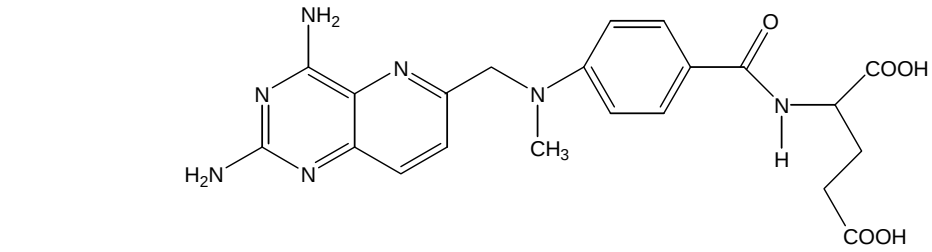
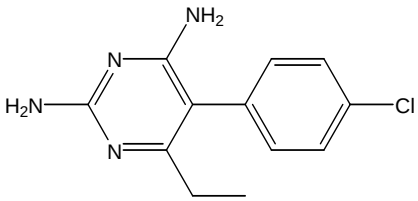


Figure 5 Structures of the main Type-1 antifolates.

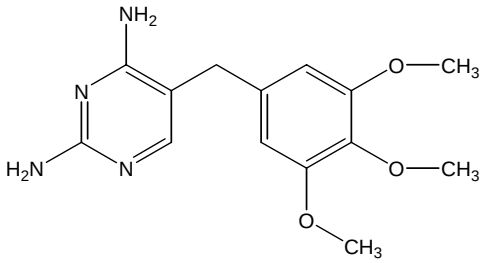
Source: Delfino (2002)



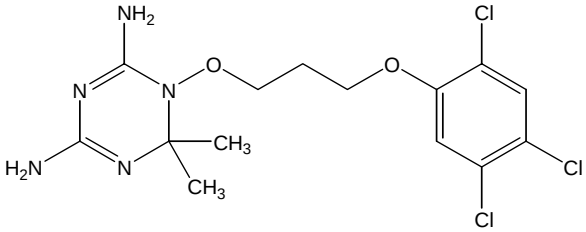
Methotrexate



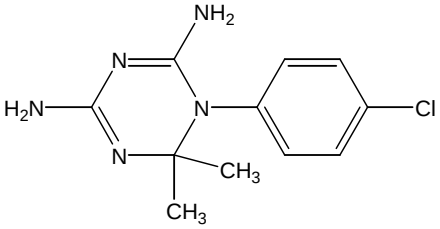
Pyrimethamine



Trimethoprim



WR99210



Cycloguanil

Figure 6 Structures of the main Type-2 antifolates.

Source: Delfino (2002)

Drug-resistance and design of new inhibitors

Drug-resistance can be caused by different modifications in the parasite cell. For *P. falciparum*, it is apparent that a gene is mutated, generating a mutant protein with lower drug affinity. Sirawaraporn and co-worker made a kinetic study of several mutant *Pf*DHFR, naturally occurring or not, and proposed a model for the contribution of each mutation to antifolate resistance (Sirawaraporn *et al*., 1997). They also proposed an evolutionary tree, reproduced in Figure 7. According to this model, *Pf*DHFR S108N, the only single mutant found in nature, has inhibition constants approximately ten times greater than the wild-type enzyme. Additional mutations in residues 51 and 59 of this mutant generate double mutants N51I+S108N and C59R+S108N, which are 10 to 50 times more resistant to PYR and CYC than the wild-type enzyme. From these mutations, the triple mutants N51I+C59R+S108N and C59R+S108N+I164L are generated; they are 40 to 200 times more resistant than the wild-type *Pf*DHFR. Finally, the quadruple mutant N51I+C59R+S108N+I164L is highly resistant to both drugs (Yuvaniyama *et al*., 2003). But WR99210 is the most effective inhibitor againsts both wild type and mutant *Pf*DHFR enzyme in vitro (Yuvaniyama *et al*., 2003).

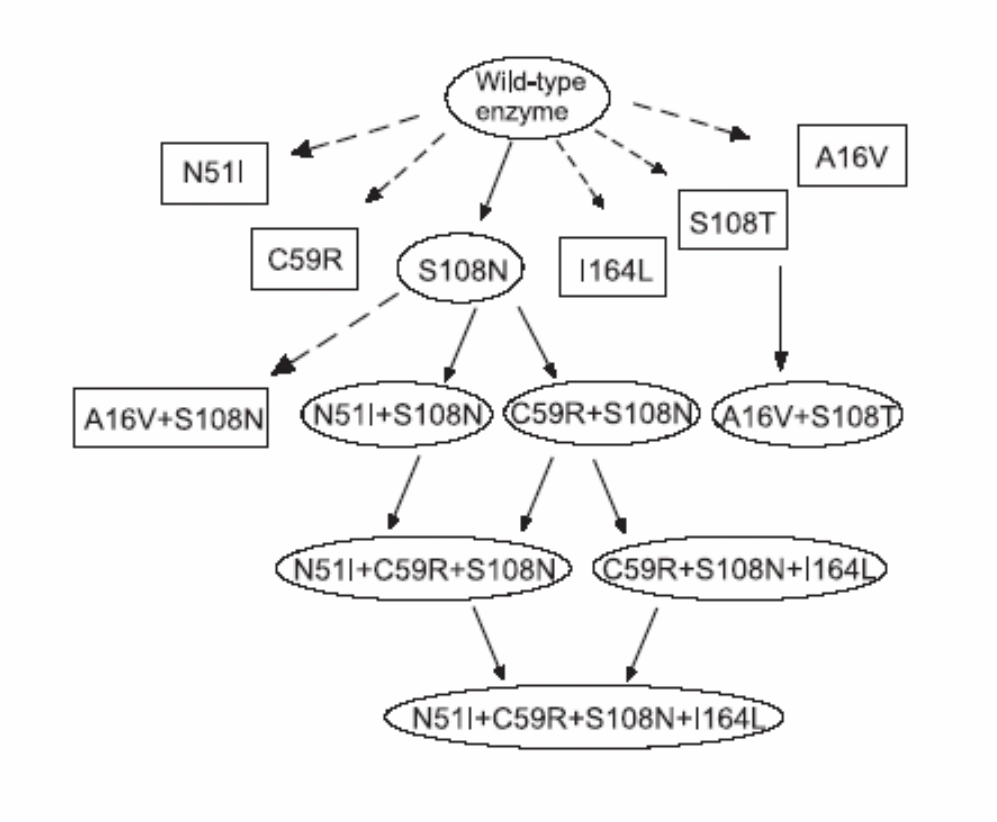


Figure 7 Evolutionary tree for the development of mutations in *Pf*DHFR, proposed by Sirawaraporn and co-workers (*Proc. Natl. Acad. Sci*. **94**, 1124-1129, 1997). Solid arrows indicate probable trajectories for mutation development, while dashed arrows indicate unlikely trajectories. Ellipse-surrounded mutants are naturally-occurring and rectangle-surrounded mutants have not been found in nature.

Source: Sirawaraporn (1997)

In this study, we have focused on the system of WR99210 and wild type *Pf*DHFR enzyme. A series of synthetic WR99210 derivatives has been synthesized and tested antimalarial activities by National Center for Genetic Engineering and Biotechnology, National Science and Technology Development Agency (BIOTEC), Thailand. Therefore, to be better understanding the structural properties of WR99210 derivatives and their activities within the binding site, the molecular modeling is required.

Computer-aided molecular modeling allows predicting the biological activity, toxicity and the nature of the pharmacophore. Nowadays, many medicinal chemists use Quantitative Structure Activity Relationships (QSAR) methods because it hoped that the procedure will minimize the number of compounds that synthetic chemists should prepare and the time needed to discover new drug candidates. Actually, the correlation of physicochemical properties with biological activity is believed to offer a useful tool for the design of new drugs.

Quantitative structure–activity relationship (QSAR) enables the investigators to establish a reliable quantitative structure–activity and structure–property relationships to derive an in silico QSAR models to predict the activity of novel molecules prior to their synthesis. The overall process of QSAR model development can be divided into three stages namely; the data preparation, data analysis, and model validation, representing a standard practice of any QSAR modeling. Three-dimentional quantitative structure-activity relationship (3D-QSAR) methods (Cramer *et al.,* 1996; Oprea *et al.*, 1997 and Greco *et al.,* 1997), comparative molecular field analysis (CoMFA), was applied to investigate the local physicochemical properties involved in the interaction between ligand and receptor. The widely used CoMFA calculates steric and electrostatic properties according to Lennard-Jones and Coulomb potentials (Crammer *et al*., 1988; Kubinyi, 1997). The contour maps derived from the CoMFA models permitted an understanding of the steric and electrostatic requirements for ligand binding. As a consequence, the structural variations in the training set that gives rise to variations in the molecular fields at particular regions of the space are

correlated to biological activities serving as a guide to the design of novel inhibitors. Form the literatures, 3D-QSAR methods have been successfully used to generate models for various antimalarial chemotherapeutic agents, such as Artemisinin (Cheng *et al*., 2002), Sulfonamides (Agrawal *et al*., 2001), Tryptanthrins (Bhattacharjee *et al*., 2004) and Chalcones (Xue *et al*., 2004).

A series of WR99210 analogues has been designed to have two substituents at the the position 2 on the triazine ring, and five substituents on the aromatic ring as shown in Figure 8. The substituent at position 2 has the stereochemistry in *(R)*- or *(S)*-configuration, therefore, probing which enantiomeric form, *(R)*- or *(S)*-form, has higher binding affinity. Molecular docking calculations were performed to investigate the orientation of ligands in the binding pocket and estimated free energies of binding between ligands and bound enzyme. Then, the best configuration from docking result was selected and constructed to perform the comparative molecular fields analysis (CoMFA). Also, the interaction between WR99210 and the active site of the *Pf*DHFR based on quantum calculation have been investigated to understand the nature of the particular interactions. The correlation of these results obtained from molecular docking calculations, CoMFA analysis and quantum calculations validate each other. All investigations in this work can be integrated to be a beneficial guideline to design and predict new and more potent compounds which active against wild type *Pf*DHFR.

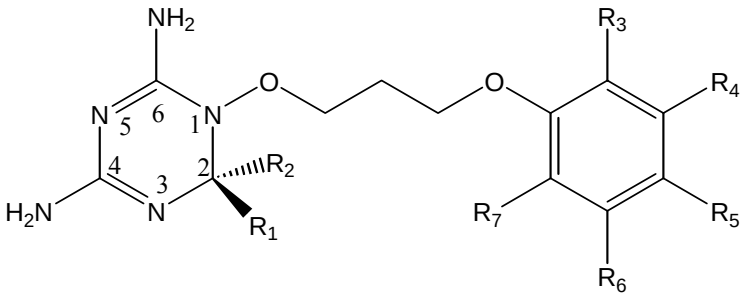


Figure 8 Structure of WR99210 analogues and the numbering atoms.

Objectives

In the present work, structure-based drug design and ligand-based drug design approaches using Molecular docking, Quantum chemical calculations and 3D-QSAR have been applied to WR99210 derivatives with the aims of:

1. To investigate the interactions between *(R)*- and (*S)*-WR99210 derivative inhibitors and active site of *Pf*DHFR enzyme and to probe which enantiomeric form *(R)- or (S)*- form, having higher binding affinity by Molecular Docking.
2. To study the nature of particular interactions between WR99210 and amino acids in the active site of *Pf*-DHFR by using quantum chemical calculations
3. To set up three dimensional quantitative structure-activity relationships of WR99210 derivatives against wild-type of *Pf*DHFR enzyme.
4. To design of potent wild type of *Pf*DHFR inhibitor.

LITERATURE REVIEW

QSAR and 3D-QSAR in drug design

In 1964 Free and Willson derived a mathematical model that describes the presence and absence of certain structural features. Also in 1964, the linear free-energy-related Hansch model (sometimes called the "extra-thermodynamic approach") was published. Classical quantitative structure-activity relationships (QSAR) method describe structure-activity relationships in terms of biological activities of drug and physiochemical properties or indicator variables which encode certain structural features (Free Wilson analysis) or with aromatic group or molecular properties, such as lipophilicity, polarizability, electronic and steric properties (Hansch analysis). Although relationship between lipophilicity and unspecific biological properties, have been know since the turn of our century, the independent publications of the Free Willson method and of Hansch analysis, mark a milestone in the development of QSAR.

Since then, QSAR equations have been used to describe thousands of biological activities within different series of drugs and drug candidates. Especially enzyme inhibition data have been successfully correlated with physicochemical properties of the ligands (Ramsden, 1990; Kubinyi, 1993; Hansch and Leo, 1995; Kubinyi, 1997). In certain case, where X-ray structures of the proteins became available, the results of QSAR regression models could be interpreted with the additional information from the three-dimensional (3D) structure.

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

1. 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.

2. In 1988, a key publication appeared in the *Journal of the American Chemical Society* (Cramer *et al.*) and the method was called comparative molecular field analysis (CoMFA). Cramer *et al.* developed a promising new approach to structure/activity correlation. Its characteristic features were (1) representation of ligand molecules by their steric and electrostatic fields, sampled at the intersections of a three-dimensional lattice, (2) a new "field fit" technique, allowing optimal mutual alignment within a series, by minimizing the RMS field differences between molecules, (3) data analysis by partial least-squares (PLS), using cross-validation to maximize the likelihood that the results had predictive validity and (4) graphic representation of results, as contoured three-dimensional coefficient plots. CoMFA was exemplified by analyses of the affinities for 21 varied steroids to corticosteroid- and testosterone-binding otential. From these results, a set of the steroid-binding affinity values unknown during the CoMFA analysis was well predicted.

3. Approptiate software became commercially available (Sybyl Molecular Modelling Software)

This molecular field-based method constituted the first real 3D-QSAR method. Since 1988, many successful CoMFA applications proved the value of this method, especially in case where classical QSAR fail. A few hundred publications (for example in antimalarial chemotherapeutic agents, such as Artemisinin (Cheng *et al*., 2002), Sulfonamides (Agrawal *et al*., 2001), Tryptanthrins (Bhattacharjee *et al*., 2004) and Chalcones (Xue *et al*., 2004)), several reviews (e.g. Kubinyi *et al*., 1994; Kim *et al*., 1995; Green *et al*., 1995; Martin *et al.,* 1996 and Blankley, 1996) and books (Kubinyi *et al.,* 1993 and 1998) have appeared on this subject. In contrast to Hansch or Free Wilson analysis, CoMFA is better suited to describe ligand-receptor interactions, because it considers the properties of the ligand in their bioactive conformations. As the result of a CoMFA analysis, regions in space are identified that are favorable or unfavorable for ligand-receptor interaction. Now CoMFA become

widely used, especially to derive quantitative models for enzyme inhibition constants and other binding affinities.

Structure of *Pf*DHFR-TS enzyme

DHFR is an enzyme that has been studied for a long time. In the past, only the primary structure of *Pf*DHFR has been determined but the tertiary structure of the plasmodial enzyme is not available. Homology modeling can help explain structure-activity relationships and possibly help in the design of new antimalarial agents. In 1996, Lemcke and co-worker constructed a three-dimensional model of the DHFR-domain from *P. falciparum* by homology building. The initial step in the model building was a structural alignment of four X-ray structures of DHFR from different species. Subsequently the sequence of *P. falciparum* was aligned to the four structurally aligned sequences. The three-dimensional model of the plasmodial DHFR was obtained by amino acid substitution in the human DHFR template followed by the modification of five loops (three insertions, two deletions). Finally, the model was subjected to a stepwise energy minimization. And then, Santos-Filho proposed a low-resolution model for both the wild type and the pyrimethamine (Pyr)/cycloguanil (Cyc) cross-resistant mutant type *Pf*DHFR, based on homology modeling using chicken liver DHFR as a template. The built models contain five α helices, eight β sheets, eight tight turns and several loops. The Ramachandran plot for the models shows 95.3 and 100% of the amino acid residues in the favorable regions for the whole enzymes and for the active sites, respectively. Furthermore, they made a preliminary analysis of the complexes Pyr/Cyc-wild DHFR and Pyr/Cyc-mutant DHFR in order to explain the probable mechanism of resistance. Their results show that the steric factor may be the main structural cause of *P. falciparum* resistance toward antifolate drugs.

Knowledge of the structures of the wild-type and mutant DHFRs is very useful for the development of antifolates with high binding affinities for these enzymes. In the absence of information about their structures from X-ray diffraction and nuclear magnetic resonance, it is necessary to rely on modeling of the enzymes based on

homology with DHFR from other species (McKie *et al.,* 1998; Warhurst *et al.,* 1998; Lemcke *et al.,* 1999; Rastelli *et al.,* 2000). Rational design of new antifolates can be made to guide synthesis of individual compounds or combinatorial libraries. The compounds can be conveniently screened against wild-type and mutant DHFR. Promising compounds can then be screened against the parasites for biological activity, and against mammalian systems for possible toxicity *in vitro* and *in vivo.*

The recently solved crystal structures of the wild-type and mutant *P. falciparum* DHFR-TS complexed with pyrimethamine or WR99210, a dihydrotriazine derivative (Yuvaniyama *et al.*, 2003) have given the molecular description of the antifolate target, and yielded insights into the mechanisms of resistance resulting from the mutations. In addition, the structures reveal some unique features of interactions among the 3 domains of the molecule that help to explain their kinetic properties, and may be exploited for development of new inhibitors. *Pf*DHFR-TS is a dimeric enzyme, with 2 subunits of TS in extensive contact, together forming 2 active sites in the contact area. The overall structural features bear some similarities with those of Leishmania major DHFR-TS (LmDHFR-TS) (Knighton *et al.,*1994) and the recently solved DHFR-TS of Cryptosporidium hominis (ChDHFR-TS) (O'Neil *et al.*, 2003), which are the only 2 other protozoal enzymes of known structure. Here they present three crystal structures of PfDHFR-TS, from three strains of *P. falciparum*: the wild type (TM4/8.2), the double mutant (K1 CB1) and the the quadruple mutant (V1/S). The wild type and the quadruple mutant are in complex with NADPH, dUMP and WR99210 (a substituted dihydrotriazine), whereas the double mutant is in complex with NADPH, dUMP and pyrimethamine (Pyr). The parasites bearing these double mutant (C59R/S108N) and quadruple mutant (N51I/C59R/S108N/I164N) enzyme show high resistance to pyrimethamine and cycloguanil but are still sensitive to WR99210. The structures provide explanations for the observed poor binding of Pyr in contrast to the good binding of WR99210, and should allow the design of new families of active site inhibitors. In addition, they provide new strategies for the design of inhibitors that work through disruption of interdomain or intersubunit.

DHFR Inhibitors

Several drugs which inhibit the action of *Pf*DHFR thus disrupting the folate cycle, have been used against *falciparum* malaria. The most extensively used antifolates are pyrimethamine (PYR) and cycloguanil (CYC). PYR is a potent and selective inhibitor of *Pf*DHFR, used alone or in combination with other drugs, such as sulfadoxine (combination known as SP). SP is cheap, practicable and highly effective in some African countries (Winstanley, 2000); however, it is prone to the rapid emergence of resistance. CYC is the metabolic product of proguanil, a drug developed in Britain during World War II. It is also a potent drug, and can be used alone or in combination with chloroquine or atovaquone. PYR and CYC have a low level of toxicity and, if used in the correct dose, are completely tree of side-effects. But its use is also affected by increasing parasite resistance. WR99210 is an experimental drug, which is highly effective in vitro against strains of *Pf*DHFR. However, WR99210, which differ from the compromised inhibitor in having a flexible side-chain, still binds tightly with the mutant enzyme and retrain its antimalarial efficacy. In 1996, Rieckman and co-worker investigated an in vivo-in vitro model of PS-15 and WR99210 that was used to assess the antimalarial activity of PS-15 and its metabolite, WR99210, against *Plasmodium falciparum*. WR99210, an antifolate triazine compound, was given as a single oral dose of 30 mg/kg to 8 *Saimiri sciureus* monkeys and, 3 months later, the parent compound, PS-15, was given similarly to the same monkeys. Serum samples were collected at various times after drug administration, serially diluted with control serum, and their antimalarial activity in vitro was determined against the multidrug-resistant K_i isolate of *P.falciparum*. Serum concentrations of PS-15 and WR99210 were estimated by high performance liquid chromatography. The maximum dilutions of serum that inhibited parasite growth were 20- to 86-fold higher 3 and 6 h after administration of PS-15 than following WR99210 administration. Substantial serum antimalarial activity was observed even at 48 h after medication with PS-15. Serum drug concentrations provided further evidence that PS-15 was absorbed far better from the gastrointestinal tract than WR99210. The substantial and sustained activity of PS-15 suggests that a single dose, or several smaller doses given once a day, should be effective in curing drug-resistant

infections of *P.falciparum*. In 1997, Hekmat-Nejad and Rathod presentd novel antifolates which it against drug-resistant *P. falciparum*. With emerging drug resistance in *Plasmodium falciparum,* novel antifolates effective against pyrimethamine-resistant and cycloguanil-resistant dihydrofolate reductase (DHFR) are in demand. Based on structural similarity to cycloguanil, it has been proposed that WR99210, and its metabolic precursor PS-15, exerts selective antimalarial activity by binding tightly to both drug-sensitive and drug-resistant DHFR. In the present study, Linweaver–Burk plots and Ackermann–Potter plots reveal that both forms of malarial DHFR bind WR99210 at subnanomolar concentrations. It is not necessary to invoke an alternate target for WR99210 in *P. falciparum.* The present studies confirm that malarial DHFR offer potential binding interactions in the folate binding pocket distinct from those exploited by pyrimethamine and cycloguanil. These kinetic studies also provide a useful framework for the design and interpretation of future structural studies on drug-resistant DHFR from *P. falciparum.* Finally, Fidock and Wellems found that transformation with human dihydrofolate reductase renders malaria parasites insensitive to WR99210 but does not affect the intrinsic activity of proguanil. Increasing resistance of *Plasmodium falciparum* malaria parasites to chloroquine and the dihydrofolate reductase (DHFR) inhibitors pyrimethamine and cycloguanil have sparked renewed interest in the antimalarial drugs WR99210 and proguanil, the cycloguanil precursor. To investigate suggestions that WR99210 and proguanil act against a target other than the reductase moiety of the *P. falciparum* bifunctional DHFR–thymidylate synthase enzyme, we have transformed *P. falciparum* with a variant form of human DHFR selectable by methotrexate. Human DHFR was found to fully negate the antiparasitic effect of WR99210, thus demonstrating that the only significant action of WR99210 is against parasite DHFR. Although the human enzyme also resulted in greater resistance to cycloguanil, no decrease was found in the level of susceptibility of transformed parasites to proguanil, thus providing evidence of intrinsic activity of this parent compound against a target other than DHFR. The transformation system described here has the advantage that *P. falciparum* drug-resistant lines are uniformly sensitive to methotrexate and will complement transformation with existing pyrimethamine-resistance markers in functional studies of *P. falciparum* genes. This system also provides an approach for

screening and identifying novel DHFR inhibitors that will be important in combined chemotherapeutic formulations against malaria.

Resistance to antifolates occurred soon after their deployment as antimalarials. Although antifolate resistance in cancer and antibacterial chemotherapy can occur through a variety of mechanisms, in malaria this has been shown to be due to a change in DHFR which resulted in a decrease in binding strengths of the inhibitors. Early studies with *P. chabaudi* (Sirawaraporn and Yuthavong, 1984) that showed that the enzyme from resistant parasites was a mutant form which was not as susceptible to these inhibitors as those from the wild-type parasites, were followed by discovery of changes in codons of the gene for DHFR of antifolate-resistant *P. falciparum*, which strongly indicated that the amino acid changes led to the observed resistance (Cowman et al., 1988; Peterson *et al.,* 1998; Hyde *et al.,* 1990; Foote *et al.,* 1990). Conclusive proof of the importance of these mutations in development of resistance was given by transfection of wild-type parasites with constructs bearing mutant forms of DHFR (Wu *et al.,*) 1996. Recombinant mutant DHFR showed a decrease in binding with pyrimethamine and cycloguanil in an expected manner (Sirawaraporn *et al.,* 1997). A scheme for evolution of resistance could be derived as a result of stepwise mutations starting with the Ser108Asn mutation, which was shown to be the optimal mutation in leading to both decreased binding affinity for inhibitors and retention of enzyme activity (Sirawaraporn *et al.,* 1997). This led to a moderate level of resistance, followed by mutations at other positions leading to higher levels of resistance. In addition to 108, mutations associated with resistance have been shown as Ala16Val, Asn51Ile, Cys59Arg and Ile164Leu (Cowman et al., 1988; Peterson *et al.,* 1998; Hyde *et al.,* 1990; Foote *et al.,* 1990; Wu *et al.,* 1996; Sirawaraporn *et al.,* 1997), and Cys50Arg (Plowe *et al.,* 1997; Cortese *et al.,* 1998). In addition, a five amino acid insertion after codon 30, which is a repeat of the preceding five, was also observed in areas of high resistance to pyrimethamine–sulfadoxine, although its role in conferring resistance is unclear (Plowe *et al.,* 1997; Cortese *et al.,* 1998). With multiple mutations at various sites, increased resistance to pyrimethamine was shown to be correlated to a decrease in binding with recombinant DHFR, both for those found in the field and those that have not been found and were artificially created

(Sirawaraporn *et al.,* 1997). Resistance to pyrimethamine due to these mutations was mostly correlated with resistance to cycloguanil. Exceptions were found for the Ala16Val+Ser108Thr mutant, found previously in the field to be associated only with cycloguanil resistance, and the artificially created Ala16Val mutant, both of which showed decreased binding to cycloguanil, while retaining high affinity to pyrimethamine. The difference in loci of resistance has raised the possibility that new antifolates can be developed that would be effective against parasites resistant to conventional ones. The possibility of developing effective antifolates against resistant parasites is also raised from the fact that they apparently have limitations in mutation: many mutant DHFRs created in the laboratory are not found in nature, and were shown to be either poorly resistant to the drugs or to have insufficient catalytic activity (Sirawaraporn *et al.,* 1997). Other factors, including inefficient expression and insufficient stability of the enzyme, can also contribute to limitation in mutation possibilities.

In addition to mutation in the coding region of the DHFR gene, there may be other mechanisms of resistance to antifolates of *P. falciparum*. Increase in the amount of the DHFR-TS gene has been raised as a mechanism of pyrimethamine resistance generated in the laboratory through increased drug exposure (Thaithong *et al.,* 1992). Although gene amplification has been ruled out (Thaithong *et al.,* 1992; Cowman *et al.,* 1988; Peterson *et al.,* 1998; Hyde *et al.,* 1990; Foote *et al.,* 1990; Wu *et al.,* 1996; Sirawaraporn *et al.,* 1997; Plowe *et al.,* 1997; Cortese *et al.,* 1998; Lin *et al.,* 2000), increase in DHFR-TS can occur through a variety of mechanisms. It has been shown that human TS can regulate its own expression by binding with its mRNA (Lin *et al.,* 2000). It is possible that parasite DHFR-TS can similarly auto regulate its own expression by binding with its own mRNA and the regulation can be perturbed by drug binding with the enzyme, offering another mechanism for drug resistance (P. Rathod, personal communication). The possibility of many strains of the parasite to salvage folates can also modify its response to antifolates.

Drug resistance of the malaria parasites is one of the most important problems in malaria control. A major part of this problem is the resistance of the parasites to

antifolates, inhibitors of dihydrofolate reductase (DHFR), a validated drug target which is a part of the bifunctional enzyme dihydrofolate reductase-thymidylate synthase (DHFR-TS) (Olliaro and Yuthavong, 1999; Hyde, 2002; Yuthavong, 2002). Not only have such inhibitors as pyrimethamine, cycloguanil and their derivatives been compromised by widespread resistance, but also their synergistic combinations with sulfa-drugs, which act on another enzyme, dihydropteroate synthase (DHPS) in the folate de novo synthesis pathway, are also under threat. Resistance to DHFR inhibitors is explained by the occurrence of point mutations in the enzyme, as shown by the changes in the base sequences of its gene, leading to changes in amino acids in certain positions of the enzyme (Cowman *et al.,* 1988; Peterson, Walliker and Wellems, 1988). In order to develop new effective antifolates, it is important to know the molecular structure of DHFR-TS, how it interacts with inhibitors and substrates, and how mutations affect the interactions. DHFR-TS from malarial parasites consists of the DHFR domain and the TS domain, joined through the junction region (Bzik *et al.,* 1987). The bifunctional enzyme from *Plasmodium falciparum* (*Pf*) has 608 amino acid residues, with 231 residues of the DHFR domain at the N-terminus and 288 residues of the TS domain at the C-terminus joined by an 89-residue junction region. The amino acid identity of *Pf*DHFR with DHFR from bacterial and mammalian species ranges from 24 to 42%, while the identity of *Pf*TS with TS from the other species ranges from 42 to 63%. Significant sequence differences, especially around the active site regions, between DHFR of Plasmodium and the human host have allowed the development of differential inhibitors as antifolate drugs, in contrast with the more conserved nature of TS. However, in spite of the conserved sequence of *Pf*TS, the fact that its activity depends on the integrity of the DHFR and the junction region (Shallom *et al.*, 1999; Wattanarangsan *et al.,* 2003) is a unique feature that may be exploited in the development of new differential TS inhibitors.

The structures of the wild-type and mutant enzymes bound with inhibitors give insight into the signicance of the S108N mutation in reducing the binding affinity of pyrimethamine (Yuvaniyama *et al.*, 2003). The modes of inhibitor binding as found from the X-ray structures generally confirm earlier predictions from modeling studies (McKie *et al.*, 1998; Warhurst, 1998; Lemcke *et al.*, 1999; Rastelli *et al.,* 2000;

Warhurst, 2002; Sardarian *et al*., 2003). Inhibitors like pyrimethamine and cycloguanil have a rigid *p*-chlorophenyl group in which the Cl atom lies very close to the side-chain of residue 108. The crystal structures show that changing the side-chain from that of Ser to Asn would create steric constraint to the binding of the *p*-chlorophenyl group, as well as dislocation of the nicotinamide ring of NADPH, which in turn a.ects the binding of the inhibitor. This prediction was supported by the .nding that a pyrimethamine analogue with Cl in the *m*- rather than the *p*-position binds much better with the mutant enzymes carrying the S108N mutation, and indeed has much better antimalarial activity against the mutant parasites than pyrimethamine (McKie *et al*., 1998; Tarnchompoo *et al*., 2002; Sardarian *et al.,* 2003; Kamchonwongpaisan *et al.,* 2004). It had also been earlier noted that the reduced sensitivity to pyrimethamine and cycloguanil of DHFR residue-108 mutants (Asn, Thr, Gln, Cys) tends to be correlated with the side-chain length and bulk (i.e. molecular weight, molar refractivity, volume and surface area), while the charge on the non-H atom distal to the backbone and the lipophilicity of the side-chain apparently have no impact (Warhurst, 1998).

The general validity of this conclusion was further supported by more recent extensive binding data for these plus other mutants with bulky side-chains (Tarnchompoo *et al*., 2002; Sardarian *et al*., 2003; Kamchonwongpaisan *et al*., 2004). The S108N mutant also binds poorly with other pyrimethamine derivatives with bulky groups in place of the *p*-Cl, and the binding was generally progressively poorer for the double (S108N+C59R) mutant. Such cumulative effects on inhibitor binding are seen more clearly with the triple (S108N+N51I+C59R) and the quadruple (S108N+N51I+C59R+I164L) mutants, leading to increasingly poorer antimalarial effects of pyrimethamine and cycloguanil (Kamchonwongpaisan *et al.,* 2004). Structural explanation of this cumulative effect is given by the observation that the N51I mutation causes a substantial main-chain movement of residues 48–51 by 0.5–2.2 Å, with a 48–49 peptide flip, and the I164L mutation causes a minor shift (0.3–0.5 Å) of residues 164–167, which together open up the active site gap and weaken the binding of small inhibitors like pyrimethamine (Yuvaniyama *et al*., 2003).

Removal of the *p*-Cl or replacement with *m*-Cl led to better binding with the mutant DHFR. A number of other inhibitors with flexible groups which can avert the potential steric constraint around the side-chain of residue 108 also proved to have better binding a.nities with mutant enzymes. These include WR99210 and other cycloguanil analogues, and some trimethoprim analogues, which show good antimalarial activities with the mutant parasites, reflecting the binding affinities with the mutant enzymes which they carry (Kamchonwongpaisan *et al.,* 2004; Sirichaiwat *et al*., 2004). In contrast to rigid inhibitors like pyrimethamine and cycloguanil, WR99210 with a flexible side-chain can adopt a conformation that fits well in the active site of the mutant enzymes, explaining its effectiveness as an antimalarial against resistant parasites. The structure of WR99210 bound with the quadruple mutant as compared with the wild-type DHFR-TS shows that its side-chain can avert steric hindrance caused by the S108N and subsequent mutations, and can further interact with other residues in the active site, explaining its retention of its binding affinity (Yuvaniyama *et al*., 2003).

The importance of the side-chain of residue 108 in inhibitor binding was assessed by site-speci.chain mutation (Sirawaraporn *et al.,* 1997; Tarnchompoo *et al.,* 2002). Significant correlations were found between the K_i values for pyrimethamine and cycloguanil and the length and bulk of the side-chain of this residue. Furthermore, pyrimethamine derivatives with bulky groups in place of the *p*-Cl bind even more poorly with the S108N and multiple mutant enzymes in the series, indicating that the binding a.nities are reduced with increased steric interference (Tarnchompoo *et al.*, 2002; Sardarian *et al.,* 2003; Kamchonwongpaisan *et al.,* 2004). Removal of the *p*-Cl or replacement with *m*-Cl led to better binding with the mutant enzymes as expected.

Modelling of binding of cycloguanil and its derivatives has helped to understand the specific reduction in its binding with another mutant DHFR, namely, the A16V+S108T mutant (Rastelli *et al.,* 2000). Cycloguanil carries two methyl groups at the 2-position, in contrast to pyrimethamine that has only one ethyl group in the equivalent position. The binding model shows that one of the two methyl groups is in steric conflict with the enlarged side-chain of Val16. In support of this model, it

was found that 2-desmethyl cycloguanil and other derivatives with only one substituent in the 2-position generally have better binding to the mutant relative to the wild-type enzyme than the disubstituted derivatives (Yuthavong *et al.,* 2000).

Although the crystal structures of the complex of cycloguanil with the enzyme and its mutants are not known, that of the cycloguanil analogue WR99210 has been solved (Yuvaniyama *et al.,* 2003), and found to be similar to the modeled structure (Rastelli *et al.,* 2000), although there are significant differences. Modeling of new analogues into the active site should therefore help in designing effective inhibitors against resistant parasites.

METHODS OF CALCULATIONS

Molecular Docking Calculations

1. Overview of the Molecular Docking

Molecular docking has become a useful tool in drug discovery efforts. The screening of large databases for possible lead compounds is becoming a routine procedure. Early approaches to the docking problem, such as the original DOCK algorithm (Kuntz *et al*., 1982), considered both a rigid ligand and receptor and used shape complementarity to identify the native-like orientation of the ligand. While advantageous in terms of computational cost, rigid docking has very limited applicability since the large majority of small ligands are flexible. To circumvent the constraints of rigid-body docking, several low-energy conformations of the ligand can be pre-generated, then docked, and the best solution (s) can be chosen according to a scoring/energy function. However, this approach quickly becomes inefficient for ligands with multiple bonds. To avoid the combinatorial 'explosion' associated with ligands containing many rotatable bonds, several incremental construction algorithms, such as Hammerhead, DOCK4.0 (Kuntz *et al*., 1982) and FlexX (Rarey *et al*., 1996) have been implemented. At the same time, improvement in computing power and advances in the energy calculation techniques such as the introduction of a grid-based receptor field representation and the use of internal coordinates, make the simulations of continuously flexible ligands computationally feasible.

A number of docking programs such as AutoDock (Garrett *et al*., 1998), MCDOCK (Liu *et al*., 1999), ICM-dock (Totrov *et al*., 1997), QXP (McMartin *et al*., 1997) and GOLD (Jones *et al.,* 1995 utilize this observation and implement Metropolis Monte-Carlo or genetic algorithms to search for the global minimum of the energy function in the continuous conformational space of the ligand.

In this work, AutoDock (Morris *et al.,* 1998*)* has been used exclusively along with one of its search methods called Lamarckian genetic algorithm (LGA). The

macromolecule is rigid and fixed while the ligand is flexible and can both translate and rotate. Rapid grid-based method was used for finding the lowest binding energy of the bound conformation. These grids are calculated in advanced (i.e. before the actual docking) and one for each atom type present. The size of the grid box can be set manually and placed at a certain position. These boxes create maps over the molecules that are used during the docking, this to exclude atoms of no interest and also to speed up the docking calculations. The final results from the docking simulations provide the specific position and orientation of the best binding (in term of lowest free energy) between the protein and the ligand.

2. Algorithms

2.1 AutoDock and GOLD

AutoDock explores the conformational space of the ligand using the Lamarkian genetic algorithm (Figure 9), which is a hybrid of a genetic algorithm (GA) with an adaptive local search (LS) method (Morris *et al*., 1998). In this approach, the ligand's state is represented as a chromosome, which is composed of a string of real-valued genes describing the ligand location (three coordinates), orientation (four quaternions) and conformation (one value for each torsion). The simulation is started by creating a random population of individuals. It is followed by a specified number of generation cycles, each consisting of the following steps: mapping and fitness evaluation, selection, crossover, mutation and elitist selection. Each generation cycle is followed by a local search. The solutions are scored using an energy-based scoring function, which includes terms accounting for short-ranged van der Waals and electrostatic interactions, loss of entropy upon ligand binding, hydrogen bonding and solvation.

Similarly, GOLD employs a genetic algorithm. The ligand's state is encoded by a chromosome, representing its conformation and hydrogen bonding. The conformation of the ligand is represented by a binary string, in which every byte encodes for one torsional angle. Each torsions is allowed to vary between -180° and

180° in step-sizes of 1.4 Å. Two integer strings encode mappings suggesting possible hydrogen bonds between the protein and the ligand. The first of these strings encodes a mapping of acceptors in the ligand to the donor hydrogens in the protein. The second string encodes a mapping of donor hydrogens in the ligand to the acceptor atoms in the protein. On decoding a chromosome, GOLD utilizes least-squares fitting to form as many of these hydrogen bonds as possible. In the evolutionary development of the ligand conformations the program employs an island model, in 757 which several subpopulations of chromosomes are created at the beginning instead of one large population. The genetic operations include migration of individual chromosomes between the subpopulations, crossover and mutation. In order to preserve diversity within the population GOLD employs a niching technique, namely, when adding a new individual to the population, the number of individuals in the population that inhabit the same niche as the new chromosome is determined. If there are more than a specified number of individuals in the niche, then the new individual replaces the worst member of the niche rather than the worst member of the total population. Two individuals share the same niche if the RMSD between their donor and acceptor coordinates is less than 1.0 Å. The fitness of a new individual is assessed using a scoring function, which includes energy terms accounting for hydrogen bonding, short-ranged van der Waals interaction between the ligand and protein, and the ligand internal energy. The latter is a sum of ligand steric and torsional energies.

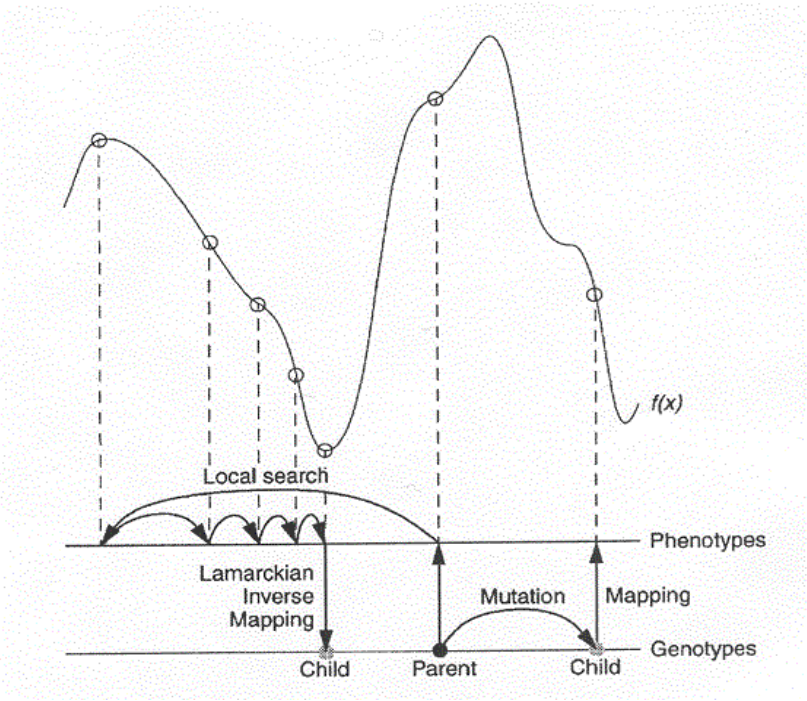


Figure 9 Diagram of Lamarckian Genetic Algorithm. Source: Garett, 1998.

2.2 DOCK and FlexX

Both DOCK (Ewing *et al*., 1997) and FlexX (Rarey *et al*., 1996) employ an incremental reconstruction algorithm. In this algorithm rigid anchor (DOCK) or base (FlexX) fragments are identified first. At the next step, the selected fragment is placed into the active site of the receptor using a sphere-matching procedure (DOCK) or a hashing technique (FlexX). The complete ligand is constructed by adding the remaining components step by step. At each step of reconstruction a specified number of optimal partial solutions are selected for the next extension step. In DOCK the solutions are scored using energy, contact or chemical scoring functions. The energy scoring function, which was used in this study, includes van der Waals and electrostatic components. In FlexX the scoring is done using a modified Böhm scoring function, which includes the following terms: entropic, which accounts for loss of entropy upon ligand binding; hydrogen bonding; ionic, accounting for electrostatic interactions; aromatic, which accounts for interactions between aromatic

groups; and lipophilic, which accounts for hydrophobic interactions. All terms, except the entropic term, are scaled by a corresponding heuristic distance and an angle dependent penalizing function.

2.3 ICM

ICM performs flexible ligand docking via global optimization of the energy function (Abagyan *et al.*, 1994). The energy terms include the internal energy of the ligand based on the ECEPP/3 force field, as well as van der Waals, hydrogen-bonding, electrostatic and hydrophobic ligand/receptor interaction terms pre-calculated on the grid for computational efficiency. A Monte-Carlo Minimization (MCM) procedure in the internal coordinate space is employed to find the global minimum of the energy function. Each step of the algorithm consists of a random conformational change of two types, torsional or positional, followed by the local minimization. Torsional move involves complete randomization of a single arbitrarily chosen torsion angle. A positional move involves a pseudo-Brownian random translation and rotation of the ligand as a whole. ICM uses an analytical gradient minimizer, which finds the local minima of the energy function more rapidly than the simplex minimizer or the stochastic search alone. To improve convergence, multiple MC runs from several starting positions are performed. The VLS (virtual library screening) scoring function used in ICM consists of the internal force-field energy of the ligand and the ligand/receptor interaction energy. The latter includes van der Waals terms, a hydrophobicity term based on the solvent accessible surface buried upon binding, an electrostatic solvation term calculated using a boundary-element solution of the Poisson equation, hydrogen-bond interaction terms and an entropic term proportional to the number of flexible torsions in the ligand.

3. System Studied

In this present work, the crystal structure (PDB code: 1J3I) from Brookhaven Protein Data Bank (http://www.rcsb.org/pdb/), *Pf*DHFR complexed with WR99210, was chosen to be the starting geometry (Yuvaniyama *et al*. 2003). The initial structures of ligands were generated by molecular modeling software Sybyl 7.0 program and constructed in to two groups: *(R)*-configurations and *(S)*-configurations. The geometries of these compounds were subsequently optimized using the PM3 method. The general structures of WR99210 analogues and numbering atoms used in this study were shown in Figure 8 and the chemical structures of WR99210 derivatives was shown in Table 1

To investigate the interaction between WR99210 derivatives and *Pf*DHFR enzyme, the system was performed by the automated molecular docking which is a module provided in the advanced docking program AutoDock 3.0. The Lamarckian genetic algorithm (LGA) was applied to deal with the inhibitor-enzyme interactions as shown in Figure 12. Briefly, the LGA described the relationship between the inhibitor and the enzyme by the translation, orientation, and conformation of the inhibitors. These so-called "state variables" were the inhibitors' genotype, and the resulting atomic coordinates together with the interaction and the intramolecular energies were the inhibitors' phenotype. The environmental adaptation of the phenotype was reverse-transcribed into its genotype and became heritable traits. Each docking cycle, or generation, consisted of a regimen of fitness evaluation, crossover, mutation, and selection. A Solis and Wets local search performed the energy minimization on a user-specified proportion of the population. The docked structures of the inhibitors were generated after a reasonable number of evaluations. The whole docking operation could be stated as follows.

First, the enzyme molecules were checked for polar hydrogens and assigned for partial atomic charges then, the PDBQs file was created, and the atomic solvation parameters were also assigned for the macromolecules. Meanwhile, all of the torsion

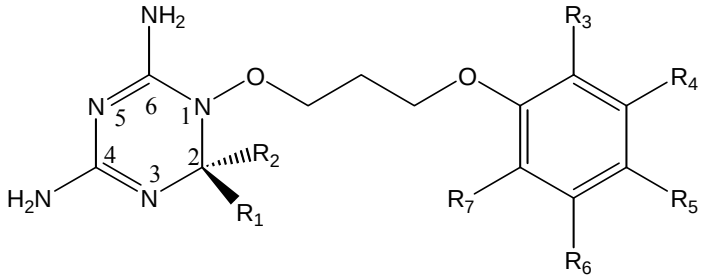
angles of the inhibitors in order to be explored during molecular docking were defined. This allowed the conformational search of inhibitors during the process of docking.

Second, the 3D grid was generated by the AutoGrid algorithm to evaluate the binding energies between inhibitor and enzyme. In this stage, the *Pf*DHFR was embedded in the 3D grid and a probe atom was placed at each grid point. The affinity and electrostatic potential grid were calculated for each type of atom in the inhibitor. The energies of a particular inhibitor configuration were found by trilinear interpolation of affinity values and electrostatic interaction of the eight grid points surrounding each of atoms in an inhibitor.

Third, a series of the docking parameters were set on. Not only the atom types but also the generations and the number of runs for the LGA algorithm were edited and properly assigned according to the requirement of the Amber force field. The number of generations, energy evaluations, and docking runs were set to 27 000, 250 000, and 50, respectively. The kinds of atomic charges were taken as Kollman-all-atom (Weiner *et al*., 1984) for *Pf*DHFR and Gasteiger-Huckel (Gasteiger *et al*., 1980) for inhibitors.

Finally, the docked complexes of inhibitor-enzyme were selected according to the criteria of interacting energy combined with geometrical matching quality. These complexes were used as the starting conformation for further energetic minimization and geometrical optimization before the final model was achieved.

Table 1 Structures of 22 WR99210 derivatives and experimental biological activities against wild type *Pf*DHFR.



Cpds. No	R_1	R_2	R_3	R_4	R_5	R_6	R_7
34	H	CH_3	Cl	H	Cl	Cl	H
237	H	C_6H_4-4-OC_3H_7	Cl	H	Cl	Cl	H
238	H	C_6H_4-4-OC_4H_9	Cl	H	Cl	Cl	H
396	H	C_6H_5	H	H	H	H	H
399	H	C_6H_5	H	H	Cl	H	H
401	H	C_6H_5	H	H	$NCOCH_3$	H	H
402	H	C_6H_5	H	H	$COOCH_3$	H	H
404	H	C_6H_5	H	H	COOH	H	H
411	H	C_6H_5	H	OCH_3	H	H	H
412	H	C_6H_5	OCH_3	H	H	H	OCH_3
413	H	C_6H_5	CH_2-C_6H_5	H	H	H	H
414	H	C_6H_5	$COCH_3$	H	H	H	H
415	H	C_6H_5	H	H	$COOCH_2$-C_6H_5	H	H
427	H	C_6H_5	F	F	F	F	F
430	H	C_6H_5	Cl	H	Cl	H	H
431	H	C_6H_5	H	Cl	Cl	H	H
441	H	C_6H_5	Cl	H	F	H	H
442	H	C_6H_5	H	H	Br	H	H
443	H	C_6H_5	Cl	H	H	H	H
446	H	C_6H_5	H	NO_2	H	H	H
447	H	C_6H_5	NO_2	H	H	H	H
454	H	C_6H_5	H	H	COOH	H	H

4. Overview of the Free Energy Function

AutoGrid and AutoDock were introduced a new kind of scoring function that is used during and at the end of the dockings. It is based on the principles of QSAR (quantitative structure-activity relationships) and was parameterized using a large number of protein-inhibitor complexes for which both their structures and inhibition constants, or K_i, were known.

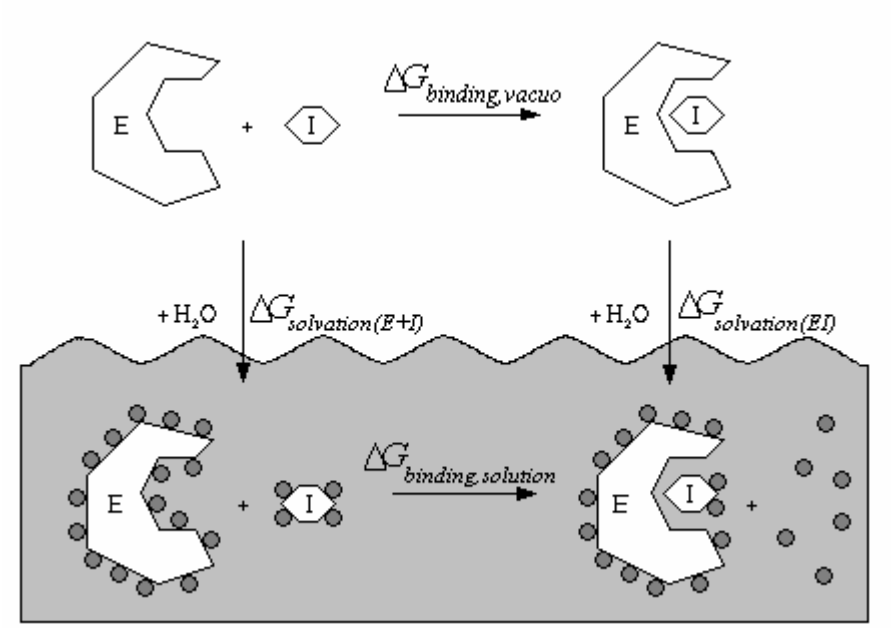


Figure 10 The thermodynamic cycle for the binding of an enzyme. Source: Garrett, 1999.

The above diagram (Figure 10) shows the thermodynamic cycle for the binding of an enzyme, *E*, and an inhibitor, *I*, in both the solvated phase and *in vacuo*. Note the solvent molecules are indicated by filled circles: they tend to be ordered around the larger molecules, but when *E* and *I* bind, several solvent molecules are liberated and become disordered. This is an entropic effect and is the basis of the hydrophobic effect. The solvent ordering around *E* and *I*, when both bound and unbound, is strongly influenced by the hydrogen bonding between these molecules. These hydrogen bonds between solvent and *E,* and solvent and *I,* contribute enthalpic stabilization, and is something we can estimate in our new free energy function.

According to Hess's law of heat summation, the change in free energy between two states will be the same, no matter what the path. So we can calculate the free energy of binding in solvent by the following equation:

$$G_{binding,solution} = G_{binding,vacuo} + G_{solvation(EI)} - G_{solvation(E+I)} \quad (1)$$

Since we can calculate $G_{binding,vacuo}$ from our docking simulation, and can estimate the free energy change upon solvation for the separate molecules *E* and *I*, and for the complex, *EI*, $G_{solvation(EI)}$ and $G_{solvation(E+I)}$ respectively, then it is also possible to calculate the free energy change upon binding of the inhibitor to the enzyme in solution, $G_{binding,solv}$. Thus, we can estimate the inhibition constant, K_i, for the inhibitor, *I*.

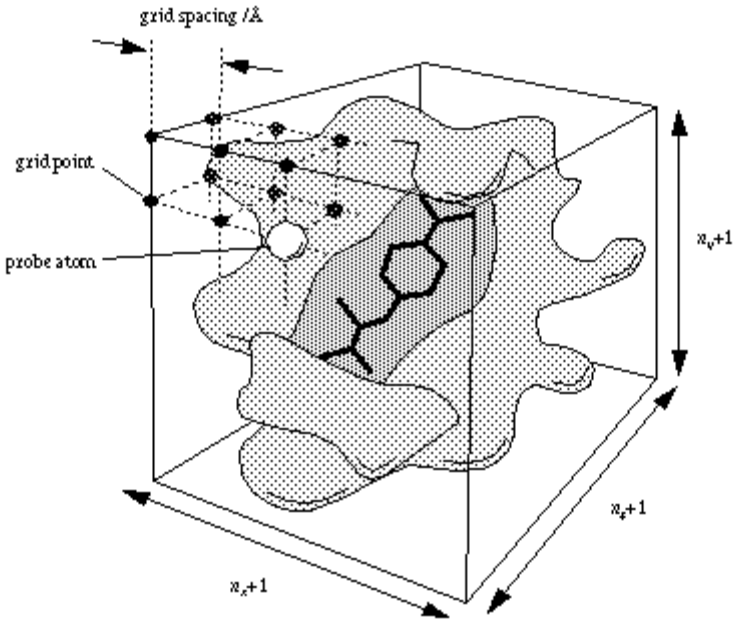


Figure 11 Grid space and probe atoms of enzyme around the ligand. Source: Garrett, 1999.

The ligand can be seen in the centre of the grid map as shown in Figure 11, buried inside the active site of the protein. In this case, the grid map encompasses the whole protein. The grid spacing is the same in all three dimensions.

The pairwise potential energy, *V(r)*, between two non-bonded atoms can be expressed as a function of internuclear separation, *r*, as follows,

$$V(r) = \frac{Ae^{-br}}{r} - \frac{C_6}{r^6} \tag{2}$$

Graphically, if r_{eqm} is the *equilibrium internuclear separation*, and ε is the *well depth* at r_{eqm} (Figure 12), then:

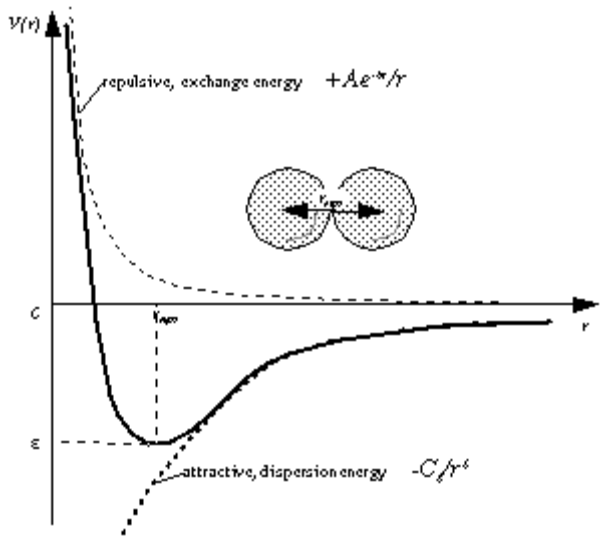


Figure 12 Diagram of Van der waals. Source: Garett, 1999.

The exponential, repulsive, exchange energy is often approximated thus,

$$\frac{A}{r}e^{-br} \approx \frac{C_{12}}{r_{12}} \tag{3}$$

In addition to the atomic affinity grid maps, AutoDock requires an electrostatic potential grid map. Polar hydrogens must be added, if hydrogen-bonds are being modeled explicitly. Partial atomic charges must be assigned to the macromolecule. The electrostatic grid can be generated by AutoGrid, or by other programs such as MEAD or DELPHI, which solve the linearized Poisson-Boltzmann equation. AutoGrid calculates Coulombic interactions between the macromolecule and a probe of charge ***e***, $+1.60219 \times 10^{-19}$ C; there is no distance cutoff used for electrostatic interactions. A sigmoidal distance-dependent dielectric function is used to model solvent screening, based on the work of Mehler and Solmajer,

$$\varepsilon(r) = A + \frac{B}{l + ke^{-\lambda Br}} \tag{4}$$

where as: B = the dielectric constant of bulk water at 25 °C = 78.4; A = -8.5525, λ = 0.003627 and k = 7.7839 are parameters.

Quantum Chemical Calculations

In order to investigate the inhibitor-enzyme interactions of the complex structure of *Pf*DHFR/WR99210, quantum chemical calculations were applied to study the nature of partial interactions between the WR99210 inhibitor and the active site of *Pf*DHFR.The understanding of this particular interaction will be helpful to design of new potent inhibitors.

1. System Studied

The structural model system was obtained from the 2.33 Å resolved crystal structure of WR99210 bound to wild type (PDB code 1J3I) of *Pf*DHFR. Based on this structure, we adopted the model system consisting of nine chains surrounding the active site of WR99210/*Pf*DHFR complex at least one atom interacting with any of the atoms of the WR99210 inhibitor within the interatomic distance of 7.0 Å. The additional residues which are not interacting within 7.0 Å were also included into the studied system to conserve the connection between the amino acids within the chain of the active site. Thus, the investigated system containing the WR99210 inhibitor bound into the active site with 21 residues was generated (see Figure 13). All amino acids, set to be in their neutral form, were terminated by the N- and C-terminal ends of cut residues capped with H atom from the adjacent residues. Bond lengths and torsion angles of model system were assumed to be the same as in the X-ray structure. Due to the lack of hydrogen atoms in crystal structure, addition of hydrogen atoms provides the complete structure of the model and their positions were optimized by the semiempirical (PM3) method. The optimization was carried out taking into account the approximation of the heavy atoms fixed (HAF). Optimized structure was used as the starting geometry for all higher accuracy of calculations.

The amino acid residues surrounding the active site included in this study consist of 23 residues; Ile14, Cys15, Ala16, Val45, Leu46, Trp48, Cys50, Asn51, Asp54, Met55, Tyr57, Phe58, Cys59, Met104, Ser108, Ser111, Ile112, Pro113 Phe116, Leu119, Ile164, Gly165 and Thr185.

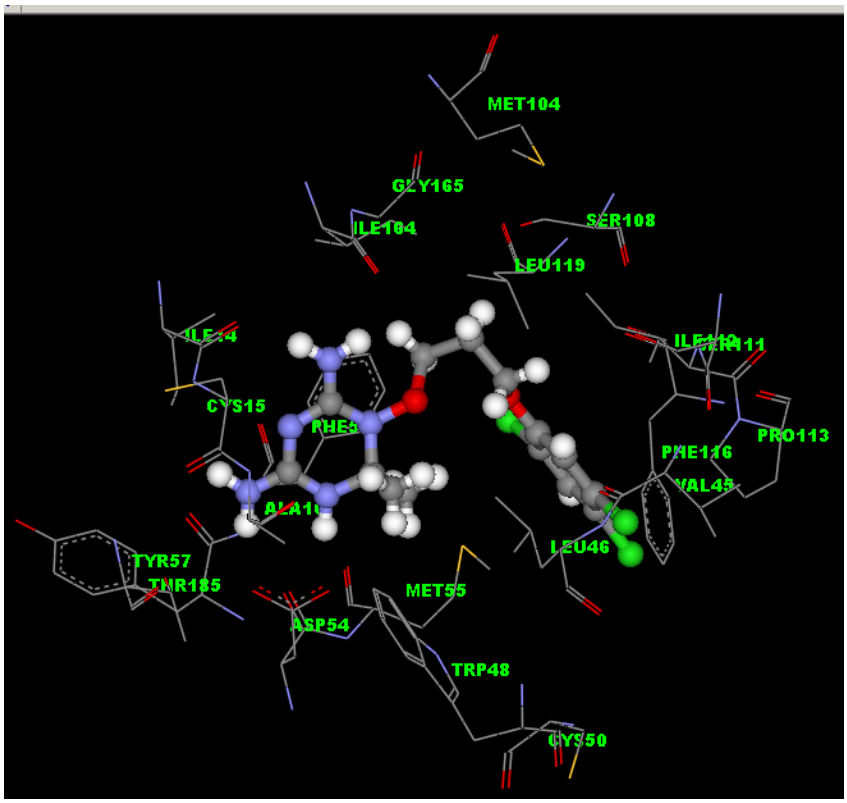


Figure 13 3D Schematic representation of the WR99210 inhibitor with residues surrounding the active site having at least one atom interacts to any atom of inhibitor within 7.0 Å.

2. Interaction Energy

The interaction energy (INT) between WR99210 inhibitor and the individual residues (X_i) calculated by quantum chemical calculations at MP2/6-31G(d) level were performed by using the geometry optimization complex structures as explained from previous section. The interaction energy (INT) of each WR99210-X_i pair is defined as following:

$$INT_{[WR99210+Xi]} = E_{[WR99210+Xi]} - E_{[WR99210]} - E_{[Xi]}, \quad (5)$$

where $E_{[WR99210+Xi]}$ is the pair energy of each WR99210 and residue X_i, $E_{[WR99210]}$ and $E_{[Xi]}$ are the energies of WR99210 and each individual residue, respectively.

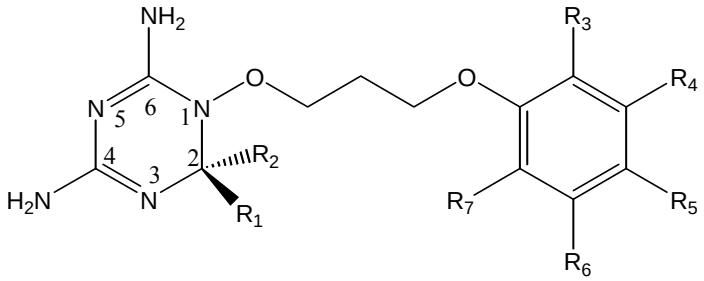
Three Dimentional Quantitative Structure-Activity Relationships (3D-QSAR) Analysis

1. Structures and Biological Data of *Pf*DHFR Inhibitors

The general structure of WR99210 and its numbering atoms used in this study was shown in Figure 8. The chemical structures of WR99210 derivatives and their biological activities against wild type of *Pf*DHFR were obtained from the National Center for Genetic Engineering and Biotechnology National Science and Technology Development Agency (BIOTEC), Pathumthani Thailand. The potency has been expressed as the negative logarithm of measured K_i (nM) (log1/K_i) where K_i is the inhibition constant of the dissociation interaction between WR99210 derivatives and *Pf*DHFR receptor. The log1/K_i was used as a dependent variable in the 3D-QSAR study. The whole data set of WR99210 derivatives against wild type of *Pf*DHFR, based on the range of these biological data, was divided into two subsets: the training set 35 compounds (Table 2) and the test set 10 compounds (Table 3). The traning set was used for 3D-QSAR analysis. In addition, the test set, the selected compounds from various structures of WR99210, was kept for testing the actual prediction of model.

In the previous studies, 3D-QSAR approaches have been successful to apply to obtain structural requirements of another antimalarial drug such as, Artemisinin (Cheng *et al*., 2002), Sulfonamides (Agrawal et al., 2001), Tryptanthrins (Bhattacharjee et al., 2004) and Chalcones (Xue et al., 2004). In this study, 3D-QSAR based on Comparative Molecular Field Analysis (CoMFA) was treated on WR99210 derivatives of *Pf*DHFR inhibitors to study the electrostatic and steric interactions between inhibitor and binding pocket which may help to design new potent inhibitors.

Table 2 Structures of 35 WR99210 derivatives, used in the training set and experimental biological activities against wild type of *Pf*DHFR.



Cpds. No	R_1	R_2	R_3	R_4	R_5	R_6	R_7	Experimental Biological Activities ($\log 1/K_i$)
34	H	CH_3	Cl	H	Cl	Cl	H	9.52
35	CH_3	CH_3	Cl	H	Cl	Cl	H	9.30
237	H	C_6H_4-4-OC_3H_7	Cl	H	Cl	Cl	H	8.77
238	H	C_6H_4-4-OC_4H_9	Cl	H	Cl	Cl	H	8.49
396	H	C_6H_5	H	H	H	H	H	9.09
399	H	C_6H_5	H	H	Cl	H	H	9.46
402	H	C_6H_5	H	H	$COOCH_3$	H	H	9.49
404	H	C_6H_5	H	H	COOH	H	H	8.93
405	CH_3	CH_3	H	H	$COOCH_2$-C_6H_5	H	H	9.60
411	H	C_6H_5	H	OCH_3	H	H	H	8.75
412	H	C_6H_5	OCH_3	H	H	H	OCH_3	8.53
413	H	C_6H_5	CH_2-C_6H_5	H	H	H	H	8.30
414	H	C_6H_5	$COCH_3$	H	H	H	H	8.94
415	H	C_6H_5	H	H	$COOCH_2$-C_6H_5	H	H	8.24
416	CH_3	CH_3	H	OCH_3	H	H	H	9.32
417	CH_3	CH_3	OCH_3	H	H	H	OCH_3	9.12
419	CH_3	CH_3	H	H	NO_2	H	H	9.14

Table 2 (cont'd)

Cpds. No.	R_1	R_2	R_3	R_4	R_5	R_6	R_7	Experimental Biological Activities (log1/K_i)
420	CH_3	CH_3	H	H	Cl	H	H	9.23
421	CH_3	CH_3	H	H	$COOCH_3$	H	H	9.15
423	CH_3	CH_3	H	H	H	H	H	9.15
424*	CH_3	CH_3						9.28
426	CH_3	CH_3	H	H	C_6H_5	H	H	9.19
428	CH_3	CH_3	H	H	COOH	H	H	9.31
442	H	C_6H_5	H	H	Br	H	H	9.12
443	H	C_6H_5	Cl	H	H	H	H	9.19
444	CH_3	CH_3	H	NO_2	H	H	H	9.19
445	CH_3	CH_3	NO_2	H	H	H	H	9.24
446	H	C_6H_5	H	NO_2	H	H	H	9.17
447	H	C_6H_5	NO_2	H	H	H	H	9.09
448	CH_3	CH_3	Cl	Cl	H	H	H	9.21
449	CH_3	CH_3	Cl	H	Cl	H	H	9.10
451	CH_3	CH_3	Cl	H	H	Cl	H	9.14
452	CH_3	CH_3	Cl	H	F	H	H	9.36
453	CH_3	CH_3	H	H	Br	H	H	9.12
454	H	C_6H_5	H	H	COOH	H	H	9.06

*Compound 424

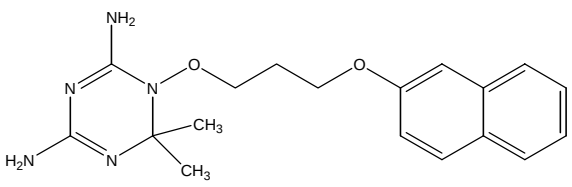


Table 3 Structures of 10 WR99210 derivatives, used in the test set and experimental biological activities against wild type of *Pf*DHFR

Cpds. No	R_1	R_2	R_3	R_4	R_5	R_6	R_7	Experimental Biological Activities ($\log 1/K_i$)
196	H	C_6H_5	Cl	H	Cl	Cl	H	9.22
401	H	C_6H_5	H	H	$NCOCH_3$	H	H	9.46
418	CH_3	CH_3	CH_2-C_6H_5	H	H	H	H	9.26
422	CH_3	CH_3	H	H	$HNCOCH_3$	H	H	9.16
427	H	C_6H_5	F	F	F	F	F	8.32
430	H	C_6H_5	Cl	H	Cl	H	H	9.22
431	H	C_6H_5	H	Cl	Cl	H	H	9.18
440	CH_3	CH_3	Cl	H	H	H	H	9.55
441	H	C_6H_5	Cl	H	F	H	H	9.26
450	CH_3	CH_3	H	Cl	Cl	H	H	9.10

2. Molecular Modeling and Quantum Chemical Calculations

Quantum chemical calculations were used to calculate geometries and chemical properties of all WR99210 derivatives. CoMFA results can be related to their biological activities. Therefore, molecular modelling is an important tool to predict the biological activity which is the function of geometries and chemical properties.

All WR99210 derivative structural geometries were initially constructed in *(R)*-configuration form and modified by using SYBYL 7.0 (SYBYL Molecular Modelling Softwares, Version 7.0, Tripos Associates, Inc., St. Louis, MO, 63144, USA, 2005.). The lowest energy conformer of each derivative was obtained from a conformational search analysis of the side chain substitutions. Then this lowest energy conformer of all WR99210 derivatives were performed by full optimization based on PM3 semiempirical molecular orbital method implemented in the GAUSSIAN 03 program. As results, the mulliken partial atomic charges were derived. Finallly, the optimized geometries of all WR99210 derivatives were used as the starting structure in next step, 3D-QSAR analysis.

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

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

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 this present study, 3D-QSAR CoMFA mothod was applied on the series of WR99210 derivatives and derived the contour maps to reveal the significance of the steric and electrostatic interactions. The structural variations in the molecular fields in their particular regions were also investigated. 3D-QSAR models can give an insight into the design of potent WR99210 inhibitors against wild type of *Pf*DHFR.

3.1 Comparative Molecular Field Analysis (CoMFA)

Comparative Molecular Field Analysis (CoMFA) was developed by Cramer *et a*l. in 1988. CoMFA is a powerful 3D-QSAR technique providing further insight into the relationships between the structure and function of these WR99210 analogues. This methodology is based on assumption that non-covalent forces dominate receptor-drug interactions and that these forces can be described in terms 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 *Pf*DHFR inhibitory activity.

3.1.1 CoMFA Set Up

3.1.1.1 Alignment Rule

A major requirement in CoMFA study is the alignment of each of all compounds relative to one another, so all of them have a comparable conformation and orientation in space. The relative interaction energies depend strongly on relative molecular positions. The compound with the highest inhibitory activity in WR99210 analogues was used as template for RMS-fit molecular alignments. All geometries of WR99210 derivatives were aligned in a 3D lattice by fitting them with template shown in Figure 14. All calculations were performed in SYBYL 7.0 molecular modeling software.

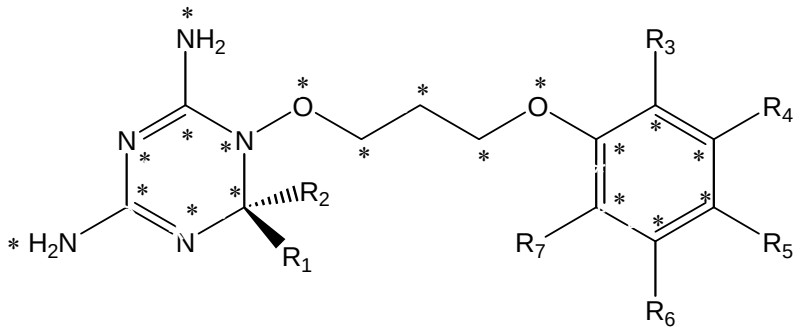


Figure 14 General structure of WR99210 derivatives, stars indicate the atom selected as the template for alignment rule.

3.1.1.2 Calculations of Interaction Energy

CoMFA cubic lattice was generated around these molecules based on the molecular volume of the structures. In this investigation, three different atoms, sp^3 carbon atom with +1 charge (default probe atom in SYBYL), sp^3 oxygen atom with –1 charge and H atom 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 in molecule were all calculated with CoMFA standard scaling and then put in a CoMFA QSAR table. In order to speed up the analysis and reduce the amount of 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.

a) 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:

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

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)

b) Electrostatic Fields

Electrostatic interactions are usually calculated 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. The general form of electrostatic interaction between two molecules is given by Equation 7.

$$E = \sum_i \sum_j \frac{q_i q_j}{r_{ij}} \quad (7)$$

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)

3.1.2 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) countour 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 setric 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.

3.2 Partial Least Squares Analysis (PLS) and Validations

Partial least squares (PLS) methodology was used for all 3D-QSAR analyses. The CoMFA and CoMSIA descriptors were used as independent variables and log ($1/K_i$) values were used as dependent variables in partial least squares regression analyses to derive 3D-QSAR models using the standard implementation in the SYBYL 7.0 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 15. 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^2_{cv} was calculated using equation 8:

$$q^2 = 1 - \frac{\sum (Y_{observed} - Y_{predicted})^2}{\sum (Y_{observed} - Y_{mean})^2} \qquad (8)$$

where $Y_{predicted}$, $Y_{observed}$ and Y_{mean} are predicted, actual and mean values of the target property (log $1/K_i$), respectively. $\Sigma(Y_{predicted} - Y_{observed})^2$ 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^2_{cv} and number of components, the conventional correlation coefficient r^2 and its standard errors (SEE) were also computed.

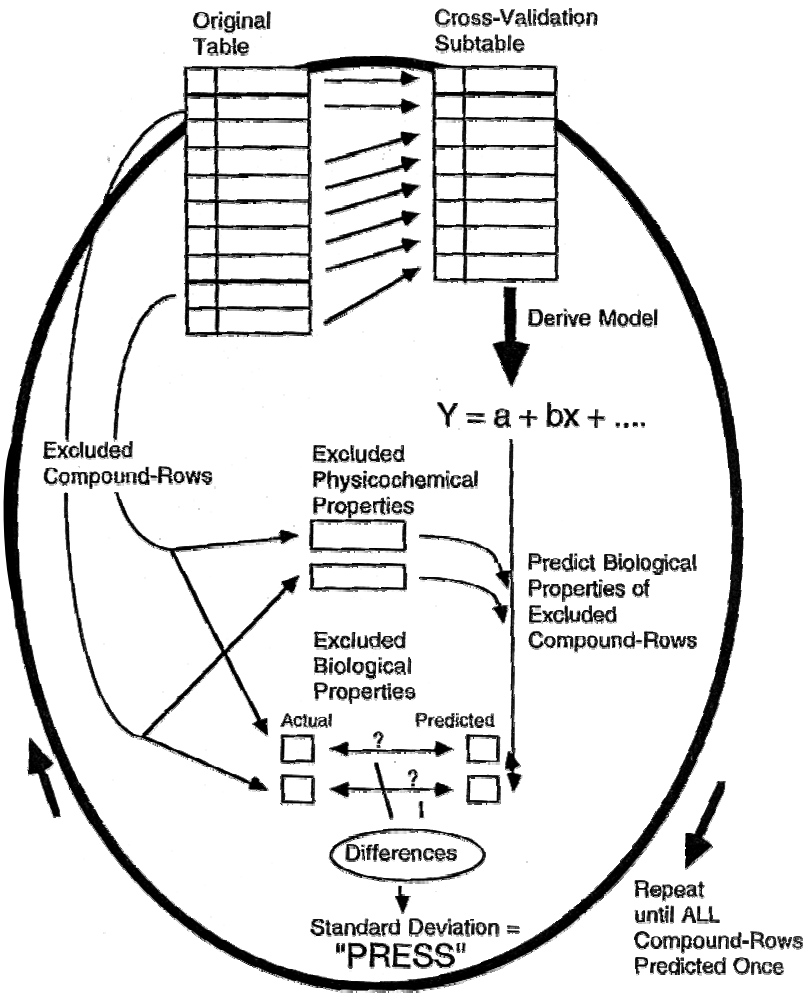


Figure 15 Cross-validated procedure

Source: Kubinyi (1993)

3.3 Predictive Ability

Q^2 or r^2_{cv} values were used to evaluate the overall predictive ability of the model. R^2_{cv} is calculated according to the following equation:

$$r_{cv}^2 = 1.0 - \frac{PRESS}{SSY} \tag{9}$$

where $$PRESS = \sum_{y} (Y_{predicted} - Y_{observed})^2 \tag{10}$$

$$SSY = \sum_{y} (Y_{observed} - Y_{mean})^2 \tag{11}$$

and the uncertainty of the prediction is defined as

$$S_{press} = \sqrt{\frac{PRESS}{n-k-1}} \tag{12}$$

The optimum number of components to be used in the non-cross-validation (conventional) analyses was defined as that yielding the highest r^2_{cv} value.

RESULTS AND DISCUSSION

Molecular Docking Investigation on *(R)*- and *(S)*- WR99210 Derivatives bound to Wild Type *Pf*DHFR enzyme

The autodock 3.0 program was used to perform an automatic docking exploration for conformation of *(R)*- and *(S)*-WR99210 derivatives in the active site of *Pf*DHFR. Two set of experiment were performed. The first set was designed to test autodocks' ability for wild type *Pf*DHFR/WR99210 complexes. Another, molecular docking calculations were performed to investigate the orientation of ligands in the binding pocket and estimate the free energies of the bound enzyme.

1. Comparison between X-ray and Docking Orientation

Docking simulations were applied to reproduce the crystal structures of WR99210 bound to *Pf*DHFR enzyme which is shown in Figure 16. The Autodock was used to predict the conformation of the ligands and compared with the X-ray crystallographic structure. The superposition and docked orientation of WR99210 are shown in Figure 17 and 18. In this study, the flexible WR99210 and fixed-torsion WR99210 were docked. The root mean square deviation (RMSD) of flexible WR99210 and fixed-torsion WR99210 with *Pf*DHFR are 0.98 and 0.58 Å, respectively, And the estimated free energies of binding are -11.83 and -10.06 kcal/mol, respectively. Bound conformation of WR99210 with the amino acids in the *Pf*DHFR active site can be used to explain interaction between ligands and the amino acids at the entrance of the *Pf*DHFR active site with orientation and hydrogen bonding interactions.

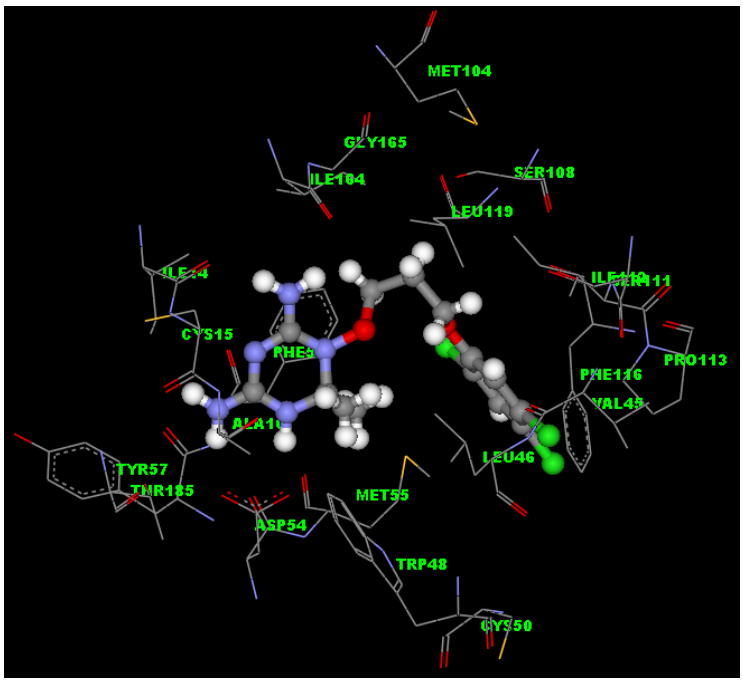
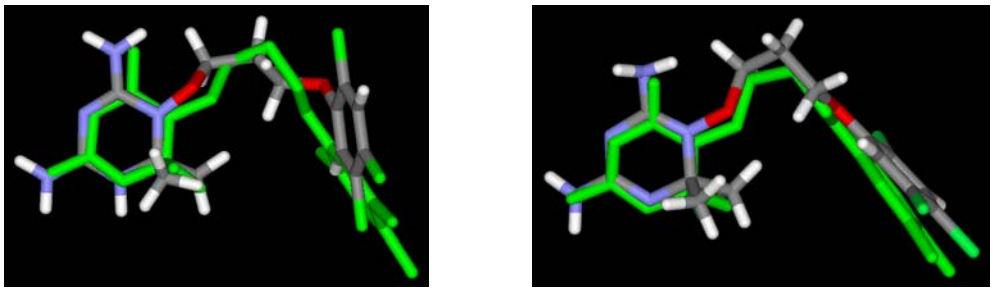


Figure 16 Bound conformation of WR99210 with the amino acids in the *Pf*DHFR active site, surrounding within 7.0 Å. Representing ball and stick of WR99210 is shown with the following atom colors: carbon, dark; nitrogen, blue; oxygen, red; sulfur, yellow; hydrogen, white; chlorine, green. Protein side chains are shown and the type of atoms color is similar to that of the WR99210 atom types.



(a) (b)

Figure 17 Superposition of WR99210 obtained from X-ray crystallographic structure (green) and docking calculations (a) flexible WR99210 and (b) fixed-torsion WR99210.

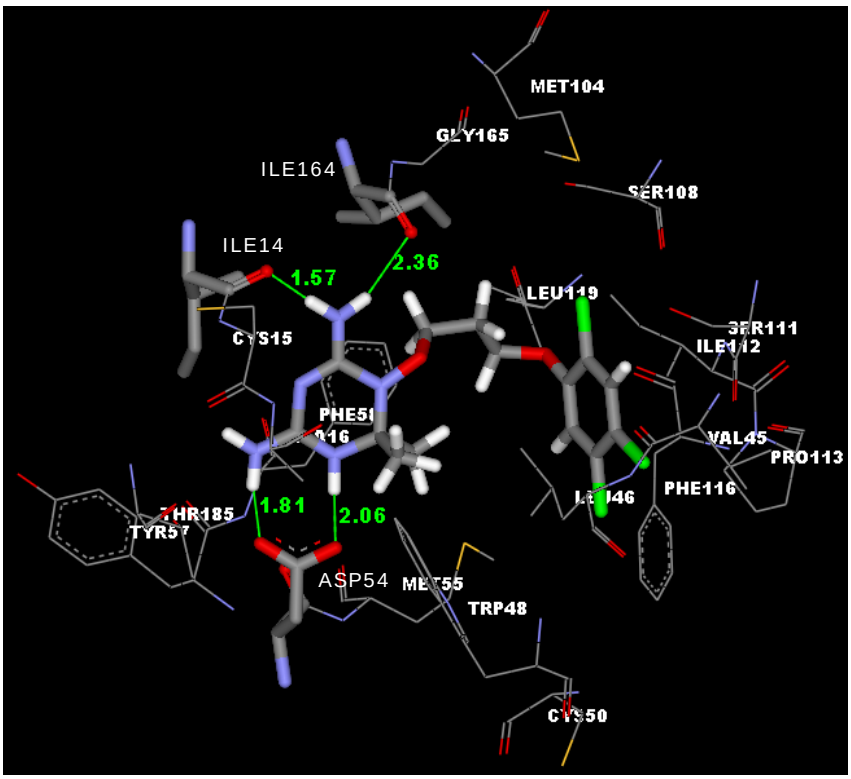


Figure 18 Bound conformation of flexible WR99210 with the amino acids in the *Pf*DHFR active site, surrounding within 7.0 Å. The hydrogen bonding interactions are shown in green lines.

The root mean square deviation (RMSD) values are less than 1.0. Therefore, the Autodock method and the set parameter could be used to search the enzyme binding conformations for further step.

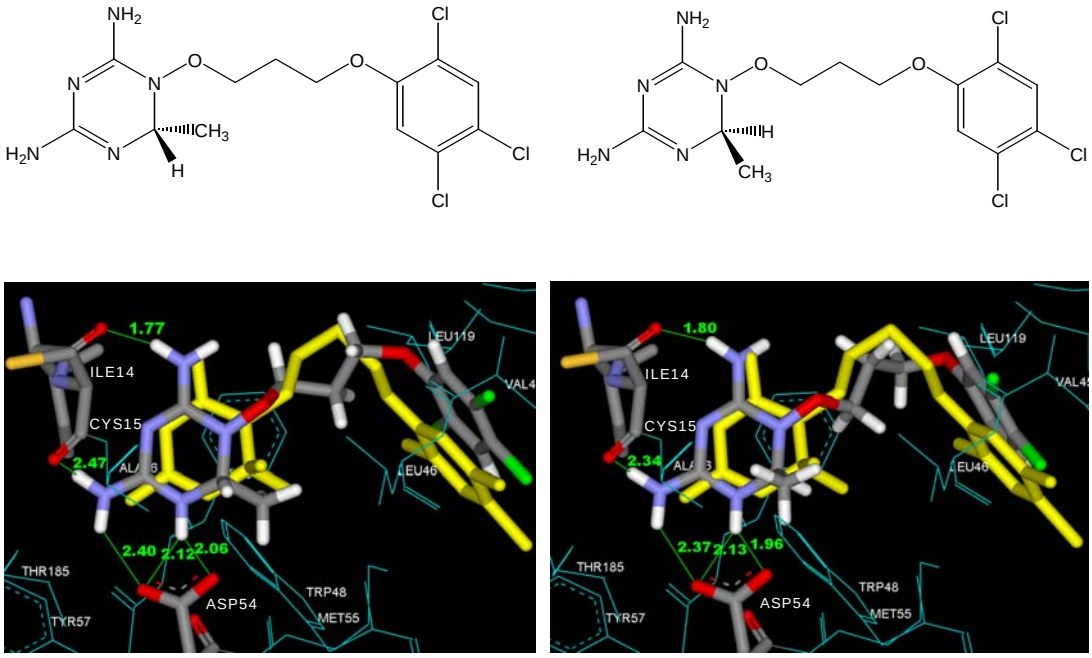
From docking results, it can be seen that the major residue contributors to the binding energy are Ile14, Asp54, Phe58 and Ile164. The Ile14, Asp54 and Ile164 residues form H-bond contacts with the carboxylate groups. Moreover, the interaction of the WR99210 with Phe58 found to be π-π interaction.

2. *(R-)* and *(S)*-Configurations of WR99210 Derivatives

Docking simulations were performed to calculate flexible *(R)*-and *(S)*-configuration of WR99210 derivatives bound to the *Pf*DHFR enzyme. The estimated free energies of binding results are presented in Table 4. The *(R)*-configuration of WR99210 derivatives were found to form strong hydrogen bonding in the *Pf*DHFR active site than *(S)*-configuration ones, except the comparison between (*(R)*-WR34) and ((*S*)-WR34), as expressed in Figure19. Moreover, if the substitution group at C_2 on triazine ring is a bulky group such as aromatic group, it shows clearly the different configuration of *(R)*- and *(S)*-configurations interacted with amino acids in the *Pf*DHFR active site because *(S)*-configuration has steric clash with Ala16 of the enzyme (Yuthavong *et al*, 2002) as shown in Figure 20. In addition, hydrogen bondings between WR99210 analogues and carboxylate group of amino acid Ile14, Cys15, Asp54 and Ile164 were also observed and presented in Figure19 and 20. It can be seen that the major residue contributors to the binding energy are Ile14, Cys15, Asp54 and Ile164. Docking results of the complex structures show estimated free energies of binding of analogs. The *(R)*-configurations energies were lower than those of the analogs with *(S)*-configurations, except compounds 238 and 412, suggesting that the enzyme preferentially binds the inhibitors with *(R)*-configurations. Therefore, *(R)*-configuration of WR99210 derivatives was selected and used in the CoMFA analysis.

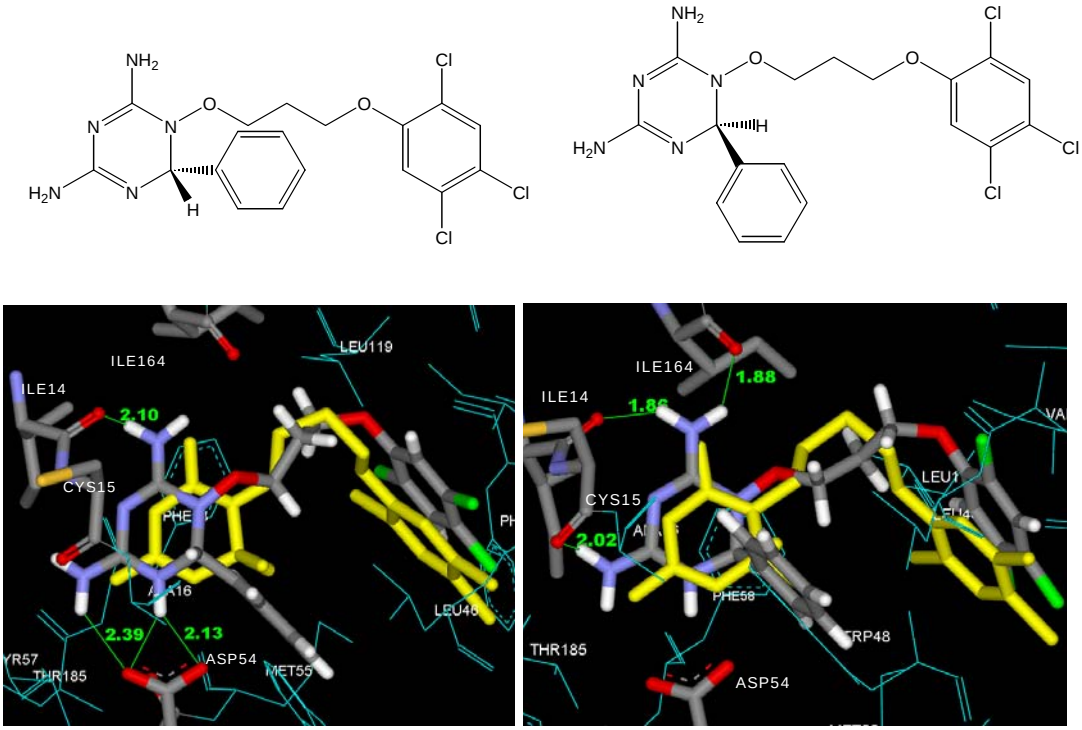
Table 4 Estimated free energies of binding (kcal/mol) of *(R)*- and *(S)*-WR99210 derivatives in the wild type *Pf* DHFR active site.

Compounds	*(R)*-configuration	*(S)*-configuration
34	-11.48	-10.83
196	-12.06	-11.61
237	-12.81	-11.17
238	-9.06	-10.78
396	-11.31	-10.08
399	-10.96	-10.73
401	-11.45	-9.70
402	-10.57	-9.98
404	-11.00	-8.73
411	-9.94	-9.65
412	-9.04	-9.80
413	-9.37	-9.17
414	-10.20	-10.16
415	-10.33	-7.74
427	-11.97	-8.91
430	-12.52	-11.44
431	-11.96	-10.71
441	-12.15	-12.08
442	-11.00	-10.49
443	-11.63	-10.79
446	-11.48	-10.83
447	-11.36	-10.60



(a) *(R)*-WR34 (b) *(S)*-WR34

Figure 19 Superposition of WR99210 obtained from X-ray crystallographic structure (yellow) and bound conformations of *(R)*- and *(S)*-WR34 derivatives with amino acids in the *Pf*DHFR active site, surrounding within 7.0 Å. The hydrogen bonding interactions are shown in green lines. (a) *(R)*-WR34 derivative and (b) *(S)*-WR34 derivative.



(a) *(R)*-WR196 (b) *(S)*-WR196

Figure 20 Superposition of WR99210 obtained from X-ray crystallographic structure (yellow) and bound conformations of *(R)*- and *(S)*-WR196 derivatives with amino acids in the *Pf*DHFR active site, surrounding within 7.0 Å. The hydrogen bonding interactions are shown in green lines (a) *(R)*-WR196 derivative and (b) *(S)*-WR196 derivative.

Particular Interaction Energies between WR99210 and Residues in *Pf*DHFR Active Site

The starting geometry was obtained from the X-ray structure and optimized by semiempirical method (PM3). In order to determine the position of hydrogen atoms due to the missing hydrogen atom in the X-ray structure, optimization was carried out taking into account the approximation the heavy atoms fixing (HAF). This structure (Figure 21) was used as the starting geometry for quantum chemical calculations.

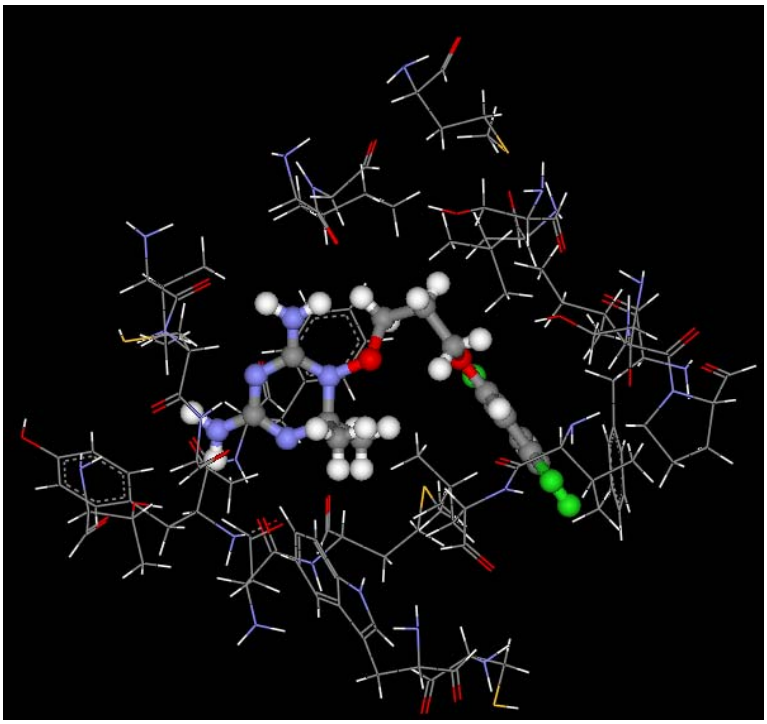
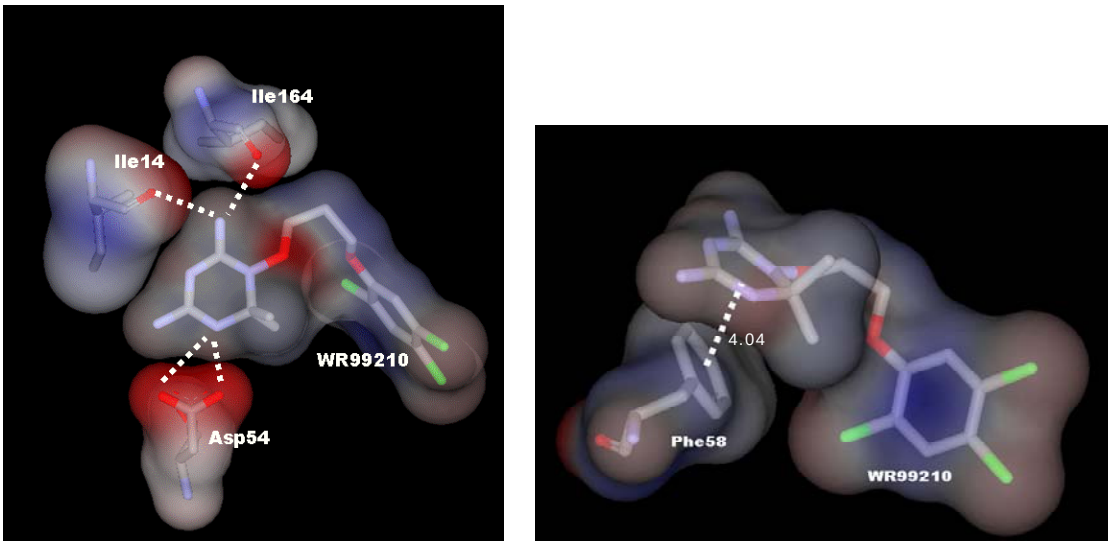


Figure 21 Optimized WR99210/*Pf*DHFR active site by PM3 method (HAF). Representing ball and stick of WR99210 is shown with the following atom colors: carbon, dark; nitrogen, blue; oxygen, red; sulfur, yellow; hydrogen, white; chlorine, green. Protein side chains are shown and the type of atoms color is similar to that of the WR99210 atom types.



(a) (b)

Figure 22 Electron density map covering WR99210 and residues (Ile14, Asp54, Phe58 and Ile164) in the wild type *Pf*DHFR active site (a) residues Ile14, Asp54 and Ile164 (b) residue Phe58. Dash line shows the van der Waals interaction between N atom of triazine ring and the carboxyl group of residue in (a) and the π–π interaction between triazine ring and aromatic side chains of Phe58 in (b).

Figure 22 shows the electron density map covering WR99210 and residues (Ile14, Asp54, Phe58 and Ile164) in the wild type *Pf*DHFR active site. The binding mode of WR99210 to the wild type *Pf*DHFR allows a lot of contacts between the residues and the inhibitor. The contact distances from the N atom on triazine ring and carboxyl oxygen atom to the inhibitor are all optimal van der Waals distance. The electron density maps of the WR99210/*Pf*DHFR complexes clearly demonstrate the orientation and conformation of the inhibitor in the active site. The mainly two consequences for the binding of WR99210 were observed, that is hydrophobic and electrostatic properties of the *Pf*DHFR active site.

The interaction energies between WR99210 and individual residues are given in Table 5. The graphic picture presented by the attractive and repulsive interactions is shown in Figure 23. It is clearly seen that there are more attractive interactions between WR99210 and the residues surrounding the active site of the *Pf*DHFR. Interaction energies of Ile14, Asp54, Ile164 and Phe58 are the main contributors (-6.19, -9.34, -4.84 and -4.43 kcal/mol, respectively). Therefore, these results can be helpful for the design of new potent inhibitors**.** Furthermore, we determined the orientation of WR99210 in the bound *Pf*DHFR. It was found that the triazine ring positioned on the top surrounded by the aromatic side chains of Phe58. Based on distant investigation between heteroatom in the triazine ring and carbon atoms (C) in the aromatic ring, this implies that WR99210 can form a π–π interaction via the triazing ring with the aromatic ring (Figure 22). The interaction energies of Asp54 clearly indicated that its interaction is stronger than the other residues and this is the most important contribution. These results as shown in Figure 24 supported the previous report that the hydrogen bonded between the carboxylate side chain of Asp54 and both N3 and 4-amino group of triazine ring (Yuvaniyama *et al*., 2003). Furthermore, hydrogen bonding between the backbone carbonyl oxygens of Ile14 Ile164 and 6-amino group of triazine ring were observed. This structural interaction results show a good agreement comparing to the previous data of protein ligand interactions: WR99210-DHFR domain (Yuvaniyama *et al*., 2003) and results from docking calculation.

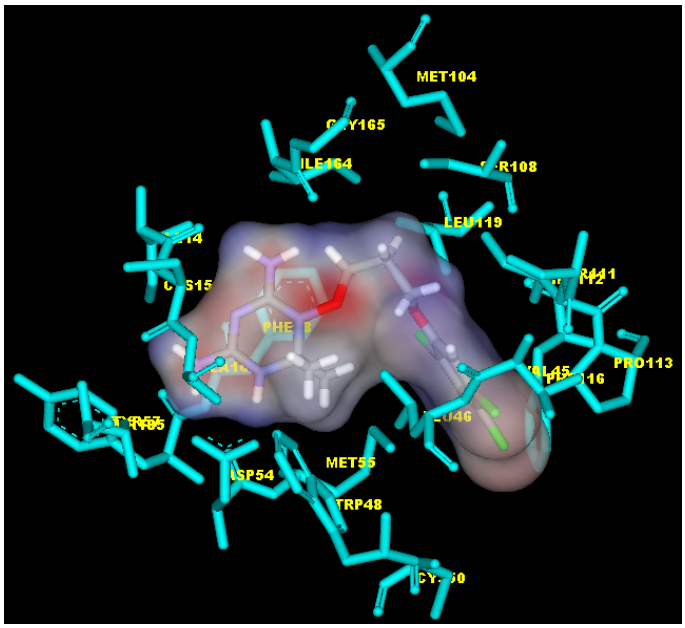


Figure 23 Graphical representation of the attractive (red) and repulsive (blue) interactions between WR99210 and individual residues of wild type *Pf*DHFR

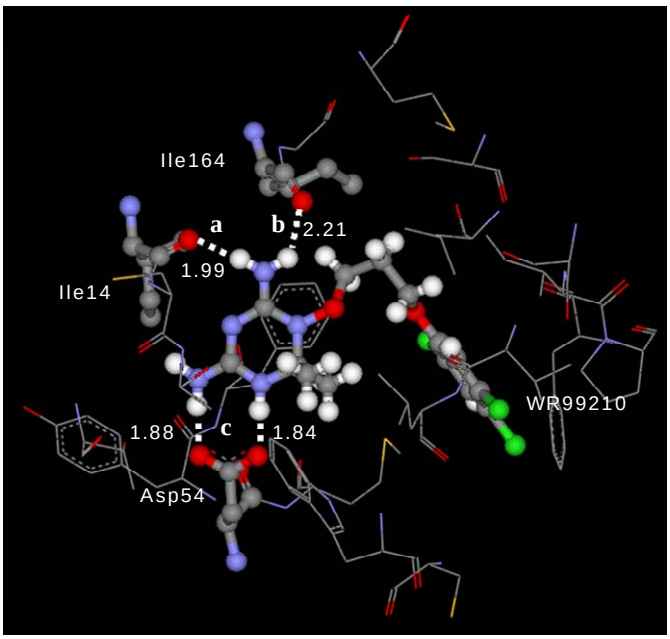


Figure 24 Show hydrogen bonds formed between WR99210 and wild type *Pf*DHFR

(a) triazine ring $-NH_2$, and the backbone carbonyl of Ile14

(b) triazine ring $-NH_2$, and the backbone carbonyl of Ile164

(c) both triazine ring $-NH_2$ and -H-potonated at N3 with carbonyl oxygen of Asp54.

Table 5 Interaction energies (kcal/mol) of WR99210 with individual residues of wild type *Pf*DHFR (1J3I) calculated at the MP2/6-31G(d) level of theory.

Residues	Interaction energies (kcal/mol)
Ile14	**-6.19**
Cys15	-0.88
Ala16	-0.24
Val45	-0.34
Leu46	-1.27
Trp48	-1.04
Cys50	-0.43
Asn51	-2.47
Asp54	**-9.34**
Met55	-1.68
Tyr57	-0.52
Phe58	**-4.84**
Cys59	0.56
Met104	-0.11
Ser108	-0.2
Ser111	0.14
Ile112	-2.95
Pro113	-0.72
Phe116	-1.29
Leu119	0.68
Ile164	**-4.43**
Gly165	0.42
Thr185	0.47

3D-QSAR Investigation on Wild Type *Pf*DHFR Inhibitors

1. CoMFA Results

The CoMFA study results of the 35 training sets are summarized in Table 6. The obtained analysis included both steric and electrostatic fields is presented. The CoMFA analysis with probed atom (sp^3 C, +1) setting default produced a model with low predictive ability, r^2_{cv} = 0.187 (model 1). Considering the other models, sp^3 O (-1) and H (+1), which show lower predictive abilities are obtained in models 2 and 3. Elimination of outlier compounds from the CoMFA analysis yields significantly improvement of the model with r^2_{cv} = 0.658 (model 4). Therefore, the final model is satisfied in both predictive ability and the residual value of maximum outlier. The first reason for eliminating the outliers is the maxima outlier compounds, 404, 413 and 415. These outlier have much higher lipophilicity values (the lipophilicity value of compounds 404, 413 and 415 are 5.94, 5.51 and 7.74, respectively), compared with the overall WR99210 derivatives. The other reason for eliminating the compounds 34, 399, 416, 419 and 446 is the different structures from the series of WR99210 derivatives. Apparently, model 4, obtained from including both steric and electrostatic fields in the analysis, shows the medium predictive ability QSAR model with r^2_{cv} = 0.658, S_{press} = 0.162 and noc = 5. The steric and the electrostatic contributions of this model are 70.0% and 30.0%, respectively. The steric interaction is shown twice times larger than the electrostatic interaction. The obtained statistical characteristics show that the conventional r^2 = 0.977, the standard error of estimate = 0.042, F = 177.816 and the probability (P) of obtaining this value of F (probability of r^2 = 0) is lower than 0.001. The plot between predicted and experimental inhibitory affinities of the non-cross-validated analysis of model 4 is illustrated in Figure 25. This graph has the variance in range of log ($1/K_i$) from 8.00 to 10.00.

Table 6 Summary of CoMFA models for wild type *Pf*DHFR inhibition with 27 WR99210 compounds at different probe atoms.

Model	Probe atom	Field type	noc	r^2_{cv}	S_{press}	r^{2a}	s	F	Steric contb[b]
1	$sp^3C(+1)$	both	4	0.187	0.303	0.882	0.115	56.149	75.8
		st	1	0.119	0.301	0.479	0.231	30.370	
		el	4	0.144	0.311	0.915	0.098	88.785	
2	$sp^3O(-1)$	both	4	0.114	0.316	0.877	0.118	53.637	73.3
		st	1	0.140	0.297	0.489	0.229	31.552	
		el	4	0.144	0.311	0.915	0.098	80.756	
3	H(+1)	both	4	0.131	0.313	0.875	0.119	52.493	71.3
		st	1	0.150	0.295	0.489	0.229	31.605	
		el	4	0.144	0.311	0.915	0.098	80.756	
4[c]	$sp^3C(+1)$	both	5	0.658	0.162	0.977	0.042	177.816	70.0
		st	4	0.400	0.210	0.952	0.059	108.595	
		el	3	0.397	0.206	0.897	0.092	55.584	

[a] conventional r^2, [b] steric contribution in %, [c] elimination of compounds 34, 399, 404, 413, 415, 416, 419, 446

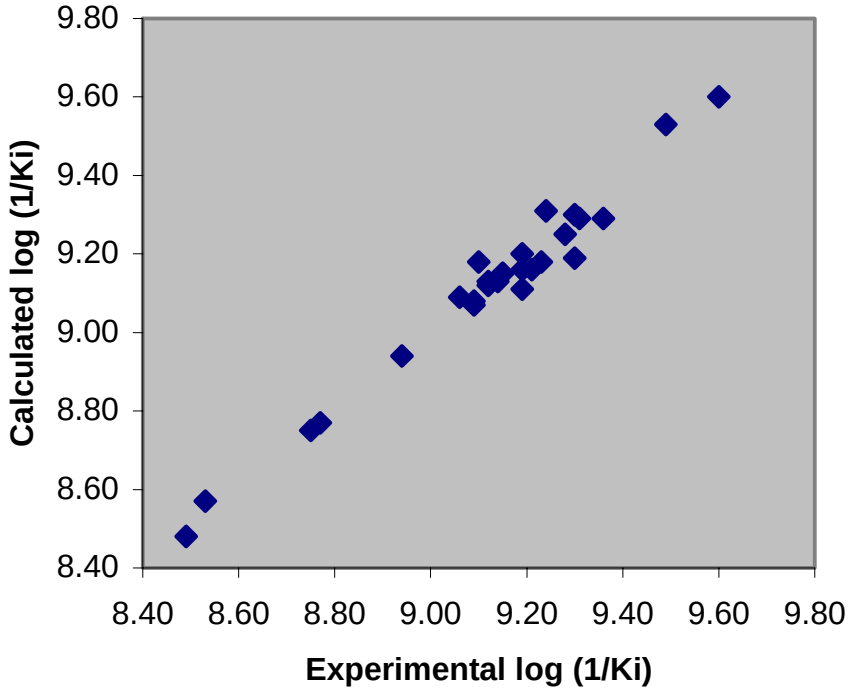


Figure 25 Plot of calculated versus experimental wild type *Pf*DHFR inhibitory affinities obtained from non-cross-validation of CoMFA model 4 for remaining 27 training set compounds.

2. Prediction for Compounds in Test Set

The CoMFA model 4, the best model, was then used to predict the inhibitory activity of the compounds in the test set. The comparison of calculated activities and experimental activities of 10 compounds are given in Table 7. In particular, this model can be useful to predict the activities for compounds 196, 401, 418, 422, 427, 430, 431, 440, 441 and 450 because these compounds have log ($1/K_i$) in range of medium to high. Compound 427 shows the largest residual because this compound has a fluorine substituent at all positions of aromatic ring these substituents are different from the other structures.

Table 7 Predicted log ($1/K_i$) of wild type *Pf*DHFR inhibitory affinity of the tested WR99210 compounds.

Compounds	log ($1/K_i$)		Residues
	Experimental	Calculated [a]	
196	9.22	8.92	0.30
401	9.46	8.97	0.49
418	9.26	8.93	0.33
422	9.16	9.05	0.11
427	**8.32**	**9.04**	**-0.72**
430	9.22	9.01	0.21
431	9.18	9.00	0.18
440	9.55	9.20	0.35
441	9.26	9.10	0.16
450	9.10	9.12	-0.02

[a] calculated by CoMFA model 4.

3. CoMFA Coefficient Contour Maps

The QSAR produced by CoMFA analysis, with its hundreds or thousands of terms, was usually represented as STDEV*COEFF contour plots. The CoMFA steric and electrostatic fields for the analysis based on alignment of binding conformation are presented as contour plots as shown in Figure 26-28. X-ray crystal structures of *Pf*DHFR complexes with various inhibitors are available, and the important residues around the inhibitor in the structure of the interaction complex were merged into all figures.

CoMFA steric contour map (in Figure 26) indicates that areas in which molecular steric bulk might have a favorable (green) or unfavorable (yellow) effect on the activity of an analogue. There are favorable steric regions corresponding to the location around the *o*-position and *p*-position of aromatic ring and R_1 of triazine ring. This result show a good agreement with the limited volume of active site between WR99210 and *Pf*DHFR as shown in Figure 29. This is supported by analyzing compound 423, lacking substituents at the *o*- and *p*-positions, which show less activity than compounds 405, 418, 422, 426, 428 and 445. In particular, the green region display surrounding R_1 side chain which is explained that the substituents larger than H atom will increase the activity. However the size of this group should not be too large because of the limited volume of active site (Ala16) as shown in Figure 29. Furthermore, the large distribution of steric contours around *p*-position in this model would explain why compound 405, used as the template, is better DHFR inhibitor than WR99210 inhibitor. The unfavorable steric contour region is close to *m*-positions and it indicates that a less bulky substituent has higher activity. It can be explained by the fact that the compounds 444, 446 and 411 show lower activities comparing to the unsubstituted compound 445.

CoMFA electrostatic contour map (in Figure 27) reveals that blue contours refer to positive charge favoring area and red contours indicate negative charge favoring areas. The red electrostatic contour regions closed to *p*-position (R_5) of aromatic ring shows the high electron density in this area. An electron rich substituent

would be more favorably placed in the *p*-positions and this can explain why substitution with H (compounds 396) shows less activity than $COOCH_3$ (compounds 402) and substitution with Ph (compounds 426) shows less activity than COOH (compounds 428) substitutions, respectively. Moreover, the red electrostatic regions around the *p*-position in this model would explain why compound 405 ($COOCH_2$-Ph substituent) and compound 428 (COOH) have higher activities than compound 426 (Ph substituent). A large blue contour surrounding the *o*-position on aromatic ring, suggests that high positive charges or low electron density in this areas are more preferable. The substituent of aromatic ring should be the donating electron groups; it will increase the activity of the inhibitors. This is further explained by comparing CH_2-Ph and NO_2 substituents at *o*-position between compounds 418 and 445 and Cl and $COCH_3$ substituents between compounds 414 and 443, respectively.

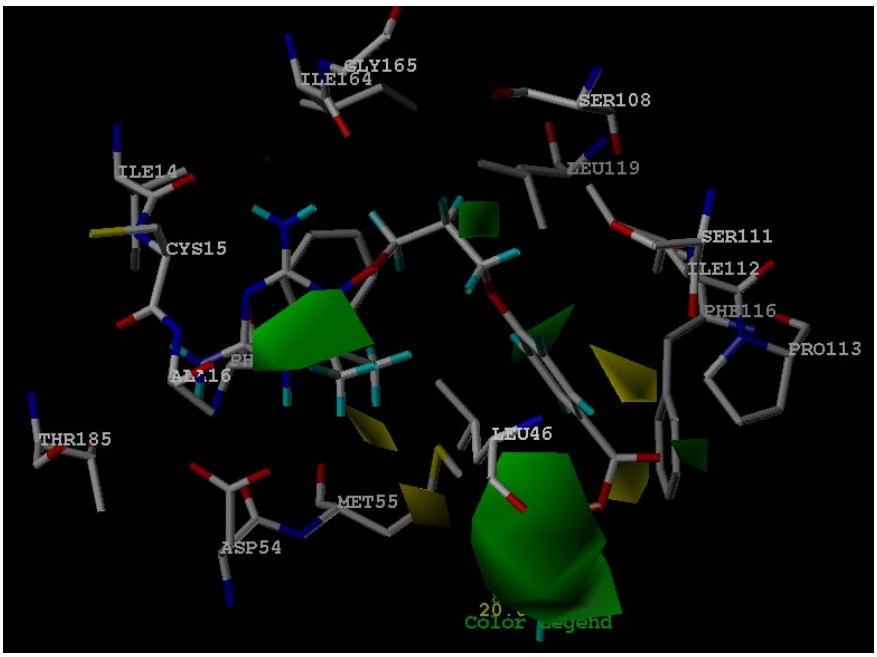


Figure 26 Stereoview of CoMFA steric STDEV*COEFF contour plots from the analysis of the 3D-QSAR model 4 with non-cross-validation based on wild type *Pf*DHFR inhibition.

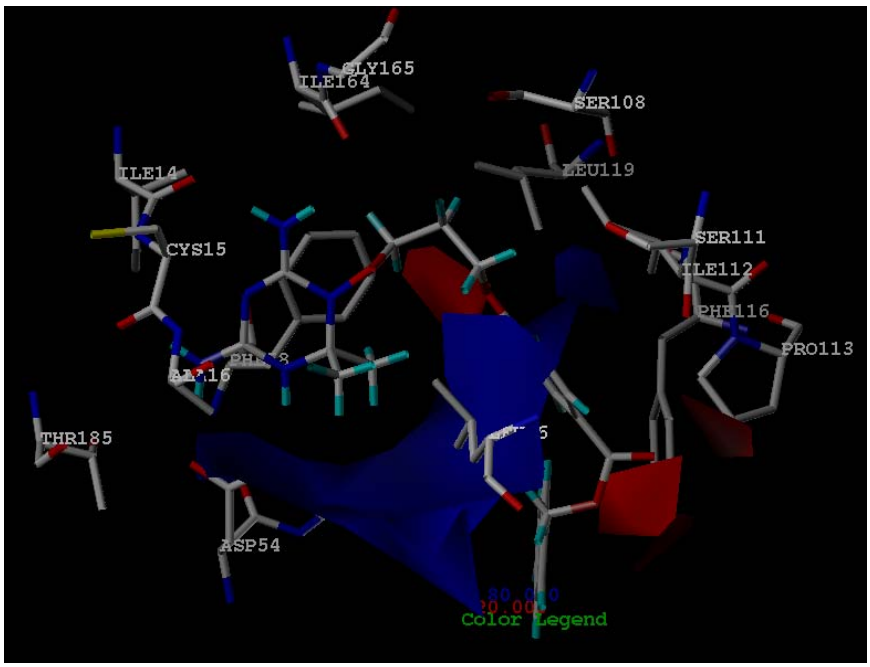


Figure 27 Stereoview of CoMFA electrostatic STDEV*COEFF contour plots from the analysis of the 3D-QSAR model 4 with non-cross-validation based on wild type *Pf*DHFR inhibition.

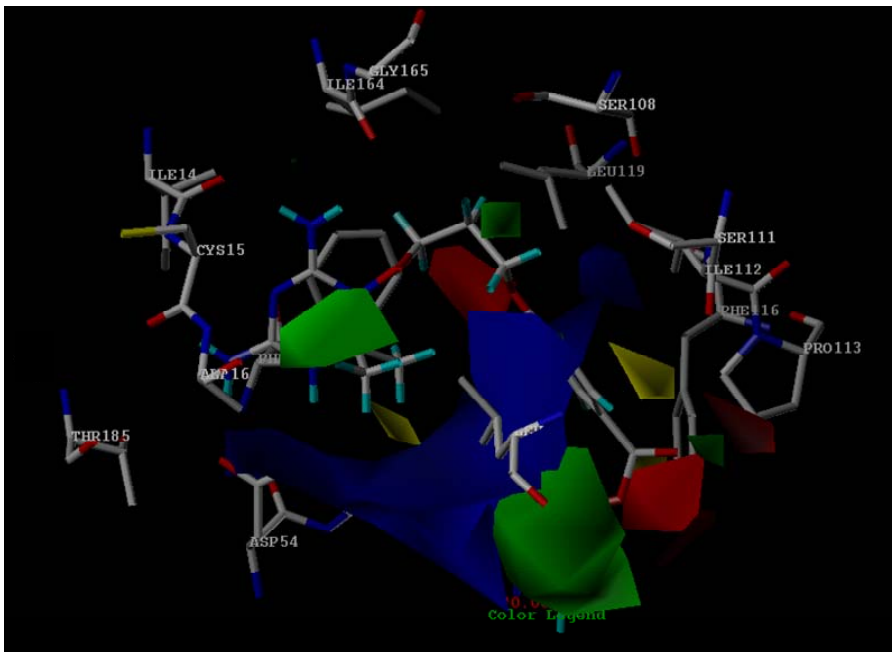
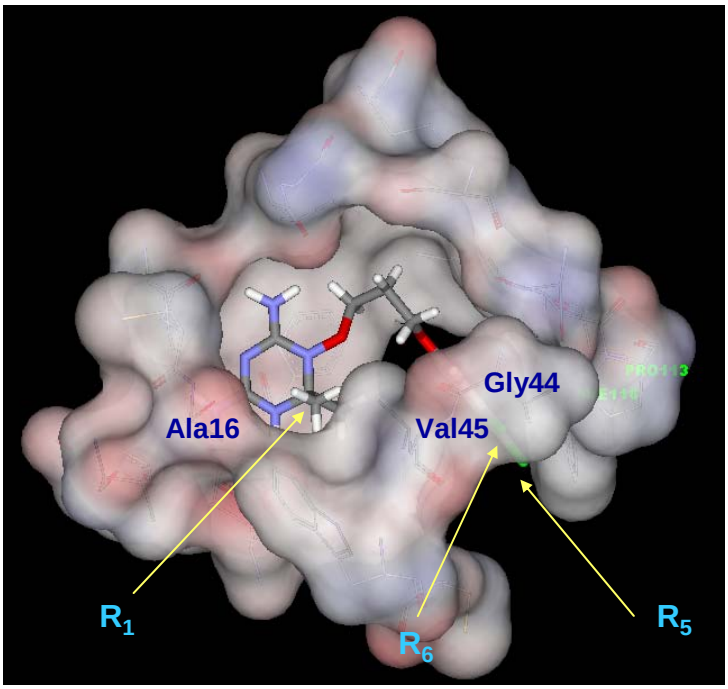


Figure 28 Stereoview of CoMFA steric and electrostatic STDEV*COEFF contour plots from the analysis of the 3D-QSAR model 4 with non-cross-validation based on wild type *Pf*DHFR inhibition.



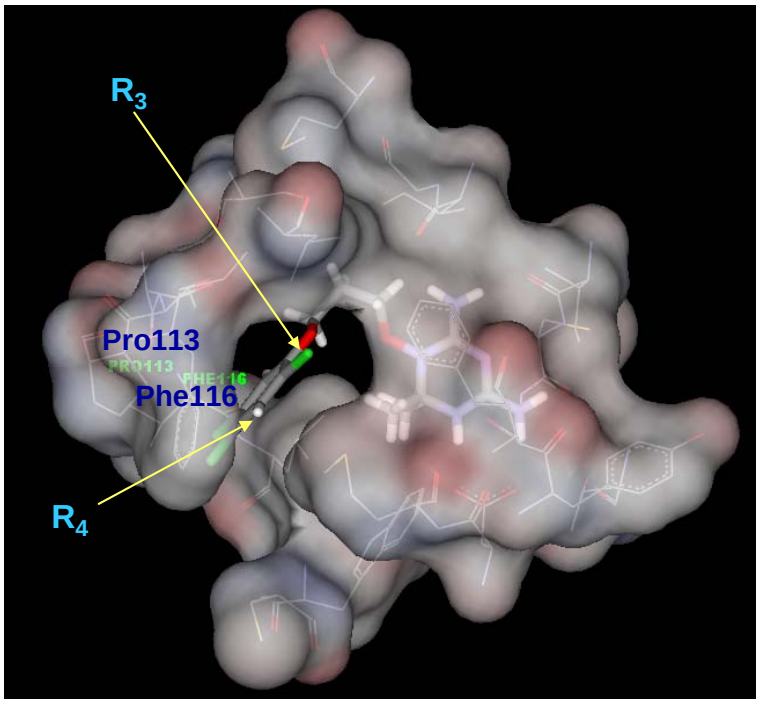


Figure 29 Graphical representation of the attractive (red) and repulsive (blue) interactions between WR99210 and individual residue of wild type *Pf*DHFR.

Design of New WR99210 Inhibitors against Wild type of *Pf*DHFR

In the present study, some new biologically active compounds have been designed based on the derived 3D-QSAR models. The obtained knowledge was employed for proposing new compounds with enhanced activity and selectivity profile for *Pf*DHFR wild type inhibition.

Based on reasonable models obtained from 3D-QSAR by using CoMFA it was found that all models reinforce each other to provide helpful insights of relationships between their structural parameters and antifolate antimalarial inhibitory activity in this class of WR99210 derivatives. Moreover, the derived principle from all obtained results was integrated to provide basic guideline for modification of existing compounds to enhance the potency against wild type *Pf*DHFR. The modification of new potent inhibitors is constructed in three dimensional of template structure in wild type *Pf*DHFR. Therefore, *o*-positions are separated into R_3 and R_7 and *m*-positions are separated into R_4 and R_6. For *p*-position, it is renamed to R_5 position.

According to wild type inhibition, the designed structures and their predicted antimalarial inhibitory activities were proposed on the basis of contour plots of CoMFA model 4 presented in Table 6.

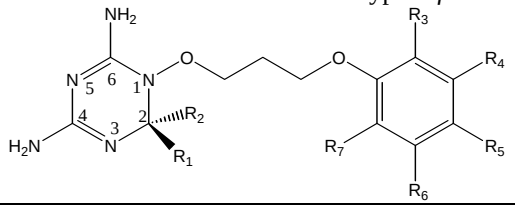
In this investigation, attempts to design novel compounds for wild type of *Pf*HDFR inhibitors have been performed in steps:

As compounds 405 show the highest potency against wild type *Pf*DHFR, this compound was used as the starting compound for modification. First, in order to examine the R_1 substituent on the triazine ring, compounds (WR1-1) and (WR1-2) were constructed and their inhibitory affinities were predicted. It was found that (WR1-1) and (WR1-2) showed low potency relative to the parent compound (compound 405) The results suggest that the large substituent and the small substituent more than CH3 group can decrease the activity in CoMFA model because limited volume of active site around R1 position. The variation of the groups

appended to the R_3 position was further evaluated; therefore, compounds (WR3-1)–(WR3-9) were constructed by displacement H atom type with CH_3, C_2H_5, C_3H_7, NH_2, OH, Cl, Br and F, respectively. These subsituents show good activities in CoMFA model. For R_7 position, compounds (WR7-1)-(WR7-9), these compounds show predicted antimalarial inhibitory activities in the same range with R_3 position for wild type *Pf*DHFR.

For the last step, the modifications on the R_3 and R_7 were performed as show in compounds (WR37-1)-(WR37-6). However, in the synthetic experimental, R_3 and R_7 can be considered as *o*-substitution, therefore, the *o*-substitution can be synthesized from modification structure on R_3 or R_7 position.

Table 8 Proposed WR99210 derivatives and calculated wild type *Pf*DHFR inhibitory affinities.



Cpds. No	R_1	R_2	R_3	R_4	R_5	R_6	R_7	Calculated ($\log 1/K_i$)[a]
WR405	CH_3	CH_3	H	H	$COOCH_2$-C_6H_5	H	H	9.60
WR1-1	H	CH_3	H	H	$COOCH_2$-C_6H_5	H	H	9.56
WR1-2	C_2H_5	CH_3	H	H	$COOCH_2$-C_6H_5	H	H	9.50
WR3-1	CH_3	CH_3	CH_3	H	$COOCH_2$-C_6H_5	H	H	9.63
WR3-2	CH_3	CH_3	C_2H_5	H	$COOCH_2$-C_6H_5	H	H	9.61
WR3-3	CH_3	CH_3	C_3H_7	H	$COOCH_2$-C_6H_5	H	H	9.54
WR3-4	CH_3	CH_3	NH_2	H	$COOCH_2$-C_6H_5	H	H	9.61
WR3-5	CH_3	CH_3	OH	H	$COOCH_2$-C_6H_5	H	H	9.60
WR3-6	CH_3	CH_3	Cl	H	$COOCH_2$-C_6H_5	H	H	9.60
WR3-7	CH_3	CH_3	Br	H	$COOCH_2$-C_6H_5	H	H	9.60
WR3-8	CH_3	CH_3	F	H	$COOCH_2$-C_6H_5	H	H	9.60
WR3-9	CH_3	CH_3	CH_2-C_6H_5	H	$COOCH_2$-C_6H_5	H	H	9.57
WR7-1	CH_3	CH_3	H	H	$COOCH_2$-C_6H_5	H	CH_3	9.53
WR7-2	CH_3	CH_3	H	H	$COOCH_2$-C_6H_5	H	C_2H_5	9.50
WR7-3	CH_3	CH_3	H	H	$COOCH_2$-C_6H_5	H	C_3H_7	9.50
WR7-4	CH_3	CH_3	H	H	$COOCH_2$-C_6H_5	H	NH_2	9.54
WR7-5	CH_3	CH_3	H	H	$COOCH_2$-C_6H_5	H	OH	9.59
WR7-6	CH_3	CH_3	H	H	$COOCH_2$-C_6H_5	H	Cl	9.51
WR7-7	CH_3	CH_3	H	H	$COOCH_2$-C_6H_5	H	Br	9.50
WR7-8	CH_3	CH_3	H	H	$COOCH_2$-C_6H_5	H	F	9.60
WR7-9	CH_3	CH_3	H	H	$COOCH_2$-C_6H_5	H	CH_2-C_6H_5	9.50
WR37-1	CH_3	CH_3	CH_3	H	$COOCH_2$-C_6H_5	H	CH_3	9.54
WR37-2	CH_3	CH_3	NH_2	H	$COOCH_2$-C_6H_5	H	NH_2	9.52
WR37-3	CH_3	CH_3	OH	H	$COOCH_2$-C_6H_5	H	OH	9.55
WR37-4	CH_3	CH_3	Cl	H	$COOCH_2$-C_6H_5	H	Cl	9.49
WR37-5	CH_3	CH_3	Br	H	$COOCH_2$-C_6H_5	H	Br	9.48
WR37-6	CH_3	CH_3	F	H	$COOCH_2$-C_6H_5	H	F	9.56

[a] calculated by CoMFA model 4

We performed the superimposition between the X-ray structure of the wild type (pdb code 1J3I) and quadruple mutant type (N51I, C59R, S108N and I164L (pdb code 1J3K)) *Pf*DHFR in complex with WR99210. Structural analysis of WR99210 in complex with the quadruple mutant type revealed that overall position of the inhibitor as well as the residues lining the WR99210 inhibitor binding pocket corresponding to the wild type complex. Comparison between the two complexes shows a RMSD value of the backbone atoms about 0.45 Å. The only subtlety changes in the overall positioning and orientation of residues lining the active site can be observed among these complexes as shown in Figure 30. Additionally, considering structure of I164L substitution, replaced isoleucine (I164) into leucine (L164) side chain, it seems that the less steric bulk, but the H-bonding still occurred between backbone carbonyl of leucine and $-NH_2$ group of triazine ring as shown in Figure 31. Considering residues C51I and C59R, It was found that these residues far from the binding pocket of WR99210. Therefore, C51I and C59R have less effect to the binding of WR99210 and quadruple mutant type. For S108N mutation, shown steric bulky group in this mutation, but WR99210 has high flexibility so as to avoid possible steric clash with Asn108. Moreover, when considering the interaction energies between WR99210 and individual residues of quadruple mutant type (as shown in Table 9), it is clearly seen that there are more attractive interactions between WR99210 and the residues surrounding the active site of the *Pf*DHFR. Interaction energies of Ile14, Asp54, Leu164 and Phe58 are the main contributors (-6.00, -8.87, -3.83 and -4.30 kcal/mol, respectively). This structural interaction results show a good agreement comparing to the the interaction energies between WR99210 and individual residues of wild type.

Figure 30 Superimposition of the backbone atoms of WR99210 bound to the *Pf*DHFR active site between wild type (green) and quadruple mutant type (pink).

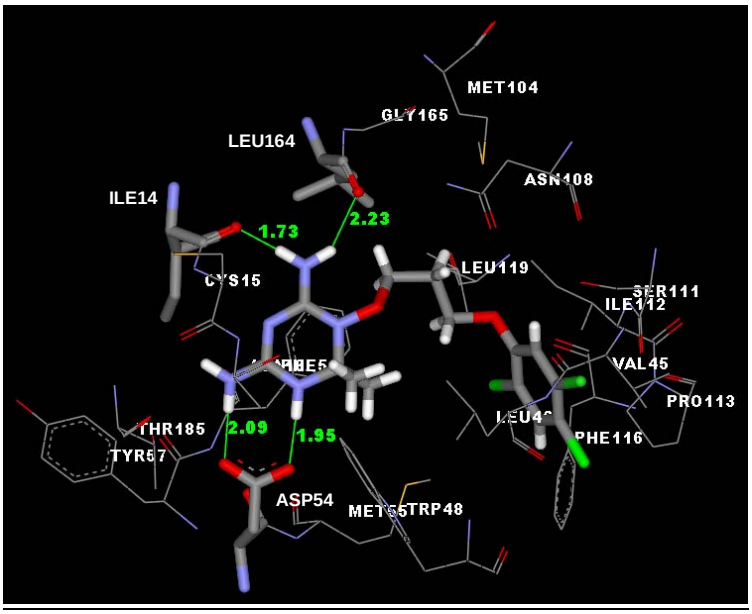
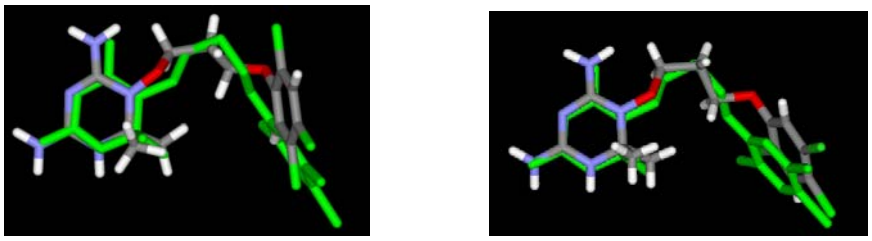


Figure 31 Bound conformation of WR99210 with residues in the quadruple mutant (N51I, C59R, S108N and I164L) type *Pf*DHFR active site, surrounding within 7.0 Å. The hydrogen bonding interactions are shown in green lines.

Table 9 Comparison Interaction energies (kcal/mol) of WR99210 with individual residues between wild type and quadruple mutant type *Pf*DHFR calculated at the MP2/6-31G(d)//PM3 level of theory.

Residues	Interaction energies (kcal/mol)	
	Wild type	Quadruple mutant type
Ile14	**-6.19**	**-6.00**
Cys15	-0.88	-0.19
Ala16	-0.24	-1.11
Val45	-0.34	-0.67
Leu46	-1.27	-1.08
Trp48	-1.04	-0.87
Cys50	-0.43	-0.30
Asn51(Ile)	-2.47	-2.17
Asp54	**-9.34**	**-8.87**
Met55	-1.68	-1.31
Tyr57	-0.52	-0.14
Phe58	**-4.84**	**-3.83**
Cys59(Arg)	0.56	5.20
Met104	-0.11	-0.31
Ser108(Asn)	-0.2	2.78
Ser111	0.14	-0.46
Ile112	-2.95	-2.98
Pro113	-0.72	0.38
Phe116	-1.29	-1.52
Leu119	0.68	-0.42
Ile164(Leu)	**-4.43**	**-4.30**
Gly165	0.42	0.05
Thr185	0.47	-2.63

Therefore, the proposed WR99210 derivatives from CoMFA model 4 of wild type DHFR should be against quadruple mutant type. Consequently, to investigate the potential binding modes of proposed WR99210 derivatives, they were docked to wild type (pdb code 1J3I) and quadruple mutant type (N51I, C59R, S108N and I164L) (pdb code 1J3K) by using Autodock program. At first, this program was applied to the WR99210/*Pf*DHFR complexes of wild type and quadruple mutant type to determine the ability of this program for reproducing the WR99210's X-ray orientation in its binding pocket. Autodock is a docking method using a genetic algorithm (GA) for docking flexible ligands into protein binding sites. The superposition and docked orientation of WR99210 are shown in Figure 32. In this study, the RMSD of WR99210 with wild type and quadruple mutant type *Pf*DHFR are 0.98 and 0.97 Å, respectively. These values are less than 1.0. Therefore, the Autodock method and the set parameter could be used to search the enzyme binding conformations for further step.



(a) **(b)**

Figure 32 Superposition of WR99210 obtained from X-ray crystallographic structure (green) and docking calculations in (a) wild type and (b) quadruple mutant type.

Docking calculations were performed to calculate proposed WR99210 derivatives bound to the wild type and quadruple mutant type *Pf*DHFR enzyme. The estimated free energies of binding results of wild type and quadruple mutant type are estimately in the same range. These results are presented in Table 10. From docking results, it can be seen that the major residue contributors to the binding energy of both wild type and quadruple mutant type are Ile14, Asp54, Ile164(Leu) and Phe58, as expressed in Figure 31.

Table 10 Estimated free energies of binding (kcal/mol) of proposed WR99210 derivatives in the wild type and quadruple mutant type (N51I, C59R, S108N and I164L) *Pf* DHFR active site.

Cpds. No.	Wild type (kcal/mol)	Quadruple mutant type (kcal/mol)
WR405	-14.56	-14.98
WR1-1	-12.00	-12.89
WR1-2	-11.96	-12.00
WR3-1	-12.51	-13.89
WR3-2	-12.31	-13.56
WR3-3	-14.87	-15.01
WR3-4	-13.39	-13.10
WR3-5	-12.96	-13.56
WR3-6	-13.00	-13.35
WR3-7	-12.95	-13.29
WR3-8	-14.91	-15.36
WR3-9	-11.23	-11.99
WR7-1	-11.62	-11.56
WR7-2	-11.58	-11.79
WR7-3	-13.94	-14.20
WR7-4	-14.85	-15.00
WR7-5	-11.05	-11.65
WR7-6	-12.87	-13.12
WR7-7	-12.50	-13.00
WR7-8	-14.37	-14.89
WR7-9	-11.10	-11.40
WR37-1	-12.18	-12.79
WR37-2	-11.57	-11.96
WR37-3	-11.98	-12.35
WR37-4	-12.65	-13.20
WR37-5	-13.00	-13.96
WR37-6	-14.12	-14.85

CONCLUSION

In this study, from the molecular docking calculation, the estimated free energies of binding of WR99210 analogs with (*R*)-configuration were lower than those of the analogs with (*S*)-configuration suggesting that the enzyme preferentially binds the inhibitors with (*R*)-configuration. Therefore, (*R*)-configuration was selected to perform CoMFA analysis. In addition, hydrogen bonding between WR99210 analogs and amino acid Ile14, Cys15, Asp54 and Ile164 were also observed. The results suggest that the appropriate configuration of WR99210 derivatives prefers *(R)*-configuration upon binding.

Investigation of the contributing interactions of the inhibitor/enzyme interactions for the inhibitory activity of WR99210 against wild type *Pf*DHFR in the active site showed that there is attractive interaction between WR99210 and surrounding residues. The energetic results as obtained from the interaction between WR99210 and the active site indicated that Ile14, Asp54, Phe58 and Ile164 are the main contributors. Asp54 demonstrates a stronger interaction than the others. The understanding of this particular interaction energy will be helpful for the design of new potent inhibitors.

The CoMFA analysis was successfully applied to predict the antimalrial activity of a set of WR99210 derivatives active against the wild type of *Pf*DHFR. The CoMFA models produced satisfactory predictivity results in terms of r^2_{cv} and r^2 values. The CoMFA model provided the most significant correlation of steric and electrostatic fields with biological activities. The steric hindrance is majority field distribution in the wild type. The results obtained from CoMFA contour maps gave an insight information in identifying structural requirements of WR99210 inhibitors for wild type of *Pf*DHFR inhibition. The steric properties are very important in the substitutions occurring at *o*-position, *p*-position and R_1 position. The other positions (*m*-position and R_2) on consist in non-steric adjustments of the inhibitor to the protein. Whereas, negative electrostatic forces surrounded *p*-positions and positive electrostatic closed to *o*-position.

The information obtained from CoMFA guides the further structural modifying and synthesizing new potent antimalarial agents with enhanced *Pf*DHFR inhibitory activities for wild type inhibition.

LITERATURE CITED

Abagyan, R., M. Totrov and D. Kuznetsov. 1994. ICM: A new method for structure modeling and design: Applications to docking and structure predictive from the distorted native conformation. **J. Comp. Chem**. 15: 488-506.

Agrawal, V. K., R. Srivastava and P. V. Khadikar. QSAR Studies on some antimalarial sulfonamides. 2001. **Bioorg. Med. Chem.** 9: 3287-3293.

Basco, L. K., P. E. Pecoulas, C. M. Wilson, J. L. Bras and A. Mazabraud. 1995. Point mutations in the dihydrofolate reductase-thymidylate synthase gene and pyrimethamine and cycloguanil resistance in *Plasmodium falciparum*. **Mol. Biochem. Parasitol**. 69: 135-138.

Blaney, J. M., C. Hansch, C. Silipo and A. Vittoria. 1984. Structure activity relationship of dihydrifolate reductase inhibitors. **Chem. Rev**. 84: 333-407.

Bhattacharjee, A. K., M. G. Harll, D. A. Nichols, R. P. Hicks, B. Stanton, J. E. Hamont and W. K. Milhous. 2004. Structure-activity relationship study of antimalarial indolo [2,1-b]quinazoline-6,12-diones (tryptanthrins). Three dimensional pharmacophore modeling and identification of new antimalarial candidates. **Euro. J. Med. Chem.** 39: 59-67.

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.

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.

_____, R. D. Clark, D. E. Patterson and A. M. Ferguson. 1996. Bioisosterism as a Molecular Diversity Descriptor: Steric Fields of Single "Topomeric" Conformers **J. Med. Chem**.39: 3060-3069.

Cheng, F., J. Shen, X. Luo, W. Zhu, J. Gu, R. Ji, H. Jiang and K. Chen. 2002. Molecular docking and 3D-QSAR studies on the possible antimalarial mechanism of artemisinin analogues. **Bioorg. Med. Chem**. 10: 2883-2891.

Chusacultanachai, S., P. Thiensathit, B. Tarnchompoo, W. Sirawaraporn and Y. Yuthavong. 2002. Novel antifolate resistant mutations of *Plasmodium falciparum* dihydrofolate reductase selected in Escherichia coli. **Mol. Biochem. Parasitol**. 120: 61-72.

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

Cortese J.F. and C.V. Plowe. 1998. Antifolate resistance due to new and known *Plasmodium falciparum* dihydrofolate reductase mutations expressed in yeast. **Mol. Biochem. Parasitol.** 94: 205–214.

Cowman, A.F., M.J. Morry, B.A. Biggs, G.A.M. Cross and S. Foote.1988. Amino acid changes linked to pyrimethamine resistance in the dihydrofolate reductase-thymidylate synthase gene of *Plasmodium falciparum*. **Proc. Natl. Acad. Sci. USA** 85: 9109–9113.

Czaplinski, K-H., W. Hansel, M. Wiese and JK. Seydel. 1995. New benzyl pyrimidines: inhibition of DHFR from various species. QSAR, CoMFA and PC analysis. **Eur. J. Med. Chem**. 30: 779-787.

Delfino, R. T., O. A. Santos-Filho and J. D. Figueroa-Villar. 2002. Molecular modeling of wild-type and antifolate resistant mutant *Plasmodium falciparum* DHFR. **Biophys. Chem**. 98: 287-300.

_____, O. A. Santos-Filho and J. D. Figueroa-Villar. 2002. Type 2 antifolates in the chemotherapy of falciparum malaria. **J. Braz. Chem. Soc**. 13: 727-741.

Dewar, M.J.S., E.G. Zoebisch, E.F. Healy and J.J.P. Stewart. 1985. AM1: A new general purpose quantum mechanical molecular model. **J. Am. Chem. Soc**. 107: 3902-3909.

Edstein, M. D., A. E. T. Yeo, G. D. Shanks and K. H. Rieckmann. 1997. Ex vivo antimalarial activity of proguanil combined with dapsone against cycloguanil-resistant *Plasmodium falciparum* isolates. **Acta. Tropica**. 66: 127-135.

Ewing, T. J. A. and J. Kuntz. 1997. Critical evaluation of search algorithm used in automated molecular docking. **Comput. Chem**. 18:1175-1189.

Ferlan, J.T., S. Mookherjee, I.N. Okenzie, L. Fulgence, C.H. Sibley. 2001.Mutagenesis of dihydrofolate reductase from *Plasmodium falciparum*: analysis in *Saccharomyces cerevisiae* of triple mutant alleles resistant to pyrimethamine or WR99210. **Mol. Biochem. Parasitol**. 113: 139-150.

Foote, S.J., D. Galatis, A.F. Cowman.1990. Amino acids in the dihydrofolate reductase-thymidylate synthase gene of *Plasmodium falciparum* involved in cycloguanil resistance differ from those in pyrimethamine resistance. **Proc. Natl. Acad. Sci. USA.** 87: 3014-3017

Free, S.M., Jr and J.M. Wilson.1964. Free and Wilson analysis. **J. Med. Chem.** 7: 395-399

Frisch M.J., G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, V.G. Zakrzewski, J.A. Montgomery, R.E., Jr. Stratmann, J.C. Burant, S. Dapprich, J.M. Millam, A.D. Daniels, K.N. Kudin, M.C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G.A. Petersson, P.Y. Ayala, Q. Cui, K. Morokuma, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J. Cioslowski, J.V. Ortiz, A. G. Baboul, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, 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, J.L. Andres, C. Gonzalez, M. Head-Gordon, E.S. Replogle, J. A. Pople. 2003. **Gaussian 03**, Revision B.03, Gaussian, Inc., Pittsburgh, PA.

Gasteiger, J. and M. Marsili. 1980. Iterative partial egualization of orbital electronegativity arapid acess to atomic charges. **Tetrahedron.** 36: 3219-3228.

Garrettt, G., G. M. Morris, D. S. Goodsell, R. S. Hallidays and S. J. Olson. 1998. Auutomated docking using a Lamarckian genetics algorithm and an empirical binding free energy function. **J. Comp. Chem.** 19: 1639-1662

Gilbert, I. H. 2002. Inhibitors of dihydrofolate reductase in leishmania and trypanosomes. **Biochimica et Biophysica Acta**. 1587: 249-257.

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.

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

Hansch, C and A. Leo. 1979. Substituent Constants for Correlation Analysis in Chemistry and biology. Wiley Interscience. New York. NY.

_____ and A. Leo.1995. Exploring QSAR Fundamentals and Applications in Chemistry and Biology. In **American Chemistry Societry**, Washington, DC.

Hekmat-Nejad, M. and P. K. Rathod. 1997. *Plasmodium falciparum*: kinetic interactions of WR99210 with pyrimethamine-sensitive and pyrimethamine-resistant dihydrofolate reductase. **Exp. Parasitol**. 87: 222-228.

Hyde, J.E. 1990. The dihydrofolate reductase-thymidylate synthase gene in the drug resistance of malaria parasites. **Pharmacol. Ther.** 48: 45–59.

Hyde, J. E. 2002. Mechanisms of resistance of Plasmodium falciparum to antimalarial drugs. **Microbes Infect.** 4: 165–174.

Jones, G., P. Willett and R. C. Glen. 1995. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. **J. Mol. Biol.** 245: 43-53.

Jung, M. and H. Kim. 2001. CoMFA of artemisinin derivatives: effect of location and size of lattice. **Bioorg. Med. Chem. Lett**. 11: 2041-2044.

Kamchonwongpaisan, S., R. Quarrell, N. Charoensetakul, R. Ponsinet, T. Vilaivan, J. Vanichtanankul, B. Tarnchompoo, W. Sirawaraporn, G. Lowe and Y. Yuthavong. 2004. Inhibitors of multiple mutants of *plasmodium falciparum* dihydrofolate reductase and their antimalarial activities. **J. Med. Chem**. 47: 673-680

_____, J. Vanichtanankul, B. Tarnchompoo, J. Yuvaniyama, S. Taweechai and Y. Yuthavong. 2005. Stoichiometric Selection of Tight-Binding Inhibitors by Wild –Type and Mutant Forms of Malarial (Plasmodium falciparum) Dihydrofolate Reductase. **Anal. Chem**. 77(5): 1222-1227.

Kim, K.H. 1995. Molcular Similarity in Drug Design**.** pp. 291-331. In P.M. Dean, eds. Chapman and Hall, New York.

Knighton, D. R., C. C. Kan, E. Howland, C. A. Janson, Z. Hostomska, K.M.Welsh and D. A. Matthews. 1994. Structure of and kinetic channelling in bifunctional dihydrofolate reductase-thymidylate synthase. **Nat. Struct. Biol.** 1: 186–194.

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.

_____1997. QSAR and 3DQSAR in drug design Part1: methodology. **Drug Discovery Today**. 2: 457-467.

_____1997. QSAR and 3DQSAR in drug design Part2: applications and problems. **Drug Discovery Today**. 2: 538-546.

Kuntz, I.D., J.M. Blaney, S.J. Oatley, R. Langridge and T.E. Ferrin. 1982. A geometric approach to macromolecule-ligand interactions. **J. Mol. Biol**. 161: 269–288.

Kuno M., S. Hannongbua, 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.

Leartsakulpanich, U., M. Imwong, S. Pukrittayakamee, N. J. White, G. Snounou, W. Sirawaraporn and Y. Yuthavong. 2002. Molecular characterization of dihydrofolate reductase in relation to antifolate resistance in *Plasmodium vivax*. **Mol. Biochem. Parasitol.** 119: 63-73.

Lemcke T., I.T. Christensen, F. Jorgensen. 1999. Towards an understanding of drug resistance in malaria: Three-dimensional structure of *Plasmodium falciparum* dihydrofolate reductase by homology building. **Bioorg. Med. Chem.** 7: 1003–1011.

Li, L., K. Z. Maoshuang and A. Tian. 2001. Semi-empirical quantum chemical study on structure-activity relationship in monocyclic-B-lactam antibiotics. **J. Mol. Struct. (Theochem).** 545: 1-5.

Lin X., L.A. Parsels, D.M. Voeller, C.J. Allegra, G.F. Maley, E. Chu.2000. Characterization of a cis-acting regulatory element in the protein coding region of thymidylate synthase mRNA. **Nucleic Acids Res.** 28:1381–1389.

Liu, M. and S. Wang. 1999. MCDOCK: a Monte Carlo simulation approach to the molecular docking problem. **J. Comput. Aided. Mol. Des**. 13: 435-451.

Macreadie, I., H. Ginsburg, W. Sirawaraporn and L. Tilley. 2000. Antimalarial drug development and new targets. **Parasitology Today**. 16: 438-444.

Mattioni, B. E. and P. C. Jurs. 2003. Prediction of dihydrofolate reductase inhibition and selectivity using computational neural network and linear discriminant analysis**. J. Mol. Graph. Mod.** 21: 391-419.

McKie, J.H., K.T. Douglas, C. Chan, S.A. Roser, R. Yates, M. Read, J.E. Hyde, M.J. Dascombe, Y. Yuthavong, W. Sirawaraporn.1998. Rational drug design approach for overcoming drug resistance: Application to pyrimethamine resistance in malaria. **J. Med. Chem.** 4: 1367–1370.

McMartin, C. and R. S. Bohacek. 1997. QXP: Powerful, rapid computer algorithms for structure based drug design. **J.Comput. Aided. Mol. Des**. 11: 333-344

Militello, K., M. Dodge, L. Bethke and D. F. Wirth. 2004. Identification of regulatory elements in the *Plasmodium falciparum* genome. **Mol. Biochem. Parasitol**. 134: 75-88.

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

Numrium, 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.

Olliaro, P. L. and Y. Yuthavong. 1999. An overview of chemotherapeutic targets for antimalarial drug discovery. **Pharmacol. Ther**. 81: 91-110.

_____. 2001. Mode of action and mechanisms of resistance for antimalarial drugs. **Pharmacol. Ther.** 89: 207-219.

O'Nell, R. H., R. H. Lilien, B. R. Donald, R.M. Stroud and A. C. Anderson. 2003. Phylogenetic classification of protozoa based on the structure of the linker domain in the bifunctional enzyme, dihydrofolate reductasethymidylate synthase. **J. Biol. Chem.** 278: 52980–5298.

Oprea, T. I. and C. T. Waller. 1997. Theorretical and practical aspects of three-dimensional quantitative structure-activity relationship, pp. 127-182. In K. B. Lipkowitz and D. B. Boyd, eds. **Review in computational chemistry**. Wiley-VCH, New York.

Patrick, G.L. 1995. An introduction to medicinal chemistry. p. 336. Oxford University Press Inc., New York.

Peterson, D.S., D. Walliker and T.E. Wellems. 1988. Evidence that a point mutation in dihydrofolate reductase-thymidylate synthase confers resistance to pyrimethamine in falciparum malaria. **Proc. Natl. Acad. Sci**. **USA**. 85: 9114–9118.

_____, W.K. Milhous and T.E. Wellems. 1990. Molecular basis of differential resistance to cycloguanil and pyrimethamine in *Plasmodium falciparum* malaria, **Proc. Natl. Acad. Sci. USA**. 87: 3018–3022.

Plowe C.V., J.F. Cortese, A. Djimde, O.C. Nwanyanwu, W.M. Watkins, P.A. Winstanley, J.G. Estrada-Franco, R.E. Mollinedo, J.C. Avila, J.L. Cespedes, D. Carter, O.K. Doumbo.1997 Mutations in *Plasmodium falciparum* dihydrofolate reductase and dihydropteroate synthase and epidemiologic patterns of pyrimethamine-sulfadoxine use and resistance. **J. Infect. Dis.** 176 :1590–1596.

Prapunwattana, P., W. Sirawaraporn, Y. Yuthavong and D. V. Santi. 1996. Chemical synthesis of the *Plasmodium falciparum* dihydrofolate reductase-thymidylate synthase gene. **Mol. Biochem. Parasitol**. 83: 93-106.

Ramsden, C. A., C. Hansch, P.G. Sammes and J.B. Taylor. 1990. Quantitative Drug Design, vol. 4. In **Comprehensive Medicinal Chemistry**. Pergamon, Oxford.

Rastelli, G., S. Sirawaraporn, P. Sompornpisut, T. Vilaivan, S. Kamchonwongpaisan, R. Quarrell, G. Lowe, Y. Thebtaranonth and Y. Yuthavong. 2000. Interaction of pyrimethamine, cycloguanil, WR99210 and their analogues with *Plasmodium falciparum* dihydrofolate reductase: structural basis of antifolate resistance. **Bioorg. Med. Chem**. 8: 1117-1128.

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

Reynolds M.G and D.S. Roos.1998. A biochemical and genetic model for parasite resistance to antifolates. *Toxoplasma gondii* provides insight into pyrimethamine and cycloguanil resistance in *Plasmodium falciparum*. **J. Biol. Chem.** 273: 3461–3469.

Santos-Filho, O. A., R. B. Alencastro and J. D. Figueroa-Villar. 2001. Homology modeling of wild type and pyrimenthamine/cycloguanil-cross resistant mutant type *Plasmodium falciparum* dihydrofolate reductase. A model for antimalarial chemotherapy resistance. **Biophys. Chem**. 91: 305-317.

Shallom, S., K. Zhang, L. Jiang and P. K. Rathod.1999. Essential protein-protein interactions between Plasmodium falciparum thymidylate synthase and dihydrofolate reductase domains. **J. Biol. Chem**. 274:37781–37786.

Sardarian, A., K. T. Douglas, M. Read, P. F. G. Sims, J. E.Hyde, P. Chitnumsub, R .Sirawaraporn and W. Sirawaraporn. 2003. Pyrimethamine analogs as strong inhibitors of double and quadruple mutants of dihydrofolate reductase in human malaria parasites. **Organic Biomolecular Chemistry**. 1: 960–964.

Sirawaraporn, W., T. Sathitkul, R. Sirawaraporn, Y. Yuthavong and D.V. Santi. 1997. Antifolate-resistant mutants of *Plasmodium falciparum* dihydrofolate reductase. **Proc. Natl. Acad. Sci. USA**. 94: 1124–1129.

_____, S. Yongkiettrakul, R. Sirawaraporn, Y. Yuthavong and D. V. Santi. 1997. *Plasmodium falciparum*: asparagine mutant at residue 108 of dihydrofolate reductase is an optimal antifolate-resistant single mutant. **Exp. Parasitol**. 87: 245-252.

_____ and Y. Yuthavong. 1984. Kinetic and molecular properties of dihydrofolate reductase from pyrimethamine-sensitive and pyrimethamine-resistant *Plasmodium chabaudi*, **Mol. Biochem. Parasitol**.10: 355–367.

Sirawaraporn, S., R. Sirawaraporn, S. Yongkiettrakul, A. Anuwatwora, G. Rastelli, S. Kamchonwongpaisan and Y. Yuthavong. 2002. Mutational analysis of *Plasmodium falciparum* dihydrofolate reductase: the role of aspartate 54 and phenylalanine 223 on catalytic activity and antifolate binding. **Mol. Biochem. Parasitol**. 39: 127-34.

Sirichaiwat, C., C. Intaraudom, S. Kamchonwongpaisan, J. Vanichtanankul, Y. Thebtaranonth and Y. Yuthavong. 2004. Target guided synthesis of 5- benzyl-2,4-diaminopyrimidines: their antimalarial activities and affinities to wild type and mutant dihydrofolate reductases from plasmodium falciparum. **J. Med. Chem**. 47: 345-354.

SYBYL Molecular Modelling Softwares, Version 7.0, Tripos Associates, Inc., St. Louis, MO, 63144, USA, 2005.

Tarnchompoo, B., C. Sirichiwat, W. Phupong, C. Intaraudom, W. Sirawaraporn, S. Kamchonwongpaisan, J. Vanichtanankul, Y. Thebtaranth and Y. Yuthavong. 2002. Development of 2,4-diaminopyrimidines as antimalarials based on inhibition of the S108N and C59R+S108N mutants of dihydrofolate reductase from pyrimethamine-resistant Plasmodium falciparum. **J. Med Chem**. 45: 1244–1252.

Thaithong S., S.W. Chan, S. Songsomboon, P. Wilairat, N. Seesod,T. Sueblinwong, M. Goman, R. Ridley, G. Beale. 1992. Pyrimethamineresistant mutations *Plasmodium falciparum*. **Mol. Biochem. Parasitol.** 52: 149–157.

Tonmunphean, S., V. Parasuk and S. Kokpol. 2000. QSAR study of antimalarial activities and artemisinin-heme binding properties obtained from docking calculations. **Quant. Struct. -Act. Relat**. 19: 475-483.

Topliss, J. G. and R.J. Costello. 1972. Chance correlations in structure-activity studies using multiple regression analysis. **J. Med. Chem**. 15: 1066-1068.

Totrov, M. and R. A. Abagyan. 1997. Flexible protein-ligand docking by global energy optimization in internal coordinates. Proteins Suppl. 1: 215-220.

Toyoda T., R.K.B. Brobey, G. Sano, T. Horii, N. Tomioka and A. Itai. 1997. Lead discovery of inhibitors of the dihydrofolate reductase domain of *Plasmodium falciparum* dihydrofolate reductase-thymidylate synthase. **Biochem. Biophys. Res. Commun.** 235: 515–519.

Verloop, A., W. Hoogenstraaten and J. Tipfer. 1976. Development and application of new steric substituent parameters. **Drug Design.** 7: 165-207.

Vilaivan, T., N. Saesaengseerung, D. Jarprung, S. Kamchonwongpaisan, W. Sirawaraporn and Y. Yuthavong. 2003. Synthesis of solution-phase combinatorial library of 4,6-diamino-1,2-dihydro-1,3,5-triazine and identification of new leads against A16V + S108T mutant dihydrofolate reductase of *Plasmodium falciparum*. **Bioorg. Med. Chem**. 11: 217-224.

Vieth, M., J. D. Hirst, B. N. Dominy, H. Daigler and III. C. L. Brooks. 1998. Assessing search strategies for flexible docking. **J. Comp.Chem.** 19: 1623-1631.

Warhurst, D.E. 1998. Antimalarial drug discovery: Development of inhibitorsof dihydrofolate reductase active in drug resistance. **Drug Disc. Today.** 3: 538–546.

Warhurst, D. C. 2002. Resistance to antifolates in *Plasmodium falciparum*, the causative agent of tropical malaria. **Science Progress**. 85: 89-111.

Wade, R.C., S. Henrich and T. Wang. 2004. Using 3D protein structures to derive 3D-QSARs. **Drug Discovery Today: Technologies**.

Wattanarangsan, J., S. Chusacultanachai, J. Yuvaniyama, S. Kamchonwongpaisan and Y. Yuthavong. 2003. Effect of N-terminal truncation of *Plasmodium falciparum* dihydrofolate reductase on dihydrofolate reductase and thymidylate synthase activity. **Mol. Biochem. Parasitol**. 126: 97-102.

Weiner, S. J., P. A. Kollman, D. A. Case, S. U. Chandra, C. Ghio, G. Alagona, S. Profeta and P. Weiner. 1984. A new force field for molecular mechanical simulation of nucleic acids and proteins. **J. Am. Chem. Soc**. 106: 756-784.

Winstanley, P.A. 2000. Chemotherapy for *falciparum* malaria: the armoury, the problems and the prospects. **Parasitol. Today** 16: 146-153.

_____. 2002. Clinical status and implications of antimalarial drug resistance. **Microbes infect**. 4: 157-164

Winter, G., Q. Chen and M. Wahlgren. 2004. Meeting report: the molecular background of severe and complicated malaria. **Mol. Biochem. Parasitol**. 134: 37-41.

Wold, S., E. Johansson, M. Cocchi and H. Kubinyi. 1993. 3D-QSAR in Drug Design: Theory, Methods and Applications. pp. 523-549. Leiden. ESCOM.

Wu Y., L.A. Kirkman and T.E. Wellems. 1996. Transformation of *Plasmodium falciparum* malaria parasites by homologous integration of plasmids that confer resistance to pyrimethamine. **Proc. Natl. Acad. Sci. USA.** 93: 1130–1134.

Xue, C.X., S.Y. Cui, M.C. Liu, Z.D. Hu and B.T. Fan. 2004. **Eur. J. Med. Chem**. 39: 745-753.

Yuthavong, Y., T. Vilaivan, N. Chareonsethakul, S. Kamchonwongpaisan, W. Sirawaraporn, R. Quarrell and G. Lowe. 2000. Development of a lead inhibitor for the A16V+S108N mutant of dihydrofolate reductase from the cycloguanil-resistant strain (T9/94) of *plasmodium falciparum*. **J. Med. Chem**. 43: 2738-2744.

_____. 2002. Basic for antifolate action and resistance in malaria. **Microbes Infect**. 4: 175-182.

_____, J. Yuvaniyama, P. Chitnumsub, J. Vanichtanankul, S Chusacultanachai, B. Tarnchompoo, T. Vilaivan and S. Kamchonwongpaisan. 2005. Malarial (*Plasmodium falciparum*) dihydrofolate reductase-thymidylate synthase: structural basis for antifolate resistance and development of effective inhibitors. **Parasitology.** 130: 249-259.

Yuvaniyama, J., P. Chitnumsub, S. Kamchonwongpaisan, J. Vanichtanankul, S. Sirawaraporn, P. Taylor, M. D. Walkinshaw and Y. Yuthavong. 2003. Insights into antifolate resistance from malarial DHFR-TS structures. **Nat**. **Struct. Biol**. 10: 357-365.

APPENDIX

APPENDIX A: Theoretical Background

Theoretical Background in Quantum Chemistry

1. 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

$$H\Psi = E\Psi \tag{13}$$

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

$$H = T + V \tag{14}$$

Then, rewrite equation (13) is;

$$\left\{ \frac{-\bar{h}^2}{2m} \nabla^2 + V \right\} \Psi = E\Psi \tag{15}$$

where

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \tag{16}$$

ħ is Plank's constant divided by 2π. Ψ is the wavefunction which characterizes the particle's properties. E is the energy of the particle.

1.1 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

$$\Psi_i = \sum_{\mu} c_{\mu i} \phi_{\mu} \tag{17}$$

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

$$\sum_{\mu\nu} c^{*}_{\mu i} c_{\nu i} S_{\mu\nu} = \delta_{ij} \tag{18}$$

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

$$S_{\mu\nu} = \int \phi_{\mu}(1)\phi_{\nu}(1)d\tau \tag{19}$$

1.2 Solving for the Molecular Orbital : LCAO-MO- SCF

Introducing Eq. (17) and (18) into Eq. (13), the equation takes the final form generally known as the Roothaan equations as

$$\sum_{\nu} \left(F_{\mu\nu} - \varepsilon_i S_{\mu\nu}\right) c_{\nu i} = 0 \tag{20}$$

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

$$F_{\mu\nu} = H_{\mu\nu}^{core} + \sum P_{\lambda\sigma}\left[(\mu\nu/\lambda\sigma) \quad {}^{1}\!/_{2}(\mu\lambda/\nu\sigma)\right] \tag{21}$$

and density matrix defined as

$$P_{\mu\nu} = 2\sum_{i} c_{\mu i}^{*} c_{\nu i} \tag{22}$$

$$(\mu\nu/\lambda\sigma) = \iint \phi_{\mu}(1)\phi_{\nu}(1)\frac{1}{r_{12}}\phi_{\lambda}(2)\phi_{\sigma}(2)d\tau_{1}\,d\tau_{2} \tag{23}$$

and one-electron orbital energy is

$$\varepsilon_{i} = H_{i}^{(1)} + \sum_{j} 2J_{ij} - K_{ij} \tag{24}$$

where

$$\text{Coulomb integral, } J_{ij} = \sum_{\mu\nu\lambda\sigma} c_{\mu i}^{*} c_{\lambda j}^{*} c_{\nu i} c_{\sigma j} (\mu\nu/\lambda\sigma) \tag{25}$$

$$\text{and Exchange integral, } K_{ij} = \sum_{\mu\nu\lambda\sigma} c_{\mu i}^{*} c_{\lambda i}^{*} c_{\nu i} c_{\sigma i} (\mu\lambda/\nu\sigma) \tag{26}$$

The total electronic energy (ε) is

$$\varepsilon = \sum_{i} \left(\varepsilon_{i} + H_{ii}^{(1)}\right) \tag{27}$$

Therefore, Eq. (20) can be written in matrix form as

$$\mathbf{FC = SCE} \tag{28}$$

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. (28) 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

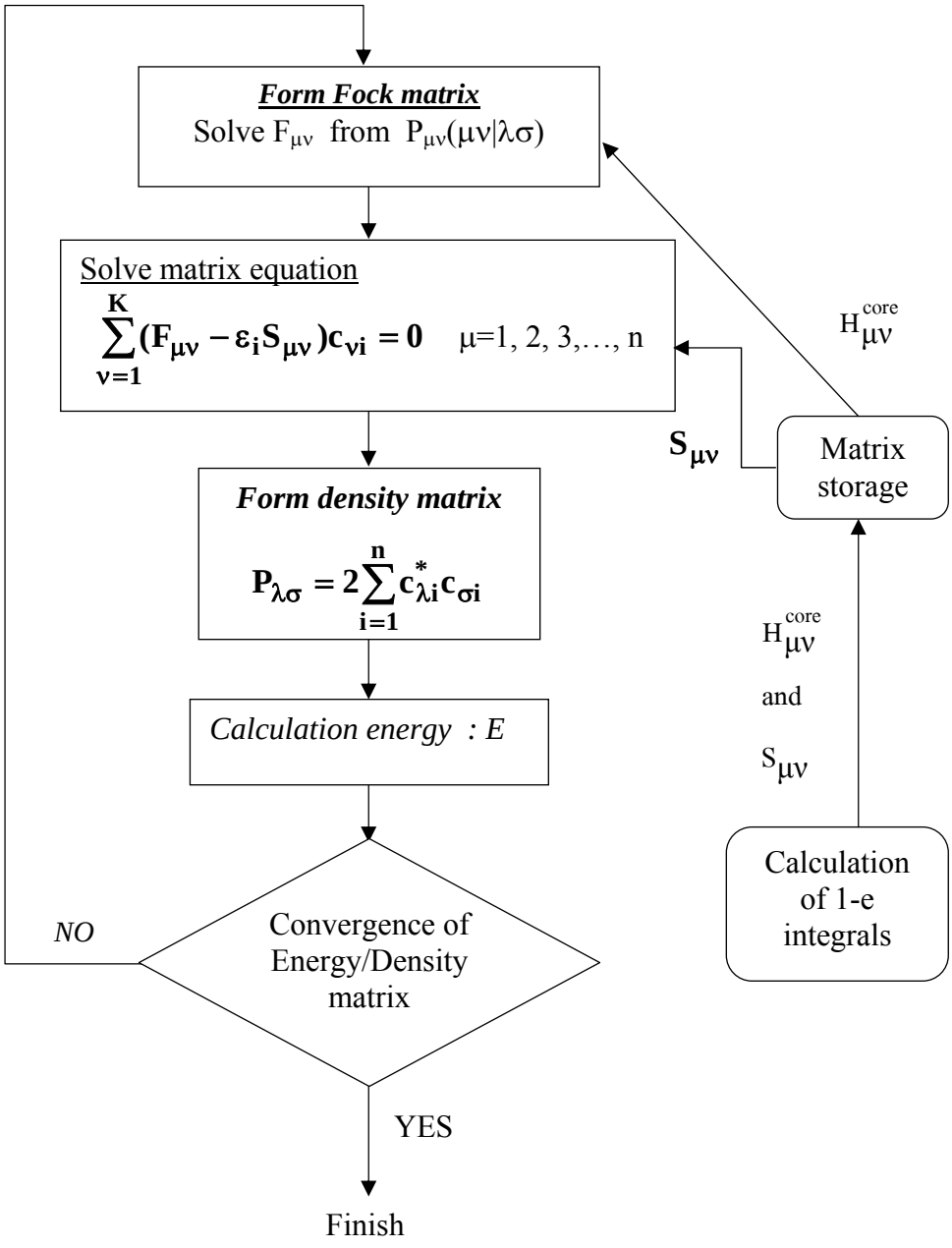
$$F^{\tau} = S^{1/2} F S^{-1/2} \tag{29}$$

$$C^{\tau} = S^{1/2} C \tag{30}$$

then obtain

$$F^{\tau} C^{\tau} = C^{\tau} E \tag{31}$$

Matrix equation (31) 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 assuming 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 Rothaan-Hall equations are outlined in Figure A1.



Appendix Figure A1 Sequence of program step required to solve the Roothaan-Hall equations, Self Consistent Field procedure.

2. Semiempirical Calculations

There are six distinct methods available within MOPAC: MINDO/3, MNDO, AM1, PM3, PM5, and MNDO-*d*. All are semiempirical, and have roughly the same structure. A complete knowledge of these methods is not necessary in order to use MOPAC; however, a superficial understanding of these methods and their relationship to *ab initio* methods is important for using MOPAC and particularly for interpreting the results.

The six methods within MOPAC have many features in common. They are all self-consistent field (SCF) methods, they take into account electrostatic repulsion and exchange stabilization, and, in them, all calculated integrals are evaluated by approximate means. Further, they all use a restricted basis set of one *s* orbital and three *p* orbitals (p_x, p_y, and p_z) per atom (except MNDO-*d*, which has five *d* orbitals in addition to the *s*-*p*basis set) and ignore overlap integrals in the secular equation. Thus, instead of solving

$$\left|H - ES\right| = 0 \tag{32}$$

which H is the secular determinant, S is the overlap matrix, and E is the set of eigenvalues, is solved. These approximations considerably simplify quantum mechanical calculations on systems of chemical interest. As a result, larger systems can be studied. Computational methods are only models, and there is no advantage in rigorously solving Schrödinger's equation for a large system if that system has had to be abbreviated in order to make the calculations tractable. Semiempirical methods are thus seen to be well balanced: they are accurate enough to have useful predictive powers, yet fast enough to allow large systems to be studied.

All the semiempirical methods contain sets of parameters. For MINDO/3 and PM5 atomic and diatomic parameters exist, while MNDO, AM1, PM3, and MNDO-*d* use only single-atom parameters. Not all parameters are optimized for all methods; for example, in MINDO/3, MNDO and AM1 the two electron one center integrals are

normally taken from atomic spectra. In the list given in, parameters optimized for a given method are indicated by '*'. A '+' indicates that the value of the parameter was obtained from experiment (not optimized). Where neither symbol is given, the associated parameter is not used in that method.

All five semiempirical methods also use two experimentally determined constants per atom: the atomic mass of the most abundant isotope and the heat of atomization.

The greatest proportion of the time required to perform an ab-initio Hartree-Fock SCF calculation is invariably calculating and manipulating integrals. The most obvious way to reduce the computational effort is therefore to neglect or approximate some of these integrals. Semiempirical methods achieve this in part by explicit considering only the valence electrons of the system; the core electrons are assumed into the nuclear core. The overlap matrix, S (in equation (31)), is set equal to the identity matrix, I. The main implication of this is that the Roothaan-Hall equations are simplified: FC = SCE becomes FC = CE.

2.1 MNDO

Dewar and Thiel introduced the modified neglect of diatomic overlap (MNDO) method which was based on the neglect of diatomic differential overlap (NDDO); this theory only neglects differential overlap between atomic orbitals on different atoms. The Fock matrix elements in MNDO were as follows:

$$F_{\mu\mu} = H_{\mu\mu}^{core} + \sum_{\nu on A}\left[P_{\nu\nu}(\mu\mu/\nu\nu) - \frac{1}{2}P_{\nu\nu}(\mu\nu/\mu\nu)\right] + \sum_{B\neq A}\sum_{\lambda on B}\sum_{\sigma on B}P_{\lambda\sigma}(\mu\mu/\lambda\sigma) \qquad (33)$$

where $$H_{\mu\mu}^{core} = U_{\mu\mu} - \sum_{B\neq A}V_{\mu\mu B} \qquad (34)$$

$$F_{\mu\nu} = H_{\mu\nu}^{core} + \frac{3}{2}P_{\mu\nu}(\mu\nu/\mu\nu) - \frac{1}{2}P_{\mu\nu}(\mu\mu/\nu\nu) + \sum\sum\sum P_{\lambda\sigma}(\mu\nu/\lambda\sigma) \qquad (35)$$

μ and ν both on A

where $H_{\mu\mu}^{core} = -\sum_{B \neq A} V_{\mu\mu B}$ (36)

$$F_{\mu\nu} = H_{\mu\nu}^{core} + \frac{3}{2} P_{\mu\nu}(\mu\nu / \mu\nu) - \frac{1}{2} \sum_{\lambda on B} \sum_{\sigma on A} P_{\lambda\sigma}(\mu\sigma / \nu\lambda) \quad (37)$$

where $H_{\mu\nu}^{core} = \frac{1}{2} S_{\mu\nu}(\beta_\mu + \beta_\nu)$ μ on A and ν on B (38)

$V_{\mu\mu B}$ and $V_{\mu\nu B}$ are two-center, one-electron attractions between an electron distribution $\phi_\mu \phi_\mu$ or $\phi_\mu \phi_\nu$, respectively, on atom A and the core of atom B. These are expressed as follows.

$$V_{\mu\mu B} = Z_B(\mu_A \mu_A / S_A S_A) \quad (39)$$

$$V_{\mu\nu B} = Z_B(\mu_A \nu_A / S_A S_A) \quad (40)$$

$$E_{AB} = Z_A Z_B (S_A S_A / S_B S_B)\{1 + \exp(-\alpha_A R_{AB}) + \exp(-\alpha_B R_{AB})\} \quad (41)$$

2.2 AM1 and PM3

The Austin Model 1 (AM1) and PM3 are based on MNDO (the name derives from the fact that PM3 is the third parametrization of MNDO, AM1 being considered the second.) AM1 and PM3 modified the core-core repulsions just outside bonding distances. With this modification the expression for the core-core term was related to the MNDO expression by

$$E_{AB} = E_{MNDO} + \frac{Z_A Z_B}{R_{AB}} \left\{ \begin{array}{l} \sum_I K_{Ai} \exp\left[-L_{Ai}(R_{AB} - M_{Ai})^2\right] \\ + \sum_j K_{Bj} \exp\left[-L_{Bj}(R_{AB} - M_{Bj})^2\right] \end{array} \right\} \quad (42)$$

The extra terms define spherical Gaussian functions, the L, M and K parameters were optimized for each atom. PM3 has two Gaussians per atom, while AM1 has two - four Gaussians per atom. In AM1 a sum of Gaussians is employed to better represent the core repulsion behaviour at van der Waals distances.

One advantage of methods parameterised using experimental data is their implicit inclusion of electron correlation effects. However, dependence on experimental data means that semi-empirical methods would not be expected to perform well on unusual types of molecules for which no data are available from which to construct parameters.

Statistical Analysis for 3D-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 variables (x):

$$y = ax_1 + bx_2 ... + k \qquad (43)$$

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:

$$r^2 = \frac{\sum_{i=1}^{n} \left(y_{i,calc} - y_{i,mean} \right)^2}{\sum_{i=1}^{n} \left(y_{i,obs} - y_{i,mean} \right)^2} = 1 - \frac{\sum_{i=1}^{n} \left(y_{i,calc} - y_{i,obs} \right)^2}{\sum_{i=1}^{n} \left(y_{i,obs} - y_{i,mean} \right)^2} \qquad (44)$$

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:

$$F = \frac{n - p - 1}{p} \frac{\sum_{i=1}^{n} \left(y_{i,calc} - y_{i,mean} \right)^2}{\sum_{i=1}^{n} \left(y_{i,obs} - y_{i,calc} \right)^2} \qquad (45)$$

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) degrees of freedom at a chosen confidence level (for instance 95%).

Good statistics 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 inappropriate.

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 interpretation of the model and the identification of outliers, as well as non-linear relationships.

Regression coefficients in terms of original variables can also be computed, so that the PLS solution can be still reported in the traditional form (Eq. 43). Unlike MLR PLS can simultaneously handle more activity.

The optimal number of LVs of a PLS model is usually estimated by cross-validation (CV). CV means that the objects are divided in n groups, a model is derived with n-1 groups and due 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:

$$r_{cv}^{\ 2} = 1 - \frac{\sum_{i=1}^{n} \left(y_{i,pred} - y_{i,obs}\right)^2}{\sum_{i=1}^{n} \left(y_{i,obs} - y_{i,mean}\right)^2} \qquad (46)$$

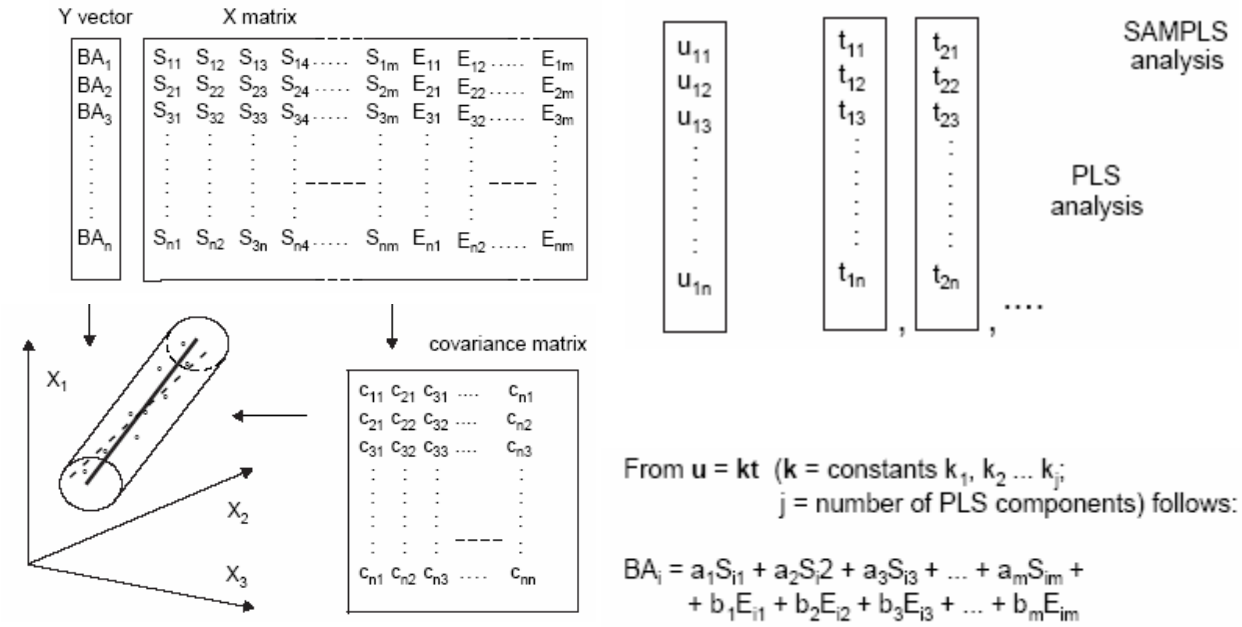
The above formula is analogous to r^2 (Eq. 44) where calculated y values are replaced by predicted values. Other statistical parameters commonly calculated are the SDEP and the S_{PRESS}

$$s = \sqrt{\frac{\sum_{i=1}^{n} (y_{i,pred} - y_{i,obs})^2}{n}} \qquad (47)$$

$$s_{PRESS} = \sqrt{\frac{\sum \left(y_{i,pred} - y_{i,obs}\right)^2}{n - a - 1}} \qquad (48)$$

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 prediction, 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 can be easily understood by a geometrically interpretation as shown in Appendix Figure A2.



Appendix Figure A2 PLS analysis derives vectors ***u*** and ***t*** from the **Y** block (or ***y*** vector; BAi D logarithms of relative affinities or other biological activities) and the **X** block (Sij D steric field variable of molecule i in the grid point j; Eij D electrostatic field variable of molecule i in the grid point j) that are related to principal components. These 'latent variables' are skewed within their confidence hyperboxes to achieve a maximum intercorrelation (diagram). SAMPLS is a PLS modification which first derives the covariance matrix of the **X** block and then the PLS result from this covariance matrix. Especially in cross-validation (see below), SAMPLS analysis is much faster than ordinary PLS analysis

APPENDIX B: Presentation and Proceedings

Oral Presentation and Proceedings

1. Wanwimon Mokmak, Phornphimon Maitarad, Supa Hannongbua, Sumalee Kamchonwongpaisan, Jarunee Vanichtanankul, Tirayut Vilaivan and Yongyuth Yuthavong. Molecular Docking Study on (*R*)- and (*S*)-Configuration of WR99210 Derivatives Against Wild-type *Pf*DHFR. In the proceeding of The 9th Annual National Symposium, On Computational Science and Engineering (ANSCSE 2005), Faculty of science, Mahidol University, Bangkok, Thailand, 23-25 Mar 2005, page 515-521.

2. Wanwimon Mokmak, Phornphimon Maitarad, Supa Hannongbua, Sumalee Kamchonwongpaisan, Jarunee Vanichtanankul, Tirayut Vilaivan and Yongyuth Yuthavong. Comparative Molecular Field Analysis of WR99210 Derivative of *Plasmodium falciparum* Dihydrofolate reductase Inhibitors. In the proceeding of 3rd International Symposium Computational Methods in Toxicology and Pharmacology Integrating Internet Resources (CMTPI2005), Shanghai, China, 29 Oct-1 Nov 2005, page 36.

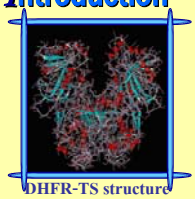
Poster Contribution to Conferences

1. Phornphimon Maitarad, Wanwimon Mokmak, Supa Hannongbua, Sumalee Kamchonwongpaisan, Jarunee Vanichtanankul, Tirayut Vilaivan and Yongyuth Yuthavong. 3D-QSAR Study on Cycloguanil Derivatives Against Quadruple Mutant Type (N51IC59RS108NI164L) of Plasmodium falciparum Dihydrofolate Reductase. 2nd Asian Pacific Conference on Theoretical and Computational Chemistry, 2005, Chulalongkorn University, Bangkok, Thailand.

3D-QSAR Study on Cycloguanil Derivatives Against Quadruple Mutant Type (N51IC59RS108NI164L) of *Plasmodium falciparum* Dihydrofolate Reductase

Phornphimon Maitarad,[1] Wanwimon Mokmak, [1] Supa Hannongbua,[1] Sumalee Kamchonwongpaisan,[2] Jarunee Vanichtanankul,[2] Tirayut Vilaivan [3] and Yongyuth Yuthavong [2]
[1] Department of Chemistry, Faculty of Science, Kasetsart University, Bangkok 10900, Thailand
[2] National Center for Genetic Engineering and Biotechnology, National Science and Technology Development Agency, Pathumthani 12120, Thailand
[3] Department of Chemistry, Faculty of Science, Chulalongkorn University, Bangkok 10330, Thailand

Introduction


DHFR-TS structure

Malaria disease is mostly caused by protozoan parasites of the *Plasmodium falciparum*. The dihydrofolate reductase (DHFR) of *Plasmodium falciparum* is one of the target in malarial chemotherapy. One of the potent and selective antifolate drugs is the cycloguanil (4,6-diamino-1,2-dihydrotriazine) (Cyc). Unfortunately, the emergence of resistant parasites has reduced the utility of this drug in the treatment of malaria disease. Mutations of residues at amino acid positions 16, 51, 59, 108 and 164 of *Pf*DHFR have been reported. However, the review of the literatures has reported that the quadruple mutant (N51I+C59R+S108N+I164L) is highly resistant to Cyc drugs. Therefore, understanding of the structural requirements for antimalarial activity in the Cyc derivatives against quadruple mutant type DHFR is important in guiding and designing for new and more potent inhibitors.

Material & Methods

All Cyc derivatives were synthesized and tested biological activity in term of *Ki*. In the present analysis, based on the range of these biological data, the structures of 120 Cyc compounds serve as a training set. Twelve additional inhibitors were used as a test set to evaluate the predictive ability of the resulting models.

Cyc derivatives were built using the HYPERCHEM version 7.0 software.

Structural energy minimization was performed by using AM1 semiempirical molecular orbital method implemented in the GAUSSIAN 98 program.

Alignment in SYBYL6.91

Comparative Molecular Filed Analysis (CoMFA) calculated the steric and electrostatic fields within 4 Å in all direction (+1 sp^3 C probe atom in (2.0 Å) grid, cutoff $\pm$ 30 kcal/mol).
Comparative Molecular Similarity Indices Analysis (CoMSIA) (steric, electrostatic, hydrophobic, hydrogen bond donor and hydrogen bond acceptor) calculated in the same cubic lattice as CoMFA.

3D-QSAR analyses were calculated by using the **Partial least squares (PLS)** method. Cross validation by **leave-one-out (LOO)** method were performed to determine the optimum number of components.

r^2_{cv} value

contour maps

Conclusion

3D-QSAR analyses were successfully to discriminate structural requirements of Cyc derivatives against quadruple mutant type (N51IC59RS108NI164L) of *Pf*DHFR. The information of 3D-QSAR results provides a helpful guideline to design and predict the affinity of novel Cyc compounds with enhanced quadruple mutant type of *Pf*DHFR inhibitory activities prior to synthesis.

Acknowledgement

This work was supported by the Thai Graduate Institute of Science and Technology (TGIST) Funds (TG-B-11-22-11-609M), the Postgraduate Education and Research on Petroleum and Petrochemical Technology (MUA-ADB), and the Kasetsart University Research and Development Institute (KURDI). We would like to thank the high performance-computing center of the National Electronics and Computer Technology (NECTEC), National Center for Genetic Engineering and Biotechnology (BIOTEC), LCAC and computing center of KU for research facilities.

Results and Discussion

The statistical results of the CoMFA and CoMSIA analyses are summarized in Table 1. The best predictions were obtained with CoMFA standard model and CoMSIA combined steric, electrostatic and hydrophobic fields. The relative contributions of the CoMFA and CoMSIA are shown in Table 2. Both models can be used to predict the activities of test set compounds with $r^2_{prediction}$ value 0.686 and 0.738 of CoMFA and CoMSIA, respectively.

	CoMFA	CoMSIA
Number of components	6	6
Crossvalidation r^2_{cv}	0.737	0.703
S_{press}	0.551	0.593
Conventional r^2	0.943	0.954
Standard error of estimate	0.257	0.225
F value	265.871	362.046
Probe atom	sp^3 C atom	

Table 1. Summary of CoMFA and CoMSIA statistical results

Relative contributions	
CoMFA – steric	0.770
CoMFA - electrostatic	0.230
CoMSIA - steric	0.175
CoMSIA - electrostatic	0.381
CoMSIA - hydrophobic	0.444

Table 2. Relative contributions of the CoMFA and CoMSIA

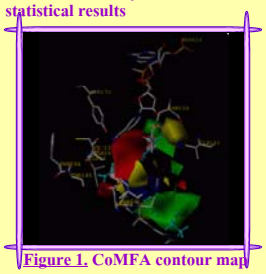


Figure 1. CoMFA contour map

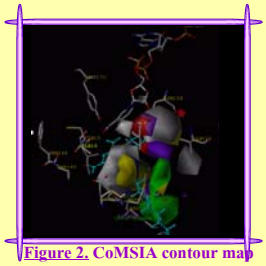
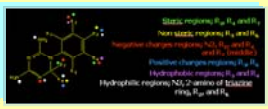


Figure 2. CoMSIA contour map



CoMFA and CoMSIA contour maps are shown in Figure1 and 2, respectively. Sterically favored green region are found surrounding R_2, R_4 and R_7 positions where the bulkly groups are well tolerated. In contrast, the sterically unfavored yellow regions are found near R_3 and R_6 positions where large subsituents would decrease the activity. The red regions indicate that high electron density groups would increase the activity. In the other hand, low electron density groups are suitable in blue regions. The hydrophobic favorable in purple regions are closed to R_2 and R_4 positions which mean that more hydrophobic subsituents at these positions would increase activity.

CURRICULUM VITAE

NAME : Miss Wanwimon Mokmak

BIRTH DATE : February 20, 1980

BIRTH PLACE : Chainat, Thailand

NATIONALITY : Thai

RELIGION : Buddhism

EDUCATION

: YEAR	**INSTITUTION**	**DEGREE**
1998-2002	Thammasat University	B.Sc. (Chemistry)

SCHOLARSHIPS

: YEAR	**GRANT SOURCE**
1998 - 2002	National Science and Technology Development Agency Scholarship National Science and Technology Development Agency Scholarship, THAILAND
2003-2004	The Postgraduate Education and Research on Petroleum and Petrochemical Technology, MUA-ADB, THAILAND
2004-2005	Thailand Graduate Institute of Science and Technology, TGIST, THAILAND

ORALS PRESENTATIONS AND PROCEEDINGS

1. Wanwimon Mokmak, Phornphimon Maitarad, Supa Hannongbua, Sumalee Kamchonwongpaisan, Jarunee Vanichtanankul, Tirayut Vilaivan and Yongyuth Yuthavong. Molecular Docking Study on (*R*)- and (*S*)-Configuration of WR99210 Derivatives Against Wild-type *Pf*DHFR. In the proceeding of The 9th Annual National Symposium, On Computational Science and Engineering (ANSCSE 2005), Faculty of science, Mahidol University, Bangkok, Thailand, 23-25 Mar 2005, page 515-521.

2. Wanwimon Mokmak, Phornphimon Maitarad, Supa Hannongbua, Sumalee Kamchonwongpaisan, Jarunee Vanichtanankul, Tirayut Vilaivan and Yongyuth Yuthavong. Comparative Molecular Field Analysis of WR99210 Derivative of *Plasmodium falciparum* Dihydrofolate reductase Inhibitors. In the proceeding of 3rd International Symposium Computational Methods in Toxicology and Pharmacology Integrating Internet Resources (CMTPI2005), Shanghai, China, 29 Oct-1 Nov 2005, page 36.

PARTITIONING CONGRESS

9-13 Nov, 2003	Future trends in modeling of catalytically active surfaces, Rama Garden Hotel, Bangkok, Thailand.
24-28 May, 2004	The second Thai Summer School of Computational chemistry, TS_2C_2, Department of Chemistry, Faculty of Science, Chang Mai University.
21-23 July, 2004	The 8th Annual National Symposium, On Computational Science and Engineering (ANSCSE 2004), Suranaree University of Technology, Thailand.
19-21 Oct, 2004	30th Congress on Science and Technology of Thailand (STT 2004), Impact Exhibition and Convention Center, Muang Thong Thani, Bangkok, Thailand.
23-25 Mar, 2005	The 9th Annual National Symposium, on Computational Science and Engineering (ANSCSE 2005), Mahidol University, Bangkok Thailand.
28-30 Mar, 2005	NSTDA Annual Conference 2005 (S&T in Thailand: Towards the Molecular Economy), National Science and Technology Development Agency, Pathumthani,Thailand.
2-6 May, 2005	2nd Asian Pacific Conference on Theoretical & Computational Chemistry(APCTCC), Chulalongkorn University, Bangkok, Thailand.

22-23 Sep, 2005	2nd Protein Research Network Symposium on " Protein: Structure, Function, and Proteomics", Chulabhorn Research Institute Conference Center, Bangkok, Thailand.
29 Oct-1 Nov, 2005	3rd International Symposium Computational Methods in Toxicology and Pharmacology Integrating Internet Resources(CMTPI2005), Shanghai, China.