



# **THESIS**

**MOLECULAR MODELLING AND QUANTUM CHEMICAL  
CALCULATIONS STUDY ON ANTIFOLATE ANTIMALARIAL  
INHIBITORS**

**PHORNPHIMON MAITARAD**

**GRADUATE SCHOOL, KASETSART UNIVERSITY**

**2008**

**THESIS**

**MOLECULAR MODELLING AND QUANTUM CHEMICAL  
CALCULATIONS STUDY ON ANTIFOLATE ANTIMALARIAL  
INHIBITORS**

**PHORNPHIMON MAITARAD**

**A Thesis Submitted in Partial Fulfillment of  
the Requirements for the Degree of  
Doctor of Philosophy (Chemistry)  
Graduate School, Kasetsart University  
2008**

Phornphimon Maitarad 2008: Molecular Modelling and Quantum Chemical Calculations Study on Antifolate Antimalarial Inhibitors. Doctor of Philosophy (Chemistry), Major Field: Physical Chemistry, Department of Chemistry. Thesis Advisor: Associate Professor Supa Hannongbua, Dr.rer.nat. 125 pages.

Comparative molecular field analysis (CoMFA) was performed on twenty-three pyrimethamine (Pyr) derivatives acted against quadruple mutant type (Asn51Ile, Cys59Arg, Ser108Asn, Ile164Leu) *plasmodium falciparum* dihydrofolate reductase (*Pf*DHFR). The best model ( $r_{cv}^2 = 0.702$ ,  $S_{press} = 0.608$ ,  $r_{nv}^2 = 0.980$ ,  $s = 0.156$ , and  $r_{testset}^2 = 0.698$ ) with combined three types of probe atoms,  $C_{sp^3}$  (+1),  $O_{sp^3}$  (-1) and H (+1), can be used to explain steric and electrostatic structural requirements for Pyr compounds. In addition, the CoMFA method was also performed on twenty-five cycloguanil (Cyc) derivatives derived both wild type ( $r_{cv}^2 = 0.727$  and  $r^2 = 0.985$ ) and quadruple mutant ( $r_{cv}^2 = 0.786$  and  $r^2 = 0.979$ ) *Pf*DHFR models. These two models can be determined the different structural requirements for the potency of inhibiting between the wild type and the quadruple mutant *Pf*DHFRs very well.

Deeply in molecular details, an understanding of particular interaction energy between antifolate inhibitors and surrounding residues in the binding pocket was performed by using MP2/6-31G(d,p) accurate quantum chemical calculations. The obtained results clearly demonstrate that Asn108 is the cause of Pyr and Cyc resistances. Furthermore, we investigated the different binding energy between the potent WR99210 inhibitor and the poor inhibitors represented by Pyr and Cyc, active against the quadruple mutant type of *Pf*DHFR. The AMBER molecular dynamics simulations are well bimolecular force fields for constructing and optimizing the complexes of three antifolates and the mutant *Pf*DHFR. Consequently, the binding energy of all complexes was extrapolated using the ONIOM3 (B3LYP/6-31G(d,p):PM3:UFF) calculations. The WR99210/*Pf*DHFR gave highest total binding energy. Basically, the loss of ligand/enzyme binding is the key point to clear that any mutations lead to unstable of ligand in the binding pocket. The obtained molecular dynamics minimizations and ONIOM3 binding energy extrapolations indicate that the mutated residue Asn108 is the main cause of the less potency  $K_i$  for Pyr and Cyc drugs which due to the loss of high quantum chemical energy for small region (antifolate inhibitor + Asp54 + Asn108 + Leu164). Hence, this appearance accorded to experimental biological activity.

The goal of the research is the demonstrating the cause of the antifolate drugs resistance which came from the changing of Ser108 to Asn108 using the CoMFA and quantum chemical calculations studies. Moreover, the obtained CoMFA models can be used to guide the new designed potent antifolate inhibitors for the mutant enzyme.

---

Student's signature

---

Thesis Advisor's signature

/ /

## ACKNOWLEDGEMENT

I would like to express my deep and sincere gratitude to my supervisor, Associate Professor Dr. Supa Hannongbua. Her wide knowledge and her logical way of thinking have been of great value for me. Her understanding, encouragement and guidance have provided a good basis for the present research works. My supervisor not only guides me in aspect of scientific research but also teaches me many great things, such as, positive thinking, good performance, active coordination and kind attitude.

I am deeply grateful to my co-supervisor, Dr. Sumalee Kamchonwongpaisan, Senior Researcher at National Center for Genetic Engineering and Biotechnology (BIOTEC), for her detailed, valuable advice, kind help and constructive comments, and for her important collaboration throughout this work.

I owe my most sincere gratitude to Professor Dr. Yongyuth Yuthavong, who gave me a great an opportunity to work with his malaria research unit at National Center for Genetic Engineering and Biotechnology (BIOTEC) and his very kind help.

I warmly thank Dr. Chak Sangma for his program supporting around my work and interesting explorations in operations have been very helpful for this study.

My warm thanks are due to Professor Jumras Limtrakul, who provided facility on the supercluster computer calculations and gave me a great experience for PTT projects. I also wish to thank the Large Scale Laboratory, National Electronic and Computer Technology Center (NECTEC), for supporting Sybyl program throughout my research works.

I wish to thank Mr. Ben Parslew and Mr. Andy for revising the English of my manuscripts. I also thank to Dr. Thanyada Rungrotmongkol, Dr. Rungtiwa Chid tong, Dr. Pensri Srivub, Dr. Patchreenart Saparpakorn, Mr. Witcha Treesuwan, Mr. Bundet Boekfa, Ms. Pornthip Boonsri, Ms. Natcha Injan and Ms. Wanwimon Mokmak as

well as LCAC members whose names were not mentioned here for their friendships, valuable assistances and excellent helps.

During this work I received full supported from the Thai Graduate Institute of Science and Technology scholarship (TGIST), National Science and Technology Development Agency. I also have partial supported from the Commission on Higher Education (CHE), Graduate School of Kasetsart University and the Postgraduate Education and Research Program in Petroleum Petrochemical Technology and Advance Material.

I owe my loving thanks to my best friend Mr. Pakasit Jirask in manner of his kind assistance and his time spent in my thesis preparation behind the scenes without ever receiving and he always besides me when I fell down.

Finally, I would like to express my most profound gratitude to my lovely parents and two sisters for their love, sincere care, constant supports and great encouragement throughout the duration of my educations.

Phornphimon Maitarad

April, 2008

## TABLE OF CONTENTS

	<b>Page</b>
TABLE OF CONTENTS	i
LIST OF TABLES	ii
LIST OF FIGURES	iii
LIST OF ABBREVIATIONS	vii
INTRODUCTION	1
LITERATURE REVIEW	12
METHODS OF CALCULATIONS	18
RESULTS AND DISCUSSION	39
CONCLUSION	72
LITERATURE CITED	76
APPENDIX	92

## LIST OF TABLES

Table		Page
1	Data set used for CoMFA analysis with $K_i$ (nM) and $pK_i$ values in the quadruple mutant of <i>Pf</i> DHFR.	19
2	Data set used for CoMFA analysis with their $K_i$ (nM) and $pK_i$ (M) values in the wild type (wt) and mutant (mut) <i>Pf</i> DHFR models.	20
3	Interaction energy (kcal/mol) between three antifolate and four mutated residues of <i>Pf</i> DHFR using MP2/6-31G(d,p) level of calculations	35
4	Details of ONIOM3 layer partitions and level of calculations	36
5	PLS statistical results of CoMFA models for quadruple mutant <i>Pf</i> DHFR	40
6	Actual (Act) and predicted (Pred) $pK_i$ values and the residuals ( $\Delta$ ) of the training set and test set molecules for the mutant <i>Pf</i> DHFR models	41
7	Particular interaction energy (kcal/mol) of <b>1</b> (Pyr) and <b>6</b> with individual residues, calculated by MP2/6-31G(d,p) and MP2/6-31G(d,p) with BSSE-CP.	46
8	PLS statistical results of CoMFA models of Cyc derivatives active against wild type and quadruple mutant type <i>Pf</i> DHFRs	49
9	Actual (Act) and predicted (Pred) $pK_i$ values and the residuals ( $\Delta$ ) of the training set and test set molecules for the wild type models	50
10	Actual (Act) and predicted (Pred) $pK_i$ values and the residuals ( $\Delta$ ) of the training set and test set molecules for the quadruple mutant type models.	51
11	ONIOM3 Energy terms in kcal/mol for antifolate ligands bound in the same binding pocket in quadruple mutant type <i>Pf</i> DHFR.	68

## LIST OF FIGURES

Figure		Page
1	Alphonse Laveran was the first to notice parasites in the blood of a patient suffering from malaria.	1
2	Ronald Ross was the first to demonstrate that a mosquito could transmit a malaria parasite.	2
3	Biosynthetic cycle of dTMP (R = pABA + L-glutamate)	3
4	The contribution of each naturally mutations of <i>PfDHFR</i>	4
5	Chemical structures of proguanil and its derivatives	6
6	Chemical structures of pyrimethamine	7
7	Biological activity of Pyr, Cyc and WR99210 antifolates against to wild type and quadruple mutant type (Asn51Ile, Cys59Arg, Ser108Asn, Ile164Leu) <i>PfDHFR</i> s	8
8	Example of a three-layer partitioning scheme	15
9	The energy extrapolation scheme of the three-layer ONIOM method	16
10	Template structure of 5-phenyl-2,4-diaminoPyrimidine derivatives	18
11	Template structure of 1,3,5 dihydrotriazine derivatives	20
12	Example of graphic representation of the binding site with pyrimethamine (compound <b>4</b> ), the cofactor NAD was also presented and caused limited space for position 5 substitution.	22
13	Template alignment of Pyr derivatives shown with the asterisks	23
14	Template alignment of Cyc derivatives displayed with the asterisks	24
15	The backbone superimposition between the X-ray <i>PfDHFR</i> /WR99210 complex structures of quadruple mutant (PDB entry 1J3K shown in green color) and <i>PfDHFR</i> /Pyr double mutant (PDB entry 1J3J shown in orange color), indicating the similarity	26

binding position of both two inhibitors.

### LIST OF FIGURES (Continued)

Figure		Page
16	The 2D scheme of the adopted model system of Pyr inhibitor bound to the quadruple mutant type (Asn51Ile, Cys59Arg, Ser108Asn, and Ile164Leu) of <i>Pf</i> DHFR binding site.	28
17	The 2D scheme of the adopted model system of Cyc bound to the wild type and quadruple mutant type <i>Pf</i> DHFR binding sites.	30
18	The 2D scheme of the adopted model system of flexible potent Cyc derivative, WR99210, bound to the <i>Pf</i> DHFR binding site.	31
19	Schematic representation of the three-layer ONIOM; region of high level represented by ball and stick, medium level represented by pink stick and whole enzyme using low level of calculations	36
20	Extrapolation schematic representation of the three-layer ONIOM partitions and energy calculations.	37
21	Plot of the predicted and actual $pK_i$ values of the training set and the test set molecules with CoMFA model IV	41
22	CoMFA (stdev.*coeff.) sterically favored areas are represented by green regions. Sterically unfavored areas are represented by yellow regions (level of steric contour contribution = 80%) and compound <b>12</b> is represented by ball and stick.	43
23	CoMFA (stdev.*coeff.) negative charge favored area is represented by the red region. Positive charge favored area is represented by the blue region (level of electrostatic contour contribution = 80%) and compound <b>12</b> is represented by ball and stick.	44
24	H-bond distances between inhibitor and residues in the binding pocket; (A) compound <b>1</b> and (B) compound <b>6</b> (in Å).	46

## LIST OF FIGURES (Continued)

Figure		Page
25	The electrostatic potential is shown on the solvent accessible surface as red for negative and blue for positive values for Asn108 interacted with (A) compound <b>1</b> (B) and compound <b>6</b> .	47
26	Plotted graph of the predicted vs actual activities of the training set and the test set molecules (A) wild type-Model III (B) mutant-Model VII	52
27	The steric contour maps of wild type <i>Pf</i> DHFR CoMFA model. The favorable steric areas (contribution levels 80%) with more bulk are indicated by green regions, where as the disfavorable steric areas (contribution levels 20%) are shown by yellow regions. The compound <b>14</b> is represented by ball and stick.	54
28	The electrostatic contour maps of wild type <i>Pf</i> DHFR CoMFA model. The favorable electrostatic areas (contribution levels 80%) with positive charges are indicated by blue regions, whereas the favorable electrostatic areas (contribution levels 20%) with negative charges are show by red regions. The compound <b>14</b> is represented by ball and stick.	55
29	The steric contour maps of quadruple mutant type <i>Pf</i> DHFR CoMFA model. The favorable steric areas (contribution levels 80%) with more bulk are indicated by green regions; where as the disfavorable steric areas (contribution levels 20%) are shown by yellow regions. The compound <b>14</b> is represented by ball and stick.	57
30	The electrostatic contour maps of quadruple mutant type <i>Pf</i> DHFR CoMFA model (st.dev. $\times$ coeff.). The favorable electrostatic areas (contribution levels 80%) with positive charges are indicated by blue regions, whereas the favorable electrostatic areas (contribution levels 20%) with negative charges are show by red regions. The compound <b>14</b> is represented by ball and stick.	58

## LIST OF FIGURES (Continued)

Figure		Page
31	The obtained MP2/6-31G(d,p) with BSSE-CP interaction energies of Cyc and individual amino acids surrounding the binding pocket of wild type (A) and quadruple mutant type (B) <i>Pf</i> DHFRs	60
32	H-bond distances between Cyc inhibitor and residues in the binding pocket; (A) wild type and (B) quadruple mutant type DHFRs (in Å).	61
33	The electrostatic potential is shown on the van der Waal surface as red for negative and blue for positive values for Cyc and Ser108 (A) and Asn108 (B).	62
34	The MP2/6-31G(d,p) with BSSE-CP energetic comparison between Cyc and WR99210 in case of (A) wild type and (B) quadruple mutant type <i>Pf</i> DHFR binding sites	64
35	RMSD of the three antifolates: Cyc (blue line), Pyr (orange line) and WR99210 (purple line) in Å	65
36	RMSD of the back bond amino acids atoms from the quadruple mutant <i>Pf</i> DHFR obtained from the complex structures: Cyc (pink line), Pyr (green line) and WR99210 (brown line)	66
37	H-bond distances between small regions of quadruple mutant <i>Pf</i> DHFR binding pocket and ligands (A) Cyc (B) Pyr and (C) WR99210 (in Å).	69
38	The electrostatic potential is shown on the van der Waal surface as red for negative and blue for positive values for Asn108 and three types of antifolates, Pyr, Cyc and WR99210.	70

## LIST OF ABBREVIATION

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

## LIST OF ABBREVIATION (Continued)

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

## LIST OF ABBREVIATION (Continued)

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