

**COMPUTATIONAL MODELING OF ARTEMISININ AND  
ITS STRUCTURAL DERIVATIVES  
- INTERACTION, ACTIVITY AND MECHANISM**

by

Mani Srivastava

**A THESIS SUBMITTED IN FULFILLMENT OF THE REQUIREMENTS  
FOR THE DEGREE OF**

**DOCTOR OF PHILOSOPHY**

**IN**

**BIOINFORMATICS**



**JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY  
WAKNAGHAT, SOLAN 173215, HIMACHAL PRADESH, INDIA**

**DECEMBER 2009**

Dr. Pradeep Kumar Naik  
Assistant Professor  
Biotechnology & Bioinformatics



Jaypee University of  
Information Technology  
Waknaghat-173215  
Solan, Himachal Pradesh  
Phone No.: 91-1792-239227  
Fax No.: 91-1792-245362

## CERTIFICATE

This is to certify that the thesis entitled "Computational modeling of artemisinin and its structural derivatives - interaction, activity and mechanism" submitted by Mani Srivastava to the Jaypee University of Information Technology, Waknaghat in fulfillment of the requirement for the award of the degree of **Doctor of Philosophy in Bioinformatics (Science)** is a record of bona fide research work carried out by her under my supervision and guidance and no part of this work has been submitted for any other degree or diploma.

(Dr. P. K. Naik)

## DECLARATION

I hereby declare that the work presented in this thesis has been carried out by me under the supervision of Dr. Pradeep Kumar Naik, Department of Biotechnology & Bioinformatics, Jaypee University of Information Technology, Waknaghat, Solan-173215, Himachal Pradesh, and has not been submitted for a degree or diploma of any other university. All assistance and help received during the course of the investigation has been duly acknowledged.

Mani Srivastava

## ACKNOWLEDGEMENT

***“To speak gratitude is courteous and pleasant, to enact gratitude is generous and noble, but to live gratitude is to touch Heaven”***

*Completing a PhD is truly a marathon event, and I would not have been able to complete this journey without the aid and support of countless people over the past three years. I thank my parents for instilling in me confidence and a drive for pursuing my PhD. The great desire to acquire higher qualifications and pursue the research drove me to promise my parents, teachers, brothers and my sisters, a useful research work. My promise was to live up to their expectations. And I kept my word.*

*I would like to express my heartfelt gratitude to all those who have contributed directly or indirectly towards obtaining my doctorate degree. I am highly indebted to my esteemed supervisor, Dr. Pradeep Kumar Naik, who has guided me through thick and thin. I could not have imagined having a better advisor and mentor for my Ph.D study. His positive attitude towards research and zest for high quality research work has prompted me for the timely completion of the work. I deem it a privilege to be a scholar doing research under Dr. P. K. Naik who has endeared himself to his students and scholars.*

*I cherish the years spent in the department of Bioinformatics and Biotechnology. This project would not have been possible without the guidance and active support of the teaching and administrative staff of JUIT. Therefore, in so many ways, I feel indebted to Dr. Y. Medury (Vice Chancellor, JUIT), Brigadier (retd). Balbir Singh (Registrar, JUIT) and Prof. R. S. Chauhan (Head of the department). They have throughout helped me in providing all kinds of facilities to carry out my research work. Also, my sincere thanks go to Dr. Gyan Prakash Mishra (Scientist DRDO, Leh), Harvindar Singh and all the faculty members of the Dept. of Biotechnology and Bioinformatics for keeping their faith in me and backing me with their moral support.*

*Further, I would personally like to thank Mr. Shree Ram, library in charge of JUIT, for his support in providing me with journals and articles during research work.*

*Besides, I can never ever forget my seniors Dr. Md Afroz Alam, Dr. Hemant Sood and colleagues Arti, Anu, Payal, Nidhi, Sree Krishna and friends Shweta, Priyanka, Akriti, Karnika, Ashu, Poonam who have always been a constant source of inspiration and helped me in a numerous ways. I also thank Technical staff of Department of Biotechnology and Bioinformatics Ms. Somlata Sharma, Ismail, Ravikant, Baleshwar, Mrs. Mamta for their assistance and valuable contributions.*

*Sincere thanks to Dr. S. B. S. Mishra (Department of Chemistry, NIT Jalandhar) who has always been inspirational for putting in hard work.*

*I shall remain ever grateful to Retd. Reader in English – cum - Principal, Sadananda Naik and his wife Mrs. Naik for going through the manuscript and making valuable suggestions for fine tuning the Language aspect.*

*Finally, I wish to thank the respondents of my study and all the individuals whose comments, reviews and suggestions have helped my project in developing an informative insight. Lastly, I extend my apologies to anyone I failed to mention.*

(Mani Srivastava)

# CONTENTS

Abstract	i-iv
Chapter 1	1-19
<b>Introduction</b>	
1.1 Epidemiology of malaria	
1.2 Life cycle of <i>P. falciparum</i>	
1.3 Treatment of malaria – Current antimalarials	
1.3.1 Quinine related drugs	
1.3.2 Drug resistance in <i>P. falciparum</i> malaria	
1.4 Artemisinin and its derivatives as antimalarial drugs	
1.4.1 The plant <i>Artemisia annua</i>	
1.4.2 Chemical structure and metabolism	
1.4.3 Antimalarial activity	
1.4.4 Artemisinin derivative	
1.4.5 Pharmacology	
1.4.6 Artemisinin drugs in the treatment of <i>Plasmodium falciparum</i> malaria in India	
1.5 Mechanism of action	
1.5.1 PfATP6, a target for artemisinin	
1.6 Computational studies and its utility in drug design	
1.6.1 Structure-based virtual screening methods for computer-aided drug discovery	
1.6.2 Overview of the works	
Chapter 2	
<b>Computational and molecular modeling evaluation of the antimalarial activity of artemisinin analogues: Molecular Docking and rescoring using prime/MM-GBSA approach</b>	<b>20-56</b>
Abstract	
2.1 Introduction	
2.2 Materials and methods	
2.2.1 Receptor preparation	
2.2.2 Atomic charge calculations	
2.2.3. Preparation of ligands	
2.2.4 Docking of the ligands	
2.2.5. Rescoring using Prime/MM-GBSA	
2.3 Results and Discussion	
2.3.1. Effect of atomic charges	
2.3.2. Effect of artemisinin structure	
2.3.3. Effect of heme structure	
2.3.4 Docking of artemisinin derivatives	
2.3.5 Building models for prediction of pIC <sub>50</sub> using Glide score and binding free energy	
2.4. Conclusion	

Chapter 3	
<b>Application of linear interaction energy method for binding affinity calculations of artemisinin analogues using continuum solvent model and prediction of antimalarial activity</b>	<b>57-87</b>
Abstract	
3.1 Introduction	
3.2 Materials and Methods	
3.2.1 LIE Methodology	
3.2.2 Computational details	
3.2.3. Receptor preparation	
3.2.4. Preparation of ligands	
3.2.4. Docking of the ligands	
3.2.5. LIE calculations	
3.3 Results and Discussions	
3.4 Conclusions	
Chapter 4	
<b>Quantitative structure-activity relationship (QSAR) of the artemisinin: the development of predictive in vitro antimalarial activity models</b>	<b>88-116</b>
Abstract	
4.1 Introduction	
4.2 Materials and Methods	
4.2.1. Data set	
4.2.2. Descriptor calculation	
4.2.3. Regression analysis	
4.2.4. Validation test	
4.3. Results and Discussion	
4.4. Conclusion	
Chapter 5	
<b>The Binding Modes and Binding Affinities of Artemisinin Derivatives with <i>Plasmodium falciparum</i> Ca<sup>2+</sup>-ATPase (PfATP6)</b>	<b>117-157</b>
Abstract	
5.1. Introduction	
5.1.2. Overview of the methodologies tested in this study	
5.1.3. Glide Docking	
5.1.4. Multi-Ligand Bimolecular Association with Energetics (eMBrAcE)	
5.1.5. Prime MM-GBSA	
5.2. Materials and Methods	
5.2.1. Sequence analysis	
5.2.2. Homology model construction	
5.2.3. Ligand binding site prediction	
5.2.4. Preparation of the ligands	
5.2.5 Docking of the ligands	
5.2.6 Ligand & Structure-Based Descriptors (LSBD) protocol	
5.2.7 MM and Binding Free Energies	

5.3 Results and Discussion	
5.3.1 Building models for prediction of Log RA using Glide score and Prime/MM-GBSA energy	
5.3.2 Linear optimization of energy parameters vs Activity	
5.4 Conclusions	
Conclusion	<b>159-161</b>
Bibliography	<b>163-179</b>
List of publications	<b>180</b>

## LIST OF TABLES

Table 2.1.	Artemisinin analogues with antimalarial activities against the drug-resistant malarial strain <i>P. falciparum</i> (W-2 clone) used in the work.	29-36
Table 2.2	Experimental and theoretical values of the 1,2,4-trioxane ring parameters in artemisinin (bond lengths in Å; bond angles and torsional angles in degrees).	37
Table 2.3	Results for docking of heme-pdb with different atomic charges and the artemisinin with HF/3-21G charge.	39
Table 2.4	Results for docking of heme-pdb with HF/6-311G** charge and the artemisinin with different atomic charges.	39
Table 2.5	Atomic charges of four oxygen atoms in artemisinin for all charge schemes.	40
Table 2.6	Results for docking of haem-pdb with artemisinin (QHS) analogues as well as computed activity using Glide score as a descriptor.	44-47
Table 2.7	Results for rescoring using Prime/MM-GBSA of haem-pdb with artemisinin analogues as well as computed activity using $\Delta G_{\text{bind}}$ energy as a descriptor.	50-53
Table 3.1	Artemisinin analogues with antimalarial activities against the drug-resistant malarial strain <i>P. falciparum</i> (W-2 clone) used in the work.	64-72
Table 3.2	Experimental and theoretical values of the 1,2,4-trioxane ring parameters in artemisinin (bond lengths in Å; bond angles and torsional angles in degrees). The atom numbers are with respect to Figure 3.1.	73
Table 3.3	Average electrostatic (ele), van der Waals (vdw) and cavity (cav) energy terms as well as binding affinity model calculations for the Training set inhibitors using SGB-LIE method.	78-80
Table 3.4	Average electrostatic (ele), van der Waals (vdw) and cavity (cav) energy terms as well as binding affinity model calculations for the Test set inhibitors using SGB-LIE method.	81-82
Table 4.1a	Artemisinin analogues with antimalarial activities against the drug-resistant malarial strain <i>P. falciparum</i> (W-2 clone) used in the work.	93
Table 4.1b	Deoxy-artemisinin derivatives with antimalarial activity against the drug-resistant malarial strain <i>P. falciparum</i> (W-2 clone) used in the work.	94

Table 4.1c	Seco-artemisinin derivatives with antimalarial activity against the drug-resistant malarial strain <i>P. falciparum</i> (W-2 clone) used in the work.	94
Table 4.1d	10-Substituted artemisinin derivatives with antimalarial activities against the drug-resistant malarial strain <i>P. falciparum</i> (W-2 clone) used in the work.	95
Table 4.1e	11-Aza-artemisinin derivatives with antimalarial activities against the drug-resistant malarial strain <i>P. falciparum</i> (W-2 clone) used in the work.	96
Table 4.1f	Artemisinin derivatives lacking the D-ring with antimalarial activity against the drug-resistant malarial strain <i>P. falciparum</i> (W-2 clone) used in the work.	97
Table 4.1g	Miscellaneous Artemisinin derivatives with antimalarial activity against the drug-resistant malarial strain <i>P. falciparum</i> (W-2 clone) used in the work.	98
Table 4.1h	9-Substituted Artemisinin derivatives with antimalarial activity against the drug-resistant malarial strain <i>P. falciparum</i> (W-2 clone) used in the work.	99
Class 4.1i	Dihydroartemisinin derivatives with antimalarial activity against the drug-resistant malarial strain <i>P. falciparum</i> (W-2 clone) used in the work.	99
Table 4.1j	Tricyclic 1.2.4 – Trioxanes derivatives with antimalarial activity against the drug-resistant malarial strain <i>P. falciparum</i> (W-2 clone) used in the work.	100
Table 4.1k	N-Alkyl-11-aza-9-desmethylartemisinins derivatives with antimalarial activity against the drug-resistant malarial strain <i>P. falciparum</i> (W-2 clone) used in the work.	100
Table 4.1l	3C- substituted artemisinin derivatives with antimalarial activity against the drug-resistant malarial strain <i>P. falciparum</i> (W-2 clone) used in the work.	101
Table 4.1m	Various derivatives of artemisinin and artemether with antimalarial activity against the drug-resistant malarial strain <i>P. falciparum</i> (W-2 clone) used in the work.	101
Table 4.2	Experimental and theoretical values of the 1,2,4-trioxane ring parameters in artemisinin (bond lengths in Å; bond angles and torsional angles in degrees).	102

Table 4.3	List of descriptors used in the study.	103
Table 4.4	Statistical assessment of QSAR equations with varying number of descriptors.	107
Table 4.5	Observed and predicted activity against the drug-resistant malarial strain <i>P. falciparum</i> (W-2 clone) of Training set of artemisinin derivatives.	109-110
Table 4.6	Observed and predicted activity against the drug-resistant malarial strain <i>P. falciparum</i> (W-2 clone) of Test set of artemisinin derivatives.	111
Table 4.7	Correlation matrix of the descriptors used in the QSAR model.	114
Table 4.8	Observed and predicted activity against the drug-resistant malarial strain <i>P. falciparum</i> (W-2 clone) of validation set of artemisinin derivatives.	115
Table 5.1	Artemisinin analogues with anti-malarial activities against the drug resistant malarial strain <i>P. falciparum</i> (W-2 clone) used in this work.	127-136
Table 5.2	The RMSD and docking score from the docking simulation of 10 lowest configurations of crystal structure of TG with PfATP6.	142
Table 5.3	Predicted antimalarial activities of (a) analogues using Glide score (XP) and Prime/MM-GBSA energy as a descriptor and experimental activity.	145-148
Table 5.4	Predicted antimalarial activities of artemisinin derivatives based on linear response scheme of energy parameters and experimental activities for selected analogues.	153-156
Table 5.5	Regression properties of energy parameters with experimental activities (Log RA).	156

## LIST OF FIGURES

Figure 1.1	Various stages in the life cycle of malaria parasite involving human host and the vector, mosquito.	2
Figure 1.2	Global status of resistance to chloroquine and sulfadoxine/pyrimethamine.	4
Figure 1.3	Artemisinin and its derivatives.	6
Figure 1.4	Mechanism of action of artemisinin. Formation of free radicals.	12
Figure 2.1	The structure of (a) artemisinin and (b) deoxyartemisinin with atom numbering.	23
Figure 2.2	The structure of heme.	23
Figure 2.3	The proposed mechanism of action of artemisinin.	24
Figure 2.4	The structures of three heme compounds: (a) heme-pdb, (b) heme-deoxy and (c) heme-oxy.	27
Figure 2.5	Docking configuration between artemisinin and (a) heme-pdb, (b) heme-deoxy and (c) heme-oxy structure.	42
Figure 2.6(a-d)	A linear relationship between Fe-O distance and Glide score as well as Fe-O distance and binding free energy of the (a & b) Artemisinin (QHS) derivatives and (c & d) Deoxyartemisinin (DQHS) derivatives.	49
Figure 2.7(a-d)	Models for predicting antimalarial activity ( $pIC_{50}$ ) of the (a&b)Artemisinin (QHS) analogues and (c&d) Deoxy-artemisinin analogues based on Glide score and Binding free energy ( $\Delta G_{bind}$ ) as descriptor.	55
Figure 3.1	Stereochemistry and atomic numbering scheme of artemisinin.	74
Figure 3.2	The structure of the heme compound.	74
Figure 3.3	LIE binding energies for the Training set from MD sampling.	77
Figure 3.4	LIE binding energies for the Test set from MD sampling.	83
Figure 3.5	LIE binding energies for the Training set from HMC sampling.	84
Figure 3.6	LIE binding energies for the Test set from HMC sampling.	85

Figure 4.1	Stereochemistry and atomic numbering scheme of artemisinin.	102
Figure 4.2	Relationship between predicted and experimental activities before removal of outliers.	111
Figure 4.3	Relationship between predicted and experimental activities after removal of outliers (deoxy-artemisinin derivatives)..	112
Figure 4.4	Relationship between predicted and experimental activities.	115
Figure 5.1	The 2D structure of (a) Thapsigargin (TG) and (b) Artemisinin showing the similarity between sesquiterpene moieties.	119
Figure 5.2	Overall structure of SERCA. The thapsigargin molecule (shown in blue and yellow spheres) binds to SERCA in a cavity formed by residues on transmembrane helices 3, 5, and 7.	121
Figure 5.3	Close up view of TG binding site formed by mostly hydrophobic residues on M3 and M5 helices lining the channel.	121
Figure 5.4	Alignment of PfATP6 sequence with IIWO as reference protein.	125
Figure 5.5	Schematic display of SERCA (above) and PfATP6 (below) generated using DS Visualizer for windows. Helices and sheets are represented as red cylinders and cyan arrows, respectively. The ligand TG (in brown stick) is included in the structure.	140
Figure 5.6	Ligplot of (a) PfATP6- artemisinin binding site and (b) SERCA-TG binding site.	141
Figure 5.7	Superposition of all docked configurations of TG on crystal structure ( <i>red-stick</i> ).	144
Figure 5.8	Binding mode of artemisinin and TG within the binding site of PfATP6. TG: brown stick model and Artemisinin: green stick model.	144
Figure 5.9	Models for predicting antimalarial activity (Log RA) of the artemisinin derivatives based on (a) Glide score and (b) Prime/MM-GBSA energy ( $\Delta G_{\text{bind}}$ ).	149
Figure 5.10	Models for predicting antimalarial activity (Log RA) of the artemisinin derivatives based on free energy equation.	156

## **ABSTRACT OF THE DISSERTATION**

Artemisinins are the most potent antimalarials available, rapidly killing all asexual stages of *P. falciparum*. They are sesquiterpene lactones, widely used to treat multidrug-resistant malaria. Thus it has been the objective of numerous studies to prepare better and safer anti-malarial drugs. However, the mode of action of this antimalarial is not fully understood. Artemisinins act via mechanisms that are distinct from other antimalarial classes, including those that inhibit well defined targets such as enzymes of folate biosynthesis, the DOXP reductase pathway or the cytochrome electron transport system. The peroxide within the 1,2,4-trioxane system of artemisinins is essential for antimalarial activity. Therefore, the peroxide structure becomes a focus for considerable chemical analysis aimed at trying to understand how artemisinins work. There have been two mechanism of action of artemisinin proposed. One the activated artemisinins form adducts with heme and leads to inhibition of heme polymerization. Secondly as shown recently, artemisinins, but not quinine or chloroquine, inhibit the sarco/endoplasmic reticulum  $\text{Ca}^{2+}$  ATPase (SERCA) orthologue (PfATP6) of *P. falciparum*. PfATP6 is essential for *P. falciparum* calcium homeostasis. However the mechanism of interaction as well as the binding affinity of artemisinin with heme and PfATP6 has not yet known. In this regards the molecular modeling study could be very helpful to explore the mode of interaction. Docking, binding free energy and quantitative structure activity relationship (QSAR) are computational ways to explore the binding structure, binding affinity, interaction of ligand/receptor and development of activity model. In the work, several computational approaches were used to explore the binding of artemisinin and its structural derivatives in heme and PfATP6.

An automated docking protocol has been established in this study to explore the binding interaction of artemisinin with heme. Several ab initio atomic charge schemes (HF/6-21G, HF/6-21G\* and HF/6-21G\*\*) for both artemisinin and heme structures were applied in the simulations and their effects on the docking results were investigated. Artemisinin structures taken from various optimization methods and three heme models were employed for this purpose. The docking results depended on the structures of both artemisinin and heme. Moreover, the atomic charges of heme have a significant effect on the docking configurations. A library of artemisinin analogues has been designed consisting of 144

analogues. The combined approaches of docking-molecular mechanics based on generalized Born/surface area (MM-GB/SA) solvation model showed that artemisinin and its structural derivatives approaches heme by pointing O1 and O2 at the endoperoxide linkage toward the iron center, a mechanism that is controlled by steric hindrance. A linear correlation was observed between the Fe-O distance and Glide score & binding free energy with correlation coefficient ( $R^2$ ) of 0.658 and 0.707. Quantitative structure activity relationships were developed between the antimalarial activity ( $pIC_{50}$ ) of these compounds and molecular descriptors like docking score and binding free energy. Using Glide score and binding free energy the  $R^2$  were in the range of 0.714 to 0.763 and 0.718 to 0.763 indicating that the predictive capabilities of the models were acceptable. Low level of root means square error for the majority of inhibitors which establish the docking and prime MM-GB/SA based prediction model as an efficient tool for generating more potent and specific inhibitors of heme by testing rationally designed lead compounds based on artemisinin derivatives.

The antimalarial activity of artemisinin derived drugs appears to be mediated by an interaction of the drug's endoperoxide bridge with intraparasitic heme. The binding affinity of artemisinin analogues with heme were computed using a linear interaction energy (LIE) method with a surface generalized Born (SGB) continuum solvation model. A training set of 101 artemisinin analogues with known in vitro antimalarial activity was used to build the SGB-LIE model utilizing molecular dynamics (MD) and hybrid Monte Carlo (HMC) sampling techniques. For the test set of 57 compounds the SGB-LIE model was able to predict their activity with an overall root mean square (RMS) error of 0.348 and 0.415 kcal/mol respectively with respect to experimental data. Low levels of RMS error establish the structure-based LIE method as an efficient tool for generating more potent inhibitors of heme by testing rationally designed lead compounds based on artemisinin derivatization. The developed LIE method demonstrates to be a powerful tool to estimate binding affinity of a large set of ligands within a reasonable computer time and is a promising approach in computeraided rational drug design.

A quantitative structure-activity relationship (QSAR) analysis has been performed on a data set of 194 artemisinin analogues for antimalarial activity. Several types of descriptors

including topological, spatial, thermodynamics, information content, lead likeness and E-state indices have been used to derive a quantitative relationship between antimalarial activity and structural properties. A systematic approach of zero tests, missing value test, simple correlation test, multicollinearity test and genetic algorithm method of variable selection was used to generate the model. Statistically significant model ( $r^2 = 0.845$ ,  $q^2_{cv} = 0.799$ ,  $F$ -test = 53.40) was obtained with the descriptors like molecular connectivity indexes, E-state index, length-to-breadth ratio of compounds, MLog P, HOMO, electron density, Balabans topological index and strain energy of the molecules. The robustness of the QSAR models was characterized by the values of the internal leave one out cross-validated regression coefficient ( $q^2_{cv}$ ) for the training set and determination coefficient in prediction,  $q^2_{test}$  for the test set. The value of  $q^2_{test} = 0.876$  for the test set; revealed good external predictability of the QSAR model. Also for an external data set (validation set) of 4 artemisinin analogues the QSAR model was able to predicts the antimalarial activity with very well in comparison to experimental values. The model was also tested successfully for external validation criteria. QSAR model developed in this study shall aid further design of novel potent artemisinin derivatives.

In addition to heme bio-molecular system, docking simulations and binding free energy calculation were carried out to explore the binding properties of artemisinin with PfATP6 of *Plasmodium falciparum*. The 3D structure of PfATP6 was constructed by homology modeling. A library of artemisinin analogues has been designed consisting of 154 analogues. Their molecular interactions and binding affinities with modeled PfATP6 protein have been studied using the docking, molecular mechanics based on generalized Born/surface area (MM-GBSA) solvation model and eMBrAcE. Docking and binding free energies scores show good relation with in vitro antimalarial activities. The main binding source of artemisinins to the PfATP6 is hydrophobic interaction and biologically important peroxide bonds were exposed to outside of the binding pocket. The study suggests binding of artemisinin to PfATP6 precedes activation of peroxide bond by  $Fe^{2+}$  species. Quantitative structure activity relationships were developed between the antimalarial activity (log RA) of these compounds and molecular descriptors like docking score and binding free energy. For both the cases the  $r^2$  was in the range of 0.538–0.688 indicating good data fit and  $r^2_{cv}$  was in

the range of 0.525–0.679 indicating that the predictive capabilities of the models were acceptable. In addition, a scheme similar to Linear Response was used to develop a free energy of binding (FEB) relationship based electrostatic ( $\Delta G_{ele}$ ), van der Waal ( $\Delta G_{vdw}$ ) and surface accessible surface area (SASA), which can express the activity of these artemisinin derivatives. It can be seen that  $\Delta G_{vdw}$  has most significant correlation to the activity (log RA) and electrostatic energy ( $\Delta G_{ele}$ ) has less significant correlation to the activity. It indicates that the binding of these artemisinin derivatives to PfATP6 is almost hydrophobic.  $\Delta G_{vdw}$  may be a major drive force to their binding and contribution to their activity. Low levels of root mean square error for the majority of inhibitors establish the docking, Prime/MM-GBSA and eMBrAcE based prediction model as an efficient tool for generating more potent and specific inhibitors of PfATP6 by testing rationally designed lead compounds based on aremisinin derivatization. Docking score as well as binding free energy calculated based on EmBrace and Prime MM-GB/SA show good relation with in viro antimalarial activities. The main binding source of artemisinins to the PfATP6 is hydrophobic interaction and biologically important peroxide bonds were exposed to outside of the binding pocket. This study suggests binding of artemisinin to PfATP6 precedes activation of peroxide bond by  $Fe^{+2}$  species.

# **C**HAPTER 1

## **Introduction**

---

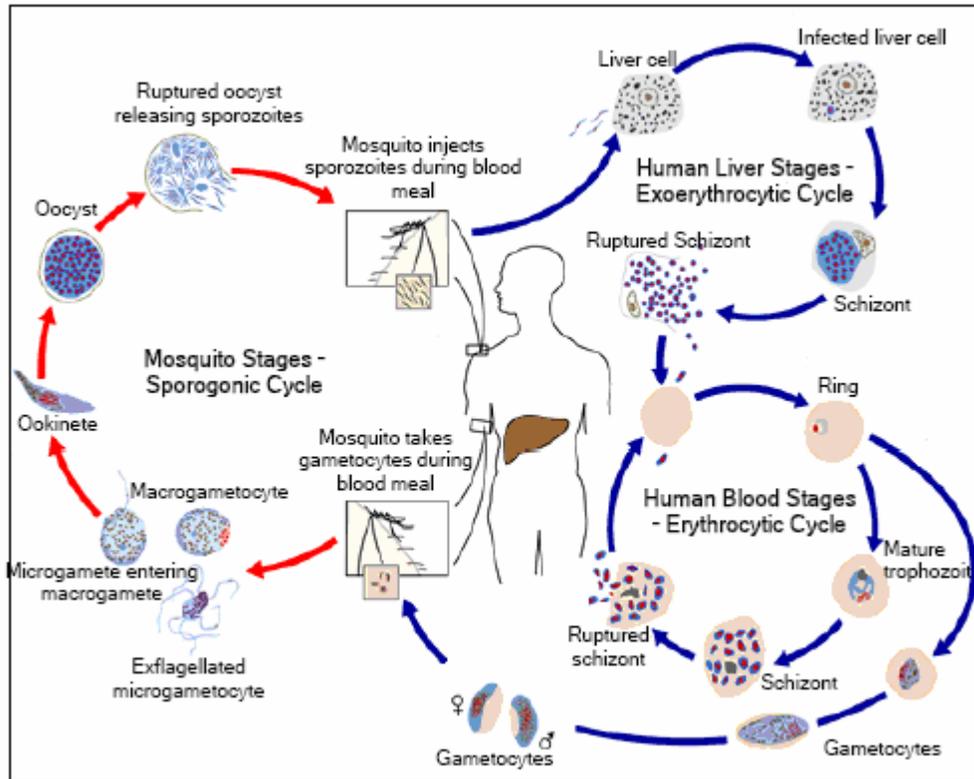
## **1.1 Epidemiology of malaria**

Malaria is a deadly disease threatening half of the global population. It is caused by protozoan parasites of the phylum Apicomplexa and the genus *Plasmodium*. There are more than 100 *Plasmodium* species that can infect mammals, birds and reptiles. Five *Plasmodium* species can infect humans: *Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium malariae*, *Plasmodium ovale* and *Plasmodium knowlesi*. *P. knowlesi*, whose natural host is macaque monkeys, was recently suggested to be the fifth species that can infect humans (White, 2008). It has been shown to be a major cause of malaria in Malaysia (Cox-Singh et al., 2008). *P. falciparum* and *P. vivax* are the most common human species, while *P. falciparum* causes most severe disease and death. It has been estimated that malaria transmission occurs in 109 countries putting 3.3 billion people at risk. In 2006 there was an estimated 247 million malaria cases, resulting in nearly 900000 deaths. This makes *P. falciparum* one of the leading causes of death worldwide, from a single infectious agent. The malaria burden is greatest in African children as 90% of the deaths occur in sub-Saharan Africa and 85% of the mortality affects children under five years. Although the number of deaths is similar to that of 2004, amelioration of the malaria situation has been achieved in seven African regions, where malaria control measures have resulted in 50% or more reduction in malaria incidence and mortality since 2000, and further 22 countries in the world have reached similar effects (WHO, 2008a-c). Morbidity and mortality are not the only consequences of malaria infection. The disease is estimated to be responsible for an average annual reduction of 1.3% in economic growth for countries with the heaviest malaria burden (Sachs & Malaney, 2002). The great variation in malaria burden between different geographical regions can be driven by several factors. Moreover, the parasite is becoming resistant to commonly used antimalarial drugs.

## **1.2 Life cycle of *P. falciparum***

For *Plasmodium falciparum*, mosquito is always the vector, and is always a female *Anopheles* mosquito. There are 380 species of *Anopheles* mosquitoes, but only 60 can transmit malaria. With the exception of *P. malariae* (which may affect the higher primates), the other three species of *Plasmodium* that infect man are obligate parasites of human beings. The life cycle of *Plasmodium* parasite is very complex and requires two different hosts, a

vertebrate host (man) and an invertebrate host, the female *Anopheles* mosquito (vector) (Figure 1.1). The different stages were named on the basis of their morphology such as merozoite, trophozoite, gametocyte found in humans and zygote, ookinete and sporozoite found in mosquitoes (Wirth, 2002).



**Figure 1.1** Various stages in the life cycle of malaria parasite involving human host and the vector, mosquito (Adapted from <http://www.cdc.gov/malaria/lifecycle>).

The cycle in man begins with the bite of a female *Anopheles* mosquito harboring sporozoites in its salivary gland during its blood meal. The sporozoites travel through the punctured skin into the blood stream. The sporozoites in the blood stream travel to the liver and invade hepatocytes within 30 minutes of being released by the mosquito. In the liver cells they reside for 9-16 days and then start multiplying asexually within the cells. Asexual reproduction (exoerythrocytic schizogony) in the liver releases thousands of merozoites, which are the first stage of the 48-hour asexual reproduction cycle in the red blood cells (erythrocytic schizogony). The blood stages constituting this cycle, studied by light and electron microscopy (Bannister et. al., 2000) are the merozoite, the ring, the trophozoite and the schizont (Figure 1.1). Schizont ruptures to release around 8-32 merozoites which are ready to

invade fresh red blood cells. The erythrocytic schizogony is the time when the human host suffers periodic cycles of clinical symptoms like fever and chills. While the merozoites continue invading fresh RBCs and continue asexual reproduction, some of them exit the asexual reproduction cycle and mature to male and female gametocytes by a process known as gametogenesis. Upon a subsequent mosquito bite, these gametocytes are taken up into its gut, where sexual reproduction takes place. In the gut the gametes undergo exflagellation and the macrogametocytes are fertilized. The resulting ookinete penetrates the wall of a cell in the midgut, where it develops into an oocyst. Sporogony within the oocyst produces many sporozoites and when the oocyst ruptures, the sporozoites migrate to the salivary gland, for injection into another host and thus the cycle continues (Sherman, 1998).

### **1.3 Treatment of malaria – Current antimalarials**

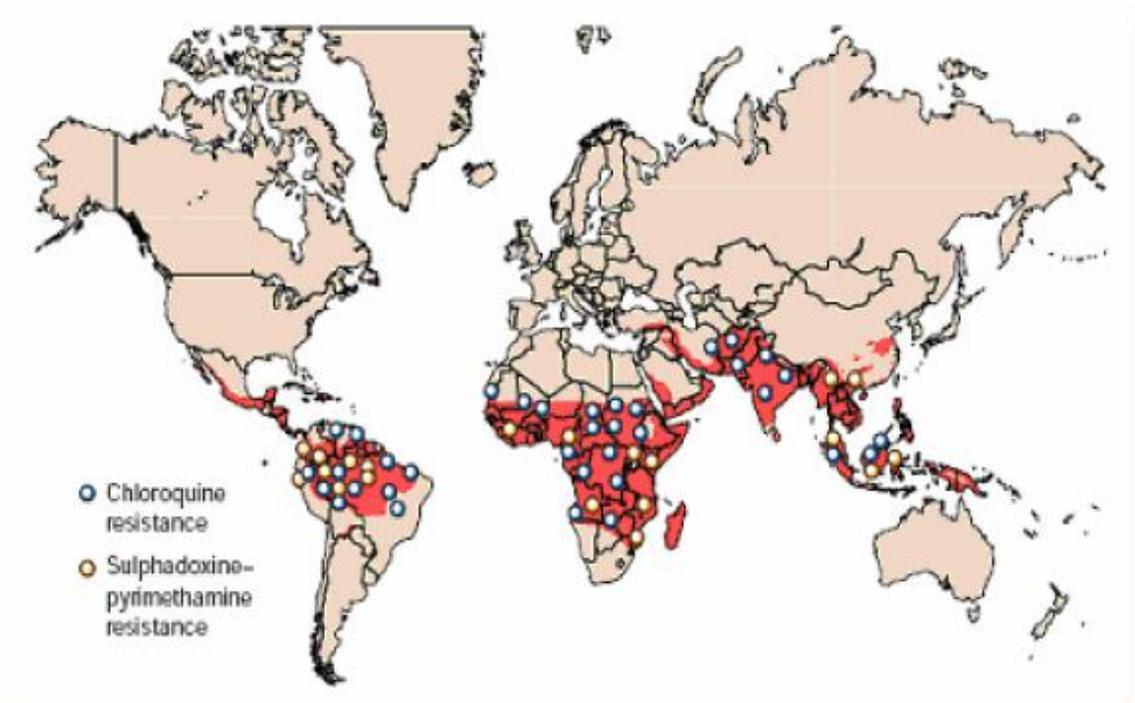
#### **1.3.1 Quinine related drugs**

Quinine is one of the four main alkaloids found in the bark of the Cinchona tree. Till date the actual mechanism of action of quinine has not been unequivocally worked out. Over many decades, different variants of quinine (most often, the structural analogues) such as chloroquine (4-amino-quinoline), mefloquine (quinoline methanol), primaquine have been used for malaria treatment (Surolia et. al., 2002). But the anti-malarial efficacy of each of these has been far from satisfactory due mainly to two major factors: (1) these drugs act on the targets whose biochemical structure/function overlaps with that of the human host (Milhous and Kyle, 1998) (2) evolution of resistant strains of the parasite within the last two decades due to indiscriminate usage of the drugs (Padmanaban and Rangarajan, 2001). Chloroquine is a lysosomotropic drug. It is a weak base, uncharged at neutral pH and gets positively charged at acidic pH. Owing to this property, chloroquine is selectively accumulated inside lysosomes. The uncharged compound rapidly diffuses through the plasma and lysosomal membranes, while once charged the compound becomes trapped inside the acidic lysosomal compartment of the parasite (Homewood et. al., 1972). This may lead to the generation of concentration gradient of several orders of magnitude. The intracellular trophozoite feeds on the hemoglobin of the red blood cell that serves as a source of amino acids. Digestion of the globin protein takes place inside the *Plasmodium* lysosome resulting in the generation of free heme (ferriprotoporphyrin IX, FP). The latter is insoluble and precipitates in the form of a black malaria pigment inside the lysosomes. Chloroquine in the

lysosome interferes with pigment formation and the FP-chloroquine complex is highly toxic to the parasite.

### 1.3.2 Drug resistance in *P. falciparum* malaria

Globally the control of malaria is deteriorating. Key factor contributing to the increasing malaria mortality and morbidity is the wide spread resistance of *P. falciparum* to the conventional antimalarial drugs such as chloroquine, sulfadoxine-pyrimethamine (SP) and mefloquine (Figure 1.2) (Ridley, 2002; Ronn, 1996; Sowunmi et. al., 1998; vvn Agtmael et. al., 1999). Multidrug-resistant *P. falciparum* malaria is prevalent in Southeast Asia and



**Figure 1.2.** Global status of resistance to chloroquine and sulfadoxine/pyrimethamine. Data are from the WHO.

South America. Now Africa, with the highest burden of malaria, is also being affected (Wensdorfer, 1994; Wesdorfer et. al., 1991). Antimalarial drug resistance is usually a result either of changes in drug accumulation or efflux (chloroquine, quinine, amodiaquine, mefloquine, halofantrine resistance) (Whit, 1998) or reduced affinity of the drug target resulting from point mutations in the respective genes encoding the target (pyrimethamine, cycloguanil, sulphonamides, atovaquone resistance) (Foote et. al., 1994; Ward et. al., 1995).

The resistance occurs when the drug concentrations are sufficient to reduce the susceptible parasite population, but inhibit less or do not inhibit multiplication of the mutants (Chawira et. al., 1987). Resistance causes drug failures when, because of reduced susceptibility, drug levels that would normally eliminate the infection can no longer do so. However, fully drug-sensitive parasites can still cause a recrudescence infection if the plasma concentrations of the drug are insufficient (White, 1999a). Increasing multidrug resistant *P. falciparum* in many parts of the world has aggravated the problem of deciding which antimalarial to use, particularly in countries where *P. falciparum* has developed resistance to chloroquine, mefloquine primaquine, antifolates such as Fansidar (Sulphadoxine-Pyrimethamine) and, to some extent, quinine which previously was effective in the treatment of severe and complicated malaria (Olliaro et. al., 1995).

Further proliferation of drug resistance is closely related to (Kondrachine, 1997):

- Massive population movements.
- Inadequate health services.
- Improper use of antimalarial drugs.
- Limited resources and operational difficulties in implementing malaria control activities.

## **1.4 Artemisinin and its derivatives as antimalarial drugs**

### **1.4.1 The plant *Artemisia annua***

The Chinese herb Qing Hao (blue green herb) has been used for two millennia in Traditional Chinese Medicine (1979; 1992; van Agtmael et. al., 1999). The earliest reference to the plant goes back to “52 prescriptions” found in the Mawangudi Tomb in an era dating back to 206 BC-AD23. The first prescription of Qing Hao for treatment of related symptoms is found in “The handbook of Prescriptions for Emergencies” by Ge Hong, who lived during AD 281- 340 (Wu and Li, 1995). Active moiety Artemisinin (qinghaosu) was isolated by Chinese scientists in 1972 from the aerial parts of *Artemisia annua* L (Journal Report; Klayman, 1985; Liu, 1979). The compound showed good *in vitro* and *in vivo* antimalarial activity. Several studies showed artemisinin to be an exceptional antimalarial agent with negligible toxicity and high efficacy against human malaria parasites, including those malaria-resistant to conventional antimalarials (Li et. al., 1994).

### 1.4.2 Chemical structure and metabolism

Artemisinin is structurally different from the previously known antimalarials. The compound is an unusually stable sesquiterpene lactone with an endoperoxide ring (empirical formula  $C_{15}H_{22}O_5$ ) (Figure 1.3). Presence of the endoperoxide moiety is the key to its antimalarial activity (Brossi et. al., 1988; Klayman, 1985; Lee et. al., 1990; Luo et. al., 1987).

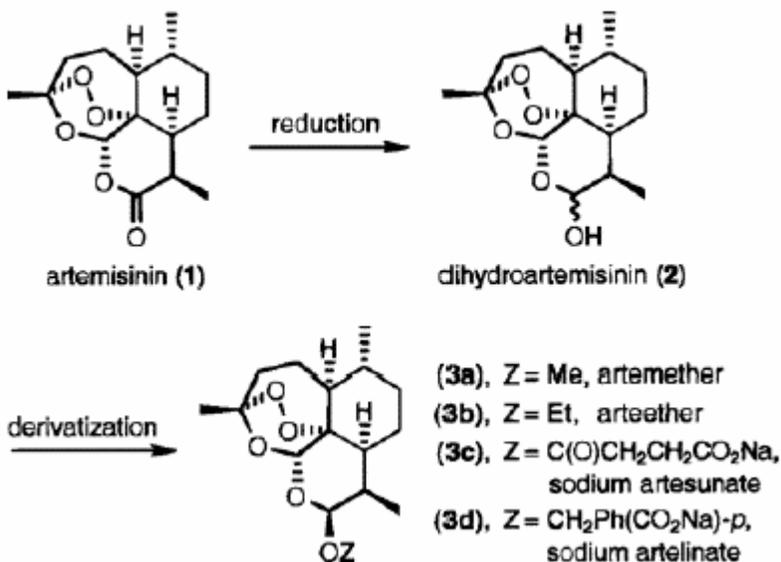


Figure 1.3. Artemisinin and its derivatives.

The white needle crystals of artemisinin are hardly soluble in water or oil therefore formulations other than oral and rectal are not in clinical use. However, since the peroxide bridge is stable under certain chemical reactions, several more soluble artemisinin derivatives, arteether, artemether, sodium artesunate, sodium artelinate and dihydroartemisinin (DHQ) have been synthesized for the treatment of malaria (Figure 1.3). DHQ is the first metabolite of artemether, arteether and artesunate. DHQ is the most effective compound of this class (Janse et. al., 1994). Artesunate can be regarded as a pro-drug of DHQ.

### 1.4.3 Antimalarial activity

Now artemisinin and its derivatives have been recognized as a new generation of powerful antimalarial drug for combating the most popular infectious disease malaria worldwide. Artemisinin and its derivatives induce more rapid reduction of parasitemia (van Agtmael et. al., 1999), decreasing the number of parasites faster than any other known drug.

As a consequence they are of special interest for severe malaria (Hien and White, 1993). The fast decline in the number of parasites is also beneficial in combination therapies. Most of the antimalarials work at late trophozoite and schizont stages of the malaria parasite but artemisinin also act at early trophozoites and ring stages (van Agtmael et. al., 1999). Artemisinins do not affect liver-stage parasites or stages within the mosquito (Price et. al., 1996). Artemether and artesunate were approved by the Chinese authority and collected in the “Essential Medicine List” by WHO. These derivatives have been successfully applied to remedy several million malaria-suffering patients since their advent. Meanwhile many research papers have been published to record rapid progress of artemisinin research from different disciplines of botany, chemistry, pharmacology, and clinical medicine etc. Qinghao has been used as a traditional medicine for at least 2000 years in China (Wallaart, 2000). Since then a series of Chinese medicine books including the most famous book “Compendium of Medical Herbs” (Bencao Gangmu) by Li Shizhen in 1596, have described the application of qinghao for fever remedy. In the phytotaxonomy qinghao is *Artemisia annua* L. Composites, so Qinghaosu is also named as artemisinin or seldom as arteannuin.

#### **1.4.4 Artemisinin derivatives**

From indigenous *Artemisia annua* L., continuous phytochemical studies by Chinese researchers in the early 1980s led to the excavation of another 10 sesquiterpenes including deoxy-artemisinin (Tu et. al., 1981), artemisinin D (Tu et. al., 1981), artemisinin F (Zhu et. al., 1984), artemisinin E (Wu et. al., 1984), artemisinin A (Tu et. al., 1981), epoxyarteannuic acid (Wu et. al., 1984), artemisinic acid (Deng et. al., 1981; Tu et. al., 1981), artemisinic acid methyl ester (Zhu, 1982) artemisinol (Zhu, 1982) and arteannuin B (Tu et. al., 1981). From biogenetic viewpoint, artemisinic acid or its 11, 13- dihydroanalogue, dihydro-artemisinic acid which was isolated later from *A. annua* is late precursors in the biogenesis of qinghaosu. By the year 1991, 16 closely related sesquiterpenes had been isolated from aerial part of *A. annua* and briefly summarized by Zaman and Sharma (1991). A bisnor-sesquiterpene, norannuic acid was reported in 1993 (Misra et. al., 1993) and three more sesquiterpenes were isolated and reported by Misra et al. (1994). Sy et. al., (1998) isolated seven new sesquiterpenes in 1998. Two amorphane sesquiterpenes, deoxyarteannuin B and dihydrodeoxyarteannuin B were introduced to sesquiterpene family isolated from aerial pars of *A. annua* in 2001.

Recently, the first phytochemical investigation of natural products from the seeds on *A. annua* led to discovery of fourteen new sesquiterpenes (Sy et. al., 2001). In addition, two sesquiterpene plant hormones, abscisic acid and its methyl ester were found in an Indian growing *A. annua* (Tewari et. al., 2003). Apart from sesquiterpenes from *A. annua*, essential oils are another active research interest as it could be potentially used in perfumery, cosmetics and aromatherapy. Depending on its geographical origin, the oil yield in *A. annua* ranges between 0.02-0.49% on fresh weight basis and 0.04-1.9% on dry weight basis (Bagchi et. al., 2003; Bhakuni et. al., 2001; Jain et. al., 2002; Rasooli et. al., 2003; Liu et. al., 1981). Other chemical compounds in *A. annua* includes carbohydrates, traces of glycosides, resins etc.

#### **1.4.5 Pharmacology**

Antimalarial drugs derived from natural *Artemisia annua* L. have many advantages: quick reduction of fevers, fast clearing parasites in blood (90% of malaria patients recovered within 48 hrs) and no significant side effects. Experimental and clinical studies reveal that artemisinin, Artemether and artesunate are not only the potent antimalarial drugs but also the useful agents for other disease, especially as antiparasitic agent. In 1970s, Artemether and artesunate were confirmed to be more active than artemisinin in animal models (Le et. al., 1980; Le et. al., 1982; Wu et. al., 1995). Some components of *A. annua* Such as qinghaosu, artemisinin B, artemisinic acid, artemisitene, flavonoids and other terpenoids, showed antitumor activities at varying concentrations against L-1210, P-388, A-549, HT-29, MCF-7 and KB *in vitro* (Zheng et. al., 1994; Jung, 1990; Jung, 1997). It was found that dihydroartemisinin can selectively kill cancer cells in presence of holotransferrin, which can increase intracellular iron concentrations, and normal breast cells (HTB 125) and lymphocytes had non-significant changes. It seems the mechanisms of anticancer action and of antimalarial activity are similar (Lai et. al., 1995; Moor, 1995; Singh, 2001). As a response to increasing levels of antimalarial resistance, WHO recommends that all countries experiencing resistance to conventional mono therapies should use combination therapies preferably those containing artemisinin derivatives (ACTs- Artemisinin based Combination Therapies) for *falciparum* malaria. WHO currently recommends the following therapeutic options:

- artemether
- artesunate + maodiaquine

- artesunate + sulphadoxine/ pyremethamine
- artesunate + mefloquine
- amodiquine + sulphadoxine-pyrimethamine

Malaria is a highly treatable disease, and very effective treatments available in the form of Artemisinin based Combination Therapies (ACTs). WHO call on all “Roll Back Malaria (RBM) “ partners to unite in a global coalition to enable countries to accelerate access to ACTs and make these lifesaving medicines affordable to the people in need (WHO, 1998).

#### **1.4.6 Artemisinin drugs in the treatment of *Plasmodium falciparum* malaria in India**

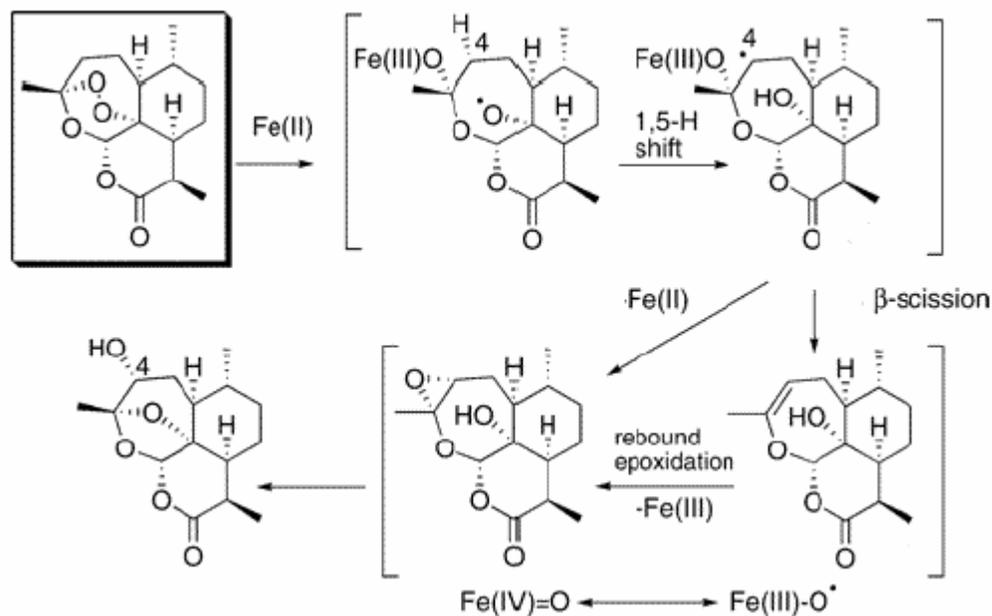
The National Anti Malaria Drug Policy envisages treatment of uncomplicated *Plasmodium falciparum* (suspected and laboratory diagnosed) with chloroquine and primaquine and in five provinces by the co-administration of artesunate and sulphadoxine pyrimethamine (<http://namp.govt.in>). In this connection, the article on artemisinin-based combination therapy (ACT) by Sushil Kumar and Srivastava is timely (Kumar et. al., 2005). In 2003, the National Vector Borne Disease Control Programme, NVBDCP in short (formerly National Anti Malaria Programme), reported 1.87 million cases of malaria (including 0.86 million *P. falciparum* cases) and 1006 deaths. In 2004, the largest number of malaria cases was reported from Orissa, followed by Gujarat, Chhattisgarh, West Bengal, Jharkhand, Karnataka, Uttar Pradesh and Rajasthan (<http://namp.govt.in>). The epidemic has been occurring with increasing frequency, killing and demoralizing the affected population; and pushing people below the poverty line. For example, in India, 90% *P. falciparum* cases occur in states below the poverty line (Sharma, 2003). Such epidemics prevent national development in all walks of life and retard the gross domestic product (GDP) (Sachs et. al., 2002). Recent examples of malaria epidemics reported by the NVBDCP include *inter alia* Rajasthan, Haryana (1976), Gujarat, Goa, West Bengal (1997), Goa, Maharashtra (1998), Andhra Pradesh, Assam, Bihar, West Bengal (1999), Uttar Pradesh, Madhya Pradesh, Karnataka (2000), and Rajasthan (2003) (<http://w3.whosea.org/EN/Section10/Section21/Section1987.htm>). Obviously, the disease burden in the country is hugely under-estimated. This is also reflected by the fact that India’s chloroquine consumption in 1976 was 61 metric tons (mt) to treat 6.45 million cases; and in 2005 cases have reduced by 70% but antimalarial usage has increased ten fold (Department of

commerce, Government of India, New Delhi, 2004). Based on environment determinants, the World Health Organization (WHO) estimates 100 million cases in the South East Asia Region (SEARO) (WHO, 2004), 70% of these contributed by India (WHO, 2004). *P. falciparum*, the killer parasite accounts for 45–50% malaria cases. Control of *P. falciparum* is important but difficult, as was evident from the failure of the Swedish International Development Agency (SIDA) supported Indian *P. falciparum* Containment Programme (1977–88) (Sharma, 1998). *P. falciparum* parasite is present all over the country, but its distribution is highly uneven. It is the major cause of infection in the Northeast, Orissa, tribal settlements across the country and forests. In the plains of India, *Plasmodium vivax* peak is followed by *P. falciparum* and in all other endemic areas *P. falciparum* predominates. *P. falciparum* abounds in communities lacking awareness, resources and suffering from endemic poverty (Sharma, 2003; Sharma, 1996; Sharma, 1999; Anon, 1987). With the national antimalaria drug policy of sequential monotherapies and serious compromises in vector control, drug resistance against chloroquine and sulphadoxine pyrimethamine (SP) is on the rise, and more areas are coming under multidrug-resistant malaria (Anon, 1987). Drug pressure is selecting for mutations, for example, Pfcrt K76T mutation, an important determinant of chloroquine resistance is present in 95% of the isolates studied (Vathsala et. al., 2004). A direct consequence of drug resistance is the rise in malaria morbidity and mortality, and steep rise in treatment cost by a factor of 40–50 compared to chloroquine (Gupta et. al., 2005).

ACT is a scientific approach to tackle this problem. It is a combination of artemisinin derivative drug with one or more long acting antimalarial drug having different modes of action and different drug targets (Lee, 2002; White, 1999). Artemisinin drugs have a short half-life of 1–4 h or so, but because of their strong anti-plasmodial activity, they reduce biomass of the existing parasites by about 10–4 or say by 95% at each dosage of administration, and also kill the sexual stages of the malarial parasite. Residual parasites, if any, and the recrudescences are eliminated by the long acting antimalarial and the host immunity (International Artemisinin group, 2004; Davis et. al., 2005).

## 1.5 Mechanism of action

Although artemisinin has been on the market for more than 30 years little is known to date about its biological targets (Wu, 2002). One of the main reasons is that artemisinin does not exert its lethal effect through the whole intact molecule, but rather, through some transient species generated after cleavage of the peroxy bond (Wu, 2002). Artemisinin is hydrophobic and passes biological membranes easily (Augustijns et. al., 1996). *In vitro* studies have suggested an uptake of artemisinin by both healthy and malaria infected red blood cells (Asawamahasakda et. al., 1994). During the blood stage phase of the parasite, more than 70% of the hemeoglobin within the infected erythrocyte is digested (Francis et. al., 1997). Heme is released which is toxic for the parasite and neutralized by polymerization into hemozoin or “malaria pigment” in the form of a crystalline, insoluble, black-brown pigment. The heme polymerization pathway is specific to the malarial parasite and offers a potential biochemical target for the design of antimalarials. Heme or iron (II) salts triggers reductive cleavage of the peroxide bond in artemisinin to form oxygen centred radicals. Oxy radicals then form carbon centred radicals (Kamchonwongpaisan et. al., 1996) (Figure 1.4). These radicals cause oxidative stress and damage to the parasite’s membrane systems like mitochondria, rough endoplasmic reticulum and plasma membranes (Asawamahasakda et. al., 1994; Cumming et. al., 1997; Maeno et. al., 1993). Recent studies have shown that artemisinin taken up by the malarial parasite growing *in vitro* was selectively concentrated in the parasite food vacuole and was associated with hemozoin (Hong et. al., 1994). Artemisinin also interacts with heme, forming covalent adducts (Hong et. al., 1994; Meshnick et. al., 1991). However, it has also been reported that the artemisinin heme complex does not possess any antimalarial activity (Meshnick et. al., 1991). Further studies related to structural and mechanistic aspects of the interaction of artemisinin with heme may yield important information for the design of better antimalarials.



**Figure 1. 4.** Mechanism of action of artemisinin. Formation of free radicals.

### 1.5.1 PfATP6, a target for artemisinin

More recently, an alternative mechanism of action for artemisinins based on inhibition of the malarial parasite's calcium ATPase (sarco endoplasmic reticulum calcium ATPase, SERCA) has been suggested. Only one *P. falciparum* SERCA orthologue, PfATP6, has been identified (Kimura et. al., 1993; Varadi et. al., 2003). It was demonstrated that the SERCA inhibitor thapsigargin could induce calcium release into the cytosol from intracellular stores, probably endoplasmic reticulum (ER), by inhibition of the PfATP6, suggesting that PfATP6 is essential for *P. falciparum* calcium homeostasis (Varadi et. al., 2003) and that PfATP6 is functionally related with higher mammal homologues. The SERCA inhibitor thapsigargin is a sesquiterpene lactone, as are artemisinin. From these structural similarities the hypothesis emerged that ART act by inhibiting PfATP6. This was supported by the demonstration that artemisinin specifically inhibited PfATP6 expressed in *Xenopus laevis*, as thapsigargin. The two drugs showed an antagonistic interaction in *P. falciparum* cultures and similar localization in the parasite. Hence PfATP6 was suggested to be a target of ART (Eckstein-Ludwig et. al., 2003). However the mechanism of interaction as well as the binding affinity of

artemisinin with PfATP6 has not yet known. In this regards the molecular modeling study could be very helpful to explore the mode of interaction.

### **1.6 Computational studies and its utility in drug design**

Using computational methods to solve chemical and biological problems has been one of important branches in current scientific research and has been approaching into more and more areas with the rapid development on computer ability and speed. The supercomputers as well as personal computers become more and more powerful. The new computer technique enables many computational researches, which were thought formidable before as they requires too many computer resources. Applications of computational methods in bio-interested problems produce a new way to explore the bioactivities of chemicals and bio-macro molecules. These insights on these biological systems, protein, nuclear acid and others, obtained from computational work help us a lot to understand these activities and to solve problems in biological area.

While they are carried out to explore some fundamental problems in the chemical and biological systems, computational researches also are applied to help on drug development. They can be used in two aspects to help drug development: One is that by exploring the biologic system, the computational research provides a lot of insights on enzyme functions and interaction of enzymes with active agents. This knowledge give research a unique chance to learn how these enzymes function in biologic condition and how an agent affect an enzyme and finally helps them to design a suitable agent to affect the function of an enzyme. Other way is to directly use computational methods to select and screen out the lead drug candidates. This will decrease the experimental work required to synthesize and test large number of possible candidates.

With the help of variety computational methods, traditional drug development has benefited a lot from the computational research. It is believed that expense and the development period for a new drug has been decreased since computational rational drug design was used in traditional drug development as computational methods can dramatically decrease the number of candidates of a drug which need to be synthesized and tested.

Computer-aided drug design has been integrated into major drug design groups (labs). It is one of important applications of computational researches. Although it has made a lot of progress in academic research and industrious drug developments, generally speaking, the computer-aided drug design is still in its early stage. There are lots of fundamental problems need to be solved before the method can produce thoroughly and accurately enough results.

### **1.6.1 Structure-based virtual screening methods for computer-aided drug discovery**

One of the major challenges in drug discovery is to identify novel compounds with biological activity. Computer-aided drug discovery technology has become an essential and powerful platform for the discovery of new lead compounds, as an alternative from, and complement to experimental approaches. As the number of high resolution structures of potential therapeutic targets and small molecules has grown, the significance of *in silico* experimental approaches has become increasingly important as demonstrated in recent studies by making use of public data (Cherkasov et. al., 2006; Cleves and Jain, 2006; Yoon et. al., 2005a; b).

Virtual high throughput screening (Klebe, 2006; Oprea and Matter, 2004), which is a method to rapidly identify biologically active compounds *in silico*, can be roughly divided into two categories; ligand centric and receptor centric. Ligand centric methods essentially focus on the comparative analysis of the structural shapes and chemical complementarities between compounds and known ligands. A knowledge of the experimentally selected active compounds is a prerequisite when using this approach (Stahura and Bajorath, 2004). Receptor centric methods predict the interaction of given compounds with a target receptor, and hence they do not require experimental data about the structure of the ligand. Molecular docking is one of the key methodologies for receptor centric virtual screening. It is a technique for predicting the best binding mode for a given compound that fits into a target receptor, and evaluating its binding affinity. The docking approach has become a primary technique used in many drug discovery programs (Kitchen et. al., 2004; Sousa et. al., 2006).

The docking process involves a conformational search for a compound which complements a target binding site, with the aim of identifying the best matching binding pose.

A common computational strategy is to use a suitable scoring function to theoretically evaluate the binding affinities of thousands of molecules in a compound library for a target protein. An accurate rank ordered prediction of the compound binding affinities using the scoring function is an invaluable step. Most of scoring functions used in docking programs are designed to predict binding affinity by evaluating the interaction between a compound and a receptor. However, it should be noted that ligand receptor recognition process is determined not only by enthalpic effects but also by entropic effects. Moreover, the scoring functions have a simplified form for the energy function to facilitate high throughput evaluation of a large number of compounds in a single docking run. These functions may be problematic when used with contemporary docking programs, and can result in a decrease of virtual screening accuracy. To overcome this problem, more precise but time consuming computational methodologies are necessary.

There have been a number of reports evaluating the efficiency of various virtual screening approaches, including the evaluation of docking programs (Warren et. al., 2006), machine learning methods for ligand based descriptors (Chen et. al., 2007) and comparison of shape matching with docking (Hawkins et. al., 2007).

Docking, modeling, molecular simulation, QSAR, virtual screening, free energy calculations and data mining etc. methods have been used directly in rational drug discovery projects to speed development and help to find good agents. These methods produce a lot of information in variety of drug related researches. They benefit basic scientific activities as well as industrious efforts. But most of these computational tools have their own limitations and they need further development on some basic, methodological, and application problems. A lot of applications have demonstrated that if a proper tool and suitable approach are chosen on a specific research, good results can be produced to solve targeted problems.

### **1.6.2 Overview of the works**

In the work, we tried to use computational methods to explore the binding structures, binding affinity and inhibition mechanism of active ligands in their corresponding receptors. Two bio-systems were used in the work: heme polymerization and PfATP6. By studying

these two systems, we want to produce more information for researcher to understand their biological function affected by inhibitors, how a ligand affect its receptor and what type of ligand will better inhibit the biosystems. Also we tried to develop approaches to calculate the activity of a set of ligands by ways of free energy of binding (FEB) and quantitative structure-activity relationship (QSAR). These ways should be convenient approaches, which can be used to normal set of compounds to benefit ligand activity evaluations in a rational drug design.

In most QSAR approaches, such as CoMFA and CoMSIA, the QSAR models are built on the variety of calculated properties of ligands alone. However, despite statistically excellent and offer good predictive performance, CoMFA and CoMSIA are inherently limited to the need to align with the database molecules correctly within 3D space. The determination of the ‘active’ conformation that each compound will retain is a critical issue due to unavailability of X-ray structure. We should have some knowledge or hypothesis regarding active conformations of the molecules under study as a prerequisite for structural alignment. Nevertheless, especially for structurally diverse molecules, unambiguous 3D alignment to initiate the CoMFA process is still a difficult task. We, as well as other researchers, were motivated to explore possible alternatives that would use alignment free descriptors derived from 2D or 3D molecular topology and thus alleviate frequent ambiguity of structural alignment typical of 3D QSAR methods. They are liganded-based computer-aided drug design approach. On the other side, in molecular modeling and docking, etc. ligands are modeled into a given active site. The shape of the active site of a receptor is the key element used to design a new ligand. These methods focus on the steric and energetic fitting of ligand into a corresponding active site. They are type of receptor-based ligand design approach. In the work, we have applied both structure based and ligand based approaches in building QSAR models.

The binding structure of a ligand in its receptor is important to understand the interaction between ligand and receptor. The binding structure of a ligand in its receptor is also a basis for many other studies, such as binding affinity calculation, new ligand improvement to fit a binding site better, and MD simulation to explore the ligand effect on

receptor dynamic properties, etc. Docking is one of the ways to explore binding structures of a ligand in a receptor. According to the way it treats a ligand and a receptor during a docking simulation, docking methods can be grouped into three categories: rigid dock, flexible dock and flexible receptor dock. A rigid dock treats both ligand and receptor rigidly during docking simulation. It is first generation of docking simulation. It searches best fit for a given conformation of a molecule. It is important to correctly predict a conformation of molecule in order to get a satisfactory result in this type of docking. The method requires the least computer time compared to other two methods. A flexible dock treats a ligand flexible to allow the molecule conformation change, but keep receptor fixed. This method is most used currently in research. Several new dock methods have been developed to treat receptor partially flexible while treat ligand flexible, in which some side chains of a receptor around active site are allowed to rotate. The method can sample more possible binding structures than other two methods. As the methods have just been developed for short time and limited application are available up to the research started to evaluate their advantages over others. So, in this work, we used flexible docking simulation (Glide, Schrodinger) to predict binding structures of ligands in their receptors.

Two bio-molecular systems, heme and PfATP6, were studied in the work. In humans, malaria parasites digest more than 70% of the hemoglobin within the infected red blood cell, giving globin and heme as the products. The globin is hydrolyzed to give amino acids, which are used in protein synthesis by the parasite. The toxic heme is mostly detoxified by a specific mechanism of heme polymerization into hemozoin. The heme polymerization is a target for antimalarials and many inhibitors have been developed, such as chloroquine that inhibits this process. The study by Peters et. al. (1986) revealed that artemisinin also inhibits heme polymerization. The chloroquine-resistant strain of *Plasmodium berghei* that lacks hemozoin, possibly because heme polymerization does not occur, is also resistant to artemisinin (Peters et. al., 1986). This supports the view that inhibition of heme polymerization is the mode of action of artemisinin. It is very possible that artemisinin interacts with heme and hence inhibits the polymerization process.

In the work, we studied artemisinin binding in heme: the binding structures and modes as well as the activity of different heme structure with artemisinin interaction. To learn the bound structures of artemisinin in heme will help to understand the interaction of artemisinin with heme and enable other studies. In the work, the bound structures of artemisinin in heme were explored using flexible docking. The results were used to study their interaction and binding affinity in heme. The activity model produced a way to evaluate the activity of interested inhibitors. Also the QSAR of a set of inhibitors was built based on the obtained binding structures using structure based approaches. Free energy of binding will be used to study a set of artemisinin analogues in heme. These computations will provide structural and energetic information of these artemisinin and also can be used to predict their binding affinity with heme.

PfATP6 an orthologue protein of SERCA is also proposed to be an important target of artemisinin. The SERCA inhibitor thapsigargin is a sesquiterpene lactone, as are ART. From these structural similarities the hypothesis emerged that ART act by inhibiting PfATP6. This was supported by the demonstration that artemisinin specifically inhibited PfATP6 expressed in *Xenopus laevis*, as thapsigargin. The two drugs showed an antagonistic interaction in *P. falciparum* cultures and similar localization in the parasite. Hence PfATP6 was suggested to be a target of ART (Eckstein-Ludwig et. al., 2003). PfATP6 functionally very important for the survival of *P. falciparum* as it regulates calcium homeostasis (Varotti et. al., 2003). (Varotti et. al., 2003). Calcium has been shown to regulate several processes in apicomplexan parasites including host cell invasion and motility (Nagamune et. al., 2008). In *Plasmodium* it has been suggested that similar mechanisms may be involved in host cell invasion (Billker et. al., 2004, Green et. al., 2008). Calcium may also be important for *Plasmodium* gametocyte differentiation (Billker et. al., 2004) and for synchronization of the parasite life cycle in response to the host melatonin production (Garcia et. al., 2008). Here in this work also attempt has been taken to study the mechanism and mode of interaction of artemisinin analogues in the binding site of PfATP6. Further prediction model of antimalarial activity has been developed based on structure based approach.

By the work, we want to explore the two bio-molecular systems to learn the binding structures and modes of variety of ligands in the proteins to elucidate the inhibition of the inhibitors to enzymes by studying the interaction between ligand and receptor. We hope the knowledge of these two systems will help to understand other similar biologic systems. The information from this work will provide helps to other related research, especially those works that try to develop new agents to cure malaria. Meanwhile, we try to develop an approach to calculate and evaluate the binding affinity of ligand in its receptor that is used to predict the activity of interested ligands in a reasonable computer requirement for a normal set of molecules.

Each of the four pieces of work has distinct characteristics as well as they are related to one another. To clearly and coherently demonstrate the goal, results, and conclusion of each piece of work, we arrange each work in each chapter in a publishing format. The format will benefit reader to clearly and well understand the idea development, conclusion coherence, and whole significance, as each one will be a consistent full manuscript for background to conclusion in publication stage. A final summary will link the four parts together and give a general conclusion of the whole work.

## CHAPTER 2

**Computational and molecular modeling evaluation of the antimalarial activity of artemisinin analogues: Molecular Docking and rescoring using prime/MM-GBSA approach**

---

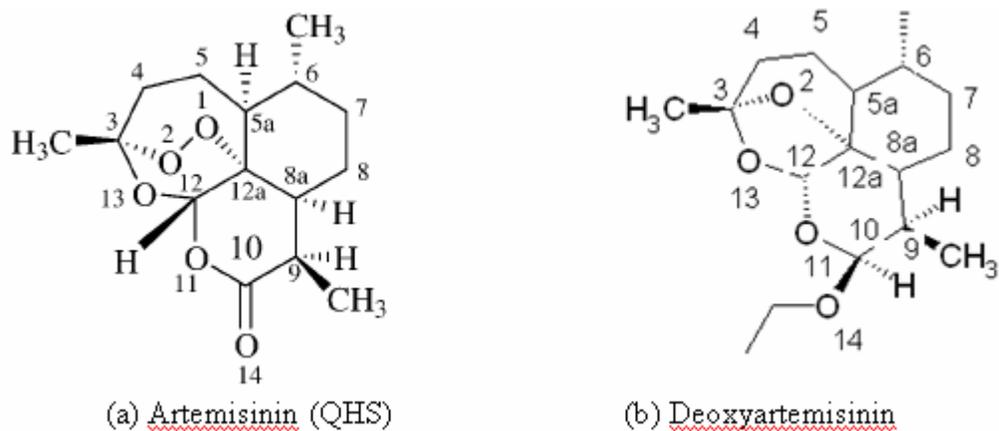
## Abstract

Artemisinin a class of sesquiterpene endoperoxide, have been the objective of numerous studies to prepare better and safer anti-malarial drugs. We report here an automated molecular docking of artemisinin to heme. The effects of atomic charges, and ligand and heme structures on the docking results were investigated. Several charge schemes for both artemisinin and heme, artemisinin structures taken from various optimization methods and X-ray data, and three heme models, were employed for this purpose. The docking results also depended on the structures of both artemisinin and heme. Moreover, the atomic charges of heme have a significant effect on the docking configurations. The combined approaches of docking-molecular mechanics based on generalized Born/surface area (MM-GB/SA) solvation model showed that artemisinin and its structural derivatives approaches heme by pointing O1 and O2 at the endoperoxide linkage toward the iron center, a mechanism that is controlled by steric hindrance. A library of artemisinin analogues has been designed consisting of 144 analogues. A linear correlation was observed between the Fe-O distance and Glide score & binding free energy with correlation coefficient ( $R^2$ ) of 0.658 and 0.707. Quantitative structure activity relationships were developed between the antimalarial activity ( $pIC_{50}$ ) of these compounds and molecular descriptors like docking score and binding free energy. Using Glide score and binding free energy the  $R^2$  were in the range of 0.714 to 0.763 and 0.718 to 0.763 indicating that the predictive capabilities of the models were acceptable. Low level of root means square error for the majority of inhibitors which establish the docking and prime MMGBSA based prediction model as an efficient tool for generating more potent and specific inhibitors of heme by testing rationally designed lead compounds based on artemisinin derivatives. The effects of atomic charges and ligand and heme structures on the docking results were investigated. Several charge schemes for both artemisinin and heme, artemisinin structures taken from various optimization methods and three heme models were employed for this purpose. The docking results also depended on the structures of both artemisinin and heme. Moreover, the atomic charges of heme have a significant effect on the docking configurations.

## 2.1 Introduction

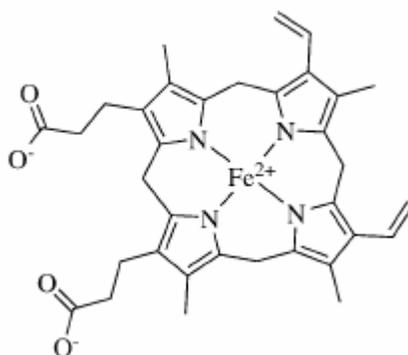
Malaria is one of the most common diseases in tropical countries. It is one of the most widespread and prevalent endemic diseases; it threatens approximately 40 percent of the world's population in more than 90 countries. Over 300 million new malaria infections and millions of deaths due to malaria occur world wide each year. This tremendous prevalence might be partly because of the resistance of malaria parasites to most antimalarial agents, e.g. chloroquine, quinine, and mefloquine (Moore et. al., 1961; Mockenhaupt 1995). Because a vaccine for malaria is not available, it is essential to find new antimalarial drugs and understand their antimalarial mechanism for treating patients. Artemisinin (qinghaosu) (Figure 2.1), a sesquiterpene endoperoxide isolated from *Artemisia annua* is however, a remarkable life saving antimalarial compound, effective against drug-resistant *Plasmodium falciparum* and cerebral malaria (Cumming et. al., 1997; Haynes and Vonwiller, 1997; Jung, 1994; Klayman, 1985). Artemisinin and its derivatives induce more rapid reduction of parasitemia, decreasing the number of parasites faster than any other known drugs. As a consequence they are of special interest for severe malaria. The first decline in the number of parasites is also beneficial for combination therapies. Artemisinin has a unique structure (Figure 2.1a) bearing a stable endoperoxide lactone (1,2,4-trioxane) totally different from previous antimalarials in its structure. Its unusual structure might be indicative of a different mode of action from those of the other antimalarial drugs and hence, the high potency against the resistant strains. This has led to tremendous interest in the mechanism of action (Cumming et. al., 1997), chemistry (Haynes and Vonwiller, 1997) and drug development (Jung, 1994) of this novel class of antimalarials. Although the mechanism of its antimalarial activity is still in doubt, there is general agreement on the significance of the endoperoxide group of artemisinin to the antimalarial activity (Klayman, 1985; Rafiee et. al., 2005). This is evident from the inactivity of the deoxyartemisinin (Fig. 2.1b) compound that lacks the endoperoxide moiety (China Cooperative Research Group 1982). In addition, in-vitro experiments revealed that iron is required for artemisinin to have antimalarial activity (Meshnick et. al., 1991; Meshnick et. al., 1993; Posner et. al., 1994).

The high selectivity in the killing of parasites by artemisinin may be due to its interaction (Meshnick et. al., 1991) with heme which accumulates in high quantities in



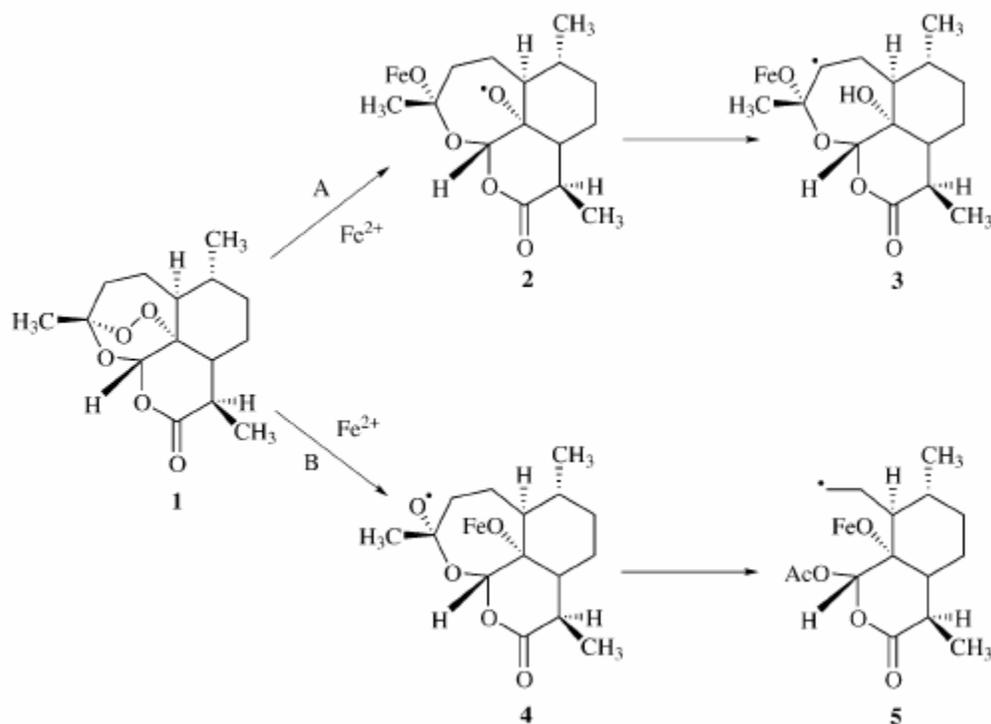
**Figure 2.1** The structure of (a) artemisinin and (b) deoxyartemisinin with atom numbering.

parasitised red blood cells as a by-product of hemoglobin lysis by the malarial parasite (Goldberg et. al., 1990). The globin is hydrolyzed to give amino acids, which are used in protein synthesis by the parasite. The toxic heme (Fig. 2.2) is mostly detoxified by a specific mechanism of heme polymerization into non-toxic and insoluble polymer called as hemozoin which accumulates as a crystalline pallet in the cytosol of the erythrocytes (Goldberg et. al., 1990). The heme polymerization is a target for some antimalarials, such as chloroquine, that inhibit this process (Slater, 1993). A recent study reported that artemisinin also inhibits heme polymerization (Pandey, 1999). The chloroquine-resistant strain of *Plasmodium berghei* that lacks hemozoin, possibly because heme polymerization does not occur is also resistant to artemisinin (Peters et. al., 1986). This supports the view that inhibition of heme polymerization is the mode of action of artemisinin. It is very possible that artemisinin interacts with heme and hence inhibits the polymerization process.



**Figure 2.2** The structure of heme.

From studies with model systems, Jefford and co-workers (1996) suggested that 1,2,4-trioxanes structurally related to artemisinin form a complex with  $\text{Fe}^{\text{II}}$  of heme and generate oxyl radicals, whereas Posner and co-workers (1996) proposed that Fe-catalyzed decomposition of artemisinin leads to reactive carbon centered free radicals, high valent iron-oxo species, and electrophiles. The chemical behavior of artemisinin in the presence of heme and non-heme iron(II) and iron(III) has been studied and artemisinin decomposition products of such reactions have been identified by Haynes and Vonwiller (1996). Two reaction schemes have been proposed that heme iron attacks the endoperoxide linkage of artemisinin either at the O1 (Jefford et. al., 1996) or O2 position (Posner et. al., 1996) (Fig. 2.3). In pathway A, heme iron attacks the compound at the O2 position and produces a free radical at the O1 position. Later it rearranges to form the C4 free radical. In pathway B, heme iron attacks the compound at the O1 position and produces a free radical at the O2 position. After that the C3–C4 bond is cleaved to give a carbon radical at C4. It has been suggested that the C4 free radical in both pathways is an important substance in antimalarial activity (Posner et. al., 1994).



**Figure 2.3** The proposed mechanism of action of artemisinin.

The mechanism of action of any drug is very important in drug development. Generally, the drug compound binds with a specific target, a receptor, to mediate its effects. Therefore, suitable drug–receptor interactions are required for high activity. Understanding the nature of these interactions is very significant and theoretical calculations, in particular the molecular docking method, seem to be a proper tool for gaining such understanding. The docking results obtained will give information on how the chemical structure of the drug should be modified to achieve suitable interactions. Hence, this could bring about the development of new and more effective drugs. However, it is quite important to have an accurate model for the heme–artemisinin complex, because this knowledge can be used to design better and more potent antimalarial drugs.

Prompted by the clinical successes of the artemisinin, significant efforts have been focused on identifying new analogues that have a similar mechanism of action yet superior activity. A consistent number of structural modifications have been introduced in the original structure of artemisinin in order to overcome the solubility as well as neurotoxic problem associated with its utilization as antimalarial drug. The study and assessment of these have permitted the clinical development and their usage in the treatment of malaria. Since the discovery of the therapeutic properties of artemisinin, new findings related to its activities, its mechanism of action and pharmacological properties have been unveiled. The great diversity of the artemisinin analogues, the huge number of assays carried out on them, and the different mechanisms of action observed in the different series make it difficult to clearly define the minimum structural requirements necessary for their biological activity. Additionally, the results available have been obtained by different authors, at different times and using different technologies and on very diverse types of cell lines. For all these reasons, greater systematization would be required to obtain definitive conclusions.

In this study, automated docking calculations were performed to eliminate the bias in selecting preferred configurations (orientations). Thus, all possible configurations between heme and artemisinin were explored. The crystallographic X-ray structure of artemisinin was used for the docking simulation. In addition, because few crystallographic X-ray structures of artemisinin derivatives are available, it is worth establishing a suitable geometry optimization

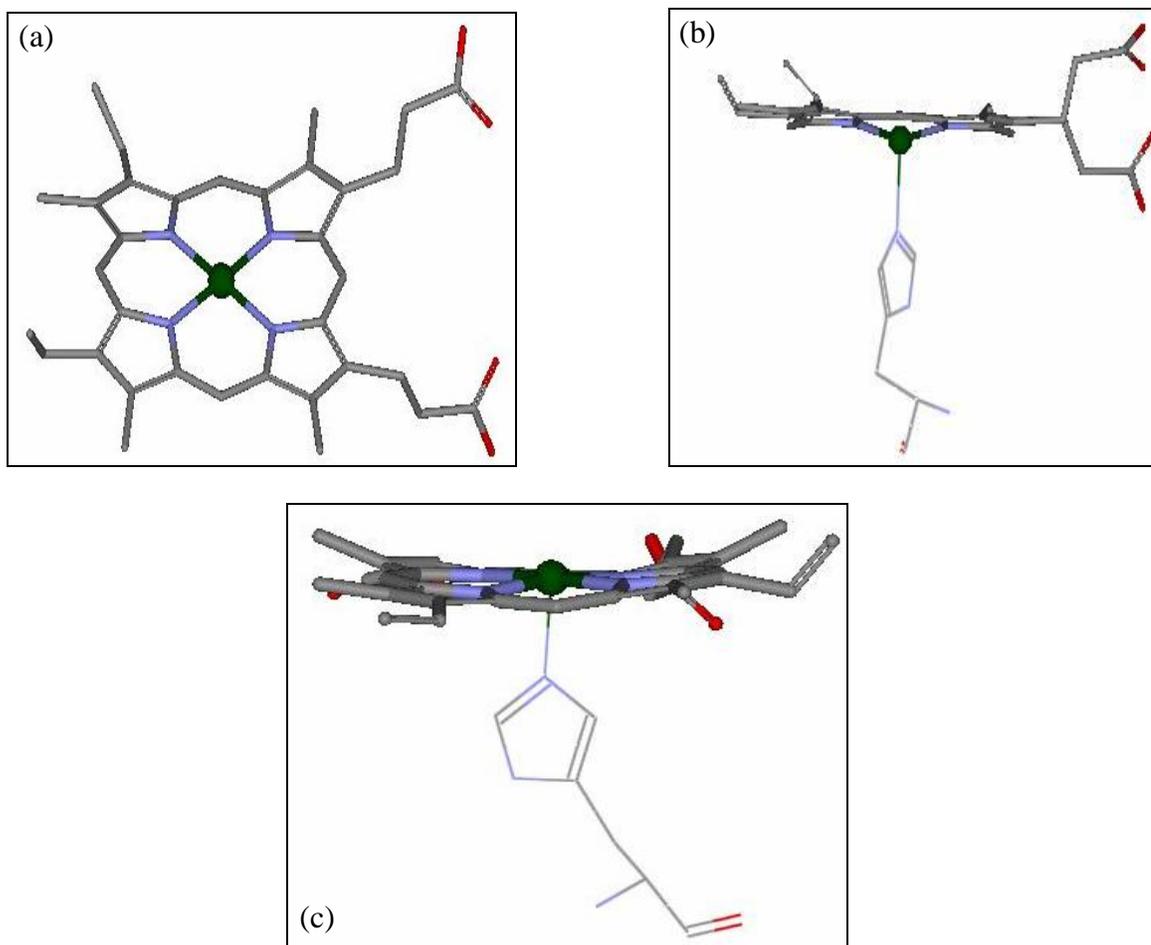
scheme to determine structures of artemisinin derivatives for further investigations (Tonmumphean et. al., 2000). For the heme iron, accurate ab initio calculations were performed to obtain its atomic charge (and those of artemisinin) instead of using a crude approximation for the charge of iron, and specific parameters for iron were used in the docking calculations. The effects of different heme structures were also considered. Thus, three heme structures taken from the literature were studied. The knowledge obtained from this study has been used as a guide for series of docking experiments between heme and artemisinin derivatives. Further in this work we created a virtual library of artemisinin analogues which are collected from different sources and screened them for heme binding based on the optimized docking protocol. Prediction models for predicting the antimalarial activity of these compounds were developed based on binding interaction with heme as descriptor. This prediction model was used for predicting the antimalarial activity of these derivative and we found a very pronounced relationship between their binding energies and antimalarial activity. We have used the molecular modeling techniques such as molecular docking and rescoring using Prime MMGB/SA.

## **2.2 Materials and methods**

### **2.2.1 Receptor preparation**

The X-ray structure of three heme structures i.e., heme-pdb, heme-deoxy and heme-oxy were taken from the Protein Data Bank. These structures are all different owing to the source of heme and the oxidation state of iron. The first structure heme-pdb (PDB ID: 1CTJ) has planar heme structure with a strong positive charge on its central iron atom, which lies slightly above the porphyrin plane (Figure 2.4a). In the process of hemoglobin degradation by the malaria parasite, the proximal ligand may possibly still be attached to the heme iron and, therefore, it is very possible that the histidine remains with the heme structure. As a result, the fourth and the fifth structures, heme-deoxy and heme-oxy, respectively, were obtained from the modifications of deoxy and oxy forms of hemoglobin which contain histidine as the proximal. Both deoxy and oxy forms of hemoglobin were taken from the Protein Data Bank (ID: 1A3 N and 1HHO, respectively). In the heme-deoxy, the histidine pulls the Fe atom to lie below the protoporphyrin plane and gives it a basin-like structure (Figure 2.4b). In the oxy hemoglobin structure, there are six coordinations for heme iron, i.e.

with four N atoms in the protoporphyrin ring, with the proximal ligand (histidine), and with O<sub>2</sub>. Thus, for docking purposes, the O<sub>2</sub> coordination was deleted while maintaining the coordinates of the rest; this modified structure was taken as the receptor structure. As in heme-deoxy, the protoporphyrin plane has a basin-like structure, because of the attraction to the heme iron by histidine. Interaction with O<sub>2</sub> causes the Fe atom to be drawn up above the plane (Figure 2.4c), however, and thus results in a structure which is markedly different from the heme-deoxy.



**Figure 2.4.** The structures of three heme compounds: (a) heme-pdb, (b) heme-deoxy and (c) heme-oxy.

Charge on the iron was assigned as +2 but the structure was kept the same. Hydrogens were added to the model automatically via the Maestro interface leaving no lone pair and using an explicit all-atom model. The multi step Schrodinger's protein preparation tool (PPrep) has been used for final preparation of receptor model. The structure was energy minimized using

OPLS\_2005 force field and the conjugate gradient algorithm, keeping all atoms except hydrogen fixed. The minimization was stopped either after 1000 steps or after the energy gradient converged below 0.01 KJ/mol.

### 2.2.2 Atomic charge calculations

To investigate the effect of the atomic charge on docked configurations, atomic charges of both artemisinin and heme obtained at various levels of theory were used. For heme the HF/3-21G and HF/6-311G\*\* atomic charges were calculated. For artemisinin, atomic charge calculations were performed at HF/3-21G, HF/6-31G\* and HF/6-311G\*\*. All quantum chemical calculations were carried out using Jaguar (Schrodinger, Inc. 2000). It has been seen that all the 3 atomic charges (HF/3-21G, HF/6-31G\* and HF/6-311G\*\*) used for artemisinin gave similar results. Thus for the sake of saving CPU times, the HF/3-21G charges were chosen for atomic charge calculation and complete geometry optimization of all the artemisinin analogues used in the study.

### 2.2.3. Preparation of ligands

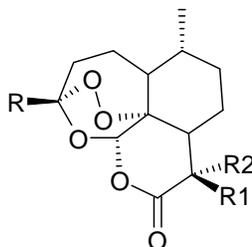
An initial dataset of 144 artemisinin analogues were collected from published data (Acton et. al., 1993; Lin et. al., 1989; Posner et. al., 1992; Avery et. al., 1995; Avery et. al., 1996; Pinheiro et. al., 2001) in which several different ring systems were represented. All of the analogues were either peroxides or trioxanes, which should act via similar mechanisms of action and were categorized into 10 classes (Table 2.1). Each of these compounds had associated in vitro bioactivity values (IC<sub>50</sub> values reported in ng/ml) against the drug resistant malaria strain *P. falciparum* (W-2 clone). The log value of the relative activity (RA) of these compounds was used for analysis and was defined as:

$$\text{Log(RA)} = \log[(\text{artemisinin IC}_{50}/\text{analogue IC}_{50})(\text{analogue MW}/\text{artemisinin MW})] \quad (1)$$

Molecular models of the artemisinin and its analogues (Table 2.1) were built using the Builder feature in Maestro (Schrodinger package) and energy minimized in a vacuum using Impact. Each structure was assigned an appropriate bond order using ligprep script shipped by Schrödinger and optimized initially by means of the OPLS 2005 force field using default setting. Complete geometrical optimization of these structures was carried out with the HF/3-21G method (in this work) using the Jaguar (Schrodinger Inc.). In order to check the

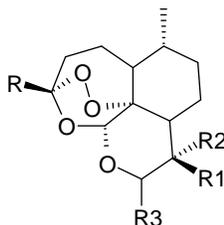
reliability of the geometry obtained, we compared the structural parameters of the artemisinin 1,2,4-trioxane ring with theoretical (Leban et. al., 1988) and experimental (Lisgarten et. al., 1998; Fersht 1984) values from the literature. All calculations reproduced most of the structural parameters of the artemisinin 1,2,4-trioxane ring seen in X-ray structures (Table 2.2). This applies especially to the bond length of the endoperoxide bridge which seems to be responsible for the antimalarial activity.

**Table 2.1.** Artemisinin analogues with antimalarial activities against the drug-resistant malarial strain *P. falciparum* (W-2 clone) used in the work.



Compounds	R	R1	R2	log (RA)	IC <sub>50</sub> (ng/ml)
1	CH <sub>3</sub>	CH <sub>3</sub>	H	1.00	0.040
2	C <sub>4</sub> H <sub>8</sub> Ph	H	H	0.45	0.194
3	CH <sub>3</sub>	H	2-Z-Butenyl	-1.10	5.750
4	CH <sub>3</sub>	H	H	0.79	0.065
5	CH <sub>3</sub>	Allyl	H	0.34	0.550
6	CH <sub>3</sub>	C <sub>4</sub> H <sub>9</sub>	H	0.17	0.311
7	C <sub>4</sub> H <sub>8</sub> Ph	C <sub>4</sub> H <sub>9</sub>	H	-0.32	1.310
8	CH <sub>2</sub> CH <sub>2</sub> CO <sub>2</sub> Et	C <sub>4</sub> H <sub>9</sub>	H	1.36	0.025
9	C <sub>4</sub> H <sub>9</sub>	C <sub>4</sub> H <sub>9</sub>	H	-0.48	1.568
10	CH <sub>3</sub>	C <sub>2</sub> H <sub>5</sub>	H	1.40	0.017
11	CH <sub>3</sub>	C <sub>6</sub> H <sub>13</sub>	H	0.86	0.069
12	CH <sub>3</sub>	i- C <sub>6</sub> H <sub>13</sub>	H	-0.04	0.547
13	CH <sub>3</sub>	i-C <sub>5</sub> H <sub>11</sub>	H	0.07	0.408
14	C <sub>3</sub> H <sub>6</sub> (p-Cl-Ph)	H	H	0.10	0.457
15	C <sub>4</sub> H <sub>9</sub>	H	H	-0.74	2.416
16	CH <sub>2</sub> CH <sub>2</sub> CO <sub>2</sub> Et	H	H	0.37	0.214
17	CH <sub>3</sub>	C <sub>3</sub> H <sub>6</sub> (p-Cl-Ph)	H	1.37	0.025
18	CH <sub>3</sub>	Br	CH <sub>2</sub> Br	-1.64	27.24
19	CH <sub>3</sub>	=CH <sub>2</sub>	-	-0.89	3.083
20	CH <sub>3</sub>	CH <sub>2</sub> CH <sub>3</sub>	-	-0.36	1.053
21	CH <sub>3</sub>	-CH <sub>2</sub> CH <sub>2</sub> -	-	-0.94	3.632
22	CH <sub>3</sub>	C <sub>5</sub> H <sub>11</sub>	H	1.02	0.046
23	CH <sub>3</sub>	C <sub>4</sub> H <sub>8</sub> Ph	H	0.63	0.133
24	CH <sub>3</sub>	C <sub>4</sub> H <sub>8</sub> Ph	H	0.12	0.400

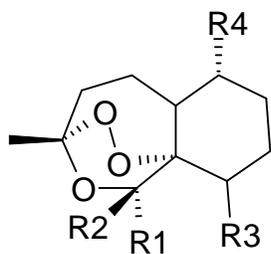
**Table 2.1(continued).** 10-Substituted artemisinin derivatives with antimalarial activities against the drug-resistant malarial strain *P. falciparum* (W-2 clone) used in the work.



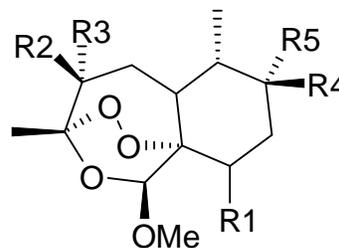
Compounds	R	R1	R2	R3	log (RA)	IC <sub>50</sub> (ng/ml)
25	CH <sub>3</sub>	CH <sub>3</sub>	H	H	0.75	0.068
26	CH <sub>3</sub>	CH <sub>3</sub>	H	OH	0.55	0.114
27	CH <sub>3</sub>	CH <sub>3</sub>	H	OEt	0.34	0.202
28	CH <sub>3</sub>	CH <sub>3</sub>	H	OH	0.96	0.051
29	CH <sub>3</sub>	H	Br	H	0.28	0.248
30	CH <sub>3</sub>	CH <sub>3</sub>	Br	NH-2-(1,3-thiazole)	0.66	0.134
31	CH <sub>3</sub>	CH <sub>3</sub>	Br	p-Cl-aniline	0.79	0.105
32	CH <sub>3</sub>	CH <sub>3</sub>	Br	aniline	0.18	0.397
33	CH <sub>3</sub>	Br	CH <sub>3</sub>	NH-2-pyridine	-0.09	0.768
34	CH <sub>3</sub>	CH <sub>3</sub>	Br	NH-2-pyridine	-0.77	3.667
35	CH <sub>3</sub>	CH <sub>3</sub>	H	$\alpha$ -OEt	0.32	0.212
36	CH <sub>3</sub>	C <sub>4</sub> H <sub>9</sub>	H	H	1.32	0.021
37	CH <sub>3</sub>	C <sub>2</sub> H <sub>5</sub>	H	H	0.67	0.086
38	CH <sub>3</sub>	C <sub>3</sub> H <sub>7</sub>	H	OEt	-0.04	0.529
39	CH <sub>3</sub>	H	H	OEt	0.43	0.157
40	CH <sub>3</sub>	CH <sub>3</sub>	H	C <sub>3</sub> H <sub>6</sub> OH	0.78	0.077
41	CH <sub>3</sub>	CH <sub>3</sub>	H	C <sub>4</sub> H <sub>9</sub>	0.06	0.400
42	CH <sub>3</sub>	CH <sub>3</sub>	H	OCH <sub>2</sub> CO <sub>2</sub> Et	0.52	0.158
43	CH <sub>3</sub>	CH <sub>3</sub>	H	OC <sub>2</sub> H <sub>4</sub> CO <sub>2</sub> Me	0.10	0.433
44	CH <sub>3</sub>	CH <sub>3</sub>	H	OC <sub>3</sub> H <sub>6</sub> CO <sub>2</sub> Me	-0.03	0.605
45	CH <sub>3</sub>	CH <sub>3</sub>	H	OCH <sub>2</sub> (4-PhCO <sub>2</sub> Me)	-0.07	0.720
46	CH <sub>3</sub>	CH <sub>3</sub>	H	(R)-OCH <sub>2</sub> CH(CH <sub>3</sub> )CO <sub>2</sub> Me	1.79	0.009
47	CH <sub>3</sub>	CH <sub>3</sub>	H	(S)-OCH <sub>2</sub> CH(CH <sub>3</sub> )CO <sub>2</sub> Me	2.25	0.003
48	CH <sub>3</sub>	CH <sub>3</sub>	H	(R)-OCH(CH <sub>3</sub> )CH <sub>2</sub> CO <sub>2</sub> Me	0.87	0.073
49	CH <sub>3</sub>	CH <sub>3</sub>	H	(S)-OCH(CH <sub>3</sub> )CH <sub>2</sub> CO <sub>2</sub> Me	1.70	0.011
50	CH <sub>2</sub> CH <sub>2</sub> CO <sub>2</sub> Et	H	H	H	0.70	0.096
51	C <sub>4</sub> H <sub>9</sub>	H	H	H	0.75	0.075
52	C <sub>4</sub> H <sub>8</sub> Ph	H	H	H	0.58	0.139
53	CH <sub>3</sub>	-OCH <sub>2</sub> -	-	OOH	-0.62	1.857
54	CH <sub>3</sub>	-CH <sub>2</sub> O-	-	OOH	-0.57	1.655
55	CH <sub>3</sub>	=CH <sub>2</sub>	-	OOH	-0.99	4.131
56	CH <sub>3</sub>	C <sub>5</sub> H <sub>11</sub>	H	H	0.16	0.318
57	CH <sub>3</sub>	C <sub>3</sub> H <sub>6</sub> Ph	H	H	1.40	0.021
58	CH <sub>3</sub>	CH <sub>3</sub>	H	OEt-C <sub>4</sub> H <sub>9</sub>	0.92	0.061
59	-	CH <sub>3</sub>	OH	$\alpha$ -OH	-0.89	3.303
60	-	CH <sub>3</sub>	H	CH <sub>2</sub> CHF <sub>2</sub>	0.11	0.366
61	-	CH <sub>3</sub>	OH	OCH <sub>2</sub> CF <sub>3</sub>	0.33	0.243
62	-	CH <sub>3</sub>	OH	OEt	-0.44	1.281

All R3 substituents are  $\beta$  except where noted.

**Table 2.1(continued).** Artemisinin derivatives lacking the D-ring with antimalarial activities against the drug-resistant malarial strain *P. falciparum* (W-2 clone) used in the work.



(63-74)

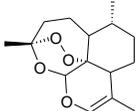
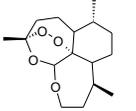
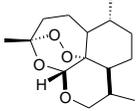
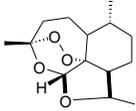
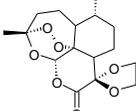
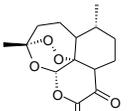
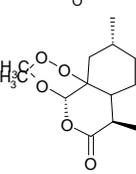
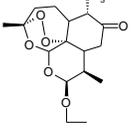
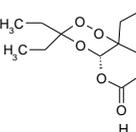
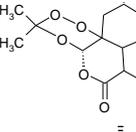
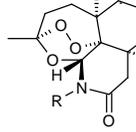


(75-79)

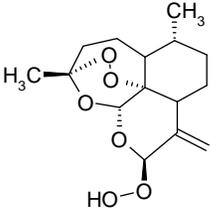
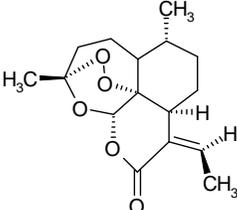
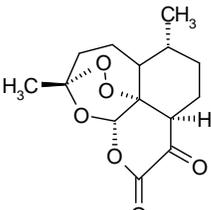
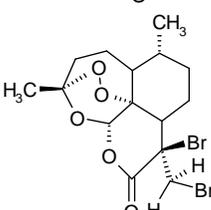
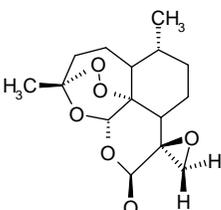
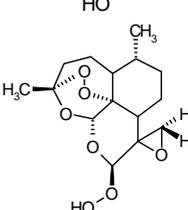
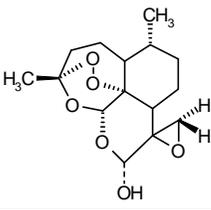
Compounds	R1	R2	R3	R4	log (RA)	IC <sub>50</sub> (ng/ml)
63	O CH <sub>3</sub> Ph	H	H	H	-0.09	0.530
64	O CH <sub>3</sub>	H	C <sub>2</sub> H <sub>4</sub> O <sub>2</sub> CNEt	H	-0.65	0.118
65	H	O CH <sub>3</sub>	C <sub>2</sub> H <sub>4</sub> OCH <sub>3</sub>	H	-0.39	0.996
66	H	O CH <sub>3</sub>	C <sub>2</sub> H <sub>4</sub> OCH <sub>2</sub> Ph	H	0.75	0.091
67	H	O CH <sub>3</sub>	C <sub>2</sub> H <sub>4</sub> O-allyl	H	0.40	0.184
68	H	O CH <sub>3</sub>	C <sub>2</sub> H <sub>4</sub> O <sub>2</sub> Ph	H	-0.59	2.086
69	H	O CH <sub>3</sub>	C <sub>2</sub> H <sub>4</sub> O <sub>2</sub> C(4-PhCO <sub>2</sub> Me)	H	0.27	0.343
70	H	O CH <sub>3</sub>	C <sub>2</sub> H <sub>4</sub> O <sub>2</sub> C(4-PhCO <sub>2</sub> H)	H	-0.81	3.856
71	-	O CH <sub>3</sub>	-	-	1.70	0.398
72	H	O CH <sub>3</sub>	C <sub>2</sub> H <sub>4</sub> O <sub>2</sub> C(4-PhCO <sub>2</sub> C <sub>2</sub> H <sub>4</sub> NMe <sub>2</sub> )	H	0.25	2.790
73	H	O CH <sub>3</sub>	C <sub>2</sub> H <sub>4</sub> O <sub>2</sub> CCH <sub>2</sub> NCO <sub>2</sub> -(t-C <sub>2</sub> H <sub>9</sub> )	H	-0.04	0.670
74	H	O CH <sub>3</sub>	C <sub>2</sub> H <sub>4</sub> OCH <sub>2</sub> (4-N-Me-pyridine)	H	-0.90	4.439

Compounds	R1	R2	R3	R4	R5	log (RA)	IC <sub>50</sub> (ng/ml)
75	C <sub>2</sub> H <sub>4</sub> OH	H	CH <sub>3</sub>	H	H	-1.80	26.849
76	C <sub>2</sub> H <sub>4</sub> OH	CH <sub>3</sub>	H	H	H	0.23	0.251
77	C <sub>2</sub> H <sub>4</sub> OH	CH <sub>3</sub>	CH <sub>3</sub>	H	H	-1.80	28.102
78	C <sub>2</sub> H <sub>4</sub> OCH <sub>2</sub> Ph	CH <sub>3</sub>	CH <sub>3</sub>	H	H	-1.80	36.157
79	C <sub>2</sub> H <sub>4</sub> OCH <sub>2</sub> (4-py)	-	-	-	-	0.14	0.373

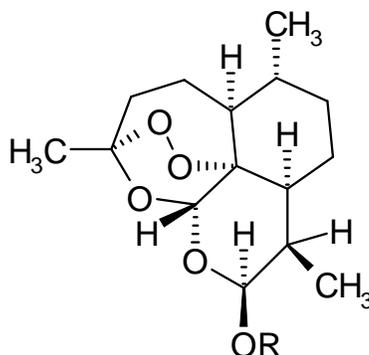
**Table 2.1(continued).** Miscellaneous Artemisinin derivatives with antimalarial activities against the drug-resistant malarial strain *P. falciparum* (W-2 clone) used in the work.

Compounds	structure	log (RA)	IC <sub>50</sub> (ng/ml)
80		0.78	0.063
81		-1.20	6.340
82		-0.79	2.344
83		-0.64	1.573
84		-2.09	56.889
85		-2.49	123.612
86		-0.80	2.309
87		0.16	0.320
88		-0.60	1.525
89		-1.27	6.762
90		0.328	0.400

**Table 2.1(continued).** 9-Substituted Artemisinin derivatives with antimalarial activities against the drug-resistant malarial strain *P. falciparum* (W-2 clone) used in the work.

Compounds	Structure	log (RA)	IC <sub>50</sub> (ng/ml)
91		-0.739	2.320
92		-0.197	0.657
93		-2.298	79.429
94		-1.487	19.143
95		-0.460	1.286
96		-0.409	1.143
97		-0.361	0.971

**Table 2.1(continued).** Dihydroartemisinin derivatives with antimalarial activities against the drug-resistant malarial strain *P. falciparum* (W-2 clone) used in the work.

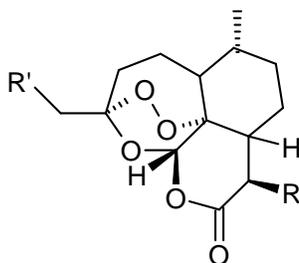


Compounds	R	log (RA)	IC <sub>50</sub> (ng/ml)
98	OR = H	0.487	0.123
99	(S)-CH <sub>2</sub> CH(CH <sub>3</sub> )COOCH <sub>3</sub>	2.104	0.004
100	(S)-CH(CH <sub>3</sub> )CH <sub>2</sub> COOCH <sub>3</sub>	0.599	0.137
101	1-adamantylmethyl	0.007	0.020
102	(S)-CH <sub>2</sub> CH(CH <sub>3</sub> )COOH	-0.658	0.603
103	(S)- CH(CH <sub>3</sub> )CH <sub>2</sub> COOH	-0.608	2.123
104	(R)-CH(CH <sub>3</sub> )CH <sub>2</sub> COOH	-0.383	2.380
105	OR= =O	-0.269	0.743
106	CH <sub>2</sub> PhCOOH	0.176	0.394
107	(R)-CH <sub>2</sub> CH(CH <sub>3</sub> )COOCH <sub>3</sub>	1.524	0.016
108	(R)-CH <sub>2</sub> CH(CH <sub>3</sub> )COOH	-0.463	1.520

**Table 2.1(continued).** Tricyclic 1.2.4 – trioxanes derivatives with antimalarial activities against the drug-resistant malarial strain *P. falciparum* (W-2 clone) used in the work.

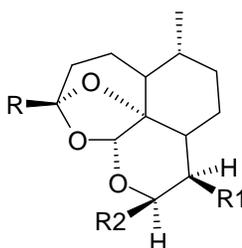
Compounds	structure	log (RA)	IC <sub>50</sub> (ng/ml)
109		-0.475	1.886
110		0.995	0.057
111		-0.413	1.771
112		0.968	0.057
113		0.905	0.057
114		0.991	0.057
115		0.660	0.143
116		0.787	0.086
117		0.717	0.057
118		0.434	0.229
119		0.129	0.314

**Table 2.1(continued).** 3C- substituted artemisinin derivatives with antimalarial activities against the drug-resistant malarial strain *P. falciparum* (W-2 clone) used in the work.



Compounds	R1	R	log (RA)	IC <sub>50</sub> (ng/ml)
120	CH <sub>3</sub>	H	0.049	0.357
121	CH <sub>3</sub> CH <sub>2</sub>	H	0.828	0.062
122	CH <sub>3</sub> CH	H	-0.347	0.977
123	EtO <sub>2</sub> CCH <sub>2</sub>	H	0.365	0.216
124	C <sub>6</sub> H <sub>5</sub> CH <sub>2</sub>	H	-2.000	50.78
125	p-ClC <sub>6</sub> H <sub>4</sub> (CH <sub>2</sub> ) <sub>2</sub>	-	0.104	0.453
126	C <sub>6</sub> H <sub>5</sub> (CH <sub>2</sub> ) <sub>3</sub>	H	0.449	0.195
127	CH <sub>3</sub>	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>3</sub>	0.410	0.187
128	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>2</sub>	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>3</sub>	-0.481	1.573
129	C <sub>6</sub> H <sub>5</sub> CH <sub>2</sub>	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>3</sub>	-2.000	58.72
130	p-ClC <sub>6</sub> H <sub>4</sub> (CH <sub>2</sub> ) <sub>2</sub>	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>3</sub>	-0.276	1.239
131	C <sub>6</sub> H <sub>5</sub> (CH <sub>2</sub> ) <sub>3</sub>	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>3</sub>	-0.319	1.306
132	EtO <sub>2</sub> CCH <sub>2</sub>	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>3</sub>	1.359	0.025

**Table 2.1(continued).** Deoxy artemisinin derivatives with antimalarial activities against the drug-resistant malarial strain *P. falciparum* (W-2 clone) used in the work.



Compounds	R	R1	R2	log (RA)	IC <sub>50</sub> (ng/ml)
133	CH <sub>3</sub>	CH <sub>3</sub>	OEt	-4	4198.58
134	CH <sub>3</sub>	CH <sub>3</sub>	OH	-4	3801.42
135	CH <sub>3</sub>	C <sub>4</sub> H <sub>8</sub> Ph	-	-4	5248.23
136	CH <sub>3</sub>	C <sub>3</sub> H <sub>7</sub>	-	-4	3971.63
137	CH <sub>3</sub>	CH <sub>3</sub>	-	-4	4567.37
138	CH <sub>3</sub>	C <sub>4</sub> H <sub>9</sub>	H	-4	4170.21
139	CH <sub>2</sub> CH <sub>2</sub> CO <sub>2</sub> Et	H	H	-4	4652.48
140	C <sub>2</sub> H <sub>4</sub> Ph	H	-	-4	3574.47
141	CH <sub>2</sub> CH <sub>3</sub>	H	-	-4	3574.47
142	CH <sub>3</sub>	C <sub>2</sub> H <sub>4</sub> Ph	-	-4	4851.06
143	CH <sub>3</sub>	C <sub>3</sub> H <sub>6</sub> Ph	-	-4	5049.64
144	CH <sub>3</sub>	CH <sub>3</sub>	-	-4	3773.05

**Table 2.2** Experimental and theoretical values of the 1,2,4-trioxane ring parameters in artemisinin (bond lengths in Å; bond angles and torsional angles in degrees).

Parameters <sup>a</sup>	Theoretical			Experimental <sup>c</sup>	Experimental <sup>d</sup>
	3-21G <sup>b</sup>	3-21G** <sup>b</sup>	6-31G <sup>b</sup>		
O1-O2	1.463	1.462	1.447	1.475(4)	1.469(2)
O2-C3	1.441	1.440	1.435	1.417(4)	1.416(3)
C3-O4	1.436	1.436	1.435	1.448(4)	1.445(2)
O4-C5	1.407	1.408	1.403	1.388(4)	1.379(2)
C5-C6	1.529	1.530	1.533	1.528(5)	1.523(2)
C6-O1	1.478	1.477	1.469	1.450(4)	1.461(2)
O1-O2-C3	106.9	107.070	108.800	107.600(2)	108.100(1)
O2-C3-O4	107.0	107.310	106.760	107.200(2)	106.600(2)
C3-O4-C5	115.6	115.700	117.300	113.500(3)	114.200(2)
O4-C5-C6	112.0	112.030	112.280	114.700(2)	114.500(2)
C5-C6-O1	111.1	111.589	110.910	111.100(2)	110.700(2)
C6-O1-O2	111.2	111.286	113.240	111.500(2)	111.200(2)
O1-O2-C3-O4	-74.9	-74.680	-71.840	-75.500(3)	-75.500(2)
O2-C3-O4-C5	31.8	32.150	33.390	36.300(4)	36.000(2)
C3-O4-C5-C6	29.4	28.400	25.320	24.800(4)	25.300(2)
O4-C5-C6-O1	-51.8	-50.769	-49.410	-50.800(4)	-51.300(2)
C5-C6-O1-O2	10.1	9.792	12.510	12.300(3)	12.700(2)
C6-O1-O2-C3	50.8	50.522	46.700	47.700	47.800(2)

<sup>a</sup> Atoms are numbered according to Figure 2.1; <sup>b</sup> This work; <sup>c</sup> Values from Ref. (Lisgarten et al., 1998) (experimental estimated standard deviations in brackets); <sup>d</sup> Values from Ref. (Fersht, 1984) (experimental estimated standard deviations in brackets)

#### 2.2.4 Docking of the ligands

All the ligands were docked to the heme receptor using Glide. After ensuring that protein and ligands are in correct form for docking, the receptor-grid files were generated using grid-receptor generation program, using van der Waals scaling of the receptor at 0.4. The default size was used for the bounding and enclosing boxes. The grid box was generated at the centroid of the heme receptor. The ligands were docked initially using the “standard precision” method and further refined using “xtra precision” Glide algorithm. For the ligand docking stage, van der Waals scaling of the ligand was set at 0.5. Of the 50,000 poses that were sampled, 4,000 were taken through minimization (conjugate gradients 1,000) and the 30 structures having the lowest energy conformations were further evaluated for the favorable Glide docking score. A single best conformation for each ligand was considered for further analysis.

### 2.2.5. Rescoring using Prime/MM-GBSA

For each ligand, the pose with the lowest Glide score was rescored using Prime/MM-GBSA approach (Lyne et. al., 2006). This approach has been used to predict the free energy of binding for set of ligands to receptor. The docked poses were minimized using the local optimization feature in Prime and the energies of complex were calculated using the OPLS-AA force field and generalized-Born/surface area (GB/SA) continuum solvent model. The binding free energy ( $\Delta G_{\text{bind}}$ ) is then estimated using the equation:

$$\Delta G_{\text{bind}} = E_{\text{R:L}} - (E_{\text{R}} + E_{\text{L}}) + \Delta G_{\text{solv}} + \Delta G_{\text{SA}} \quad (2)$$

where  $E_{\text{R:L}}$  is energy of the complex,  $E_{\text{R}} + E_{\text{L}}$  is sum of the energies of the ligand and unliganded receptor, the outcome of the use of OPLS-AA force field,  $\Delta G_{\text{solv}}$  ( $\Delta G_{\text{SA}}$ ) is the difference between GBSA solvation energy (surface area energy) of complex and sum of the corresponding energies for the ligand and unliganded protein. Corrections for entropic changes were not applied in this type of free energy calculation.

In order to explore the reliability of the proposed models we used the cross validation method. Prediction error sum of squares (PRESS) is a standard index to measure the accuracy of a modeling method based on the cross validation technique. The  $r^2_{\text{cv}}$  was calculated in accordance with the *PRESS* and *SSY* (Sum of squares of deviations of the experimental values from their mean) using the following formula.

$$r^2_{\text{cv}} = 1 - \frac{\text{PRESS}}{\text{SSY}} = 1 - \frac{\sum_{i=1}^n (y_{\text{exp}} - y_{\text{pred}})^2}{\sum_{i=1}^n (y_{\text{exp}} - \bar{y})^2}$$

## 2.3. Results and Discussions

### 2.3.1. Effect of atomic charges

In docking calculations, the electrostatic potential is built from atomic charges. Therefore, the choices for atomic charges of both the ligand and receptor would have an effect on the docking results. Using charges obtained from HF/3-21G, and HF/6-311G\*\* levels of theory for heme-pdb, the docking to the artemisinin with HF/3-21G charges was performed. The results in Table 2.3 showed that the docking configurations depend on the heme-pdb

atomic charges and especially the charge of Fe. All docking calculations agree that the heme iron binds with endoperoxide oxygens, where the O1–Fe distance is the shortest. Among these calculations, docking with HF/6-311G\*\* charges yielded the shortest O1–Fe distance of 2.04 Å. This O1–Fe distance is markedly much shorter than those predicted using HF/3-21G (2.39 Å) charges. For the binding energy, the docking with HF/6-311G\*\* charges gave the lowest energy. Thus, the employed charge scheme for heme does have a profound effect on the

**Table 2.3.** Results for docking of heme-pdb with different atomic charges and the artemisinin with HF/3-21G charge.

Heme-pdb charge	Fe charge	Docking score	O1-Fe distance (Å)	O2-Fe distance (Å)	O13-Fe distance (Å)	O11-Fe distance (Å)
HF/3-21G	1.371	-2.24	2.39	3.22	5.43	5.68
HF/6-311G**	1.589	-2.35	2.04	3.06	5.14	5.51

docking result. It is, however, quite difficult to judge which charge scheme leads to the most accurate result, because there is no supporting experimental evidence. Theoretically, HF/6-311G\*\* is the most accurate level of theory employed. It is, therefore, reasonable to choose atomic charges from HF/6-311G\*\* for heme in further docking calculations. To study the effect of atomic charges of artemisinin, the docking calculations using various charge schemes, i.e., HF/3-21G, HF/6-31G\*, and HF/6-311G\*\* for the artemisinin and HF/6-311G\*\* charges for heme-pdb structure were performed. The docking results are given in Table 2.4 and the atomic charges of four oxygen atoms in artemisinin for each charge scheme are listed in Table 2.5. From Table 2.4, the dockings with ab initio charges (HF/3-21G, HF/6-31G\*, and HF/6-311G\*\*) gave similar results. Thus, for the sake of saving CPU times, the HF/3-21G charges were chosen for artemisinin.

**Table 2.4.** Results for docking of heme-pdb with HF/6-311G\*\* charge and the artemisinin with different atomic charges.

Artemisinin charge	Docking score	O1-Fe distance (Å)	O2-Fe distance (Å)	O13-Fe distance (Å)	O11-Fe distance (Å)
HF/3-21G	-2.24	2.39	3.22	5.43	5.68
HF/6-31G*	-2.25	2.34	3.19	5.46	5.68
HF/6-311G**	-3.73	2.53	3.03	5.10	5.42

**Table 2.5.** Atomic charges of four oxygen atoms in artemisinin for all charge schemes.

Artemisinin atomic charges	O1 charge	O2 charge	O13 charge	O11 charge
HF/3-21G	-0.372	-0.348	-0.668	-0.715
HF/6-31G*	-0.410	-0.371	-0.710	-0.672
HF/6-311G**	-0.352	-0.305	-0.541	-0.463

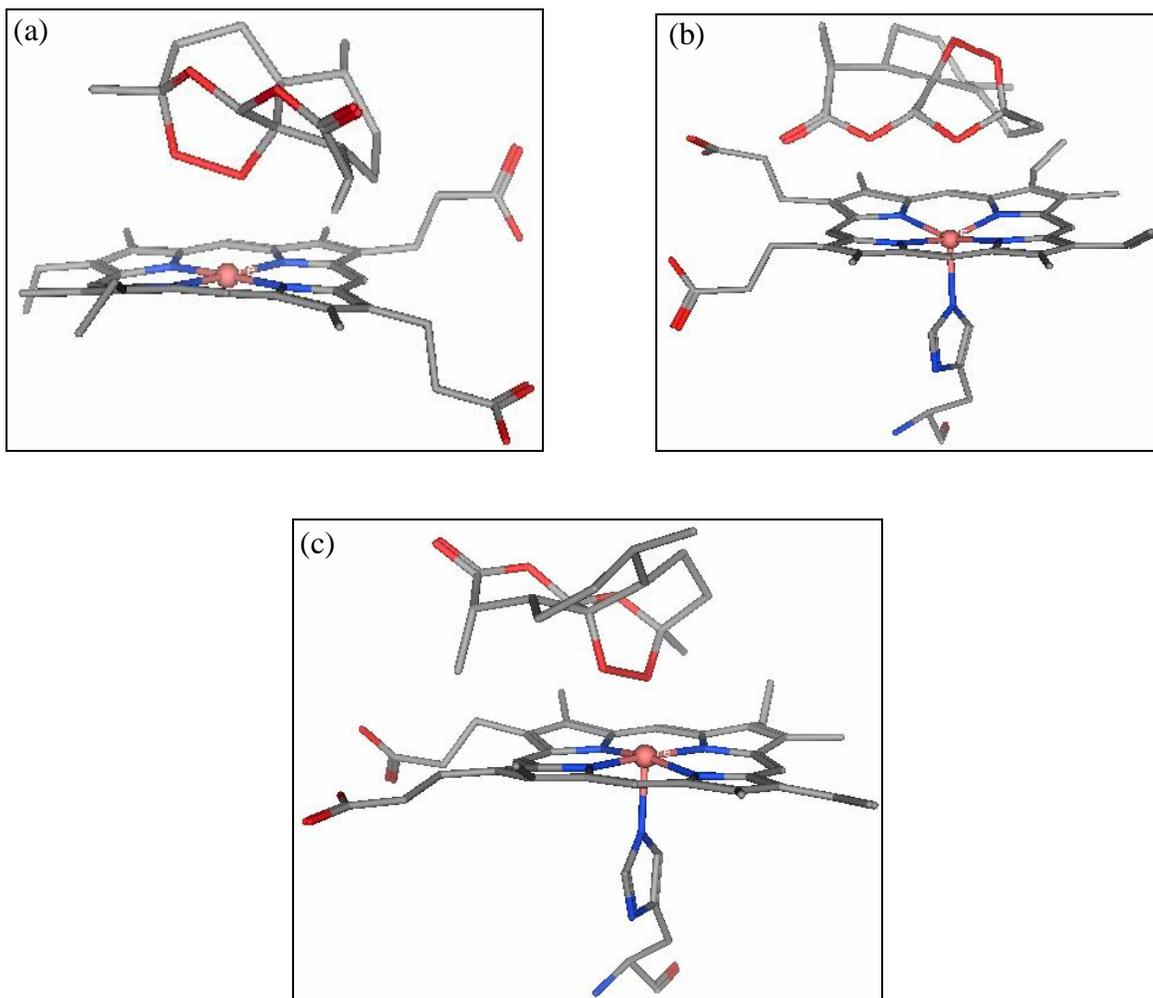
### 2.3.2. Effect of artemisinin structure

Artemisinin was geometry-optimized at various levels of accuracy using ab initio HF/3-21G, HF/6-31G\* and HF/6-31G\*\*. Comparison of these optimized geometries with the crystallographic X-ray structure (Lisgarten et. al., 1998) showed that HF/3-21G gave geometry parameters in good agreement with those of crystallographic X-ray data, especially for the bond length of the endoperoxide linkage, whereas HF/6-31G\* and HF/6-31G\*\* yielded an O–O bond distance that was too short. The HF/3-21G method is, therefore, recommended for the optimization of artemisinin derivatives. This recommendation is, however, based on geometrical criteria only, which does not necessarily guarantee good docking results. To validate the use of this optimized artemisinin structure, the docking calculations between heme-pdb with HF/6-311G\*\* atomic charges and the HF/3-21G, and HF/6-31G\* optimized structures of artemisinin were performed. The docking results are given in Table 2.4. Comparison of the configurations which occur most often reveals good agreement between the docking using the HF/3-21G optimized structure for artemisinin. HF/3-21G is, therefore, the recommended method for geometry optimization of artemisinin derivatives in further study although it has a lower level of accuracy than HF/6-31G\*. It can be argued that for artemisinin derivatives it is possible that the good agreement between the HF/3-21G and the X-ray structures no longer exists, so it would be wiser to employ the more accurate method, HF/6-31G\*. From previous calculations on artemisinin, however, and the current docking results the difference between the structures obtained from the two methods is not pronounced. Thus, the HF/3-21G method is still preferred, because of its faster computation time.

### 2.3.3. Effect of heme structure

To investigate the effect of the heme structure, three heme structures were selected as described in material and methods. The atomic charges were assigned as HF/6-311G\*\* charges for all three heme molecules. For artemisinin compounds, the HF/3-21G optimized structure and atomic charges were used. The results are shown in Table 2.6 and Fig. 2.5. The heme structure chosen does have an effect on the docking results. Although we could not observe agreement on O–Fe distances, all docking calculations with different heme structures (except heme-deoxy) suggested that artemisinin prefers to dock at endoperoxide oxygens (O1 and O2). Using heme-pdb for the heme structure, the docking results showed that artemisinin pointed its endoperoxide moiety toward the heme iron for the most occurring configuration. The O1–Fe and O2–Fe distances were measured and found to be 2.04 Å and 3.22 Å, respectively (Fig. 2.5a); the docking score obtained was –2.24. Owing to the planar structure of the heme-model, the repulsion between artemisinin and the protoporphyrin ring of heme prevents artemisinin from approaching the heme iron as closely as for heme-pdb. The O1–Fe distance of 2.04 Å is comparable with the experimental bond length between the heme iron and oxygen atom in oxyhemoglobin A (1.86 Å), taken from the Protein Data Bank (id 1HHO). For the heme-deoxy, because of its basin-like structure (see Fig. 2.5b), the binding with the endoperoxide moiety of artemisinin is less favorable and a stronger O13–Fe attraction is resulted (docking score –2.51). This could be observed from the most occurring configuration, which has the shorter O13–Fe distance of 3.28 Å, compared with 5.85 and 5.63 Å for O1–Fe and O2–Fe (Fig. 2.5b). Still, this distance is longer than those obtained from the docking with other heme structures. For heme-oxy, the most occurring configuration has O1–Fe as the shortest heme–artemisinin distance with the docking score of –3.22 (Fig. 2.5c). The O1–Fe and O2–Fe distances of 2.56 Å and 3.52 Å are comparable with those of heme-pdb. Note that hemeoxy and heme-pdb have similar structures. From the results from the three heme structures, it can be concluded that the structure of the heme molecule has a significant effect on the docking configurations. The steric hindrance at the Fe position plays an important role in the binding. The proximal ligand that increases the steric hindrance at the Fe position would significantly affect the docking results, as in heme-deoxy. If, however, the proximal ligand does not increase the steric hindrance, results similar to those without the proximal ligand, i.e. for hemeoxy and heme-pdb, would be obtained. Therefore, the heme

structures which facilitate binding between Fe and endoperoxide oxygens, such as heme-pdb and heme-oxy, are recommended for further investigation of the heme–artemisinin system. All docking calculations similarly reported O1–Fe as the shortest heme–artemisinin distance and O2–Fe as the second shortest. It could then be concluded that iron in heme interacts with O1 more preferably than O2, a preference which might arise from the more negative charge at O1 and the steric hindrance at O2. This observation is in agreement with the proposal of Posner et al (1992) (pathway B).



**Figure 2.5.** Docking configuration between artemisinin and (a) heme-pdb, (b) heme-deoxy and (c) heme-oxy structure.

### 2.3.4 Docking of artemisinin derivatives

To better understand the mechanism of interaction and antimalarial activity of artemisinin structural derivatives, computer-aided docking procedures were performed between the drug and its putative receptor. In this study heme-pdb was taken for docking of artemisinin derivatives. The mode of interaction of artemisinin analogues depends partly on the electrostatic configuration of the heme. Both the artemisinin (QHS) and deoxyartemisinin (DQHS) derivatives have similar structures with polar and nonpolar regions. The polar regions, where the oxygens are clustered are negatively charged. QHS has two prominent negative regions (endoperoxide oxygen bridge) and both may interact with the porphyrin iron bridge. Because the DQHS derivatives are lacking of peroxide bridge and are inactive, it was presumed that the main interaction in QHS derivatives is between the peroxide bridge and the heme-iron. Docking methods has been applied here in this study to test whether the peroxide bridge performs an important role in binding to heme. For QHS the atoms O1, O2, O13 and O11 were being tested for interaction with heme-iron. Similarly for DQHS the interacting oxygen atoms; O2, O13 and O11 were being tested for interaction with heme-iron. Here both the oxygen's from the peroxide bridge are in closer proximity to the positive iron than the other two. This indicates that the interaction between QHS and heme involves binding between the endoperoxide (O1 and O2) bridge. The respective distances of the oxygens with respect to the heme-iron are given in Table 2.6. The docking process was able to placed QHS derivatives at distances of 3.89 ( $\pm$  1.19) Å, 4.46 ( $\pm$  1.13) Å, 5.61 ( $\pm$ 0.64) Å and 5.53 ( $\pm$  0.63) Å with respect to O1, O2, O13 and O11 oxygen atoms respectively from heme-iron. DQHS derivatives have single oxygen instead of the peroxide bridge. The heme-iron could interact with DQHS in several ways. Form the docking results it has been seen that the heme-iron preferentially interact either at the side involving the three nonperoxide oxygens O2, O13 and O14 or the peroxide derived oxygen O11. Thus the active antimalarial QHS clearly interacts with heme in a manner different from its inactive analogue DQHS. Heme catalyzes the breakdown of artemisinin (Zhang et. al., 1992) into a free radical (Meshnick et. al., 1993) and/or electrophilic intermediate (Posner and Oh 1992). Once formed, this intermediate can alkylate heme (Hong et. al., 1994) or protein (Yang et. al., 1993). The orientation of QHS with respect to heme may be critical to formation of this intermediate and thus for drug

action. Thus molecular docking and rescoring using Prime MM-GB/SA may aid in the design of new QHS congeners.

**Table 2.6.** Results for docking of haem-pdb with artemisinin (QHS) analogues as well as computed activity using Glide score as a descriptor.

Ligands	Fe-O1 distance (Å)	Fe-O2 distance (Å)	Fe-O13 distance (Å)	Fe-O11 distance (Å)	Glide score	pIC50 <sub>expt</sub>	pIC50 <sub>Gscore</sub>
1	2.55	3.78	5.57	5.73	-2.24	1.40	1.50
2	3.06	3.26	5.53	5.83	-1.58	0.71	0.51
3	5.20	5.37	4.09	4.85	-0.76	-0.76	-0.73
4	2.91	3.30	5.73	5.56	-2.21	1.19	1.46
5	3.37	4.25	6.14	5.09	-1.62	0.26	0.57
6	3.24	3.57	5.58	5.52	-1.32	0.51	0.11
7	4.17	4.72	5.75	5.93	-1.22	-0.12	-0.04
8	2.52	3.17	5.49	5.38	-2.27	1.60	1.55
9	5.20	5.38	6.17	5.87	-1.10	-0.20	-0.22
10	2.63	3.12	5.15	5.67	-1.94	1.78	1.05
11	2.64	3.65	5.64	5.43	-2.29	1.16	1.58
12	3.01	5.08	5.60	5.65	-1.18	0.26	-0.10
13	3.34	4.56	6.61	5.80	-1.85	0.39	0.91
14	3.56	4.76	5.06	5.01	-1.21	0.34	-0.05
15	5.41	5.42	5.13	4.96	-1.08	-0.38	-0.24
16	3.76	3.40	5.79	5.46	-1.56	0.67	0.48
17	2.52	3.89	6.01	5.53	-2.29	1.59	1.58
18	5.89	5.80	4.75	5.24	-0.92	-1.44	-0.49
19	4.85	4.92	6.49	5.20	-1.50	-0.49	0.39
20	4.78	5.51	5.76	6.44	-1.04	-0.02	-0.31
21	6.63	6.32	5.82	5.13	-1.02	-0.56	-0.34
22	3.06	3.47	5.36	5.34	-2.21	1.34	1.46
23	3.18	3.27	5.57	5.70	-1.60	0.88	0.54
24	3.30	4.22	6.13	5.91	-1.50	0.40	0.39
25	2.83	3.37	5.52	5.10	-2.07	1.17	1.25
26	3.14	3.33	5.61	5.39	-1.72	0.94	0.72
27	3.16	3.74	5.70	5.57	-1.62	0.69	0.57
28	2.46	3.31	5.07	5.85	-2.22	1.29	1.47
29	3.90	3.91	5.45	5.44	-1.38	0.61	0.20
30	3.14	3.26	5.91	5.61	-1.77	0.87	0.79
31	3.11	3.18	5.01	5.82	-1.44	0.98	0.29
32	3.96	4.35	5.46	5.53	-1.24	0.40	-0.01
33	3.24	5.29	6.00	5.68	-1.11	0.11	-0.20
34	6.25	6.30	4.00	4.66	-0.46	-0.56	-1.19
35	3.56	3.68	5.18	5.17	-1.41	0.67	0.25
36	2.53	3.29	5.56	5.06	-2.32	1.68	1.62
37	2.71	3.14	5.00	5.38	-2.18	1.07	1.41
38	3.60	4.66	6.44	6.02	-1.54	0.28	0.45
39	3.15	3.16	5.57	5.93	-1.62	0.80	0.57
40	2.42	3.74	5.74	5.56	-2.19	1.11	1.43
41	3.56	4.42	5.74	5.70	-1.24	0.40	-0.01
42	3.17	3.23	4.99	5.97	-1.60	0.80	0.54

**Table 2.6** (Continued). Results for docking of haem-pdb with artemisinin(QHS) analogues as well as computed activity using Glide score as a descriptor.

Ligands	Fe-O1 distance (Å)	Fe-O2 distance (Å)	Fe-O13 distance (Å)	Fe-O11 distance (Å)	Glide score	pIC50 <sub>expt</sub>	pIC50 <sub>Gscore</sub>
43	3.38	4.53	5.71	5.41	-1.25	0.36	0.01
44	3.24	4.93	5.28	5.85	-1.11	0.22	-0.20
45	3.95	5.16	6.40	5.83	-1.50	0.14	0.39
46	2.24	3.04	5.47	5.14	-2.11	2.07	1.31
47	2.18	2.29	5.39	5.45	-2.27	2.53	1.54
48	2.60	3.55	5.86	5.66	-2.17	1.13	1.40
49	2.44	3.06	5.24	5.31	-2.32	1.96	1.62
50	2.84	3.29	5.13	5.02	-2.11	1.02	1.31
51	2.74	3.89	5.35	5.51	-2.18	1.13	1.41
52	3.65	4.67	6.39	5.32	-1.33	0.86	0.13
53	5.58	5.78	6.05	5.81	-1.06	-0.27	-0.27
54	5.40	5.66	5.19	6.40	-1.07	-0.22	-0.27
55	5.71	6.71	5.40	5.62	-0.97	-0.62	-0.42
56	3.86	3.95	5.17	6.14	-1.32	0.50	0.11
57	2.52	3.19	5.14	5.35	-2.20	1.68	1.44
58	2.50	3.22	5.84	5.70	-2.19	1.22	1.43
59	5.82	5.63	5.02	6.19	-0.90	-0.52	-0.52
60	3.54	4.41	5.39	5.99	-1.28	0.44	0.05
61	3.54	3.41	5.76	5.69	-1.65	0.61	0.61
62	4.04	5.81	7.35	6.75	-1.55	-0.11	0.46
63	3.89	4.53	5.55	6.10	-1.20	0.28	-0.07
64	3.14	4.00	6.17	5.38	-1.38	0.93	0.20
65	4.77	5.50	5.67	5.12	-1.08	0.00	-0.25
66	2.58	3.26	5.57	5.58	-2.13	1.04	1.34
67	3.47	3.58	5.54	6.38	-1.31	0.73	0.10
68	4.74	4.02	6.20	5.14	-1.24	-0.32	-0.02
69	3.68	4.72	6.39	5.97	-1.56	0.46	0.48
70	5.53	6.52	4.43	4.92	-0.68	-0.59	-0.85
71	3.36	4.52	5.43	6.41	-1.24	0.40	-0.01
72	5.79	6.02	4.41	4.65	-0.72	-0.45	-0.79
73	3.45	4.41	6.14	6.42	-1.62	0.17	0.57
74	5.70	6.51	5.10	4.74	-0.76	-0.65	-0.73
75	5.79	6.34	4.77	4.48	-0.62	-1.43	-0.94
76	3.99	4.03	6.11	6.76	-1.53	0.60	0.43
77	6.24	6.86	4.17	5.07	-0.56	-1.45	-1.03
78	6.04	7.22	4.01	4.88	-0.33	-1.56	-1.38
79	3.65	4.57	5.74	6.06	-1.29	0.43	0.07
80	2.72	3.11	5.27	5.54	-2.29	1.20	1.58
81	5.27	5.46	5.16	4.98	-1.01	-0.80	-0.35
82	4.47	4.02	5.95	4.80	-1.27	-0.37	0.04
83	5.45	6.08	5.22	5.71	-1.14	-0.20	0.45
84	6.16	6.14	5.45	5.65	-0.54	-1.76	-1.06
85	5.16	6.15	4.21	4.80	-0.51	-2.09	-1.11
86	4.75	6.08	6.84	7.28	-1.54	-0.36	0.45
87	3.43	4.19	5.94	5.73	-1.54	0.49	-0.27

**Table 2.6** (Continued). Results for docking of haem-pdb with artemisinin(QHS) analogues as well as computed activity using Glide score as a descriptor.

Ligands	Fe-O1 distance (Å)	Fe-O2 distance (Å)	Fe-O13 distance (Å)	Fe-O11 distance (Å)	Glide score	pIC50 <sub>expt</sub>	pIC50 <sub>Gscore</sub>
88	4.72	4.53	5.39	6.04	-1.15	-0.18	-0.15
89	6.44	5.03	6.24	5.38	-0.94	-0.83	-0.46
90	3.86	4.92	5.61	5.71	-1.23	0.40	-0.02
91	5.38	5.59	4.63	4.99	-1.07	-0.37	0.57
92	3.49	4.48	6.25	5.55	-1.62	0.18	-0.95
93	5.52	5.91	4.94	5.32	-0.62	-1.90	-0.40
94	5.46	5.68	5.53	4.68	-0.98	-1.28	-0.35
95	4.03	5.36	5.60	5.88	-1.01	-0.11	0.64
96	3.71	4.75	6.48	6.49	-1.67	-0.06	-0.20
97	3.34	5.35	6.05	5.91	-1.11	0.01	-0.20
98	3.17	3.19	5.34	6.86	-1.54	0.91	0.45
99	2.21	2.46	5.61	5.38	-2.38	2.37	1.71
100	3.17	3.18	4.10	5.46	-1.48	0.86	0.35
101	3.34	4.87	6.13	6.31	-1.45	0.22	0.31
102	5.72	5.48	3.73	4.73	-0.64	-0.38	-0.91
103	4.91	5.33	6.51	6.97	-1.59	-0.33	0.52
104	3.83	5.53	6.41	6.25	-1.48	-0.10	0.35
105	3.05	5.40	5.58	5.78	-1.12	0.13	-0.19
106	3.59	4.84	6.13	5.21	-1.25	0.40	0.01
107	2.88	3.08	5.49	5.96	-2.25	1.79	1.52
108	5.62	3.96	5.89	6.94	-1.58	-0.18	0.51
109	5.97	6.48	4.51	3.06	-0.46	-0.28	-1.19
110	2.81	3.68	5.16	5.36	-2.13	1.24	1.34
111	5.93	6.03	6.48	5.06	-1.13	-0.25	-0.18
112	2.41	3.22	5.65	5.61	-2.21	1.24	1.46
113	2.49	3.06	5.64	5.13	-2.23	1.24	1.49
114	2.96	3.49	5.77	5.17	-2.24	1.24	1.50
115	3.14	3.50	4.57	5.02	-1.50	0.85	0.39
116	2.44	3.19	5.44	5.30	-2.16	1.07	1.38
117	2.67	3.63	5.77	5.77	-2.22	1.24	1.47
118	3.86	3.66	5.08	6.47	-1.65	0.64	0.61
119	3.94	3.80	5.97	5.80	-1.62	0.50	0.57
120	3.63	4.28	5.54	5.87	-1.26	0.45	0.02
121	2.27	3.18	5.47	5.74	-2.18	1.20	1.41
122	3.39	4.62	5.84	6.91	-1.72	0.01	0.72
123	3.17	3.62	5.00	6.38	-1.23	0.66	-0.02
124	5.96	6.01	4.93	4.15	-0.76	-1.71	-0.73
125	3.66	4.62	5.51	6.36	-1.23	0.34	-0.02
126	3.81	3.45	5.82	7.16	-1.69	0.71	0.67
127	3.16	3.19	5.32	6.42	-1.54	0.73	0.45
128	5.24	6.53	6.14	5.87	-0.93	-0.20	-0.48
129	6.67	6.11	3.91	3.99	-0.35	-1.77	-1.35
130	4.06	4.84	6.64	6.53	-1.58	-0.09	0.51
131	4.13	4.76	5.90	6.56	-1.48	-0.12	0.35
132	2.59	3.73	5.59	5.14	-1.92	1.59	1.02

**Table 2.6** (Continued). Results for docking of haem-pdb with deoxy-artemisinin(DQHS) analogues as well as computed activity using Glide score as a descriptor.

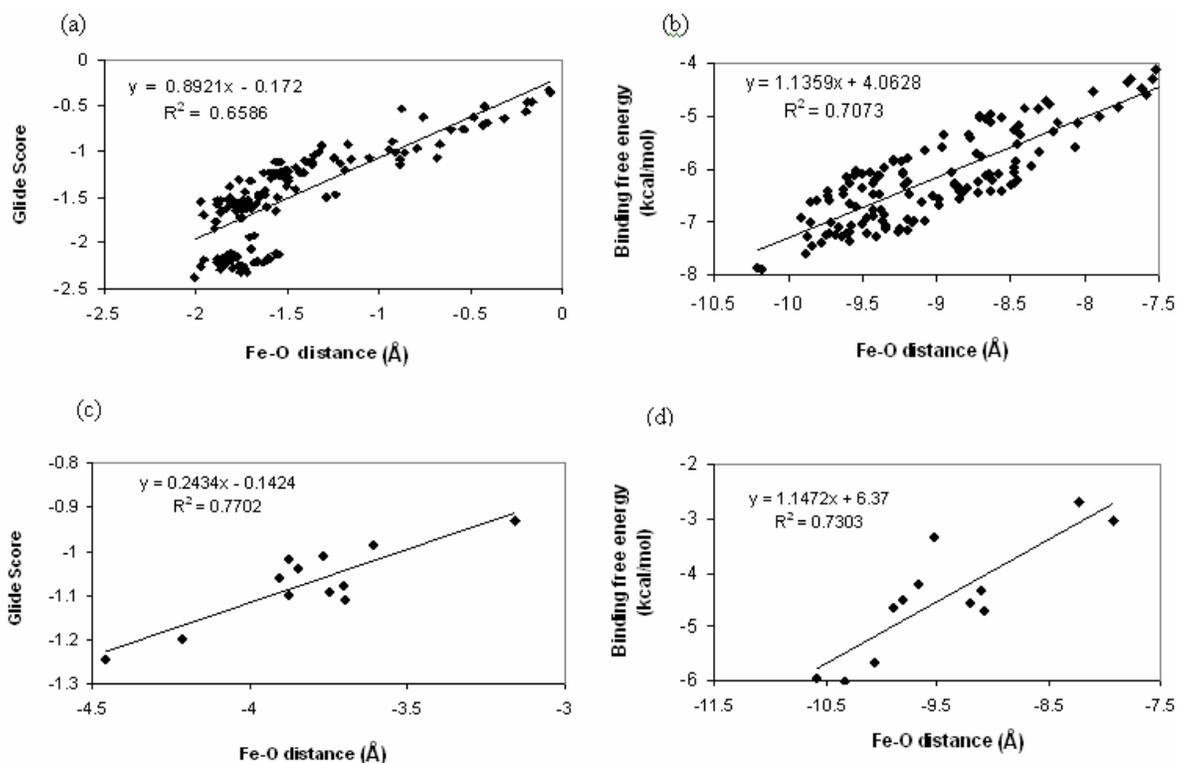
Ligands	Fe-O1 distance (Å)	Fe-O13 distance (Å)	Fe-O11 distance (Å)	Glide score	pIC50 <sub>expt</sub>	pIC50 <sub>Gscore</sub>
133	5.479	4.327	5.435	-3.61	-3.62	-3.63
134	4.141	5.778	4.916	-3.56	-3.63	-3.64
135	4.295	5.893	5.995	-3.52	-3.68	-3.65
136	4.697	6.831	6.378	-3.75	-3.60	-3.59
137	4.297	5.211	5.539	-3.49	-3.66	-3.66
138	3.934	6.186	5.892	-3.60	-3.62	-3.63
139	3.957	6.147	6.604	-3.51	-3.67	-3.65
140	5.118	3.296	4.587	-3.43	-3.66	-3.68
141	5.271	4.438	4.86	-3.59	-3.61	-3.63
142	4.005	5.671	4.578	-3.54	-3.66	-3.65
143	5.404	4.077	4.341	-3.58	-3.65	-3.64
144	4.413	6.282	5.339	-3.70	-3.59	-3.60

We applied the docking MM-GB/SA method to a data set of 144 artemisinin analogues to build a binding affinity model for evaluating antimalarial activity. The data set used for building the binding affinity model comprised nine subsets of artemisinin analogues (Table 2.1). These compounds were taken from various sources, among these are endoperoxide artemisinin analogues, 10-substituted artemisinin derivatives, artemisinin derivatives without D-ring, 9-substituted artemisinin derivatives, dihydroartemisinin derivatives, tricyclic 1,2,4-trioxanes, 3C-substituted artemisinin derivatives, deoxyartemisinin analogues and miscellaneous artemisinin derivatives. The experimental relative activity (RA) values for all those compounds were calculated against the drug resistant malarial strain *P. falciparum* (W-2 clone). The IC<sub>50</sub> value of these analogues was derived from the equation 1 and used for calculation of absolute pIC<sub>50</sub> (pIC<sub>50</sub> = - log IC<sub>50</sub>). With the wide range of difference in pIC<sub>50</sub> values and the large diversity in the structures, the combined set of 144 ligands is ideal to build the affinity binding model as the set does not suffer from bias due to the similarity of the structures. This data set compounds were docked into the haem receptor site using the Glide-XP module and rescore using Prime-MM-GB/SA (Schrodinger, Inc.). For the better understanding of the mechanism of action of the artemisinin analogues all the 144 compounds were classified into highly potent, low and inactive analogues based on the experimental pIC<sub>50</sub>.

All the active artemisinin (QHS) derivatives (1-132, Table 2.1) were found to be good binder with heme (Table 2.6). We can observe that the most potent artemisinin analogues ( $pIC_{50} > 1.0$ ) were found to be having better docking score in comparison to the analogues which are less potent ( $pIC_{50} < 1.0$ ). For the highly potent analogues the distances between O1-Fe, O2-Fe, O13-Fe and O11-Fe atom pairs between heme-iron and artemisinin were  $2.59 (\pm 0.22)$  Å,  $3.31 (\pm 0.36)$  Å,  $5.49 (\pm 0.25)$  Å and  $5.45 (\pm 0.24)$  Å respectively; the glide score obtained was  $-2.20 (\pm 0.11)$ . However, for the less potent analogues the distances were found to be  $4.31 (\pm 1.08)$  Å,  $4.83 (\pm 1.03)$  Å,  $5.54 (\pm 0.71)$  Å and  $5.67 (\pm 0.71)$  Å respectively for the O1-Fe, O2-Fe, O13-Fe and O11-Fe atom pairs; the glide score was  $-1.24 (\pm 0.36)$  (Table 2.6). A linear relationship between Glide score and optimized O-Fe distance was obtained with  $R^2$  value of 0.6586 (Figure 2.6a). The optimized O-Fe distance was obtained by linear combination of O1-Fe, O2-Fe, O13-Fe and O11-Fe atom pairs between oxygen's of artemisinin and heme-iron as explained in Figure 2.6. It has been seen that the distances between O1-Fe and O2-Fe are more important for the activity of artemisinin analogues. For the inactive artemisinin (DQHS) analogues ( $pIC_{50} < -3.0$ ) which lack the peroxide bridge the Glide score was found to be  $-1.07 (\pm 0.09)$  (Table 2.6). The distances for O2-Fe, O13-Fe and O11-Fe atom pairs were  $4.58 (\pm 0.59)$  Å,  $5.34 (\pm 1.08)$  Å and  $5.37 (\pm 0.74)$  Å respectively. Further a linear relationship with  $R^2$  value of 0.7702 was obtained between Glide score and the optimized O-Fe distance (Figure 2.6c). The interaction of the artemisinin with heme is very much dependent upon the stereochemistry of artemisinin analogues, a mechanism that is controlled by steric hindrance. The analogues which approach the heme-iron as close as possible will have better interaction and thus the good glide score.

For each ligand in the virtual library, the pose with the lowest Glide score was rescored using Prime/MM-GBSA approach. Rescoring using Prime/MM-GBSA leads to minor changes of the ligand conformations within receptor site. These changes result from minimization of the ligand in receptor's environment and consequent stabilization of receptor:ligand complex. This approach is used to predict the binding free energy ( $\Delta G_{\text{bind}}$ ) for set of ligands to receptor. Table 2.7 reveals the  $\Delta G_{\text{bind}}$  energy of artemisinin analogues. The  $\Delta G_{\text{bind}}$  energy of the highly potent QHS analogues ( $pIC_{50} > 1.0$ ) were higher ( $-6.84 \pm 0.50$

kcal/mol) then less potent analogues ( $-5.96 \pm 0.84$  kcal/mol) and inactive DQHS derivatives ( $-4.47 \pm 1.07$  kcal/mol).



**Figure 2.6(a-d).** A linear relationship between Fe-O distance and Glide score as well as Fe-O distance and binding free energy of the (a & b) Artemisinin (QHS) derivatives and (c & d) Deoxyartemisinin (DQHS) derivatives. The Fe-O distance represents the optimized value of the distances obtained by linear combination of distances between O1-Fe, O2-Fe, O13-Fe and O11-Fe atom pairs respectively between QHS analogues and haem-iron: O-Fe distance =  $\alpha$  (O1-Fe) +  $\beta$  (O2-Fe) +  $\gamma$  (O13-Fe) +  $\delta$  (O11-Fe). The  $\alpha$ ,  $\beta$ ,  $\gamma$  and  $\delta$  are fitting parameters. The values obtained for the four fitting parameters,  $\alpha$ ,  $\beta$ ,  $\gamma$  and  $\delta$  are 0.101, 0.191, -0.357, -0.129 and -0.482, 0.798, -0.520, -0.905 respectively using Glide score and binding free energy as dependent variables. The optimized equation obtained for O-Fe distance for deoxyartemisinin derivatives was: O-Fe distance =  $\alpha$  (O2-Fe) +  $\beta$  (O13-Fe) +  $\gamma$  (O11-Fe). The values obtained for the three fitting parameters  $\alpha$ ,  $\beta$  and  $\gamma$  are -0.431, -0.490, 0.144 and -1.43, -1.18, 0.74 respectively using Glide score and binding free energy as dependent variables.

The distances between O1-Fe, O2-Fe, O13-Fe and O11-Fe atom pairs for most potent analogues were  $2.77 (\pm 0.30)$  Å,  $3.23 (\pm 0.35)$  Å,  $5.13 (\pm 0.46)$  Å and  $5.57 (\pm 0.49)$  Å respectively. The binding affinity of the artemisinin derivatives with heme is very much dependent upon the proximity of O1 and O2 atoms. On the contrary for the less potent analogues the distances were  $3.69 (\pm 0.53)$  Å,  $4.03 (\pm 0.71)$  Å,  $5.25 (\pm 0.57)$  Å and  $5.26 (\pm 0.71)$  Å. For the inactive DQHS derivatives the distances for O2-Fe, O13-Fe and O11-Fe

atom pairs were 4.83 ( $\pm 0.64$ ) Å, 5.70 ( $\pm 1.09$ ) Å and 5.65 ( $\pm 0.98$ ) Å respectively. A linear relationship between linear combination of O-Fe distances and  $\Delta G_{\text{bind}}$  energy was obtained (Figure 2.5b & 2.5d) with  $R^2$  value of 0.7073 and 0.7303 for the QHS and DQHS derivatives respectively.

**Table 2.7.** Results for rescoring using Prime/MM-GBSA of haem-pdb with artemisinin analogues as well as computed activity using  $\Delta G_{\text{bind}}$  energy as a descriptor.

Ligands	Fe-O1 distance(Å)	Fe-O2 distance(Å)	Fe-O13 distance(Å)	Fe-O11 distance(Å)	$\Delta G_{\text{bind}}$ kcal/mol	pIC50 <sub>expt</sub>	pIC50 <sub><math>\Delta G_{\text{bind}}</math></sub>
1	2.58	2.661	4.407	5.842	-6.68	1.40	0.79
2	3.148	3.528	5.659	4.803	-6.55	0.71	0.68
3	3.006	2.688	5.044	4.761	-5.12	-0.65	-0.57
4	2.363	3.37	5.07	5.712	-6.63	1.19	0.75
5	3.097	3.247	5.408	4.588	-6.09	0.28	0.28
6	3.087	3.209	5.391	4.617	-6.11	0.51	0.30
7	3.286	3.795	5.758	4.704	-5.35	-0.12	-0.36
8	3.39	3.266	5.116	5.558	-6.76	1.60	0.86
9	3.162	4.943	5.586	6.33	-4.13	-2.09	-1.43
10	2.88	3.34	5.134	5.693	-7.12	1.78	1.18
11	2.956	3.245	5.5	5.429	-6.46	1.16	0.60
12	3.059	3.353	5.603	5.662	-6.49	0.26	0.63
13	4.301	4.582	5.765	4.989	-6.26	0.39	0.43
14	3.522	3.852	5.073	4.811	-6.21	0.34	0.39
15	3.15	3.909	6.15	3.74	-4.95	-0.38	-0.71
16	3.723	3.207	4.542	5.243	-6.36	0.67	0.52
17	2.995	2.898	5.117	5.58	-6.76	1.59	0.86
18	3.504	5.126	6.143	6.148	-4.53	-1.43	-1.08
19	3.518	4.453	6.101	4.808	-5.85	-0.45	0.07
20	4.11	4.836	5.241	5.891	-6.58	-0.02	0.71
21	4.89	5.264	5.588	4.913	-5.86	-0.52	0.08
22	2.558	3.396	5.19	4.874	-6.3	1.34	0.46
23	3.146	3.74	5.019	5.917	-6.92	0.88	1.00
24	2.872	2.66	4.997	4.952	-6.24	0.40	0.41
25	2.81	3.049	5.288	4.923	-6.36	1.17	0.52
26	3.284	3.412	5.704	5.624	-7.01	0.94	1.08
27	3.725	3.575	4.902	5.85	-7.17	0.69	1.22
28	2.594	3.261	4.661	5.408	-6.4	1.29	0.55
29	3.03	3.133	5.567	5.56	-7.08	0.61	1.14
30	3.295	3.211	4.867	4.625	-5.95	0.87	0.16
31	4.818	5.461	6.13	6.668	-6.97	0.98	1.05
32	3.691	3.883	4.011	5.455	-6.06	0.40	0.25
33	3.452	4.33	4.013	5.21	-5.12	0.11	-0.57
34	3.94	4.796	6.426	6.296	-4.84	-0.56	-0.81
35	3.773	3.821	4.896	5.473	-7.16	0.67	1.21
36	3.367	3.492	5.689	4.973	-7.18	1.68	1.23
37	3.273	3.336	5.65	5.509	-7.00	1.07	1.07
38	3.174	3.024	5.129	4.351	-5.67	0.28	-0.09
39	3.034	3.159	5.909	5.262	-7.26	0.80	1.30

**Table 2.7** (Continued). Results for rescoring using Prime/MM-GBSA of haem-pdb with artemisinin analogues as well as computed activity using  $\Delta G_{\text{bind}}$  energy as a descriptor.

Ligands	Fe-O1 distance(Å)	Fe-O2 distance(Å)	Fe-O13 distance(Å)	Fe-O11 distance(Å)	$\Delta G_{\text{bind}}$ kcal/mol	pIC50 <sub>expt</sub>	pIC50 <sub><math>\Delta G_{\text{bind}}</math></sub>
40	2.83	3.58	4.859	5.797	-6.97	1.11	1.05
41	3.718	3.801	4.941	5.479	-6.28	0.40	0.45
42	3.806	3.576	4.817	5.754	-7.22	0.80	1.27
43	4.158	4.716	4.897	5.537	-6.23	0.36	0.40
44	3.357	3.985	5.245	5.784	-6.04	0.22	0.24
45	4.088	4.349	4.55	6.002	-6.2	0.14	0.38
46	2.18	3.052	5.78	5.98	-7.6	2.07	1.60
47	2.392	2.42	5.39	6.39	-7.85	2.53	1.82
48	2.834	3.71	4.009	5.899	-6.09	1.13	0.28
49	2.775	2.785	5.12	6.14	-7.45	1.96	1.47
50	2.722	3.813	5.462	4.933	-6.44	1.02	0.59
51	2.496	3.421	5.079	4.955	-6.2	1.13	0.38
52	3.703	4.673	5.522	5.667	-7.21	0.86	1.26
53	4.572	4.858	5.045	5.669	-6.13	-0.25	0.32
54	4.299	4.582	5.901	3.834	-5.27	-0.20	-0.43
55	3.517	4.981	5.74	6.557	-4.58	-0.59	-1.04
56	2.902	2.882	5.11	6.138	-6.63	0.50	0.75
57	2.086	3.194	5.93	5.14	-7.01	1.68	1.08
58	2.739	3.448	5.079	6.094	-7.36	1.22	1.39
59	4.617	4.629	5.788	4.018	-5.07	-0.49	-0.61
60	3.803	3.733	5.62	5.222	-6.72	0.44	0.83
61	3.726	3.662	4.062	5.946	-6.49	0.61	0.63
62	4.489	5.67	5.332	5.625	-6.36	-0.11	0.52
63	3.963	3.53	4.542	5.278	-6.07	0.26	0.26
64	3.274	3.994	4.403	6.193	-7.13	0.93	1.19
65	4.881	5.185	5.238	5.299	-6.11	0.00	0.30
66	2.699	3.431	4.185	6.03	-6.55	1.04	0.68
67	2.852	3.44	5.565	5.449	-7.25	0.73	1.29
68	4.866	4.241	4.305	5.342	-5.58	-0.28	-0.16
69	3.794	4.878	5.535	5.219	-6.94	0.46	1.02
70	3.01	2.936	5.272	5.588	-6.09	-0.56	0.28
71	4.011	4.205	4.527	5.759	-6.48	0.40	0.62
72	4.327	4.943	5.967	6.246	-5.02	-0.38	-0.65
74	3.576	4.326	5.128	4.885	-4.96	-0.62	-0.70
75	3.523	3.477	4.646	5.155	-5.04	-1.28	-0.64
76	3.582	3.572	5.285	5.328	-7.05	0.60	1.12
77	3.606	4.604	5.319	3.67	-4.48	-1.44	-1.12
78	4.943	4.069	5.133	3.19	-4.3	-1.45	-1.28
79	3.746	4.042	5.356	5.943	-6.91	0.43	1.00
80	3.062	3.641	5.069	6.113	-7.24	1.20	1.28
81	3.223	3.194	5.387	4.59	-5.76	-0.76	-0.01
82	3.413	3.887	5.83	4.755	-5.65	-0.37	-0.10
83	4.152	4.245	5.279	4.654	-5.367	-0.20	-0.35
84	3.951	4.445	5.71	6.066	-4.83	-0.98	-0.82
86	3.967	4.579	5.864	4.964	-5.96	-0.33	0.17

**Table 2.7** (Continued). Results for rescoring using Prime/MM-GBSA of haem-pdb with artemisinin analogues as well as computed activity using  $\Delta G_{\text{bind}}$  energy as a descriptor.

Ligands	Fe-O1 distance(Å)	Fe-O2 distance(Å)	Fe-O13 distance(Å)	Fe-O11 distance(Å)	$\Delta G_{\text{bind}}$ kcal/mol	pIC50 <sub>expt</sub>	pIC50 <sub><math>\Delta G_{\text{bind}}</math></sub>
87	3.601	3.757	4.153	5.79	-6.25	0.49	0.42
88	3.682	4.339	4.828	4.627	-5.29	-0.18	-0.42
89	4.398	5.61	5.773	4.45	-5.7	-0.80	-0.06
90	2.89	3.901	5.603	5.766	-6.38	0.40	0.53
91	4.528	4.871	4.032	5.267	-4.86	-0.36	-0.79
92	3.39	3.463	5.198	5.721	-6.27	0.18	0.44
93	3.077	3.557	5.557	3.34	-4.29	-1.77	-1.29
94	3.846	5.05	5.629	5.905	-4.72	-0.83	-0.91
95	4.61	4.738	4.582	4.782	-5.35	-0.11	-0.36
96	4.17	4.732	5.085	6.051	-6.58	-0.06	0.71
97	3.486	4.449	5.128	6.038	-6.46	0.01	0.60
98	3.674	3.685	5.177	5.883	-7.39	0.91	1.41
99	2.75	2.55	5.57	6.1	-7.88	2.37	1.84
100	3.123	3.109	4.283	4.825	-5.59	0.86	-0.16
101	4.059	4.434	6.065	3.8	-5.53	0.22	-0.21
102	3.91	4.549	4.828	5.129	-5.01	-0.37	-0.66
103	4.022	4.404	5.293	5.636	-6.14	-0.32	0.32
104	4.092	4.555	4.168	5.663	-5.4	-0.10	-0.32
105	3.23	3.535	5.667	5.031	-6.09	0.13	0.28
106	3.969	3.862	5.791	5.095	-6.66	0.40	0.78
107	2.703	3.227	5.427	6.08	-7.27	1.79	1.31
108	4.051	4.253	5.274	3.736	-5.00	-0.18	-0.67
109	4.21	4.819	4.926	4.415	-5.11	-0.27	-0.57
110	2.917	3.704	4.822	5.48	-6.4	1.24	0.55
111	3.964	4.391	4.332	5.174	-5.18	-0.22	-0.51
112	2.94	2.859	4.189	5.357	-5.85	0.87	0.07
113	2.774	3.77	4.963	5.487	-6.28	1.24	0.45
114	2.978	3.07	4.537	5.122	-6.27	1.24	0.44
115	3.152	3.676	4.696	5.079	-5.97	0.85	0.18
116	2.886	3.662	4.991	5.326	-6.45	1.07	0.59
117	2.958	2.977	5.719	4.227	-6.4	1.24	0.55
118	3.961	3.814	5.241	5.525	-7.07	0.64	1.14
119	3.23	2.615	4.744	5.844	-7.03	0.50	1.10
120	3.961	3.588	5.489	5.021	-6.86	0.45	0.95
121	2.54	3.273	5.504	5.936	-7.2	1.20	1.25
122	3.42	4.37	5.42	5.372	-6.06	0.01	0.25
123	3.228	2.735	5.118	5.445	-6.46	0.66	0.60
124	3.561	4.208	5.891	5.682	-4.15	-1.56	-1.41
125	3.631	3.604	5.417	5.253	-6.11	0.34	0.30
126	3.99	4.348	5.306	5.437	-6.87	0.71	0.96
127	3.141	3.251	5.605	5.609	-7.23	0.73	1.27
128	3.644	3.2	5.387	4.956	-5.8	-0.20	0.03
129	4.117	4.361	6.79	6.095	-4.76	-1.76	-0.88
130	4.241	5.023	4.847	5.733	-5.82	-0.09	0.05
131	3.516	3.727	5.902	5.005	-6.06	-0.12	0.25
132	2.702	2.775	5.128	5.34	-6.98	1.59	1.06

**Table 2.7** (Continued). Results for rescoring using Prime/MM-GBSA of haem-pdb with deoxy-artemisinin analogues as well as computed activity using  $\Delta G_{\text{bind}}$  energy as a descriptor.

Ligands	Fe-O1 distance(Å)	Fe-O13 distance(Å)	Fe-O11 distance(Å)	$\Delta G_{\text{bind}}$ kcal/mol	pIC <sub>50</sub> <sub>expt</sub>	pIC <sub>50</sub> <sub><math>\Delta G_{\text{bind}}</math></sub>
133	5.622	4.364	5.548	-8.51	-3.62	-3.63
134	4.451	6.029	4.969	-8.29	-3.63	-3.63
135	3.806	5.92	6.099	-4.85	-3.68	-3.69
136	4.811	6.963	6.441	-9.80	-3.60	-3.61
137	4.42	6.613	6.774	-8.13	-3.66	-3.64
138	4.274	6.521	6.215	-8.35	-3.62	-3.63
139	4.845	6.935	7.543	-7.16	-3.67	-3.65
140	5.263	3.58	4.761	-6.49	-3.66	-3.67
141	5.807	5.5	5.706	-9.73	-3.61	-3.61
142	4.378	5.747	4.565	-8.02	-3.66	-3.64
143	5.693	4.293	4.488	-8.45	-3.65	-3.63
144	4.565	5.947	4.712	-10.01	-3.59	-3.60

### 2.3.5 Building models for prediction of pIC<sub>50</sub> using Glide score and binding free energy

A prediction model of antimalarial activity (pIC<sub>50</sub>) was build based on Glide score and  $\Delta G_{\text{bind}}$  as descriptors. Antimalarial activity of all the 144 analogues used in the study was generally evaluated against the drug-resistant malarial strain *P. falciparum* (W-2 clone) and were collected from different sources (Acton et. al., 1993; Lin et. al., 1989; Posner et. al., 1992; Avery et. al., 1995; Avery et. al., 1996; Pinheiro et. al., 2001) and included in Table 2.1. It has been seen that the analogues having endoperoxide linkage have significantly better activity (pIC<sub>50</sub> in the range of -2.09 to 2.53) compared to the other deoxyartemisinin derivatives (pIC<sub>50</sub> between -3.59 to -3.68). The deoxartemisinin derivatives generally showed very weak or no activity. The plot of the Glide score and experimental pIC<sub>50</sub> reveal a significant relationship ( $R^2 = 0.763$  and  $R^2 = 0.734$  for both the QHS and DQHS derivatives) between these two parameters (Figure 2.7a & 2.7c). A linear regression model for prediction of predicted pIC<sub>50</sub> of antimalarial activity based on Glide score has been developed by considering analogues with known pIC<sub>50</sub>. The equations 3 & 4 of the model and the corresponding statistics for QHS & DQHS are shown below:

$$\text{pIC}_{50} = -1.88 (\pm 0.115) - 1.51 (\pm 0.074) * \text{G-score} \quad (3)$$

(N = 132,  $r^2 = 0.763$ ,  $r^2_{\text{cv}} = 0.762$ , s = 0.428, F = 419.66)

$$\text{pIC}_{50} = -3.94 (\pm 0.061) - 0.284 (\pm 0.057) * \text{G-score} \quad (4)$$

(N = 12,  $r^2 = 0.734$ ,  $r^2_{\text{cv}} = 0.685$ , s = 0.017, F = 24.92)

Reasonably good agreement between predicted and experimental pIC<sub>50</sub> are found (root mean square error = 0.36 and 0.01 for QHS and DQHS derivatives) and suggested that the calculated pIC<sub>50</sub> based on Glide score is robust and accurate. Similar prediction model of predicted pIC<sub>50</sub> of antimalarial activity has been developed by considering ΔG<sub>bind</sub> energy as a descriptor. The equations 5 & 6 of the model and the corresponding statistics for QHS and DQHS analogues are shown below:

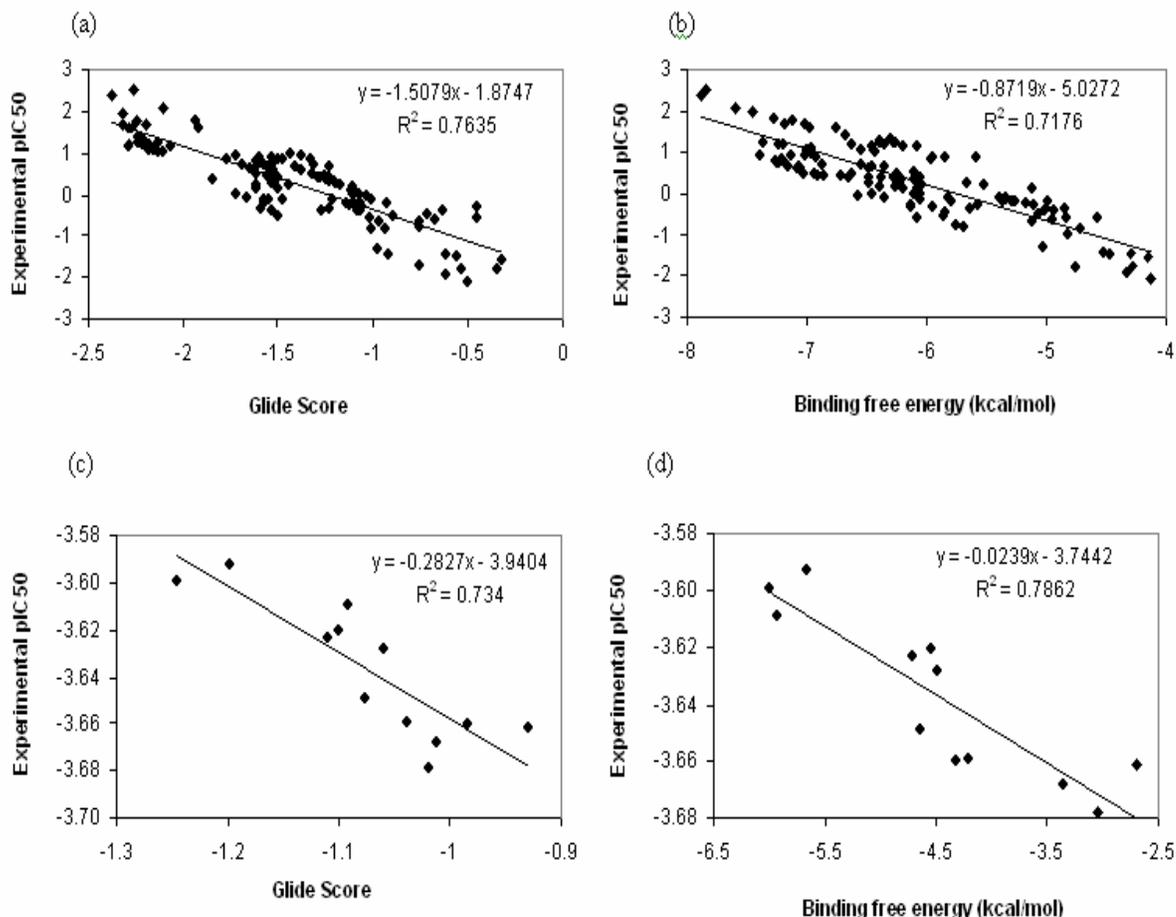
$$\text{pIC}_{50} = -5.03 (\pm 0.298) - 0.872 (\pm 0.048) * \Delta G_{\text{bind}} \quad (5)$$

$$(N = 132, r^2 = 0.718, r^2_{\text{cv}} = 0.715, s = 0.471, F = 330.22)$$

$$\text{pIC}_{50} = -3.75 (\pm 0.019) - 0.024 (\pm 0.004) * \Delta G_{\text{bind}} \quad (6)$$

$$(N = 12, r^2 = 0.786, r^2_{\text{cv}} = 0.739, s = 0.015, F = 32.20)$$

The ΔG<sub>bind</sub> energy value among the ligands of QHS library varies in between -7.88 to - 4.13 kcal/mol and the overall mean is -6.16 (± 0.859) kcal/mol. It revealed that all these ligands bind to haem iron with high affinity and showed activity (experimental pIC<sub>50</sub>) in between - 2.09 and 2.53. On the contrary for the DQHS derivatives the ΔG<sub>bind</sub> energy value varies between -5.66 to -2.69 kcal/mol; the experimental pIC<sub>50</sub> was -3.59 to -3.66. Correspondingly the plot of the binding free energy and experimental pIC<sub>50</sub> reveals a significant relationship ( $R^2 = 0.718$  and  $R^2 = 0.786$  for QHS and DQHS respectively) between these two parameters (Figure 2.7b & 2.7d). The calculated pIC<sub>50</sub> based on ΔG<sub>bind</sub> energy descriptor was in good agreement with experimental pIC<sub>50</sub> (root mean square error = 0.40 and 0.01 for QHS and DQHS derivatives) and suggested that the prediction model is robust and accurate.



**Figure 2.7(a-d).** Models for predicting antimalarial activity (pIC<sub>50</sub>) of the (a&b)Artemisinin (QHS) analogues and (c&d) Deoxy-artemisinin analogues based on Glide score and Binding free energy ( $\Delta G_{\text{bind}}$ ) as descriptor.

## 2.4. Conclusion

We have compiled a virtual library of artemisinin analogues which are built through structural modification of scaffold structure of natural artemisinin. Docking and rescoring using PRIME/MM-GBSA have been used in the work to get insights into artemisinin: haem interactions and development of prediction model for antimalarial activity. The docking result revealed that the haem-iron approaches the endoperoxide moiety at the O1 position in preference to the O2 position. Several sets of artemisinin analogues have been studied in the docking simulations. Results showed that these analogues bind in a very similar mode. The magnitude of the binding affinity can be a key factor that decides the activeness of an individual inhibitor. An energetic evaluation of the binding affinity will provide a way to estimate the activity of inhibitors. In any binding energy calculation, the correct binding

structure of each ligand has to be determined first prior to binding energy estimation. Very similar binding structures were obtained for a set of analogues. This makes a credible prediction model of the antimalarial activity ( $\text{pIC}_{50}$ ) calculation possible. The calculated Glide score and binding free energy value of a set of structural analogues demonstrate excellent linear correlation to the experimental antimalarial activity thus these models could be useful to predict the range of activity for new artemisinin analogues. We also found that refinement of poses and consequent rescoring using PRIME/MM-GBSA leads to better predictivity of  $\text{pIC}_{50}$ . The information that we have expressed in this study may lead to design (synthesis) of more potent artemisinin derivatives for inhibition of heme polymerization.

## **C**HAPTER 3

**Application of linear interaction energy method for binding affinity calculations of artemisinin analogues using continuum solvent model and prediction of antimalarial activity**

---

## **Abstract**

The antimalarial activity of artemisinin derived drugs appears to be mediated by an interaction of the drug's endoperoxide bridge with intraparasitic heme. The binding affinity of artemisinin analogues with heme were computed using a linear interaction energy (LIE) method with a surface generalized Born (SGB) continuum solvation model. A training set of 101 artemisinin analogues with known in vitro antimalarial activity was used to build the SGB-LIE model utilizing molecular dynamics (MD) and hybrid Monte Carlo (HMC) sampling techniques. For the test set of 57 compounds the SGB-LIE model was able to predict their activity with an overall root mean square (RMS) error of 0.348 and 0.415 kcal/mol respectively with respect to experimental data. Low levels of RMS error establish the structure-based LIE method as an efficient tool for generating more potent inhibitors of heme by testing rationally designed lead compounds based on artemisinin derivatization.

### 3.1 Introduction

Malaria is one of the most widespread and prevalent endemic diseases; it threatens approximately 40 percent of the world's population in more than 107 countries. This disease is estimated to cause approximately 350 to 500 million clinical illnesses and upto 3 million deaths each year (WHO, 2005). Most deaths are attributed to the parasite *Plasmodium falciparum*. The severity of the disease caused by this species results primarily from its ability to modify the surface of infected red blood cells by inserting proteins (Bowman, 1999). The enzymes in the parasite digestive vacuole (cysteine- and aspartic-proteinases) break down hemoglobin into amino acids and heme (Pandey, 1999). While all the amino acid content is used to build parasite proteins, only a small portion of the heme is incorporated into the parasite hemoproteins; the rest of the heme is detoxified (polymerized) by parasite enzymes (Kamchonwongpaisan, 1997). A number of drugs have been investigated for their use in the treatment of malaria (Olliaro, 2001; Ridley, 2002; White, 2004; Arav-Boger et. al., 2005). However, new strains of *Plasmodium falciparum* resistant to some of those drugs are causing substantial deterioration in clinical treatment (Olliaro, 2001; Ridley, 2002; White, 2004; Arav-Boger et. al., 2005). This has motivated the search for new antimalarial drugs that are effective against this form of malaria, thus having a very high priority in antimalarial drug design (Cheng et. al., 2002; Bhattacharjee et. al., 2004; Jefford, 2001). This led to Chinese researchers introducing a new compound, qinghaosu (or artemisinin, as it is known in the West), present in extracts of Qinghao or *Artemisia annua* L. that has been used in China for thousands of years (Haynes, 1997). It is a potent antimalarial drug against the multi-drug resistant strains of *Plasmodium falciparum* (Klayman, 1985; Luo et. al., 1987). The structure of artemisinin was identified as an endoperoxide containing sesquiterpene lactone and the presence of the 1,2,4-trioxane-ring system seems to be essential for its antimalarial activity (Bernardinelli et. al., 1994; Posner et. al., 1995; Haynes et. al., 1996; Rafiee et. al., 2005). Studies on the mode of action of artemisinin and its derivatives have shown that free heme could be the molecule targeted by artemisinin in biological systems and that  $\text{Fe}^{+2}$  ions interact with the peroxide when artemisinin react with heme (Cheng et. al., 2002; Jefford, 2001; Wu et. al., 1998, Meshnick, 2002; Haynes, 2004; Kannan, 2005). An initial step in the action of artemisinin includes heme-catalyzed artemisinin activation into a very reactive radical which binds covalently to the parasite proteins or heme (Olliaro, 2001; Jefford, 2001; Meshnick,

2002; Haynes et. al., 2004; Kannan et. al., 2005; Kamchonwongpaisan et. al., 1997; Hong et. al., 1994) and hemozoin (Kamchonwongpaisan et. al., 1997; Meshnick, 2002; Haynes et. al., 2004; Kannan et. al., 2005; Kamchonwongpaisan et. al., 1997; Hong et. al., 1994) It has been proposed that heme iron attacks the endoperoxide linkage of artemisinin either at the O1 (Jefford et. al., 1996) or O2 position (Posner et. al., 1995). In pathway A, heme iron attacks the compound at the O2 position and produces a free radical at the O1 position. Later it rearranges to form the C4 free radical. In pathway B, heme iron attacks the compound at the O1 position and produces a free radical at the O2 position. After that the C3–C4 bond is cleaved to give a carbon radical at C4. It has been suggested that the C4 free radical in both pathways is an important substance in antimalarial activity (Posner et. al., 1994).

The effectiveness of artemisinin and its derivatives as antimalarial drugs for the treatment of multi-drug resistant *P. falciparum* has received considerable attention in recent years. More often than not the focus of these studies has been to demonstrate antimalarial efficacy in vitro for new structural classes or modification of the natural product architecture. Since a wide variety of molecular scaffolds are available for optimization, this diversity presents a significant challenge in determining the essential features for activity. A rational approach for the discovery of a pharmaceutically acceptable, economically viable, peroxide based antimalarial awaits development of a global mechanism of action model for organic peroxides (Robert et. al., 1998; Pandey et. al., 1999) and/or a predictive quantitative structure-activity relationship. With the advent of parallel synthesis methods and technology, we might expect the number of antimalarial artemisinin analogues to be tested to grow dramatically. Combinatorial methods could also be envisioned as a semirational approach to this above discovery strategy. One method of orchestrating these strategies is to make use of linear interaction energy (LIE) models for the rapid prediction and virtual prescreening of antimalarial activity. The linear interaction energy approximation is a way of combining molecular mechanics calculations with experimental data to build a model scoring function for the evaluation of ligand-protein binding free energies. A linear interaction energy method for rational design of artemisinin analogues for inhibition of heme polymerization has not yet been determined.

The availability of X-ray structure of heme facilitates understanding the structure-activity relationships (SAR) for heme polymerization and enables molecular modeling techniques to be applied for designing novel and more potent inhibitors. In this study, we have applied a structure-based linear interaction energy method implementing a surface generalized Born (SGB-LIE) (Zhou et. al., 2001) continuum model for solvation to build a binding affinity model for estimating the free energy of binding for a diverse set of heme inhibitors. The LIE method (Aqvist et. al., 1994; Aqvist et. al., 2001) has been applied to a number of protein-ligand systems with promising results (Tominaga et. al., 2004; Leiros et. al., 2004; Ostrovsky et. al., 2003) producing small errors on the order of 1 kcal/mol for free energy prediction (van Lipzig et. al., 2003). The magnitude of free energy changes upon binding of inhibitors to heme directly correlates with the experimental potency of these inhibitors. Hence, fast and accurate estimation of binding free energies provides a means to screen the compound libraries for lead optimization and rational design. This could bring about the development of new and more effective drugs.

## **3.2. Materials and methods**

### **3.2.1 LIE Methodology**

The LIE method employs experimental data on binding free energy values for a set of ligands (referred as training set) to estimate the binding affinities for a set of novel compounds. The method is based on the linear response approximation (LRA), which dictates that binding free energy of a protein-ligand system is a function of polar and nonpolar energy components that scale linearly with the electrostatic and van der Waals interactions between a ligand and its environment. The free energy of binding for the complex is derived from two states: (1) free ligand in the solvent and (2) ligand bound to the solvated protein. The conformational changes and entropic effects pertaining to unbound receptor are taken into account implicitly and only interactions between the ligand and either the protein or solvent are computed during molecular mechanics calculations. Among the various formulations of the LIE methodology developed in the past, the SGB-LIE method implementing a surface generalized Born (SGB) model for the solvation has been shown to be 1 order of magnitude faster than the methods based on explicit solvent (Zhou et. al., 2001) with the same order of accuracy. The SGB-LIE method also offers better accuracy in treating the long-range

electrostatic interactions. The SGB-LIE method implements the original formulation proposed by Jorgensen (Carlson et. al., 1995) for the case of continuum solvent replacing the solvent accessible surface area term by a cavity term as follows

$$\Delta G = \alpha(\langle U_{vdw}^b \rangle - \langle U_{vdw}^f \rangle) + \beta(\langle U_{ele}^b \rangle - \langle U_{ele}^f \rangle) + \gamma(\langle U_{cav}^b \rangle - \langle U_{cav}^f \rangle) \quad (1)$$

where bracketed terms represent the ensemble average of the energy terms, such as van der Waals ( $U_{vdw}$ ), electrostatic ( $U_{ele}$ ), or cavity ( $U_{cav}$ ) energy. The energy terms involved can be computed using energy minimization, molecular dynamics, or Monte Carlo calculations. All the terms are evaluated for interaction between ligand, both in the free (f) and bound (b) state and its environment. The  $\alpha$ ,  $\beta$  and  $\gamma$  are LIE fitting parameters. The transferability and dependence of LIE parameters on force fields and protein-ligand system are still the subject of debate. In the Jorgensen formulation, LIE parameters are free coefficients that need to be determined by fitting the experimental data on the training set compounds. In the SGB model of solvation, there is no explicit van der Waals or electrostatic interaction between the solute and solvent. The contribution for net free energy of solvation comes from two energy terms, namely, reaction field energy ( $U_{rxn}$ ) and cavity energy ( $U_{cav}$ ):

$$U_{SGB} = U_{rxn} + U_{cav}$$

The cavity and reaction field energy terms implicitly take into account the van der Waals and the electrostatic interactions, respectively, between the ligand and solvent. The application of the SGB-LIE method for a given protein-ligand system essentially involves computing four energy components, i.e., the van der Waals and Coulombic energy between the ligand and protein and the reaction field and cavity energy between the ligand and continuum solvent. The total electrostatic energy in the SGB-LIE method is the sum of Coulombic and reaction field energy terms.

### 3.2.2 Computational details

All the computations and molecular modeling were carried out using Schrodinger package from Schrodinger Inc (Schrodinger Inc.: Portland, 2004). All the calculations for the SGB-LIE method were performed in the Liaison package. The Liaison module performs LIE

calculations in the OPLS force field with a residue-based cut off of 15 Å. The OPLS force field was also used for charge assignment and all energy calculations.

### 3.2.3. Receptor preparation

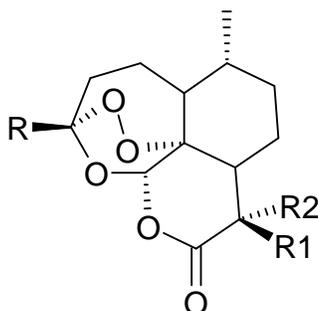
The X-ray structure of heme-pdb was taken from the Protein Data Bank (PDB ID: 1CTJ) and has been used as initial structure in the preparation of heme receptor site. Heme is a planar molecule with a strong positive charge on its central iron atom, which lies slightly above the porphyrin plane. Charge on the iron was assigned as +2 but the structure was kept the same. Hydrogens were added to the model automatically via the Maestro interface leaving no lone pair and using an explicit all-atom model. The multi step Schrodinger's protein preparation tool (PPrep) has been used for final preparation of receptor model. The complex structure was energy minimized using OPLS\_2005 force field and the conjugate gradient algorithm, keeping all atoms except hydrogen fixed. The minimization was stopped either after 1000 steps or after the energy gradient converged below 0.01 KJ/mol. Complete geometry optimization was carried out using LACVP\*\* (Hay et. al.,1985) for the iron atoms, followed by single-point calculations using LACVP\*\* for the iron atom. An unrestricted density functional theory (DFT) was employed to model effectively the open shell orbitals on the two iron atoms. The Jaguar suite of ab initio quantum chemical program (Jaguar, version 4.1: Schrodinger, Inc.: Portland, OR, 2000.) was used to carry out all quantum mechanics (QM) calculations.

### 3.2.4. Preparation of ligands

An initial dataset of 158 artemisinin analogues were collected from published data (Woolfrey et. al., 1998; Acton et. al., 1993; Lin et. al., 1989; Posner et. al., 1992; Avery et. al., 1995; Avery et. al., 1996) in which several different ring systems were represented. All of the analogues were either peroxides or trioxanes, which should act via similar mechanisms of action and were categorized into 10 classes (Table 3.1). Each of these compounds had associated in vitro bioactivity values (IC<sub>50</sub> values reported in ng/ml) against the drug resistant malaria strain *P. falciparum* (W-2 clone). The log value of the relative activity (RA) of these compounds was used for analysis and was defined as:

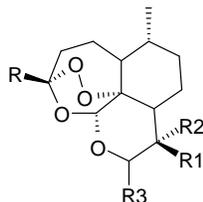
$$\text{Log(RA)} = \log[(\text{artemisinin IC}_{50}/\text{analogue IC}_{50})(\text{analogue MW}/\text{artemisinin MW})].$$

**Table 3.1.** Artemisinin analogues with antimalarial activities against the drug-resistant malarial strain *P. falciparum* (W-2 clone) used in the work.



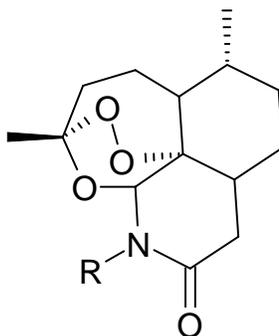
Compounds	R	R1	R2	log (RA)	IC <sub>50</sub> (ng/ml)
Training set					
1	CH <sub>3</sub>	CH <sub>3</sub>	H	1.00	0.040
2	C <sub>4</sub> H <sub>8</sub> Ph	H	H	0.45	0.194
3	CH <sub>3</sub>	H	2-Z-Butenyl	-1.10	5.750
4	CH <sub>3</sub>	H	H	0.79	0.065
5	CH <sub>3</sub>	H	2-E-Butenyl	-0.60	1.818
6	CH <sub>3</sub>	Allyl	H	-0.10	0.550
7	CH <sub>3</sub>	C <sub>4</sub> H <sub>9</sub>	H	0.17	0.311
8	C <sub>4</sub> H <sub>8</sub> Ph	C <sub>4</sub> H <sub>9</sub>	H	-0.32	1.310
9	CH <sub>2</sub> CH <sub>2</sub> CO <sub>2</sub> Et	C <sub>4</sub> H <sub>9</sub>	H	1.36	0.025
10	C <sub>4</sub> H <sub>9</sub>	C <sub>4</sub> H <sub>9</sub>	H	-0.48	1.568
11	CH <sub>3</sub>	C <sub>2</sub> H <sub>5</sub>	H	1.40	0.017
12	CH <sub>3</sub>	C <sub>6</sub> H <sub>13</sub>	H	0.86	0.069
13	CH <sub>3</sub>	i-C <sub>6</sub> H <sub>13</sub>	H	-0.04	0.547
14	CH <sub>3</sub>	i-C <sub>5</sub> H <sub>11</sub>	H	0.07	0.408
15	C <sub>3</sub> H <sub>6</sub> (p-Cl-Ph)	H	H	0.10	0.457
16	C <sub>4</sub> H <sub>9</sub>	H	H	-0.74	2.416
17	CH <sub>2</sub> CH <sub>2</sub> CO <sub>2</sub> Et	H	H	0.37	0.214
18	CH <sub>3</sub>	C <sub>3</sub> H <sub>6</sub> (p-Cl-Ph)	H	1.37	0.025
Test set					
19	CH <sub>3</sub>	Br	CH <sub>2</sub> Br	-1.64	27.244
20	CH <sub>3</sub>	=CH <sub>2</sub>	-	-0.89	3.083
21	CH <sub>3</sub>	CH <sub>2</sub> CH <sub>3</sub>	-	-0.36	1.053
22	CH <sub>3</sub>	-CH <sub>2</sub> CH <sub>2</sub> -	-	-0.94	3.632
23	CH <sub>3</sub>	C <sub>5</sub> H <sub>11</sub>	H	1.02	0.046
24	CH <sub>3</sub>	C <sub>4</sub> H <sub>8</sub> Ph	H	0.63	0.133
25	CH <sub>3</sub>	C <sub>2</sub> H <sub>4</sub> Ph	H	0.12	0.400
26	CH <sub>3</sub>	C <sub>3</sub> H <sub>7</sub>	H	1.13	0.033

**Table 3.1(continued).** 10-Substituted artemisinin derivatives with antimalarial activities against the drug-resistant malarial strain *P. falciparum* (W-2 clone) used in the work.



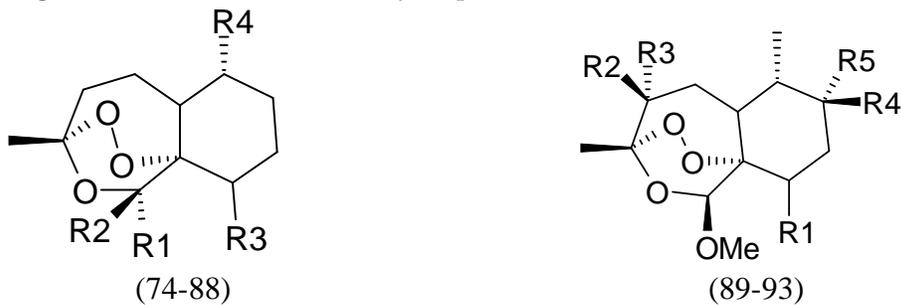
Compounds	R	R1	R2	R3	log RA	IC <sub>50</sub> (ng/ml)
Training set						
27	CH <sub>3</sub>	CH <sub>3</sub>	H	H	0.75	0.068
28	CH <sub>3</sub>	CH <sub>3</sub>	H	OH	0.55	0.114
29	CH <sub>3</sub>	CH <sub>3</sub>	H	OEt	0.34	0.202
30	CH <sub>3</sub>	CH <sub>3</sub>	H	OH	0.96	0.051
31	CH <sub>3</sub>	H	Br	H	0.28	0.248
32	CH <sub>3</sub>	CH <sub>3</sub>	Br	NH-2-(1,3-thiazole)	0.66	0.134
33	CH <sub>3</sub>	CH <sub>3</sub>	Br	p-Cl-aniline	0.79	0.105
34	CH <sub>3</sub>	CH <sub>3</sub>	Br	aniline	0.18	0.397
35	CH <sub>3</sub>	Br	CH <sub>3</sub>	NH-2-pyridine	-0.09	0.768
36	CH <sub>3</sub>	CH <sub>3</sub>	Br	NH-2-pyridine	-0.77	3.667
37	CH <sub>3</sub>	CH <sub>3</sub>	H	$\alpha$ -OEt	0.32	0.212
38	CH <sub>3</sub>	C <sub>4</sub> H <sub>9</sub>	H	H	1.32	0.021
39	CH <sub>3</sub>	C <sub>2</sub> H <sub>5</sub>	H	H	0.67	0.086
40	CH <sub>3</sub>	C <sub>3</sub> H <sub>7</sub>	H	OEt	-0.04	0.529
41	CH <sub>3</sub>	H	H	OEt	0.43	0.157
42	CH <sub>3</sub>	CH <sub>3</sub>	H	C <sub>3</sub> H <sub>6</sub> OH	0.78	0.077
43	CH <sub>3</sub>	CH <sub>3</sub>	H	C <sub>4</sub> H <sub>9</sub>	0.06	0.400
44	CH <sub>3</sub>	CH <sub>3</sub>	H	OCH <sub>2</sub> CO <sub>2</sub> Et	0.52	0.158
45	CH <sub>3</sub>	CH <sub>3</sub>	H	OC <sub>2</sub> H <sub>4</sub> CO <sub>2</sub> Me	0.10	0.433
46	CH <sub>3</sub>	CH <sub>3</sub>	H	OC <sub>3</sub> H <sub>6</sub> CO <sub>2</sub> Me	-0.03	0.605
47	CH <sub>3</sub>	CH <sub>3</sub>	H	OCH <sub>2</sub> (4-PhCO <sub>2</sub> Me)	-0.07	0.720
48	CH <sub>3</sub>	CH <sub>3</sub>	H	(R)-OCH <sub>2</sub> CH(CH <sub>3</sub> )CO <sub>2</sub> Me	1.79	0.009
49	CH <sub>3</sub>	CH <sub>3</sub>	H	(S)-OCH <sub>2</sub> CH(CH <sub>3</sub> )CO <sub>2</sub> Me	2.25	0.003
50	CH <sub>3</sub>	CH <sub>3</sub>	H	(R)-OCH(CH <sub>3</sub> )CH <sub>2</sub> CO <sub>2</sub> Me	0.87	0.073
51	CH <sub>3</sub>	CH <sub>3</sub>	H	(S)-OCH(CH <sub>3</sub> )CH <sub>2</sub> CO <sub>2</sub> Me	1.70	0.011
52	CH <sub>2</sub> CH <sub>2</sub> CO <sub>2</sub> Et	H	H	H	0.70	0.096
53	C <sub>4</sub> H <sub>9</sub>	H	H	H	0.75	0.075
Test Set						
54	C <sub>4</sub> H <sub>8</sub> Ph	H	H	H	0.58	0.139
55	CH <sub>3</sub>	-OCH <sub>2</sub> -	-	OOH	-0.62	1.857
56	CH <sub>3</sub>	-CH <sub>2</sub> O-	-	OOH	-0.57	1.655
57	CH <sub>3</sub>	=CH <sub>2</sub>	-	OOH	-0.99	4.131
58	CH <sub>3</sub>	C <sub>5</sub> H <sub>11</sub>	H	H	0.16	0.318
59	CH <sub>3</sub>	C <sub>3</sub> H <sub>6</sub> Ph	H	H	1.40	0.021
60	CH <sub>3</sub>	C <sub>3</sub> H <sub>7</sub>	H	H	0.74	0.076
61	CH <sub>3</sub>	CH <sub>3</sub>	H	OOt-C <sub>4</sub> H <sub>9</sub>	0.92	0.061
62	-	CH <sub>3</sub>	OH	$\alpha$ -OH	-0.89	3.303
63	-	CH <sub>3</sub>	H	CH <sub>2</sub> CHF <sub>2</sub>	0.11	0.366
64	-	CH <sub>3</sub>	OH	OCH <sub>2</sub> CF <sub>3</sub>	0.33	0.243
65	-	CH <sub>3</sub>	OH	OEt	-0.44	1.281

**Table 3.1(continued).** 11-Aza-artemisinin derivatives with antimalarial activities against the drug-resistant malarial strain *P. falciparum* (W-2 clone) used in the work.



Compounds	R	log RA	IC <sub>50</sub> (ng/ml)
Training set			
66	C <sub>3</sub> H <sub>6</sub> Ph	0.02	0.522
67	C <sub>2</sub> H <sub>4</sub> Ph	0.16	0.364
68	C <sub>5</sub> H <sub>11</sub>	-0.20	0.758
69	i-C <sub>5</sub> H <sub>11</sub>	-0.04	0.524
70	CH <sub>2</sub> (p-Cl-Ph)	-0.16	0.802
Test set			
71	CH <sub>2</sub> Ph	0.34	0.231
72	CH <sub>2</sub> -(2-C <sub>5</sub> H <sub>4</sub> N)	1.46	0.018
73	Acetaldehyde	1.47	0.015

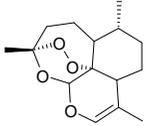
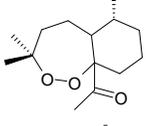
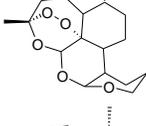
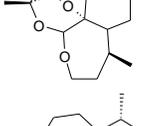
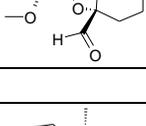
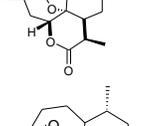
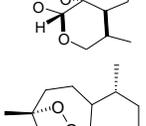
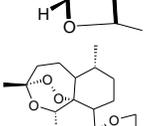
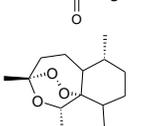
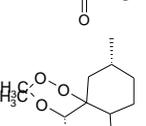
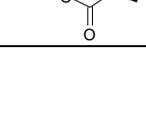
**Table 3.1(continued).** Artemisinin derivatives lacking the D-ring with antimalarial activity against the drug-resistant malarial strain *P. falciparum* (W-2 clone) used in the work.



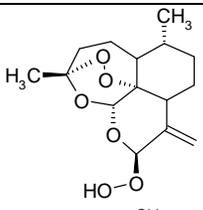
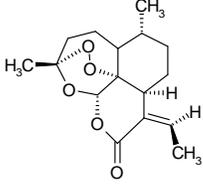
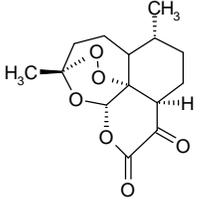
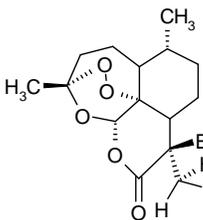
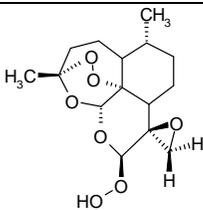
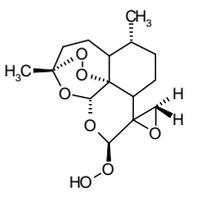
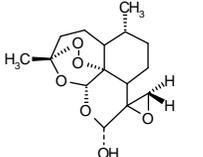
Compounds	R1	R2	R3	R4	log (RA)	IC <sub>50</sub> (ng/ml)
Training set						
74	-O <sub>2</sub> CCH <sub>2</sub> Ph	H	H	CH <sub>3</sub>	-0.51	1.648
75	H	H	H	CH <sub>3</sub>	-0.32	0.628
76	H	OCH <sub>3</sub>	H	H	-0.31	0.660
77	OCH <sub>2</sub> Ph	H	H	H	-0.09	0.530
78	OCH <sub>3</sub>	H	C <sub>2</sub> H <sub>4</sub> O <sub>2</sub> CNEt	H	-0.65	0.118
79	H	OCH <sub>3</sub>	C <sub>2</sub> H <sub>4</sub> OCH <sub>3</sub>	H	-0.39	0.996
80	H	OCH <sub>3</sub>	C <sub>2</sub> H <sub>4</sub> OCH <sub>2</sub> Ph	H	0.75	0.091
81	H	OCH <sub>3</sub>	C <sub>2</sub> H <sub>4</sub> O-allyl	H	0.40	0.184
82	H	OCH <sub>3</sub>	C <sub>2</sub> H <sub>4</sub> O <sub>2</sub> Ph	H	-0.59	2.086
83	H	OCH <sub>3</sub>	C <sub>2</sub> H <sub>4</sub> O <sub>2</sub> C(4-PhCO <sub>2</sub> Me)	H	0.27	0.343
84	H	OCH <sub>3</sub>	C <sub>2</sub> H <sub>4</sub> O <sub>2</sub> C(4-PhCO <sub>2</sub> H)	H	-0.81	3.856
85	H	OCH <sub>3</sub>	C <sub>2</sub> H <sub>4</sub> O <sub>2</sub> C(4-PhCONEt <sub>2</sub> )	H	0.230	0.398
86	H	OCH <sub>3</sub>	C <sub>2</sub> H <sub>4</sub> O <sub>2</sub> C(4-PhCO <sub>2</sub> C <sub>2</sub> H <sub>4</sub> NMe <sub>2</sub> )	H	-0.600	2.790
Test Set						
87	H	OCH <sub>3</sub>	C <sub>2</sub> H <sub>4</sub> O <sub>2</sub> CCH <sub>2</sub> NCO <sub>2</sub> -(t-C <sub>4</sub> H <sub>9</sub> )	H	-0.04	0.670
88	H	OCH <sub>3</sub>	C <sub>2</sub> H <sub>4</sub> OCH <sub>2</sub> (4-N-Me-pyridine)	H	-0.90	4.439

Compounds	R1	R2	R3	R4	R5	Log(RA)	IC <sub>50</sub> (ng/ml)
Test set							
89	C <sub>2</sub> H <sub>4</sub> OH	H	CH <sub>3</sub>	H	H	-1.80	26.849
90	C <sub>2</sub> H <sub>4</sub> OH	CH <sub>3</sub>	H	H	H	0.23	0.251
91	C <sub>2</sub> H <sub>4</sub> OH	CH <sub>3</sub>	CH <sub>3</sub>	H	H	-1.80	28.102
92	C <sub>2</sub> H <sub>4</sub> OCH <sub>2</sub> Ph	CH <sub>3</sub>	CH <sub>3</sub>	H	H	-1.80	36.157
93	C <sub>2</sub> H <sub>4</sub> OCH <sub>2</sub> (4-py)	-	-	-	-	0.14	0.373

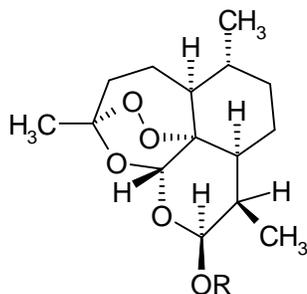
**Table 3.1(continued).** Miscellaneous Artemisinin derivatives with antimalarial activity against the drug-resistant malarial strain *P. falciparum* (W-2 clone) used in the work.

Compounds	structure	Log(RA)	IC <sub>50</sub> (ng/ml)
Training set			
94		0.78	0.063
95		-4.00	6.339
96		0.23	0.259
97		-1.20	6.340
98		-3.30	684.899
Test set			
99		-0.96	3.622
100		-0.79	2.344
101		-0.64	1.573
102		-2.09	56.889
103		-2.49	123.612
104		-0.80	2.309

**Table 3.1(continued).** 9-Substituted Artemisinin derivatives with antimalarial activity against the drug-resistant malarial strain *P. falciparum* (W-2 clone) used in the work.

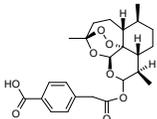
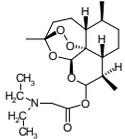
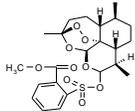
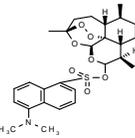
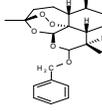
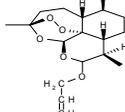
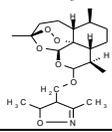
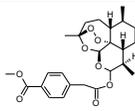
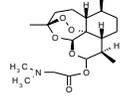
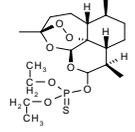
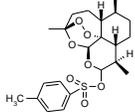
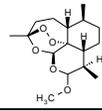
Compounds	Structure	Log (RA)	IC <sub>50</sub> (ng/ml)
Training set			
105		-0.739	2.320
106		-0.197	0.657
107		-2.298	79.429
108		-1.487	19.143
Test set			
109		-0.460	1.286
110		-0.409	1.143
111		-0.361	0.971

**Table 3.1(continued).** Dihydroartemisinin derivatives with antimalarial activity against the drug-resistant malarial strain *P. falciparum* (W-2 clone) used in the work.

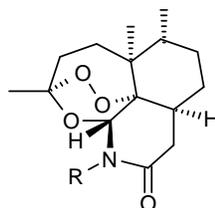


Compounds	R	Log (RA)	IC <sub>50</sub> (ng/ml)
Training set			
112	OR = H	0.487	0.123
113	(S)-CH <sub>2</sub> CH(CH <sub>3</sub> )COOCH <sub>3</sub>	2.104	0.004
114	(S)-CH(CH <sub>3</sub> )CH <sub>2</sub> COOCH <sub>3</sub>	0.599	0.137
115	(R)-CH(CH <sub>3</sub> )CH <sub>2</sub> COOCH <sub>3</sub>	1.429	0.020
116	1-adamantylmethyl	0.007	0.603
117	(S)-CH <sub>2</sub> CH(CH <sub>3</sub> )COOH	-0.658	2.380
118	(S)-CH(CH <sub>3</sub> )CH <sub>2</sub> COOH	-0.608	2.123
119	(R)-CH(CH <sub>3</sub> )CH <sub>2</sub> COOH	-0.383	1.263
Test set			
120	OR = O	-0.269	0.743
121	CH <sub>2</sub> PhCOOH	0.176	0.394
122	(R)-CH <sub>2</sub> CH(CH <sub>3</sub> )COOCH <sub>3</sub>	1.524	0.016
123	(R)-CH <sub>2</sub> CH(CH <sub>3</sub> )COOH	-0.463	1.520

**Table 3.1(continued).** Tricyclic 1.2.4 – Trioxanes derivatives with antimalarial activity against the drug-resistant malarial strain *P. falciparum* (W-2 clone) used in the work.

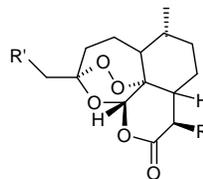
Compounds	structure	log(RA)	IC <sub>50</sub> (ng/ml)
<b>Training set</b>			
124		-0.475	1.886
125		0.995	0.057
126		-0.413	1.771
127		0.632	0.171
128		0.968	0.057
129		0.905	0.057
130		0.991	0.057
<b>Test set</b>			
131		0.660	0.143
132		0.787	0.086
133		0.717	0.057
134		0.434	0.229
135		0.129	0.314

**Table 3.1(continued).** N-Alkyl-11-aza-9-desmethylartemisinin derivatives with antimalarial activity against the drug-resistant malarial strain *P. falciparum* (W-2 clone) used in the work.



Compounds	R	Log(RA)	IC <sub>50</sub> (ng/ml)
Training set			
136	-	0.328	0.400
137	C <sub>5</sub> H <sub>11</sub> (n)	0.041	0.435
138	C <sub>5</sub> H <sub>11</sub> (i)	0.173	0.321
139	(CH <sub>3</sub> ) <sub>2</sub> NCH <sub>2</sub> CH <sub>2</sub>	-0.432	1.300
140	HO <sub>2</sub> C(CH <sub>2</sub> ) <sub>5</sub>	-0.921	4.492
Test set			
141	C <sub>6</sub> H <sub>5</sub> CH <sub>2</sub>	0.276	0.268
142	p-ClC <sub>6</sub> H <sub>4</sub> CH <sub>2</sub>	0.045	0.500
143	C <sub>6</sub> H <sub>5</sub> (CH <sub>2</sub> ) <sub>2</sub>	0.294	0.267
144	C <sub>6</sub> H <sub>5</sub> (CH <sub>2</sub> ) <sub>3</sub>	0.312	0.266

**Table 3.1(continued).** 3C- substituted artemisinin derivatives with antimalarial activity against the drug-resistant malarial strain *P. falciparum* (W-2 clone) used in the work.



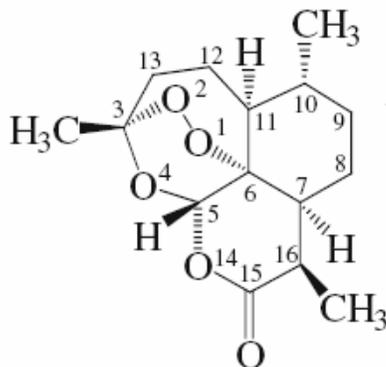
Compounds	R1	R	LogRA	IC <sub>50</sub> (ng/ml)
Training set				
145	CH <sub>3</sub>	H	0.049	0.357
146	CH <sub>3</sub> CH <sub>2</sub>	H	0.828	0.062
147	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>2</sub>	H	-0.745	2.427
148	CH <sub>3</sub> CH	H	-0.347	0.977
149	EtO <sub>2</sub> CCH <sub>2</sub>	H	0.365	0.216
150	C <sub>6</sub> H <sub>5</sub> CH <sub>2</sub>	H	-2.000	50.780
151	p-ClC <sub>6</sub> H <sub>4</sub> (CH <sub>2</sub> ) <sub>2</sub>	H	0.104	0.453
152	C <sub>6</sub> H <sub>5</sub> (CH <sub>2</sub> ) <sub>3</sub>	H	0.449	0.195
153	CH <sub>3</sub>	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>3</sub>	0.410	0.187
Test set				
154	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>2</sub>	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>3</sub>	-0.481	1.573
155	C <sub>6</sub> H <sub>5</sub> CH <sub>2</sub>	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>3</sub>	-2.000	58.723
156	p-ClC <sub>6</sub> H <sub>4</sub> (CH <sub>2</sub> ) <sub>2</sub>	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>3</sub>	-0.276	1.239
157	C <sub>6</sub> H <sub>5</sub> (CH <sub>2</sub> ) <sub>3</sub>	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>3</sub>	-0.319	1.306
158	EtO <sub>2</sub> CCH <sub>2</sub>	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>3</sub>	1.359	0.025

Molecular models of the artemisinin and its analogues (Table 3.1) were built using the Builder feature in Maestro (Schrodinger package) and energy minimized in a vacuum using Impact. Each structure was assigned an appropriate bond order using ligprep script shipped by Schrödinger and optimized initially by means of the OPLS 2005 force field using default setting. Complete geometrical optimization of these structures was carried out with the HF/3-21G method (in this work) using the Jaguar (Schrodinger Inc.). In order to check the reliability of the geometry obtained, we compared the structural parameters of the artemisinin 1,2,4-trioxane ring with theoretical (Pinheiro et. al., 2001) and experimental (Leban et. al., 1988; Lisgarten et. al., 1998) values from the literature. All calculations reproduced most of the structural parameters of the artemisinin 1,2,4-trioxane ring seen in X-ray structures (Table 3.2). This applies especially to the bond length of the endoperoxide bridge which seems to be responsible for the antimalarial activity (Bernardinelli et. al., 1994; Posner et. al., 1995; Posner et. al., 1995; Haynes et. al., 1996; Rafiee et. al., 2005 ).

**Table 3.2.** Experimental and theoretical values of the 1,2,4-trioxane ring parameters in artemisinin (bond lengths in Å; bond angles and torsional angles in degrees). The atom number are with respect to Figure 3.1.

Parameters <sup>a</sup>	Theoretical			Experimental <sup>d</sup>	Experimental <sup>e</sup>
	3-21G <sup>b</sup>	3-21G** <sup>c</sup>	6-31G <sup>c</sup>		
O1-O2	1.463	1.462	1.447	1.475(4)	1.469(2)
O2-C3	1.441	1.440	1.435	1.417(4)	1.416(3)
C3-O4	1.436	1.436	1.435	1.448(4)	1.445(2)
O4-C5	1.407	1.408	1.403	1.388(4)	1.379(2)
C5-C6	1.529	1.530	1.533	1.528(5)	1.523(2)
C6-O1	1.478	1.477	1.469	1.450(4)	1.461(2)
O1-O2-C3	106.9	107.070	108.800	107.600(2)	108.100(1)
O2-C3-O4	107.0	107.310	106.760	107.200(2)	106.600(2)
C3-O4-C5	115.6	115.700	117.300	113.500(3)	114.200(2)
O4-C5-C6	112.0	112.030	112.280	114.700(2)	114.500(2)
C5-C6-O1	111.1	111.589	110.910	111.100(2)	110.700(2)
C6-O1-O2	111.2	111.286	113.240	111.500(2)	111.200(2)
O1-O2-C3-O4	-74.9	-74.680	-71.840	-75.500(3)	-75.500(2)
O2-C3-O4-C5	31.8	32.150	33.390	36.300(4)	36.000(2)
C3-O4-C5-C6	29.4	28.400	25.320	24.800(4)	25.300(2)
O4-C5-C6-O1	-51.8	-50.769	-49.410	-50.800(4)	-51.300(2)
C5-C6-O1-O2	10.1	9.792	12.510	12.300(3)	12.700(2)
C6-O1-O2-C3	50.8	50.522	46.700	47.700	47.800(2)

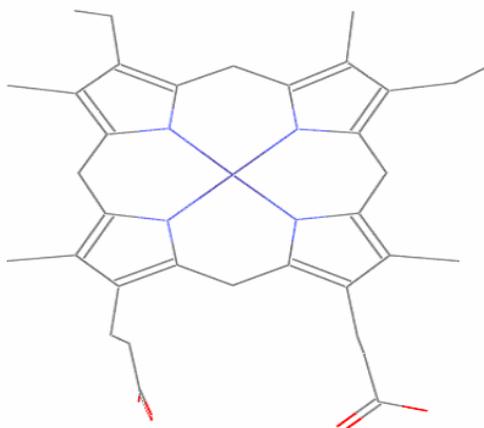
<sup>a</sup> Atoms are numbered according to Figure 3.1; <sup>b</sup> This work; <sup>c</sup> Values from Ref. (Pinheiro et. al., 2001); <sup>d</sup> Values from Ref. (Leban et. al., 1988) (experimental estimated standard deviations in brackets); <sup>e</sup> Values from Ref. (Lisgarten et. al., 1998) (experimental estimated standard deviations in brackets).



**Figure 3.1.** Stereochemistry and atomic numbering scheme of artemisinin.

### 3.2.4. Docking of the ligands

All the ligands were docked to the heme receptor using Glide. After ensuring that protein and ligands are in correct form for docking, the receptor-grid files were generated using grid-receptor generation program, using van der Waals scaling of the receptor at 0.4. The default size was used for the bounding and enclosing boxes. The grid box was generated at the centroid of the heme receptor (Figure 3.2). The ligands were docked initially using the “standard precision” method and further refined using “xtra precision” Glide algorithm. For the ligand docking stage, van der Waals scaling of the ligand was set at 0.5. Of the 50,000 poses that were sampled, 4,000 were taken through minimization (conjugate gradients 1,000) and the 30 structures having the lowest energy conformations were further evaluated for the favorable Glide docking score. A single best conformation for each ligand was considered for further analysis.



**Figure 3.2.** The structure of the heme compound.

### 3.2.5. LIE calculations

The docked complex corresponding to each analogue was transported to the LIAISON package for subsequent SGB-LIE calculations. Two sampling techniques, molecular dynamics (MD) and hybrid Monte Carlo (HMC), have been used for LIE conformation space sampling in the present work. A conjugate gradient minimization was performed first, starting from the initial docked structures and then a 15 ps MD equilibration is followed with temperature smoothly increasing from 0 to 310 K by velocity scaling and resampling. Finally, a 25 ps MD simulation was run for the SGB-LIE data collection. A residue-based cut off of 12 Å was set for the non-bonding interactions. The non-bonded pair list was updated every 10 fs. The time integration step of 1.0 fs and sampling of LIE energies in every 10 steps was used. Similarly, the average LIE energies for the ligand were obtained using the OPLS-2005 force field. The average LIE energy terms were used for building binding affinity model and free energy estimation for artemisinin analogues. The  $\alpha$ ,  $\beta$  and  $\gamma$  LIE fitting parameters were determined basing on Gaussian elimination method using Matlab 6.5 and by fitting the experimental data on the training set compounds.

In order to explore the reliability of the proposed model the cross validation method has been used. Prediction error sum of squares (*PRESS*) is a standard index to measure the accuracy of a modeling method based on the cross validation technique. The cross validation analysis performed by using the leave one out (LOO) method in which one compound removed from the data set and its activity predicted using the model derived from the rest of the data points. The cross-validated correlation coefficient ( $q^2$ ) that resulted in optimum number of components and lowest standard error of prediction were considered for further analysis and calculated using following equations:

$$q^2 = 1 - \frac{\sum_y (y_{pred} - y_{observed})^2}{\sum_y (y_{observed} - y_{mean})^2}$$
$$PRESS = \sum_y (y_{predicted} - y_{observed})^2$$

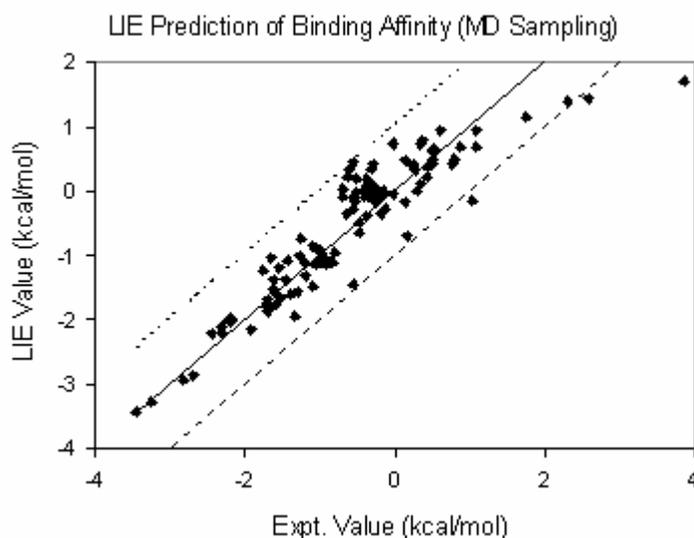
where  $y_{pred}$ ,  $y_{observed}$  and  $y_{mean}$  are the predicted actual and mean values of the inhibitory activities of the artemisinin analogues and *PRESS* is the sum of the predictive sum of squares.

### 3.3 Results and Discussions

We applied the SGB-LIE method to a training set of 101 artemisinin analogues to build a binding affinity model that was then used to compute the absolute free energy of binding and relative activity (RA) for a test set of 57 analogues. The training set used for building the binding affinity model comprised ten subsets of artemisinin analogues (Table 3.1). These compounds were taken from various sources; (Woolfrey et. al., 1998; Acton et. al., 1993; Lin et. al., 1989; Posner et. al., 1992; Avery et. al., 1995; Avery et. al., 1996) among these are endoperoxide artemisinin analogues, deoxy-artemisinin analogues, 10-substituted artemisinin derivatives, 3-substituted artemisinin derivatives, 9-substituted artemisinin derivatives, 11-Aza-artemisinin derivatives, artemisinin derivatives without D-ring and miscellaneous artemisinin derivatives. The experimental relative activity (RA) values for all those compounds in the training set were calculated against the drug resistant malarial strain *P. falciparum* (W-2 clone). The IC<sub>50</sub> value of these analogues was derived as mentioned in Table 3.1 and used for calculation of absolute  $\Delta G_{\text{binding}}$  energy. With the wide range of difference in IC<sub>50</sub> values and the large diversity in the structures, the combined set of 101 ligands is ideal to be considered as a training set as the set does not suffer from bias due to the similarity of the structures. Also, the training set of 101 analogues has enough data points not to suffer from over parameterization by the LIE model. Training set compounds were docked into the heme receptor site and the SGB-LIE calculations were performed using the LIAISON module. The simulations were performed both for the ligand-free and ligand-bound state. The various interaction energy terms described in the methods were collected and are presented in Table 3.3. Many forces are involved in the intermolecular association: hydrophobic, dispersion or van der Waals, hydrogen bonding and electrostatic interaction. The largest contribution for the binding energy comes from the van der Waals (VDW) interactions, but the specificity of the binding appears to be controlled by hydrogen bonding and electrostatic interactions (Fersht, 1984; Fersht et. al., 1985; Street et. al., 1986). This is obvious as the artemisinin analogues used in the study are mostly lipophilic molecules that interact favorably with a hydrophobic binding site.

The energy values in Table 3.3 were used to fit equation 1 using the Gaussian elimination method. The values obtained for the three fitting parameters,  $\alpha$ ,  $\beta$  and  $\gamma$  are -

0.0271, -0.0902 and -1.44 respectively. The large value of the cavity energy term signifies the fact that binding is largely driven by the ligand's ability to bury itself in the binding cavity, which is understandable given that most of the ligands are highly hydrophobic in nature. Even though the R value is low, VDW interactions contribute significantly toward the free energy of binding due to the large magnitude of the VDW interaction term. In Table 3.3, the experimental free energy values obtained from the RTIC<sub>50</sub> and the predicted free energy values estimated using fitting parameters have been presented. Figure 3.3 graphically shows the quality of fit between the SGB-LIE binding energy predictions vs the experimental values.



**Figure 3.3.** LIE binding energies for the Training set from MD sampling. The overall RMS error is 0.328 kcal/mol for 101 ligands studied here. If LIE results agree perfectly with the experimental values, the data points (represented by diamonds) should be on the diagonal line.

If a predicted binding energy agrees exactly with the experimental value, a data point (represented by diamonds) exactly in the diagonal line would be shown. To help visualize these data points, a lower and upper bound line are also plotted in the figure, with 1.0 kcal/mol below or above the experimental values. From the figures most of the data points (100 out of 101) are within or very close to these two bound lines, which means most of them have either less than or about 1.0 kcal/mol error. The only data point that shows large deviation from the experimental value is ligand 98 which has 2.187 kcal/mol error. The overall root mean square error (RMSE) between the experimental values and the values obtained by the fit was 0.328 kcal/mol, which is an indicator of the robustness of the fit. The

correlation coefficient  $r^2$  is 0.845 indicating a good correlation with experiment. The statistical significance of the SGB-LIE model is evaluated by the correlation coefficient ( $r^2$ ), standard error (s), F-test value, significance level of the model (P), leave-one-out cross-validation coefficient ( $q^2$ ) and predictive error sum of squares (PRESS).

**Table 3.3.** Average electrostatic (ele), van der Waals (vdw) and cavity (cav) energy terms as well as binding affinity model calculations for the Training set inhibitors using SGB-LIE method.

Ligand	$\langle U_{ele} \rangle^1$ kcal/mol	$\langle U_{vdw} \rangle^1$ kcal/mol	$\langle U_{cav} \rangle^1$ kcal/mol	$\Delta G_{bind,expt}^2$ kcal/mol	$\Delta G_{bind,LIE}^3$ kcal/mol	$^4RA_{expt}$	$^4RA_{pred}$
1	-2.516	-17.145	2.618	-1.906	-2.155	10.000	15.225
2	3.418	-16.108	1.619	-0.970	-0.971	2.818	2.821
3	-3.078	-18.372	1.323	1.036	-0.165	0.079	0.603
4	-3.853	-14.546	2.051	-1.620	-1.537	6.166	5.360
5	-1.981	-18.983	1.162	0.354	0.092	0.251	0.391
6	-1.958	-19.826	1.308	-0.354	-0.043	0.794	0.470
7	-2.338	-18.840	1.213	-0.692	0.016	1.479	0.447
8	-1.081	-22.970	1.939	0.160	-0.691	0.479	2.012
9	-2.327	-14.082	2.291	-2.175	-1.966	22.909	16.074
10	-5.208	-20.827	1.119	0.266	0.408	0.331	0.261
11	-0.814	-9.738	2.163	-2.423	-2.214	25.119	17.656
12	-2.665	-18.442	2.426	-1.584	-1.758	7.244	9.713
13	-0.697	-17.834	1.180	-0.357	-0.072	0.912	0.564
14	-4.569	-12.307	0.958	-0.531	-0.146	1.175	0.613
15	-1.426	-22.362	1.882	-0.463	-0.654	1.259	1.737
16	-2.462	-17.382	0.729	0.522	0.585	0.182	0.164
17	-2.614	-19.968	2.039	-0.912	-1.064	2.344	3.028
18	-2.953	-14.843	2.393	-2.175	-2.027	23.442	18.261
27	-3.078	-18.792	2.191	-1.595	-1.377	5.623	3.885
28	-3.834	-19.631	2.379	-1.288	-1.551	3.548	5.529
29	-2.707	-18.188	1.967	-0.946	-1.119	2.188	2.926
30	-2.849	-14.030	1.794	-1.766	-1.241	9.120	3.757
31	-3.244	-14.159	1.724	-0.826	-1.118	1.905	3.117
32	-3.131	-21.131	2.154	-1.191	-1.111	4.571	3.990
33	-1.162	-18.594	2.534	-1.333	-1.940	6.166	17.204
34	-2.097	-15.063	1.988	-0.546	-1.447	1.514	6.920
35	-3.044	-15.843	1.056	-0.156	-0.009	0.813	0.634
36	-2.157	-19.793	0.994	0.769	0.412	0.170	0.310
37	-1.920	-17.825	1.928	-0.919	-1.116	2.089	2.915
38	-0.706	-13.997	2.419	-2.286	-2.202	20.893	18.102
39	-3.972	-18.487	2.191	-1.456	-1.380	4.677	4.111

The data are collected from a 15 ps MD simulation after a 15 ps MD equilibration.  $^1U_{ele}$ ,  $U_{vdw}$  and  $U_{cav}$  energy terms represents the ensemble average of the energy terms calculated as the difference between bound and free state of ligands and its environment.  $^2\Delta G_{bind,expt}$  refers to free energy of binding with heme and is computed using the relationship:  $\Delta G_{binding} \approx RT \ln(IC_{50,expt})$ , where 298 K is used in the work for temperature T.  $^3\Delta G_{bind,LIE}$  refer to the absolute free energy values obtained using SGB-LIE method.  $^4RA_{expt}$  and  $RA_{pred}$  refers to the experimental and predicted relative activity and is calculated as  $RA = IC_{50}$  of artemisinin/ $IC_{50}$  of the analogue) x (MW of the analogue/MW of the artemisinin).

**Table 3.3** (Continued). Average electrostatic (ele), van der Waals (vdw) and cavity (cav) energy terms as well as binding affinity model calculations for the Training set inhibitors using SGB-LIE method.

Ligand	$\langle U_{ele} \rangle^1$ kcal/mol	$\langle U_{vdw} \rangle^1$ kcal/mol	$\langle U_{cav} \rangle^1$ kcal/mol	$\Delta G_{bind,expt}^2$ kcal/mol	$\Delta G_{bind,LIE}^3$ kcal/mol	$^4RA_{expt}$	$^4RA_{pred}$
40	-3.144	-16.665	1.376	-0.377	-0.393	0.912	0.937
41	-1.522	-16.466	1.654	-1.096	-0.855	2.692	1.791
42	-1.155	-15.515	2.148	-1.520	-1.662	6.026	7.657
43	-2.053	-18.557	0.903	-0.542	0.430	1.148	0.222
44	-2.414	-17.032	2.148	-1.091	-1.491	3.311	6.512
45	-1.365	-19.248	1.247	-0.496	-0.023	1.259	0.566
46	-3.597	-18.110	0.966	-0.298	0.340	0.933	0.318
47	-1.575	-23.515	1.558	-0.195	-0.080	0.851	0.702
48	-1.994	-11.342	2.789	-2.822	-2.939	61.660	75.041
49	-1.754	-9.730	3.019	-3.449	-3.422	177.828	169.651
50	-2.485	-18.383	2.019	-1.546	-1.182	7.413	4.007
51	-2.661	-14.790	2.954	-2.678	-2.848	50.119	66.731
52	-2.601	-16.296	2.197	-1.386	-1.623	5.012	7.476
53	-2.497	-17.867	2.303	-1.536	-1.637	5.623	6.662
66	-0.334	-21.951	1.454	-0.385	-0.105	1.047	0.652
67	-6.696	-19.333	1.102	-0.598	0.338	1.445	0.297
68	-2.908	-18.273	1.455	-0.164	-0.368	0.631	0.890
69	-1.570	-16.242	1.024	-0.383	0.033	0.912	0.452
70	-1.949	-13.471	1.085	-0.131	-0.294	0.692	0.912
74	-0.811	-14.935	0.948	0.296	0.004	0.309	0.506
75	-2.003	-15.996	0.746	-0.275	0.423	0.479	0.147
76	-3.262	-13.216	0.874	-0.246	0.021	0.490	0.312
77	-0.448	-18.855	1.058	-0.375	0.190	0.813	0.313
78	-3.355	-15.598	1.739	-1.266	-1.006	4.467	2.878
79	-1.541	-13.995	0.408	-0.002	0.716	0.407	0.121
80	-0.327	-15.482	1.739	-1.417	-1.099	5.623	3.283
81	-2.100	-16.519	1.843	-1.002	-1.107	2.512	2.998
82	-3.419	-15.003	0.852	0.435	0.219	0.257	0.370
83	-2.704	-21.131	1.634	-0.634	-0.373	1.862	1.199
84	-1.926	-12.919	0.496	0.799	0.503	0.155	0.255
85	-2.775	-11.353	0.835	-0.545	-0.103	1.698	0.805
86	-2.526	-11.029	0.090	0.607	0.934	0.251	0.145
94	-3.020	-16.659	1.824	-1.641	-1.042	6.026	2.192
95	-0.808	-13.499	0.387	1.093	0.682	0.054	0.108
96	-2.531	-15.943	1.721	-0.800	-0.972	1.698	2.268
97	-1.705	-16.422	0.406	1.094	0.943	0.063	0.081
98	-3.483	-21.475	0.244	3.866	1.680	0.001	0.020
105	-8.768	-13.714	0.696	0.498	0.472	0.182	0.190
106	-5.534	-16.037	1.213	-0.249	-0.149	0.635	0.537
107	-6.589	-17.617	0.229	2.590	1.438	0.005	0.035
108	-7.804	-15.997	0.352	1.748	1.148	0.033	0.090
112	-3.988	-13.850	1.473	-1.242	-0.764	3.071	1.370
113	-1.586	-11.374	3.017	-3.229	-3.276	127.092	137.458
114	-1.020	-20.888	2.248	-1.176	-1.325	3.972	5.106
115	-2.372	-21.331	2.839	-2.308	-2.100	26.850	18.879
116	-3.801	-21.520	1.347	-0.300	0.104	1.016	0.514
117	-2.574	-15.816	0.757	0.513	0.407	0.220	0.263
118	-2.678	-16.410	0.819	0.446	0.373	0.247	0.279

**Table 3.3** (Continued). Average electrostatic (ele), van der Waals (vdw) and cavity (cav) energy terms as well as binding affinity model calculations for the Training set inhibitors using SGB-LIE method.

Ligand	$\langle U_{ele} \rangle^1$ kcal/mol	$\langle U_{vdw} \rangle^1$ kcal/mol	$\langle U_{cav} \rangle^1$ kcal/mol	$\Delta G_{bind,expt}^2$ kcal/mol	$\Delta G_{bind,LIE}^3$ kcal/mol	$^4RA_{expt}$	$^4RA_{pred}$
119	-2.424	-15.593	1.151	0.138	-0.185	0.414	0.716
124	-2.894	-20.405	0.787	0.376	0.786	0.335	0.168
125	-2.022	-17.400	2.432	-1.695	-1.878	9.879	13.450
126	-1.210	-21.164	0.849	0.339	0.719	0.386	0.203
127	-1.761	-16.138	1.842	-1.044	-1.149	4.286	5.114
128	-2.492	-18.765	2.484	-1.695	-1.817	9.284	11.402
129	-1.684	-17.108	2.319	-1.695	-1.751	8.043	8.833
130	-3.207	-14.210	2.117	-1.695	-1.680	9.805	9.556
136	-0.640	-19.898	1.455	-0.543	-0.283	1.000	0.645
137	-3.500	-18.708	1.107	-0.493	0.188	1.100	0.348
138	-3.401	-13.899	1.004	-0.673	-0.100	1.490	0.566
139	-2.207	-16.793	0.772	0.155	0.463	0.370	0.220
140	-1.375	-19.543	0.788	0.890	0.665	0.120	0.175
145	-2.199	-10.135	0.534	-0.610	0.205	1.120	0.283
146	-3.199	-16.606	2.338	-1.643	-1.782	6.730	8.511
147	-2.036	-12.098	0.337	0.525	0.661	0.180	0.143
148	-2.081	-17.176	1.154	-0.014	-0.056	0.450	0.483
149	-2.937	-20.664	2.147	-0.906	-1.148	2.320	3.490
150	-1.725	-19.374	0.286	2.326	1.382	0.010	0.049
151	-1.522	-20.360	1.647	-0.468	-0.494	1.270	1.326
152	-1.084	-21.884	2.107	-0.968	-1.031	2.810	3.121
153	-1.788	-21.281	2.014	-0.994	-0.932	2.570	2.314

$$\Delta G = (-0.0271)\langle U_{ele} \rangle + (-0.0902)\langle U_{vdw} \rangle + (-1.44)\langle U_{cav} \rangle$$

(n = 102, r<sup>2</sup> = 0.845, S = 0.465, F = 234.1, P = 0.0001, q<sup>2</sup> = 0.844, PRESS = 21.38)

SGB-LIE model developed in this study is statistically (q<sup>2</sup> = 0.844, r<sup>2</sup> = 0.845, F = 234.1) best fitted and consequently used for prediction of antimalarial activities (pIC<sub>50</sub>) of training and test sets of molecules as reported in Table 3.3 & 3.4. The predicted activity calculated from free energy of binding is satisfactory with small deviation compared with experimental activity of training and test sets of molecules. The calculated free energy of binding (FEB) represents the experimental activity well.

**Table 3.4.** Average electrostatic (ele), van der Waals (vdw) and cavity (cav) energy terms as well as binding affinity model calculations for the Test set inhibitors using SGB-LIE method.

Ligand	$\langle U_{ele} \rangle^1$ kcal/mol	$\langle U_{vdw} \rangle^1$ kcal/mol	$\langle U_{cav} \rangle^1$ kcal/mol	$\Delta G_{bind,expt}^2$ kcal/mol	$\Delta G_{bind,LIE}^3$ kcal/mol	${}^4RA_{expt}$	${}^4RA_{pred}$
19	-3.281	-21.784	0.487	1.957	1.353	0.023	0.064
20	-2.613	-18.200	0.694	0.667	0.713	0.129	0.119
21	-2.740	-16.044	1.053	0.030	0.006	0.437	0.455
22	-3.833	-14.495	0.429	0.764	0.794	0.115	0.109
23	-2.919	-18.198	2.060	-1.826	-1.246	10.471	3.929
24	-3.139	-19.419	1.938	-1.195	-0.954	4.266	2.841
25	-0.555	-21.367	1.692	-0.542	-0.494	1.318	1.215
26	-2.136	-18.608	2.248	-2.027	-1.501	13.490	5.543
54	-3.294	-19.805	2.180	-1.168	-1.264	3.796	4.456
55	-1.837	-16.930	0.576	0.366	0.748	0.240	0.126
56	-1.815	-11.207	0.534	0.298	0.291	0.269	0.272
57	-2.520	-17.954	0.693	0.840	0.690	0.102	0.132
58	-1.891	-16.900	1.168	-0.678	-0.106	1.445	0.549
59	-2.806	-13.592	2.332	-2.287	-2.056	25.119	16.986
60	-1.133	-20.846	2.348	-1.523	-1.470	5.495	5.025
61	-2.241	-18.266	2.207	-1.659	-1.470	8.318	6.040
62	-2.837	-14.729	0.481	0.708	0.713	0.129	0.128
63	-3.090	-16.208	0.905	-0.596	0.243	1.288	0.313
64	-1.302	-17.980	1.877	-0.838	-1.046	2.138	3.035
65	-3.185	-16.301	0.936	0.147	0.209	0.363	0.327
71	-1.949	-13.565	1.463	-0.868	-0.830	2.192	2.055
72	-2.525	-9.035	1.969	-2.364	-1.952	28.840	14.354
73	-2.217	-14.830	2.495	-2.493	-2.195	29.512	17.842
87	-3.360	-21.074	1.579	-0.237	-0.282	0.912	0.983
88	-3.718	-13.614	0.490	0.883	0.623	0.126	0.195
89	-2.617	-21.764	0.342	1.948	1.542	0.016	0.032
90	-3.202	-14.273	1.356	-0.820	-0.578	1.698	1.130
91	-3.343	-20.539	0.350	1.975	1.439	0.016	0.039
92	-3.115	-20.933	0.287	2.124	1.559	0.016	0.041
93	-3.708	-18.669	1.060	-0.584	0.258	1.380	0.333
99	-2.202	-20.718	0.973	0.762	0.527	0.110	0.163
100	-1.739	-17.415	0.473	0.504	0.937	0.162	0.078
101	-2.296	-15.944	0.688	0.268	0.510	0.229	0.152
102	-3.180	-18.111	0.939	2.393	0.368	0.008	0.248
103	-3.304	-21.937	0.263	2.852	1.690	0.003	0.023
104	-2.504	-12.637	0.533	0.496	0.441	0.158	0.174
109	-6.627	-15.768	0.884	0.149	0.330	0.346	0.255
110	-7.339	-10.394	0.453	0.079	0.484	0.390	0.197
111	-6.277	-15.199	0.811	-0.017	0.374	0.435	0.225

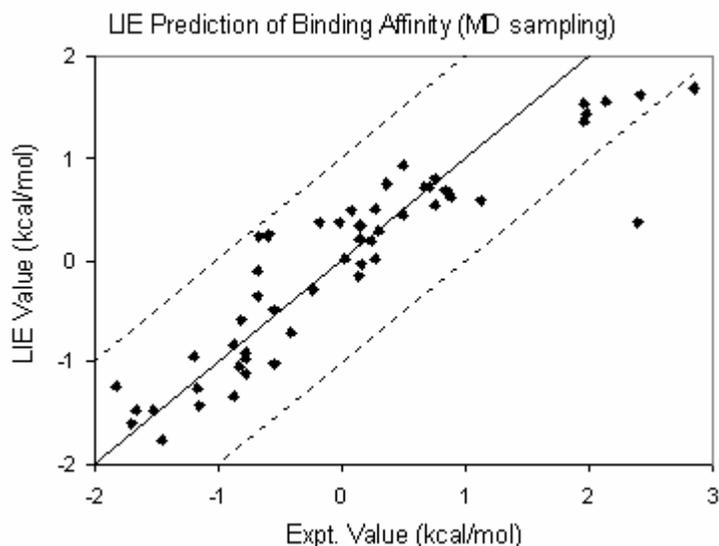
The data are collected from a 15 ps MD simulation after a 15 ps MD equilibration.  ${}^1U_{ele}$ ,  $U_{vdw}$  and  $U_{cav}$  energy terms represents the ensemble average of the energy terms calculated as the difference between bound and free state of ligands and its environment.  ${}^2\Delta G_{bind,expt}$  refers to free energy of binding with heme and is computed using the relationship:  $\Delta G_{binding} \approx RT \ln(IC_{50,expt})$ , where 298 K is used in the work for temperature T.  ${}^3\Delta G_{bind,LIE}$  refer to the absolute free energy values obtained using SGB-LIE method.  ${}^4RA_{expt}$  and  $RA_{pred}$  refers to the experimental and predicted relative activity and is calculated as  $RA = IC_{50}$  of artemisinin/ $IC_{50}$  of the analogue) x (MW of the analogue/MW of the artemisinin).

**Table 3.4 (continued).** Average electrostatic (ele), van der Waals (vdw) and cavity (cav) energy terms as well as binding affinity model calculations for the Test set inhibitors using SGB-LIE method.

Ligand	$\langle U_{\text{ele}} \rangle^1$ kcal/mol	$\langle U_{\text{vdw}} \rangle^1$ kcal/mol	$\langle U_{\text{cav}} \rangle^1$ kcal/mol	$\Delta G_{\text{bind,expt}}^2$ kcal/mol	$\Delta G_{\text{bind,LIE}}^3$ kcal/mol	$^4\text{RA}_{\text{expt}}$	$^4\text{RA}_{\text{pred}}$
120	-1.345	-12.865	0.572	-0.176	0.373	0.538	0.213
121	-1.874	-22.404	2.144	-0.551	-1.016	1.500	3.289
122	-2.973	-20.845	2.894	-2.438	-2.207	33.445	22.608
123	-2.698	-14.775	0.847	0.248	0.186	0.344	0.382
131	-3.370	-20.572	2.342	-1.152	-1.426	4.567	7.244
132	-3.772	-18.720	2.473	-1.455	-1.770	6.123	10.429
133	-3.675	-23.555	2.664	-1.695	-1.612	10.823	9.404
134	-2.567	-21.808	2.349	-0.874	-1.346	2.718	6.029
135	-2.910	-13.930	1.174	-0.685	-0.355	1.345	0.770
141	-3.131	-18.025	1.822	-0.780	-0.913	1.890	2.366
142	-2.166	-19.803	1.773	-0.411	-0.708	1.110	1.833
143	-1.677	-22.034	2.181	-0.782	-1.108	1.970	3.416
144	-1.197	-18.325	1.843	-0.783	-0.969	2.050	2.802
154	-1.597	-21.077	1.347	0.268	0.005	0.330	0.515
155	-2.852	-21.381	0.704	2.412	1.626	0.010	0.038
156	-1.402	-19.480	1.361	0.127	-0.165	0.530	0.867
157	-1.654	-22.698	1.482	0.158	-0.043	0.480	0.674
158	-2.392	-18.213	2.581	-2.174	-2.009	22.850	17.291

Satisfied with the robustness of the binding affinity model developed using the training set, we applied the LIE model to the artemisinin analogues comprising the test set. The test set includes 57 compounds categorized into ten subgroups as mentioned above (Table 3.1). The analogues comprising the test set were obtained from different sources (Woolfrey et. al., 1998; Acton et. al., 1993; Lin et. al., 1989; Posner et. al., 1992; Avery et. al., 1995; Avery et. al., 1996). Since the experimental values of  $\text{IC}_{50}$  for these inhibitors are already available, this set of molecules provides an excellent data set for testing the prediction power of the SGB-LIE method for new ligands. Table 3.4 presents the free energy values estimated for the 57 test compounds. The free energy values were estimated based on optimized SGB-LIE parameters  $\alpha$ ,  $\beta$  and  $\gamma$  from the training set. The quality of fit between the SGB-LIE binding energy predictions vs the experimental values is shown in Figure 3.4. We can see from the figure that most of the data points (56 out of 57) are within or very close to these two bound lines, which means most of them have either less than or about 1.0 kcal/mol error. The only data point that shows large deviation from the experimental value is ligand 102, which has 2.025 kcal/mol error. The reason for the big error in 102 is not very clear: it could be due to the force field parameters used since this is the only case whose R group has been substituted

with oxygen atoms. The RMSE between the experimental and predicted free energy values was 0.348 kcal/mol which is comparable to the level of accuracy achieved by the most accurate method, such as free energy perturbation. The squared correlation coefficient between experimental and SGB-LIE estimates for the free energy of the test set compounds is also significant ( $r^2 = 0.868$ ). The predicted relative antimalarial activity of artemisinin derivatives estimated using LIE free energy is also very close to experimental relative activity for the test set (Table 3.4).



**Figure 3.4.** LIE binding energies for the Test set from MD sampling. The overall RMS error is 0.348 kcal/mol for 57 ligands studied here. If LIE results agree perfectly with the experimental values, the data points (represented by diamonds) should be on the diagonal line.

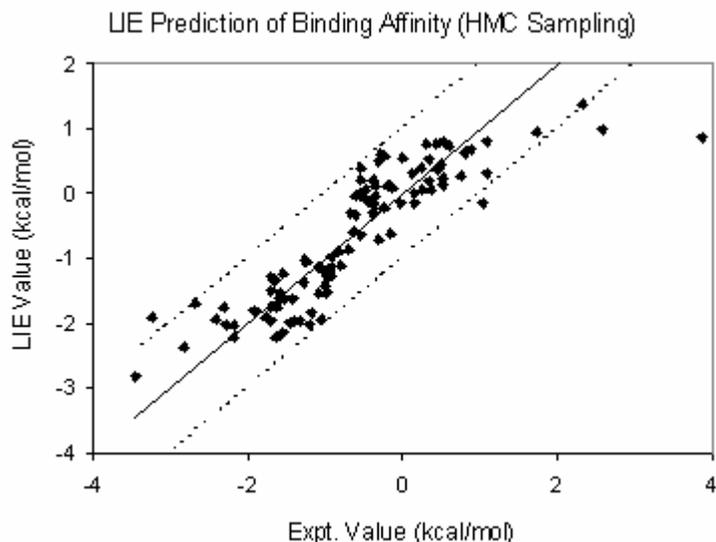
To test how sensitive the LIE method is to the underlying sampling techniques or in other words, how good the sampling techniques is in surfing the local conformation space, we also implemented LIE with the HMC sampling (Duane et. al., 1987; Zhou et. al., 1997). A Metropolis accept/reject criterion is checked every 5 steps of HMC's underlying MD simulation. The time step used in HMC's underlying MD is 3.0 fs with RESPA algorithm (Tuckerman et. al., 1992; Zhou et. al., 1995). Using the same three-parameter model, the LIE predictions are shown in Figure 4.5. Again, as we can see from the figure, most of the data points (99 out of 101) are within or very close to these two bound lines, which mean most of them have either less than or about 1.0 kcal/mol error. The data point 98 again has a large error, 2.992 kcal/mol but the data points 105 and 113 also show some significant error of

1.603 kcal/mol and 1.315 kcal/mol respectively. The reason for the big errors in these data points is not very clear. The overall RMS error is 0.415 kcal/mol and the correlation coefficient is 0.758, which are comparable to those from MD sampling. The new parameters are found to be  $\alpha = -0.0078$ ,  $\beta = -0.0735$  and  $\gamma = -1.20$ .

$$\Delta G = (-0.0078)\langle U_{ele} \rangle + (-0.0735)\langle U_{vdw} \rangle + (-1.20)\langle U_{cav} \rangle$$

(n = 103,  $r^2 = 0.758$ , S = 0.588, F = 133.6, P = 0.0001,  $q^2 = 0.757$ , PRESS = 36.51)

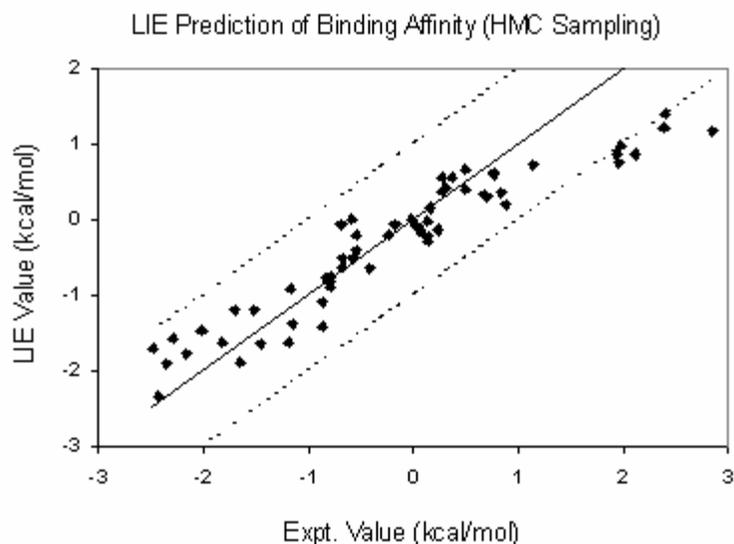
SGB-LIE model developed in this study is statistically ( $q^2 = 0.757$ ,  $r^2 = 0.758$ , F = 133.6) best fitted and consequently used for prediction of antimalarial activity (RA) of training and test sets of molecules.



**Figure 3.5.** LIE binding energies for the Training set from HMC sampling. The overall RMS error is 0.415 kcal/mol for 101 ligands studied here. If LIE results agree perfectly with the experimental values, the data points (represented by diamonds) should be on the diagonal line.

For the test set of 57 compounds the SGB-LIE model was able to predict their activity with an overall RMS error of 0.371. Figure 3.6 graphically shows the quality of fit between the SGB-LIE binding energy predictions vs the experimental values of the test set. We can see from the figure that most of the data points (56 out of 57) are within or very close to the two bound lines. The only data point that shows large deviation from the experimental value is ligand 103, which has 1.677 kcal/mol error. The squared correlation coefficient between

experimental and SGB-LIE estimates for the free energy of the test set compounds is also significant ( $r^2 = 0.891$ ).



**Figure 3.6.** LIE binding energies for the Test set from HMC sampling. The overall RMS error is 0.371 kcal/mol for 57 ligands studied here. If LIE results agree perfectly with the experimental values, the data points (represented by diamonds) should be on the diagonal line.

Except for the analogue 98 mentioned above, the LIE calculations agree with experiments quite well. A close look at the components in the LIE binding energy for each ligand reveals some important points. For example, the experiments show that analogues of artemisinin substituted at C-3 were found to be less active than those at C-9. For increasing alkyl bulk at C-3 a drop in antimalarial efficacy was noted (145, relative activity of 1.12; 147, relative activity of 0.2). Upon butyl substitution at C-9, the corresponding dual substituted analogues (3-alkyl, 9-butyl) showed a doubling of activity (153, relative activity of 2.6; 154, relative activity of 0.33). For the C-3 arylalkyl-substituted analogues alone, an increase in activity was observed with increasing chain length between ring system and the aryl ring (two carbons for 150, relative activity of 0.01; three carbons for 151, relative activity of 1.3; four carbons for 152, relative activity of 2.8). Dual substituted arylalkyl analogues (3-arylalkyl, 9-butyl) were generally less active than 3-substituted arylalkyl analogues alone (e.g., 151, relative activity of 1.3; 156, relative activity of 0.53). This has also been confirmed in our SGB-LIE predictions. The reason behind this is the fact that the binding site or acceptor for artemisinin and analogues exists with limited dimensions at both C-9 and C-3, more tolerant

of aryl and ester substitution on n-alkyl chains than of branches alkanes of any length. This can be evident from the loss of cavity energy due to burial of solvent accessible surface area. Further, dual substitution at C-3 and C-9, explored for only 9-butyl analogues, was on the whole detrimental to activity with the exception to 158. The high potency of the dual substituted analogue 158, is due to the formation of hydrogen bond between heme and 158 as suggested by (Avery et. al., 1995) In another class of analogues (Tricyclic 1,2,4-Trioxanes) it has been seen that a benzyl ether substituent is more potent as antimalarial activity than a methyl ether substituent (135, relative activity of 1.34; 128, relative activity of 9.28). Our SGB-LIE model also revealed similar conclusion. Apparently, one important factor contributing to very high antimalarial potency is the presence of lipophilic and bulky substituent capable perhaps of sterically “protecting” the trioxane moiety from biological reducing agents, thereby making the trioxane a more selective oxidizing agent (Klayman et. al., 1985). The superior antimalarial activity of the substituted esters over the substituted acid groups once again suggested that lipophilicity may play an important role in determining the antimalarial activity. For example, the ester derivatives (113-115 & 122) possess superior in vitro activity to artemisinin in comparison to their corresponding acids (117-119 & 123). The probable reason behind this may be the fact that the ester derivatives being more lipophilic have strong van der Waals interaction (SGB-LIE model) than acid derivatives with heme receptor. Also it is evident from the result that increased polarity and increased water solubility is associated with decreased antimalarial activity (e.g. analogues 105-111) (Lin et. al., 1992) and resulted in low van der Waals interaction (SGB-LIE model).

Overall, we found that the binding affinities for this binding set of artemisinin derivatives are largely coming from the van der Waals interaction between ligands and heme receptor (i.e., needs a good geometric fit) and the net loss of the cavity energy which is the same as the burial of solvent accessible surface area. These findings agree well with the new five-term model proposed by Jorgensen et al (Jorgensen et. al., 2001) based on the explicit solvents. Since the parameters  $\alpha$ ,  $\beta$  and  $\gamma$  are crucial to the LIE method, a natural question arises: How close are these parameters from one fit to another fit? If the LIE model really has the ability to predict binding affinities, one might expect that the parameters should be comparable to different fittings for the same binding set. Of course, we should not expect

them to be identical due to the “best possible fit” inside the Gaussian elimination fitting procedure. As we have already seen from above, the parameters are indeed comparable to the LIE fitting using either MD sampling or HMC sampling.

### **3.4 Conclusions**

We have demonstrated that the SGB-LIE method can be applied to estimate the free energy of binding with a high level of accuracy for a range of compounds with varying inhibition potencies. Despite the limitation imposed by the insufficient sampling inherent in the MD and HMC protocols, the methods have reproduced experimental data with reasonably small error for the majority of artemisinin analogues. A detailed study on the structure-activity relationships for artemisinin analogues can throw light on the moieties and functional groups important in determining the inhibition potency. The close estimation of inhibition potencies of a wide range of compounds has established the LIE methodology as an efficient tool for screening novel compounds with very different structures. Compared to the empirical methods, such as scoring function approaches, the LIE method is more accurate due to the semiempirical approach adopted in which experimental data are used to build the binding affinity model. The SGB-LIE method seems promising when compared to the free energy perturbation (FEP) or thermodynamic integration (TI) methods in achieving comparable accuracy with much faster speed even for structurally very different ligands.

## CHAPTER 4

### **Quantitative structure-activity relationship (QSAR) of the artemisinin: the development of predictive in vitro antimalarial activity models**

---

## Abstract

A quantitative structure-activity relationship (QSAR) analysis has been performed on a data set of 194 artemisinin analogues for antimalarial activity. Several types of descriptors including topological, spatial, thermodynamics, information content, lead likeness and E-state indices have been used to derive a quantitative relationship between antimalarial activity and structural properties. A systematic approach of zero tests, missing value test, simple correlation test, multicollinearity test and genetic algorithm method of variable selection was used to generate the model. Statistically significant model ( $r^2 = 0.845$ ,  $q^2_{cv} = 0.799$ ,  $F$ -test = 53.40) was obtained with the descriptors like molecular connectivity indexes, E-state index, length-to-breadth ratio of compounds, MLog P, HOMO, electron density, Balabans topological index and strain energy of the molecules. The robustness of the QSAR models was characterized by the values of the internal leave one out cross-validated regression coefficient ( $q^2_{cv}$ ) for the training set and determination coefficient in prediction,  $q^2_{test}$  for the test set. The value of  $q^2_{test} = 0.876$  for the test set; revealed good external predictability of the QSAR model. Also for an external data set (validation set) of 4 artemisinin analogues the QSAR model was able to predicts the antimalarial activity with very well in comparison to experimental values. The model was also tested successfully for external validation criteria. QSAR model developed in this study shall aid further design of novel potent artemisinin derivatives.

## 4.1 Introduction

Artemisinin (qinghaosu), a sesquiterpene endoperoxide isolated from *Artemisia annua* is a remarkable life saving antimalarial compound, effective against drug-resistant *Plasmodium falciparum* and cerebral malaria (Haynes et. al., 1997; Klayman, 1985; Kamchonwongpaisan et. al., 1996; Posner et. al., 1995). Artemisinin and its derivatives have many advantages: quick reduction of fevers, fast clearing parasites in blood (90% of malaria patients recovered within 48 hrs) and no significant side effects. As a consequence they are of special interest for severe malaria. The first decline in the number of parasites is also beneficial for combination therapies. Prompted by the clinical successes of the artemisinin, significant efforts have been focused on identifying new analogues that have a similar mechanism of action yet superior in activity. Subsequent research led to derivatives (Posner et. al., 1996; Robert et. al., 1998) of artemisinin such as artemether, arteether and artesunate. A consistent number of structural modifications have been introduced in the original structure of artemisinin in order to overcome the solubility as well as neurotoxic problem associated with its utilization as anti-malarial drug. The artemisinin family of molecules has been extensively studied to elucidate its mechanism of action as an antimalarial and to develop more potent and selective antimalarial agents (Kamchonwongpaisan et. al., 1996; Posner et. al., 1995; Posner et. al., 1996; Robert et. al., 1998). An essential feature of artemisinin (and analogues) activity is hypothesized to be the presence of a peroxide bridge (Figure 4.1), which forms a bond with a high valence non-heme iron molecule, leading to generation of free radicals (Posner et. al., 1995; Posner et. al., 1996).

A number of QSAR studies have also been reported for prescreening of prospective artemisinin analogues for antimalarial activity. A number of these studies (Avery et. al., 2002; Tommuphean et. al., 1998; Avery et. al., 1993) have used comparative molecular field analysis (CoMFA) (Cramer et. al., 1988; Cramer et. al., 1988) as a tool to model the activity of artemisinin analogues in terms of active site binding. Although comparative molecular field analyses (CoMFA) are statistically excellent and offer good predictive performance, they are inherently limited to the need to align with the database molecules correctly within 3D space. The determination of the 'active' conformation that each compound will retain is a critical issue due to unavailability of X-ray structure. We should have some knowledge or hypothesis

regarding active conformations of the molecules under study as a prerequisite for structural alignment. Hence, the developed models based on CoMFA may not suit to drug design, because of a false conformational hypothesis. However, we were motivated to explore possible alternatives that would use alignment free descriptors derived from 2D or 3D molecular topology and thus alleviate frequent ambiguity of structural alignment typical of 3D QSAR methods. A QSAR equation is a mathematical equation that correlates the biological activity to a wide variety of physical or chemical parameters (Hansch et. al., 2001; Livingstone, 2000). There are many examples available in literature in which QSAR models have been used successfully for the screening of compounds for biological activity (Shi et. al., 1998; Oloff et. al., 2005; Meneses-Marcel, 2005).

In this QSAR study, we have applied E-state, electronic, structural, topological quantum mechanics and physicochemical based descriptors which can be calculated without structural alignments, for the development of QSAR equation correlating *in vivo* antimalarial activity. The model developed in the present study is the first of its kind for antimalarial activity prediction of artemisinin congeners because of its high statistical quality. Further the behavior of QSAR model is examined with a variety of statistical parameters and the contribution of various descriptors are analyzed. The methodology used in the present study is in line with that has been used by Deswal and Roy (Deswal and Roy, 2006) for the development of thrombin inhibitors.

## **4.2. Materials and methods**

### **4.2.1. Data set**

An initial dataset of 194 artemisinin analogues were collected from published data (Woolfrey et. al., 1998; Acton et. al., 1993; Lin et. al., 1989; Posner et. al., 1992; Avery et. al., 1995; Avery et. al., 1996) in which several different ring systems were represented. All of the analogues were either peroxides or trioxanes, which should act via similar mechanisms of action and were categorized into different classes (Table 4.1a-m). These molecules were rationally designed as functional mimics of natural artemisinin with the goal of simplifying the chemical synthesis and improving the antimalarial activity. Structural modifications are mainly introduced at varying radicals at position R, R<sub>1</sub> and R<sub>2</sub> in artemisinin scaffold. Each of

these compounds had associated in vitro bioactivity values (IC<sub>50</sub> values reported in ng/ml) against the drug resistant malaria strain *P. falciparum* (W-2 clone). The log value of the relative activity (RA) of these compounds was used for analysis and was defined as:

$$\text{Log(RA)} = \log[(\text{artemisinin IC}_{50}/\text{analogue IC}_{50})(\text{analogue MW}/\text{artemisinin MW})]$$

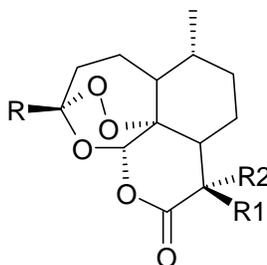
Molecular models of the artemisinin and its analogues (Table 4.1a-m) were built using the builder feature in Maestro (Schrodinger package) and energy minimized in a vacuum using Impact. Each structure was assigned an appropriate bond order using ligprep script shipped by Schrödinger and optimized initially by means of the OPLS 2005 force field using default setting. Complete geometrical optimization of these structures was carried out with the HF/3-21G method using the Jaguar (Schrodinger Inc.). In order to check the reliability of the geometry obtained, we compared the structural parameters of the artemisinin 1,2,4-trioxane ring with theoretical (Pinheiro et. al., 2001) and experimental (Leban et. al., 1988; Lisgarten et. al., 1998) values from literature. All calculations reproduced most of the structural parameters of the artemisinin 1,2,4-trioxane ring seen in X-ray structures (Table 4.2). This applies especially to the bond length of the endoperoxide bridge which seems to be responsible for the anti-malarial activity (Bernardinelli et. al., 1994; Posner et. al., 1995; Posner et. al., 1995; Haynes et. al., 1996; Rafiee et. al., 2005). These molecules were divided randomly into 156 molecules in training set and 38 molecules in test set.

#### 4.2.2. Descriptor calculation

E-state indices (Gregorio et. al., 1998), M log P (Meylan et. al., 1995), Superpendentic index (Gupta et. al., 1999), structural (Liu et. al., 1998), symmetrical, topological, lead likeness (Lipinski et. al., 2001), electronic Wang-Ford atomic charge (Avery et. al., 1995) and extended Huckel partial charge (Deswal and Roy, 2006; Eliopoulos et. al., 1996; Brenwald et. al., 1998), bulk, moments, orbital energies, molecular connectivity indexes (Kier et. al., 1976), gravitational indexes (Katritzky et. al., 1996), hydrophobicity (Livingstone et. al., 2000), steric (Shi et. al., 1998; Oloff et. al., 2005) and thermodynamic factors (Meneses-Marcel et. al., 2005) and topological descriptors were calculated using ADME Model Builder software package (version 4.5). The Superpendentic index is computed from the pendent

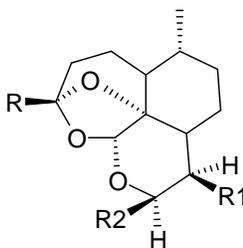
matrix. These descriptors help differentiate the molecules mostly according to their size, degree of branching, flexibility and overall shape. Some of the descriptors included in the study are listed and described in Table 4.3.

**Table 4.1a:** Artemisinin analogues with antimalarial activities against the drug-resistant malarial strain *P. falciparum* (W-2 clone) used in the work.



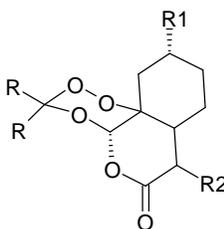
Compounds	R	R1	R2	Log RA	pIC <sub>50</sub> (ng/ml)
1	CH <sub>3</sub>	CH <sub>3</sub>	H	1.00	1.398
2	C <sub>4</sub> H <sub>8</sub> Ph	H	H	0.45	0.712
3	CH <sub>3</sub>	H	2-Z-Butenyl	-1.10	-0.760
4	CH <sub>3</sub>	H	H	0.79	1.188
5	CH <sub>3</sub>	H	CH <sub>3</sub>	-0.17	0.228
6	CH <sub>3</sub>	H	2-E-Butenyl	-0.60	-0.260
7	CH <sub>3</sub>	Allyl	H	-0.10	0.260
8	CH <sub>3</sub>	C <sub>4</sub> H <sub>9</sub>	H	0.17	0.674
9	C <sub>4</sub> H <sub>8</sub> Ph	C <sub>4</sub> H <sub>9</sub>	H	-0.32	-0.117
10	C <sub>3</sub> H <sub>6</sub> (P-Cl-Ph)	C <sub>4</sub> H <sub>9</sub>	H	-0.28	-0.097
11	C <sub>4</sub> H <sub>9</sub>	C <sub>4</sub> H <sub>9</sub>	H	-0.48	-0.195
12	CH <sub>3</sub>	C <sub>2</sub> H <sub>5</sub>	H	1.40	1.777
13	CH <sub>3</sub>	C <sub>6</sub> H <sub>13</sub>	H	0.86	1.162
14	CH <sub>3</sub>	i- C <sub>4</sub> H <sub>9</sub>	H	-0.55	-0.212
15	CH <sub>3</sub>	i-C <sub>6</sub> H <sub>13</sub>	H	-0.04	0.262
16	CH <sub>3</sub>	i-C <sub>3</sub> H <sub>7</sub>	H	-0.04	0.317
17	CH <sub>3</sub>	i-C <sub>5</sub> H <sub>11</sub>		0.07	0.389
18	C <sub>3</sub> H <sub>6</sub> (p-Cl-Ph)	H	H	0.10	0.340
19	C <sub>4</sub> H <sub>9</sub>	H	H	-0.74	-0.383
20	CH <sub>2</sub> CH <sub>2</sub> CO <sub>2</sub> Et	H	H	0.37	0.669
21	C <sub>2</sub> H <sub>5</sub>	H	H	0.05	0.448
22	i-C <sub>4</sub> H <sub>9</sub>	H	H	-0.35	0.007
23	CH <sub>3</sub>	Br	CH <sub>2</sub> Br	-1.64	-1.435
24	CH <sub>3</sub>	=CH <sub>2</sub>		-0.89	-0.489
25	CH <sub>3</sub>	CH <sub>2</sub> CH <sub>3</sub>	R <sub>1</sub> =R <sub>2</sub>	-0.36	-0.022
26	CH <sub>3</sub>	C <sub>5</sub> H <sub>11</sub>	H	1.02	1.339
27	CH <sub>3</sub>	C <sub>4</sub> H <sub>8</sub> Ph	H	0.63	0.876
28	CH <sub>3</sub>	C <sub>2</sub> H <sub>4</sub> Ph	H	0.12	0.398
29	CH <sub>3</sub>	C <sub>3</sub> H <sub>6</sub> Ph	H	0.78	1.042
30	CH <sub>3</sub>	C <sub>3</sub> H <sub>7</sub>	H	1.13	1.487

**Table 4.1b:** Deoxy-artemisinin derivatives with antimalarial activity against the drug-resistant malarial strain *P. falciparum* (W-2 clone) used in the work.



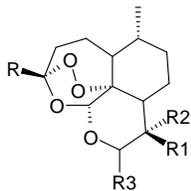
Compounds	R	R1	R2	MW	Log RA	pIC <sub>50</sub> (ng/ml)
31	CH <sub>3</sub>	CH <sub>3</sub>	OEt	296	-4	-3.623
32	CH <sub>3</sub>	CH <sub>3</sub>	OH	268	-4	-3.580
33	CH <sub>3</sub>	C <sub>4</sub> H <sub>8</sub> Ph	-	370	-4	-3.720
34	CH <sub>3</sub>	C <sub>3</sub> H <sub>7</sub>	-	280	-4	-3.599
35	CH <sub>3</sub>	C <sub>6</sub> H <sub>13</sub>	-	322	-4	-3.660
36	CH <sub>3</sub>	C <sub>4</sub> H <sub>9</sub>	H	294	-4	-3.620
37	CH <sub>3</sub>	i-C <sub>5</sub> H <sub>11</sub>	-	324	-4	-3.662
38	CH <sub>2</sub> CH <sub>2</sub> CO <sub>2</sub> Et	H	H	328	-4	-3.668
39	C <sub>2</sub> H <sub>4</sub> Ph	H	-	252	-4	-3.553
40	CH <sub>2</sub> CH <sub>3</sub>	H	-	252	-4	-3.553
41	i-C <sub>4</sub> H <sub>9</sub>	H	-	280	-4	-3.599
42	i-C <sub>4</sub> H <sub>9</sub>	H	H	280	-4	-3.599
43	CH <sub>3</sub>	C <sub>2</sub> H <sub>4</sub> Ph	-	342	-4	-3.686
44	CH <sub>3</sub>	C <sub>3</sub> H <sub>6</sub> Ph	-	356	-4	-3.703
45	CH <sub>3</sub>	CH <sub>3</sub>	-	266	-4	-3.577

**Table 4.1c:** Seco-artemisinin derivatives with antimalarial activity against the drug-resistant malarial strain *P. falciparum* (W-2 clone) used in the work.



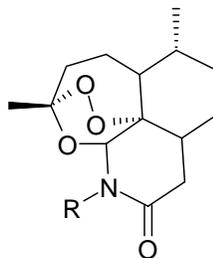
Compounds	R	R1	R2	Log RA	pIC <sub>50</sub> (ng/ml)
46	CH <sub>3</sub>	H	H	-2.37	-1.906
47	C <sub>2</sub> H <sub>5</sub>	H	H	-1.13	-0.713
48	CH <sub>3</sub>	CH <sub>3</sub>	CH <sub>3</sub>	-0.60	-0.183
49	-	-	-	-0.15	0.245
50	CH <sub>3</sub>	H	CH <sub>3</sub>	-0.86	-0.420
51	-	-	-(CH <sub>2</sub> ) <sub>4</sub> -	-0.26	0.097

**Table 4.1d:** 10-Substituted artemisinin derivatives with antimalarial activities against the drug-resistant malarial strain *P. falciparum* (W-2 clone) used in the work.



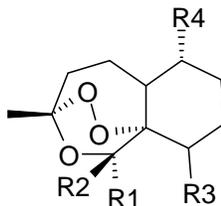
Compounds	R	R1	R2	R3	Log RA	pIC <sub>50</sub> (ng/ml)
52	CH <sub>3</sub>	CH <sub>3</sub>	H	H	0.75	1.170
53	CH <sub>3</sub>	CH <sub>3</sub>	H	OH	0.55	0.945
54	CH <sub>3</sub>	CH <sub>3</sub>	H	OEt	0.34	0.694
55	CH <sub>3</sub>	CH <sub>3</sub>	H	OH	0.96	1.295
56	CH <sub>3</sub>	CH <sub>3</sub>	H	OEt	-1.08	-0.740
57	CH <sub>3</sub>	H	Br	H	0.28	0.606
58	CH <sub>3</sub>	CH <sub>3</sub>	Br	NH-2-(1,3-thiazole)	0.66	0.874
59	CH <sub>3</sub>	CH <sub>3</sub>	Br	p-Cl-aniline	0.79	0.977
60	CH <sub>3</sub>	CH <sub>3</sub>	Br	aniline	0.18	0.401
61	CH <sub>3</sub>	Br	CH <sub>3</sub>	NH-2-pyridine	-0.09	0.115
62	CH <sub>3</sub>	CH <sub>3</sub>	Br	NH-2-pyridine	-0.77	-0.564
63	CH <sub>3</sub>	CH <sub>3</sub>	H	OMe	0.28	0.654
64	CH <sub>3</sub>	CH <sub>3</sub>	H	α -OEt	0.32	0.674
65	CH <sub>3</sub>	C <sub>4</sub> H <sub>9</sub>	H	H	1.32	1.677
66	CH <sub>3</sub>	C <sub>2</sub> H <sub>5</sub>	H	H	0.67	1.068
67	CH <sub>3</sub>	C <sub>3</sub> H <sub>7</sub>	H	OEt	-0.04	0.277
68	CH <sub>3</sub>	H	H	OEt	0.43	0.804
69	CH <sub>3</sub>	C <sub>2</sub> H <sub>5</sub>	H	OEt	0.50	0.835
70	CH <sub>3</sub>	CH <sub>3</sub>	H	C <sub>3</sub> H <sub>6</sub> OH	0.78	1.115
71	CH <sub>3</sub>	CH <sub>3</sub>	H	C <sub>4</sub> H <sub>9</sub>	0.06	0.398
72	CH <sub>3</sub>	CH <sub>3</sub>	H	OCH <sub>2</sub> CO <sub>2</sub> Et	0.52	0.800
73	CH <sub>3</sub>	CH <sub>3</sub>	H	OC <sub>2</sub> H <sub>4</sub> CO <sub>2</sub> Me	0.10	0.364
74	CH <sub>3</sub>	CH <sub>3</sub>	H	OC <sub>3</sub> H <sub>6</sub> CO <sub>2</sub> Me	-0.03	0.218
75	CH <sub>3</sub>	CH <sub>3</sub>	H	OCH <sub>2</sub> (4-PhCO <sub>2</sub> Me)	-0.07	0.143
76	CH <sub>3</sub>	CH <sub>3</sub>	H	(R)-OCH <sub>2</sub> CH(CH <sub>3</sub> )CO <sub>2</sub> Me	1.79	2.070
77	CH <sub>3</sub>	CH <sub>3</sub>	H	(R)-OCH(CH <sub>3</sub> )CH <sub>2</sub> CO <sub>2</sub> Me	0.87	1.134
78	CH <sub>3</sub>	CH <sub>3</sub>	H	(S)-OCH(CH <sub>3</sub> )CH <sub>2</sub> CO <sub>2</sub> Me	1.70	1.964
79	CH <sub>2</sub> CH <sub>2</sub> CO <sub>2</sub> Et	H	H	H	0.70	1.017
80	C <sub>3</sub> H <sub>6</sub> (p-Cl-Ph)	H	H	H	-0.55	-0.295
81	C <sub>2</sub> H <sub>5</sub>	H	H	H	-1.00	-0.580
82	C <sub>3</sub> H <sub>7</sub>	H	H	H	0.84	1.238
83	CH <sub>3</sub>	-OCH <sub>2</sub> -		OOH	-0.62	-0.269
84	CH <sub>3</sub>	-CH <sub>2</sub> O-		OOH	-0.57	-0.219
85	CH <sub>3</sub>	=CH <sub>2</sub>		OOH	-0.99	-0.616
86	-	CH <sub>3</sub>	OH	α-OH	-0.89	-0.519
87	CH <sub>3</sub>	C <sub>5</sub> H <sub>11</sub>	H	H	0.16	0.498
88	CH <sub>3</sub>	C <sub>3</sub> H <sub>7</sub>	H	H	0.74	1.117
89	-	CH <sub>3</sub>	H	CH <sub>2</sub> CF <sub>2</sub>	0.11	0.437
90	-	CH <sub>3</sub>	OH	CH <sub>2</sub> CF <sub>3</sub>	0.33	0.615
91	-	CH <sub>3</sub>	OH	OEt	-0.44	-0.108
92	-	OH	CH <sub>3</sub>	OEt	-1.13	-0.798
93	CH <sub>3</sub>	CH <sub>3</sub>	H	OEt-C <sub>4</sub> H <sub>9</sub>	0.92	1.217

**Table 4.1e:** 11-Aza-artemisinin derivatives with antimalarial activities against the drug-resistant malarial strain *P. falciparum* (W-2 clone) used in the work.



Compounds	R	Log RA	pIC <sub>50</sub> (ng/ml)
94	C <sub>3</sub> H <sub>6</sub> Ph	0.02	0.283
95	C <sub>2</sub> H <sub>4</sub> Ph	0.16	0.439
96	C <sub>5</sub> H <sub>11</sub>	-0.20	0.121
97	i-C <sub>5</sub> H <sub>11</sub>	-0.04	0.281
98	CH <sub>2</sub> (p-Cl-Ph)	-0.16	0.096
99	i-C <sub>4</sub> H <sub>9</sub>	0.02	0.359
100	CH <sub>2</sub> Ph	0.34	0.636
101	CH <sub>3</sub>	0.70	1.099
102	C <sub>3</sub> H <sub>7</sub>	0.05	0.408
103	2-Thiophene	0.17	0.458
104	2-Furan	0.11	0.418

**Table 4.1f:** Artemisinin derivatives lacking the D-ring with antimalarial activity against the drug-resistant malarial strain *P. falciparum* (W-2 clone) used in the work.



Compounds	R1	R2	R3	R4	Log RA	pIC <sub>50</sub> (ng/ml)
105	-O <sub>2</sub> CCH <sub>2</sub> Ph	H	H	CH <sub>3</sub>	-0.51	-0.217
106	H	H	H	CH <sub>3</sub>	-0.32	0.202
107	H	OCH <sub>3</sub>	H	H	-0.31	0.180
108	OCH <sub>3</sub>	H	H	H	-1.04	-0.550
109	H	H	H	-	-0.41	0.007
110	OCH <sub>2</sub> Ph	H	H	H	-0.09	0.275
111	C <sub>2</sub> H <sub>4</sub> OH	H	CH <sub>3</sub>	-	-1.80	-1.429
112	C <sub>2</sub> H <sub>4</sub> OH	CH <sub>3</sub>	H	-	0.23	0.601
113	C <sub>2</sub> H <sub>4</sub> OH	CH <sub>3</sub>	CH <sub>3</sub>	-	-1.80	-1.449
114	C <sub>2</sub> H <sub>4</sub> OCH <sub>2</sub> Ph	CH <sub>3</sub>	CH <sub>3</sub>	-	-1.80	-1.558
115	OCH <sub>3</sub>	H	C <sub>2</sub> H <sub>4</sub> O <sub>2</sub> CNEt <sub>2</sub>	H	0.65	0.929
116	OCH <sub>3</sub>	H	C <sub>2</sub> H <sub>4</sub> O <sub>2</sub> CNPh <sub>2</sub>	-	0.65	0.829
117	H	OCH <sub>3</sub>	C <sub>2</sub> H <sub>4</sub> OCH <sub>3</sub>	H	-0.39	0.002
118	H	OCH <sub>3</sub>	C <sub>2</sub> H <sub>4</sub> OCH <sub>2</sub> Ph	H	0.75	1.039
119	H	OCH <sub>3</sub>	C <sub>2</sub> H <sub>4</sub> O-allyl	H	0.40	0.735
120	H	OCH <sub>3</sub>	C <sub>2</sub> H <sub>4</sub> O <sub>2</sub> Ph	H	-0.59	-0.319
121	H	OCH <sub>3</sub>	C <sub>2</sub> H <sub>4</sub> O <sub>2</sub> C(4-PhCO <sub>2</sub> Me)	H	0.27	0.465
122	H	OCH <sub>3</sub>	C <sub>2</sub> H <sub>4</sub> O <sub>2</sub> C(4-PhCO <sub>2</sub> H)	H	-0.81	-0.586
123	H	OCH <sub>3</sub>	C <sub>2</sub> H <sub>4</sub> O <sub>2</sub> C(4-PhCONEt <sub>2</sub> )	-	0.23	0.400
124	H	OCH <sub>3</sub>	C <sub>2</sub> H <sub>4</sub> O <sub>2</sub> C(4-PhCO <sub>2</sub> C <sub>2</sub> H <sub>4</sub> NMe <sub>2</sub> )	-	-0.60	-0.446
125	H	OCH <sub>3</sub>	C <sub>2</sub> H <sub>4</sub> O <sub>2</sub> CCH <sub>2</sub> NCO <sub>2</sub> -(t-C <sub>4</sub> H <sub>9</sub> )	H	-0.04	0.174
126	OCH <sub>3</sub>	-	C <sub>2</sub> H <sub>4</sub> OCH <sub>2</sub> (4-F-Ph)	-	0.38	0.648
127	OCH <sub>3</sub>	-	C <sub>2</sub> H <sub>4</sub> OCH <sub>2</sub> (4-Py)	-	0.14	0.428
128	H	OCH <sub>3</sub>	C <sub>2</sub> H <sub>4</sub> OCH <sub>2</sub> (4-N-Me-pyridine)	H	-0.90	-0.647

**Table 4.1g:** Miscellaneous Artemisinin derivatives with antimalarial activity against the drug-resistant malarial strain *P. falciparum* (W-2 clone) used in the work.

Compound no.	Analogue structure	Log RA	pIC <sub>50</sub> (ng/ml)	Compound no.	Analogue structure	Log RA	pIC <sub>50</sub> (ng/ml)
129		0.78	1.203	136		-2.26	-1.862
130		-2.09	-1.755	137		-0.24	0.180
131		-1.27	-0.802	138		-2.59	-2.167
132		0.23	0.587	139		-0.96	-0.559
133		-0.67	-0.353	140		-0.79	-0.370
134		-4.00	-3.543	141		-0.64	-0.197
135		-4.00	-3.567	142		-0.353	0.090

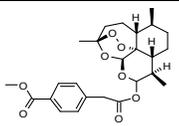
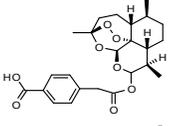
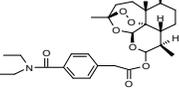
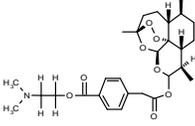
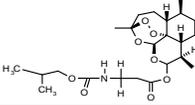
**Table 4.1h:** 9-Substituted Artemisinin derivatives with antimalarial activity against the drug-resistant malarial strain *P. falciparum* (W-2 clone) used in the work.

Compound no.	Analogue structure	Log RA	pIC <sub>50</sub> (ng/ml)	Compound no.	Analogue structure	Log RA	pIC <sub>50</sub> (ng/ml)
143		-0.729	-0.328	149		-1.487	-1.282
144		-0.739	-0.365	150		-1.926	-1.549
145		-2.447	-2.106	151		-0.460	-0.109
146		-0.198	0.182	152		-0.409	-0.058
147		-0.717	-0.325	153		-0.361	0.013
148		-2.469	-2.207				

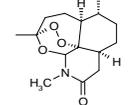
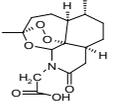
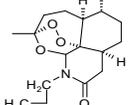
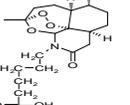
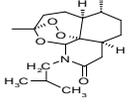
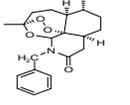
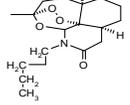
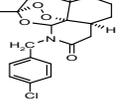
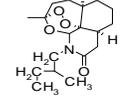
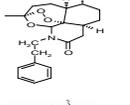
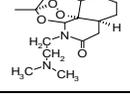
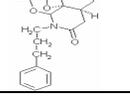
**Table 4.1i:** Dihydroartemisinin derivatives with antimalarial activity against the drug-resistant malarial strain *P. falciparum* (W-2 clone) used in the work.

Compound no.	Analogue structure	Log RA	pIC <sub>50</sub> (ng/ml)	Compound no.	Analogue structure	Log RA	pIC <sub>50</sub> (ng/ml)
154		-0.269	0.129	158		1.524	1.788
155		0.310	0.705	159		2.104	2.368
156		0.176	0.404	160		0.599	0.863
157		0.487	0.911				

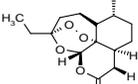
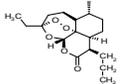
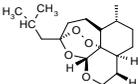
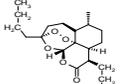
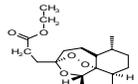
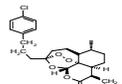
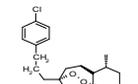
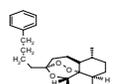
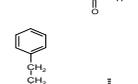
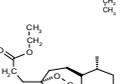
**Table 4.1j:** Tricyclic 1.2.4 – Trioxanes derivatives with antimalarial activity against the drug-resistant malarial strain *P. falciparum* (W-2 clone) used in the work.

Compound no.	Analogue structure	Log RA	pIC <sub>50</sub> (ng/ml)
161		0.660	0.845
162		-0.475	-0.275
163		0.551	0.699
164		0.205	0.340
165		0.312	0.503

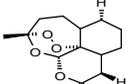
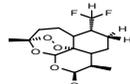
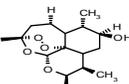
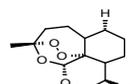
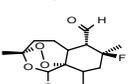
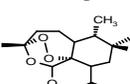
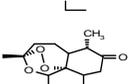
**Table 4.1k:** N-Alkyl-11-aza-9-desmethylartemisinins derivatives with antimalarial activity against the drug-resistant malarial strain *P. falciparum* (W-2 clone) used in the work.

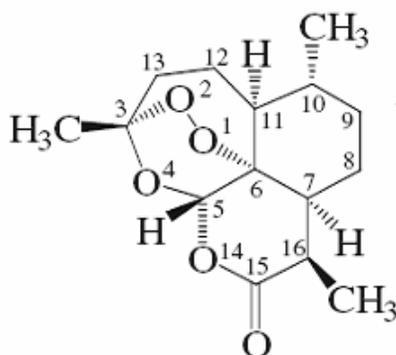
Compound no.	Analogue structure	Log RA	pIC <sub>50</sub> (ng/ml)	Compound no.	Analogue structure	Log RA	pIC <sub>50</sub> (ng/ml)
166		0.328	0.728	172		-1.222	-0.886
167		-0.125	0.233	173		-0.921	-0.652
168		0.161	0.500	174		0.276	0.572
169		0.041	0.362	175		0.045	0.301
170		0.173	0.494	176		0.294	0.573
171		-0.432	-0.114	177		0.312	0.574

**Table 4.11:** 3C- substituted artemisinin derivatives with antimalarial activity against the drug-resistant malarial strain *P. falciparum* (W-2 clone) used in the work.

Compound no.	Analogue structure	Log RA	pIC <sub>50</sub> (ng/ml)	Compound no.	Analogue structure	Log RA	pIC <sub>50</sub> (ng/ml)
178		0.049	0.447	183		0.410	0.729
179		-0.347	0.010	184		-0.481	-0.197
180		0.365	0.665	185		-0.276	-0.093
181		0.104	0.343	186		-0.319	-0.116
182		0.449	0.710	187		1.359	1.594

**Table 4.1m.** Various derivatives of artemisinin and artemether with antimalarial activity against the drug-resistant malarial strain *P. falciparum* (W-2 clone) used in the work.

Compound no.	Analogue structure	Log RA	pIC <sub>50</sub> (ng/ml)	Compound no.	Analogue structure	Log RA	pIC <sub>50</sub> (ng/ml)
188		0.437	0.083	192		2.570	0.717
189		2.188	0.672	193		0.016	-1.347
190		1.622	0.504	194		-0.120	0.192
191		1.445	0.495				



**Figure 4.1.** Stereochemistry and atomic numbering scheme of artemisinin.

**Table 4.2.** Experimental and theoretical values of the 1,2,4-trioxane ring parameters in artemisinin (bond lengths in Å; bond angles and torsional angles in degrees).

Parameters <sup>a</sup>	Theoretical			Experimental <sup>d</sup>	Experimental <sup>e</sup>
	3-21G <sup>b</sup>	3-21G** <sup>c</sup>	6-31G <sup>c</sup>		
O1-O2	1.463	1.462	1.447	1.475(4)	1.469(2)
O2-C3	1.441	1.440	1.435	1.417(4)	1.416(3)
C3-O4	1.436	1.436	1.435	1.448(4)	1.445(2)
O4-C5	1.407	1.408	1.403	1.388(4)	1.379(2)
C5-C6	1.529	1.530	1.533	1.528(5)	1.523(2)
C6-O1	1.478	1.477	1.469	1.450(4)	1.461(2)
O1-O2-C3	106.9	107.070	108.800	107.600(2)	108.100(1)
O2-C3-O4	107.0	107.310	106.760	107.200(2)	106.600(2)
C3-O4-C5	115.6	115.700	117.300	113.500(3)	114.200(2)
O4-C5-C6	112.0	112.030	112.280	114.700(2)	114.500(2)
C5-C6-O1	111.1	111.589	110.910	111.100(2)	110.700(2)
C6-O1-O2	111.2	111.286	113.240	111.500(2)	111.200(2)
O1-O2-C3-O4	-74.9	-74.680	-71.840	-75.500(3)	-75.500(2)
O2-C3-O4-C5	31.8	32.150	33.390	36.300(4)	36.000(2)
C3-O4-C5-C6	29.4	28.400	25.320	24.800(4)	25.300(2)
O4-C5-C6-O1	-51.8	-50.769	-49.410	-50.800(4)	-51.300(2)
C5-C6-O1-O2	10.1	9.792	12.510	12.300(3)	12.700(2)
C6-O1-O2-C3	50.8	50.522	46.700	47.700	47.800(2)

<sup>a</sup> Atoms are numbered according to Figure 4.1

<sup>b</sup> This work

<sup>c</sup> Values from Ref. (Pineiro et. al., 2001)

<sup>d</sup> Values from Ref. (Leban et. al., 1988) (experimental estimated standard deviations in brackets)

<sup>e</sup> Values from Ref. (Lisgarten et. al., 1998) (experimental estimated standard deviations in brackets)

**Table 4.3.** List of descriptors used in the study.

Type	Descriptors
E-state indices	Electro-topological-state indices
Electronic	Partial positive surface area, partial negative surface area, relative positive charge, relative negative charge, relative positive charged surface area, relative negative charged surface area, weighted positive charged partial surface area, weighted negative charged partial surface area, fractional negative charged partial surface area, fractional positive charged partial surface area, Huckel molecular orbital indices, highest occupied molecular orbital, lowest unoccupied molecular orbital, free valence value, nucleophilic superdelocalizability, free radical superdelocalizability, heat of formation, dipole moments, energy of the highest occupied orbital, energy of the lowest unoccupied orbital, electronegativity, hardness
Information content	Information of atomic composition index, superpendentivity index
Spatial	Radius of gyration, Jurs descriptors, shadow indices, area, density, length-to-breath ratios
Structural	Topological symmetry, geometrical symmetry, combined symmetry, conformational flexibility indices, molecular distance edge descriptors, moment of inertia indices, geometric moment indices, number of single bonds, number of aromatic bonds.
Thermodynamic	Average energy, bond strain energy, angle strain energy, non-bonded strain energy, torsional strain energy, total strain energy of molecule.
Leadlikeness	LogP (Meylan, Howard), LogS, LogP(Moriguchi, Hirono).
Topological	Wiener index, Kier and Hall molecular connectivity indices, path count and length descriptors, topological polar surface area (TPSA), Balban indices.

#### 4.2.3. Regression analysis

The total number of descriptors calculated initially was 372. A systematic search in the order of missing value test, zero test, correlation coefficient, multi-colinearity and genetic algorithm was performed to determine significant descriptors using ADME Model Builder (version 4.5) software package (Fujitsu Inc.). Any parameter which is not calculated (missing value) for any number of the compounds in the data set is rejected in the first step. Some of the descriptors were rejected because they contained a value of zero for all the compounds and have been removed (zero tests). In order to minimize the effect of colinearity and to avoid redundancy correlation matrix developed with a cut off value of 0.6 and the variables physically removed from the analysis which show exact linear dependencies between subsets of the variables and multi-colinearity (high multiple correlations between subsets of the

variables). From descriptors thus remained, the set of descriptors that would give the statistically best QSAR models were selected from the large pool using a genetic function approach implemented in ADME model Builder (version 4.5) software package (Fujitsu Inc.). The genetic algorithm starts with the creation of a population of randomly generated parameter sets. The usage probability of a given parameter from active set is 0.5 in any of the initial population sets. The sets are then compared according to their objective functions. The form of objective function favors sets that have  $r^2$  as high as possible, while minimizing the number of parameters used as descriptors. The higher the score the higher the probability of a given set will be used for the creation of the next generation of sets. Creation of a consecutive generation involves crossovers between set contents, as well as mutations. The parameters set used for genetic algorithm includes: mutation 0.1, crossover 0.9, population 300, number of generations 1000,  $R^2$  floor limit 50% and objective function was  $R^2/N_{par}$ . The form of objective function favors sets that have the  $R^2$  as high as possible, while minimizing the number of parameters used as descriptors. The higher the score, the higher the probability that a given set will be used for the creation of the next generation of sets. Creation of a consecutive generation involves crossovers between set contents, as well as mutations. The algorithm runs until the desired number of generations is reached. Equations were developed between the observed activity and the descriptors. The best equation was taken based on the statistical parameters such as regression coefficient ( $r^2$ ), adjusted regression coefficient ( $r^2_{adj}$ ), regression coefficient cross validation ( $q^2_{cv}$ ) and  $F$ -test values.

#### 4.2.4. Validation test

The predictive capability of the QSAR equation is determined using leave-one-out cross validation method. The cross validation regression coefficient ( $q^2_{cv}$ ) was calculated by following equation.

$$q^2_{cv} = 1 - \frac{PRESS}{TOTAL} = 1 - \frac{\sum_{i=1}^n (y_{exp} - y_{pred})^2}{\sum_{i=1}^n (y_{exp} - \bar{y})^2}$$

Where,  $y_{pred}$ ,  $y_{exp}$  and  $y$  are the predicted, experimental and mean values of experimental activity, respectively. Also the accuracy of the prediction of the QSAR equation was validated

by  $F$ -value,  $r^2$  and  $r_{adj}^2$ . A large  $F$  indicates that the model fit is not a chance occurrence. It has been shown that a high value of statistical characteristics need not be the proof of a highly predictive model (Golbraikh et. al., 2002; Roy et. al., 2008). Hence, in order to evaluate the predictive ability of our QSAR model, we used the method described by Golbraikh et al., (2002) and Roy et. al., (2008). The determination coefficient in prediction,  $q^2_{test}$  was calculated using the following equation (Roy et. al., 2008).

$$q^2_{test} = 1 - \frac{\sum (Y_{pred_{Test}} - Y_{Test})^2}{\sum (Y_{Test} - \bar{Y}_{Training})^2}$$

Where  $Y_{pred_{Test}}$  and  $Y_{Test}$  are the predicted value based on QSAR equation (model response) and experimental activity values, respectively, of the external test set compounds.  $Y_{Training}$  is the mean activity value of the training set compounds. Further evaluation of the predictive ability of the QSAR model for the external test set compounds was done by determining the value of  $rm^2$  by the following equation (Roy et. al., 2008):

$$rm^2 = r^2 \left( 1 - \left| \sqrt{r^2 - r_o^2} \right| \right)$$

Where  $r^2$  is the square correlation coefficient for regression ( $Y = a + bx$ ;  $a$  is referred as the y-intercept and  $b$  is the slope value of regression line) and  $r_o^2$  is the squared correlation coefficient for regression without intercept ( $Y = 0 + bx$ ; a very poor model) between experimental and predicted values for the external test set compounds. The values of  $k$  and  $k'$ , slopes of the regression line of the predicted activity vs. actual activity and vice versa, were calculated using the following equations (Jaiswal et. al., 2004):

$$k = \frac{\sum y_i \tilde{y}_i}{\sum \tilde{y}_i^2} \quad k' = \frac{\sum y_i \tilde{y}_i}{\sum y_i^2}$$

where  $\tilde{y}_i$  and  $y_i$  are the predicted and actual activities, respectively.

To further check the inter-correlation of descriptors variance inflation factor (VIF) analysis was performed. VIF value is calculated from  $1/1-r^2$ , where  $r^2$  is the multiple correlation coefficient of one descriptor's effect regressed on the remaining molecular descriptors. If VIF value is larger than 10, information of descriptor can be hidden by correlation of descriptors (Jaiswal et. al., 2004; Shapiro et. al., 1998).

### 4.3. Results and Discussion

The 194 active compounds considered as potential of W-2 strain of *P. falciparum* inhibition were randomly divided into a training set of 156 compounds and a test set of 38 compounds. The experimental IC<sub>50</sub> values against the W-2 strain of *P. falciparum* for these compounds are available from in vitro analysis. With the wide range of difference between the IC<sub>50</sub> values and the large diversity in the structures, the combined data set of 156 molecules and 38 molecules are ideal to be considered as training and test set, as both the sets do not suffer from bias, due to the similarity of the structures. The various molecular descriptors (372 in total) as described in Table 4.3 were calculated initially. By applying missing value test, zero test, correlation test with cutoff value of 0.6 and multicollinearity test with cutoff value of 0.9 we have discarded the most likely parameters that resulted in 117 parameters. Further additional parameters were discarded by applying genetic algorithm and finally 13 parameters were selected for development of QSAR equation. At first step the QSAR equation was developed using only one parameter (V7CH) which showed significant correlation with the biological activity in comparison to rest 12 parameters. Taking a brute force approach, we increased the number of parameters in the QSAR equation developed at the first step one by one and evaluated the effect of addition of new term on the statistical quality of the model. As the correlation coefficient,  $r^2$  can be easily increased by the number of terms in the QSAR equation; we took the cross-validation correlation coefficient,  $q^2_{cv}$ , as the limiting factor for a number of descriptors to be used in the final model. It was observed that the  $q^2_{cv}$  value increased till the number of descriptors in the equation reached up to 13 as shown in Table 4.4. With further addition of parameters to equation 13 (in Table 4.4), there was a decrease in  $q^2_{cv}$  value of the model. So the number of descriptors was restricted to 13 in the final QSAR model. The best significant relationship between the molecular descriptors and antimalarial activity has been deduced to be:

$$\begin{aligned} \text{pIC}_{50} = & - 1.34 - 11.3 \text{ V7CH} + 0.161 \text{ EMAX1} + 0.352 \text{ LOGP} - 0.668 \text{ GEOM3} + 0.0002 \\ & \text{STRA6} + 0.004 \text{ STRA4} + 0.054 \text{ STRA2} - 0.875 \text{ L/B2} - 9.92 \text{ FVMN} - 3.25 \text{ HOMO} - 2.53 \\ & \text{BOMX} + 0.475 \text{ MOLC9} - 9.89 \text{ V6C} \end{aligned} \quad (1)$$

$$(N = 154; r^2 = 0.777; s = 0.662; \text{PRESS} = 79.105; r^2_{adj} = 0.756; q^2_{cv} = 0.713; F\text{-test} = 37.53)$$

**Table 4.4.** Statistical assessment of QSAR equations with varying number of descriptors.

No. of descriptors	QSAR equation	$r^2$	press	$q^2$
1	$\text{pIC}_{50} = 1.78 - 6.00 \text{ V7CH}$	0.186	261.73	0.181
2	$\text{pIC}_{50} = 0.637 - 5.61 \text{ V7CH} + 0.095 \text{ EMAX1}$	0.223	252.28	0.212
3	$\text{pIC}_{50} = 0.073 - 5.75 \text{ V7CH} + 0.084 \text{ EMAX1} + 0.143 \text{ LOGP}$	0.237	251.25	0.221
4	$\text{pIC}_{50} = 0.140 - 5.77 \text{ V7CH} + 0.085 \text{ EMAX1} + 0.149 \text{ LOGP} - 0.116 \text{ GEOM3}$	0.237	253.10	0.217
5	$\text{pIC}_{50} = -0.489 - 6.09 \text{ V7CH} + 0.087 \text{ EMAX1} + 0.192 \text{ LOGP} - 0.471 \text{ GEOM3} + 0.003 \text{ STRA6}$	0.304	233.91	0.281
6	$\text{pIC}_{50} = -0.626 - 6.78 \text{ V7CH} + 0.074 \text{ EMAX1} + 0.215 \text{ LOGP} - 0.461 \text{ GEOM3} + 0.002 \text{ STRA6} + 0.004 \text{ STRA4}$	0.319	243.36	0.292
7	$\text{pIC}_{50} = -1.04 - 7.67 \text{ V7CH} + 0.072 \text{ EMAX1} + 0.216 \text{ LOGP} - 0.710 \text{ GEOM3} + 0.001 \text{ STRA6} + 0.005 \text{ STRA4} + 0.052 \text{ STRA2}$	0.424	198.28	0.397
8	$\text{pIC}_{50} = 0.134 - 8.29 \text{ V7CH} + 0.061 \text{ EMAX1} + 0.353 \text{ LOGP} - 0.716 \text{ GEOM3} + 0.001 \text{ STRA6} + 0.006 \text{ STRA4} + 0.046 \text{ STRA2} - 1.15 \text{ L/B2}$	0.467	182.97	0.438
9	$\text{pIC}_{50} = -0.736 - 7.75 \text{ V7CH} + 0.057 \text{ EMAX1} + 0.309 \text{ LOGP} - 0.696 \text{ GEOM3} + 0.001 \text{ STRA6} + 0.005 \text{ STRA4} + 0.039 \text{ STRA2} - 0.993 \text{ L/B2} - 4.95 \text{ FVMN}$	0.516	150.18	0.486
10	$\text{pIC}_{50} = -1.45 - 6.95 \text{ V7CH} + 0.058 \text{ EMAX1} + 0.348 \text{ LOGP} - 0.748 \text{ GEOM3} + 0.001 \text{ STRA6} + 0.004 \text{ STRA4} + 0.045 \text{ STRA2} - 1.05 \text{ L/B2} - 4.97 \text{ FVMN} - 1.38 \text{ HOMO}$	0.534	164.87	0.501
11	$\text{pIC}_{50} = 0.51 - 6.98 \text{ V7CH} + 0.055 \text{ EMAX1} + 0.336 \text{ LOGP} - 0.701 \text{ GEOM3} + 0.001 \text{ STRA6} + 0.005 \text{ STRA4} + 0.046 \text{ STRA2} - 1.04 \text{ L/B2} - 5.90 \text{ FVMN} - 1.38 \text{ HOMO} - 2.67 \text{ BOMX}$	0.542	163.93	0.507
12	$\text{pIC}_{50} = 3.88 - 8.40 \text{ V7CH} + 0.064 \text{ EMAX1} + 0.115 \text{ LOGP} - 0.673 \text{ GEOM3} + 0.001 \text{ STRA6} + 0.002 \text{ STRA4} + 0.041 \text{ STRA2} - 1.08 \text{ L/B2} - 4.43 \text{ FVMN} - 1.51 \text{ HOMO} - 0.73 \text{ BOMX} - 1.45 \text{ MOLC9}$	0.606	140.13	0.573
13	$\text{pIC}_{50} = 3.34 - 8.98 \text{ V7CH} + 0.088 \text{ EMAX1} + 0.172 \text{ LOGP} - 0.461 \text{ GEMO3} + 0.0001 \text{ STRA6} + 0.004 \text{ STRA4} + 0.038 \text{ STRA2} - 1.04 \text{ L/B2} - 4.78 \text{ FVMN} - 0.816 \text{ HOMO} - 1.18 \text{ BOMX} - 1.12 \text{ MOLC9} - 4.89 \text{ V6C}$	0.782	78.403	0.738

It was found that the compound numbers **31 to 45** were outliers with prediction error in between 2.233 to 3.150. The reason for those compounds being found as outliers could probably be their very low activity. The antimalarial action of artemisinin appears to be mediated by the generation of free radicals from the endoperoxide bridge of the drug (Posner *et. al.*, 1995; Jefford *et. al.*, 1996). All these compounds belong to deoxyartemisinin derivatives (Table 4.1b) that lacks the endoperoxide moiety. This group of compounds has single oxygen instead of the peroxide bridge. The interaction between heme and these analogues is mediated by three nonperoxide oxygens leading to inactivity or low activity (Meshnick *et. al.*, 1989; Shukla *et. al.*, 1995). The quality of the above QSAR model has been improved further by removing these compounds and is as follows.

$$\text{pIC}_{50} = 4.04 - 9.76 \text{ V7CH} + 0.069 \text{ EMAX1} + 0.117 \text{ LOGP} - 0.581 \text{ GEOM3} - 0.00014 \text{ STRA6} + 0.0035 \text{ STRA4} + 0.039 \text{ STRA2} - 1.09 \text{ L/B2} - 4.82 \text{ FVMN} - 1.02 \text{ HOMO} - 0.285 \text{ BOMX} - 1.36 \text{ MOLC9} - 5.24 \text{ V6C} \quad (2)$$

$$(N = 141; r^2 = 0.845; s = 0.343; \text{PRESS} = 19.386; r^2_{adj} = 0.830; q^2_{cv} = 0.799; F\text{-test} = 53.40)$$

where N is the number of compounds in the training set,  $r^2$  is the squared correlation coefficient, s is the estimated standard deviation about the regression line,  $r^2_{adj}$  is the square of adjusted correlation coefficient for degree of freedom,  $F$ -test is the measure of variance which compares two models differing by one or more variables to see if the more complex model is more reliable than the less complex one, the model is supposed to be good if the  $F$ -test is above a threshold value and  $q^2_{cv}$  is the square of the correlation coefficient of the cross-validation. The QSAR model developed in this study is statistically ( $r^2 = 0.845$ ,  $q^2_{cv} = 0.799$ ,  $F$ -test = 53.40) best fitted and consequently used for prediction of antimalarial activity ( $\text{pIC}_{50}$ ) of training and test sets of molecules as reported in Table 4.5 and Table 4.6. The quality of the prediction models for the training compounds before and after removal of outliers have been shown in Figure 4.2 & Figure 4.3. The  $r^2$  and  $q^2_{cv}$  values of 0.845 and 0.799, respectively of the model corroborates with the criteria for a QSAR model to be highly predictive (Leban *et. al.*, 1988). The standard error of estimate for the model was 0.343, which is an indicator of the robustness of the fit and suggested that the predicted  $\text{pIC}_{50}$  based on equation (2) is reliable.

**Table 4.5.** Observed and predicted activity against the drug-resistant malarial strain *P. falciparum* (W-2 clone) of Training set of artemisinin derivatives.

Compound No.	W-2 clone inhibition (pIC <sub>50</sub> )			Compound No.	W-2 clone inhibition (pIC <sub>50</sub> )		
	Observed	Predicted	Residual		Observed	Predicted	Residual
187	1.594	0.899	0.695	111	-1.429	-1.336	0.093
52	1.170	0.474	0.696	112	0.601	0.103	0.498
2	0.712	0.118	0.594	114	-1.558	-1.026	0.532
56	-0.740	0.137	0.877	115	0.929	0.687	0.242
188	0.083	0.354	0.271	116	0.829	1.133	0.304
1	1.398	1.016	0.382	31	-3.623	-3.346	0.723
4	1.188	0.518	0.670	32	-3.580	-3.079	0.299
5	0.228	0.492	0.264	41	-3.599	-2.759	0.840
6	-0.260	0.302	0.562	35	-3.660	-2.693	0.967
189	0.672	1.070	0.398	36	-3.620	-2.868	0.752
7	0.260	1.070	0.810	37	-3.662	-3.075	0.587
60	0.401	0.227	0.174	39	-3.553	-2.651	0.902
61	0.115	-0.372	0.487	40	-3.553	-2.771	0.782
63	0.654	0.446	0.208	42	-3.599	-3.407	0.308
64	0.674	0.450	0.224	43	-3.686	-3.220	0.134
8	0.508	0.295	0.213	118	1.039	0.848	0.191
9	-0.117	0.562	0.679	119	0.735	0.456	0.279
11	-0.195	0.495	0.690	120	-0.319	0.139	0.458
66	1.068	1.079	0.011	121	0.465	0.139	0.326
67	0.277	0.245	0.032	122	-0.586	0.139	0.725
68	0.804	0.379	0.425	123	0.400	-0.395	0.795
69	0.835	0.231	0.604	124	-0.446	-0.094	0.352
13	1.162	1.192	0.030	126	0.648	1.213	0.565
14	-0.212	-0.304	0.092	128	-0.647	-0.065	0.582
15	0.262	-0.270	0.532	134	-3.543	-3.060	0.483
16	0.317	0.325	0.008	135	-3.567	-3.197	0.370
17	0.389	-0.056	0.445	136	-1.862	-1.740	0.122
94	0.283	-0.121	0.404	138	-2.167	-1.693	0.474
95	0.439	0.174	0.265	139	-0.559	-0.526	0.033
96	0.121	-0.024	0.145	140	-0.370	-0.292	0.078
97	0.281	-0.243	0.524	141	-0.197	-1.011	0.814
98	0.096	-0.082	0.178	104	0.418	0.373	0.045
70	1.115	0.103	1.012	143	-0.328	0.654	0.982
129	1.203	0.023	1.180	145	-2.106	-1.313	0.793
72	0.800	0.794	0.006	146	0.182	0.253	0.071
73	0.364	0.966	0.602	147	-0.325	0.653	0.978
75	0.143	0.823	0.680	149	-1.282	-1.054	0.228
76	2.070	1.220	0.850	151	-0.109	-0.348	0.239
77	1.134	0.359	0.775	152	-0.058	-0.498	0.440
78	1.964	1.677	0.287	153	0.013	-0.295	0.308
79	1.017	0.211	0.806	154	0.129	0.320	0.191
80	-0.295	-0.380	0.085	155	0.705	0.827	0.122
18	0.340	-0.266	0.606	156	0.404	0.834	0.430
20	0.669	0.192	0.477	157	0.911	0.195	0.716
21	0.448	0.182	0.266	158	1.788	1.102	0.686

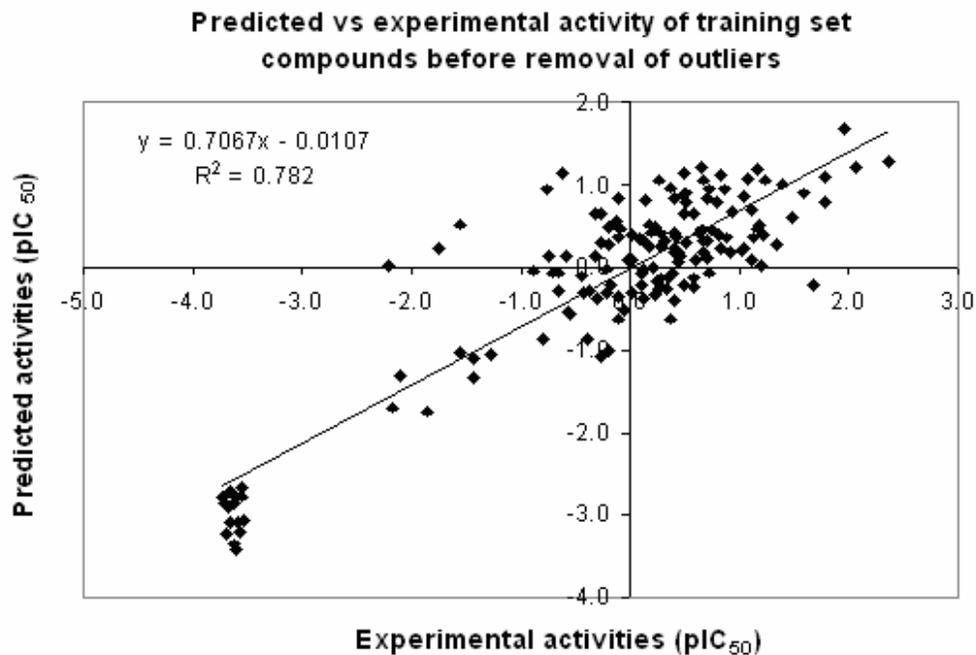
**Table 4.5.** (Continued)

Compound No.	W-2 clone inhibition (pIC <sub>50</sub> )			Compound No.	W-2 clone inhibition (pIC <sub>50</sub> )		
	Observed	Predicted	Residual		Observed	Predicted	Residual
82	1.238	1.054	0.184	159	2.368	1.276	1.092
100	0.636	0.407	0.229	160	0.863	0.940	0.077
102	0.408	0.429	0.021	162	-0.275	0.664	0.939
23	-1.435	-1.080	0.355	165	0.503	0.785	0.282
83	-0.269	-1.067	0.798	167	0.233	-0.331	0.564
25	-0.022	0.098	0.120	168	0.500	-0.226	0.726
87	0.498	0.879	0.381	169	0.362	-0.617	0.979
26	1.339	0.284	1.055	170	0.494	-0.227	0.721
27	0.876	0.358	0.518	171	-0.114	-0.612	0.498
28	0.398	0.187	0.211	172	-0.886	-0.041	0.845
29	1.042	0.242	0.800	173	-0.652	-0.284	0.368
88	1.117	0.709	0.408	174	0.572	-0.134	0.706
30	1.487	0.620	0.867	177	0.574	-0.245	0.819
194	0.192	0.427	0.235	176	-0.385	-0.875	0.490
89	0.437	0.080	0.357	179	0.010	0.107	0.097
92	-0.108	0.358	0.466	180	0.665	0.317	0.348
191	0.504	0.895	0.391	181	0.343	-0.132	0.475
192	0.495	1.142	0.647	182	0.710	0.316	0.394
193	0.717	0.938	0.221	183	0.729	-0.054	0.783
93	0.497	0.659	0.162	184	-0.197	0.284	0.481
46	1.217	0.390	0.827	185	-0.093	0.475	0.568
47	-0.713	-0.056	0.657	186	-0.116	0.833	0.949
48	-0.183	-0.226	0.043	12	1.777	0.785	0.992
49	0.245	-0.163	0.408	38	-3.668	-2.900	0.768
50	-0.420	-0.302	0.118	33	-3.720	-2.759	0.961
105	-0.217	-0.021	0.196	44	-3.703	-2.850	0.853
132	0.587	0.660	0.073	190	-0.798	-0.861	0.063
106	0.202	-0.001	0.203	148	-2.207	0.019	2.226
107	0.180	0.513	0.333	150	-1.549	0.532	2.081
51	0.097	-0.226	0.323	130	-1.755	0.242	1.997
108	-0.550	-0.542	0.008	85	-0.616	1.152	1.768
109	0.007	0.395	0.388	3	-0.760	0.942	1.702
110	0.275	0.411	0.136	65	1.677	-0.223	1.900

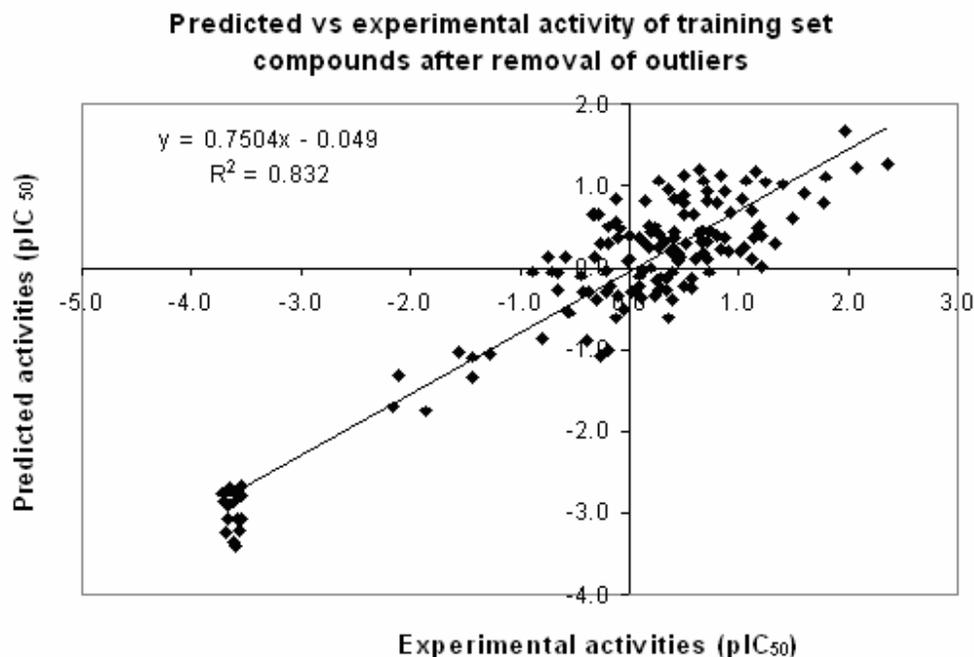
**Table 4.6.** Observed and predicted activity against the drug-resistant malarial strain *P. falciparum* (W-2 clone) of Test set of artemisinin derivatives.

W-2 clone inhibition (pIC <sub>50</sub> )			W-2 clone inhibition (pIC <sub>50</sub> )				
Compound No.	Observed	Predicted	Residual	Compound No.	Observed	Predicted	Residual
53	0.945	0.685	0.260	133	-0.353	-0.330	0.023
54	0.694	0.484	0.210	34	-3.599	-2.850	0.749
55	1.295	1.304	0.009	117	0.002	0.013	0.011
57	0.606	0.361	0.245	125	0.174	0.552	0.378
58	0.874	0.211	0.663	127	0.428	0.403	0.025
59	0.977	0.797	0.180	137	0.180	0.958	0.778
62	-0.564	-0.537	0.027	142	0.090	0.536	0.446
10	-0.097	-0.181	0.084	103	0.458	0.199	0.259
99	0.359	0.299	0.060	144	-0.365	-0.978	0.613
71	0.398	0.944	0.546	161	0.845	0.215	0.630
74	0.218	0.045	0.173	163	0.699	0.146	0.553
19	-0.383	-0.281	0.102	164	0.340	0.010	0.330
81	-0.580	-0.256	0.324	166	0.728	0.250	0.478
22	0.007	0.217	0.210	175	0.301	0.313	0.012
84	-0.219	-0.622	0.403	178	0.447	0.880	0.433
24	-0.489	-0.365	0.124	101	1.099	1.199	0.100
86	-0.519	-0.124	0.395	113	-1.449	-1.820	0.371
90	0.615	0.320	0.295	45	-3.577	-3.190	0.387
131	-0.802	-0.499	0.303	91	-0.415	-0.694	0.279

$$pIC_{50} = -\log_{10}IC_{50}$$



**Figure 4.2.** Relationship between predicted and experimental activities as per equation (1) before removal of outliers.



**Figure 4.3.** Relationship between predicted and experimental activities as per equation (2) after removal of outliers (deoxy-artemisinin derivatives).

V7CH is the 7th order chain molecular connectivity indexes. However, V6C measures 6<sup>th</sup> order cluster molecular connectivity indexes. These descriptors contain information about the size and the degree of branching in a molecule (Kier et. al., 1976). EMAX1 is the maximum atomic E-state index for each atom of each compound in the data set which is the measure of the reactivity of each atom, analogous to the concept of free valence. It provides information regarding intermolecular interactions (Kier et. al., 1997). L/B2 is the length-to-breadth ratio of compounds calculated by rotating the molecule in Z-axis in increments of N degrees. MLog P calculate Octanol/Water Partition coefficient of the molecule based on the algorithm by (Moriguchi et. al., 1992). It is the most popular and traditional. It explains one of the principal characteristics of any preparation, the lipophilicity. The higher its value, the more probable the transfer of the preparation from the aqueous medium into the biological membrane. This property is critical for medicinal preparations that are administered orally and must be absorbed through the GI tract. Log P value less than 0.5 will be absorbed appropriately. HOMO is the highest occupied molecular orbital energy (calculated using single point MOPAC (AMI)-based semiempirical quantum mechanical methods); this descriptor considers only interactions of valence  $\pi$  electrons for adjacent atoms. The

descriptor BOMX is a measure of the electron density between adjacent atoms, representative to the strength of that bond. FVMN is a measure of the available bonding capacity left in an atom. It is considered to be a measure of the likelihood that the specified atom will be the site of a radical attack. The descriptors such as STRA2, STRA4, STRA6 comprises each of the strain energy terms like bond, torsional and total energy terms of the molecules used in the molecular mechanics force field. The descriptor MOLC9 includes Balabans topological index J. It measure the degree of branching in structures (Kier et. al., 1976). GEOM3 is the mass weighted thickness descriptor. The calculation involves diagonalization of the covariance matrix formed from the (x,y,z) coordinates of the atoms, translated to the center of mass of the structure. The contribution of each atom is weighted by its mass.

The inter-correlation of the descriptors used in the final model (Equation 2) was very low (below 0.6) which is in conformity to the study that for a statistically significant model, it is necessary that the descriptors involved in the equation should not be inter-correlated with each other (Deswal et. al., 2006). The correlation matrix for the used descriptors is shown in Table 4.7. To further check the inter-correlation of descriptors variance inflation factor (VIF) analysis was performed. In this model, the VIF values of these descriptors are 1.70 (V7CH), 1.152 (EMAX1), 2.079 (Log P), 1.271 (GEOM3), 2.252 (STRA6), 2.331 (ATRA4), 1.344 (STRA2), 1.398 (L/B2), 1.441 (FVMN), 1.479 (HOMO), 1.402 (BOMX), 2.257 (MOLC9) and 1.155 (V6C). Based on VIF analysis it has been found that the descriptors used in the final model have very low inter-correlation.

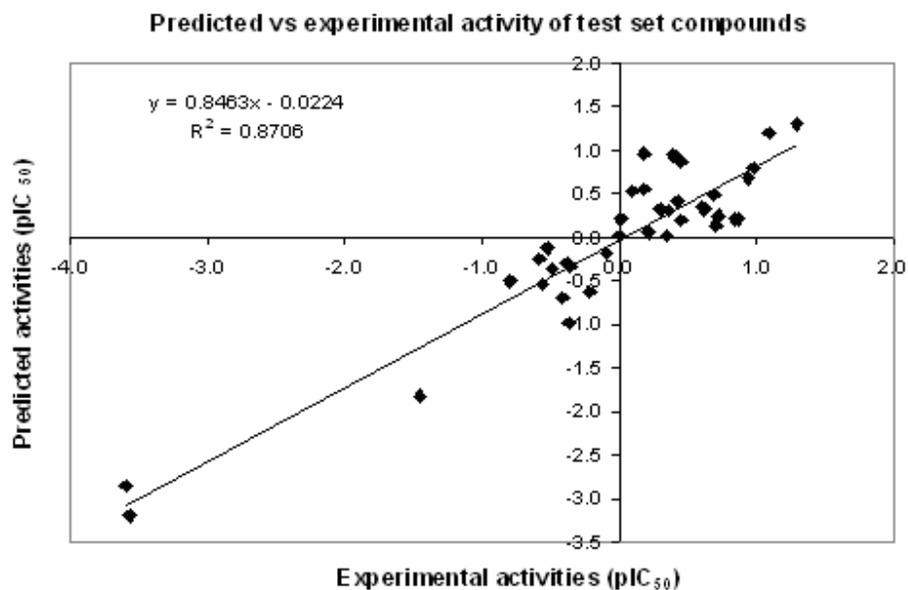
Satisfied with the robustness of the QSAR model developed using training set, we have applied the QSAR model to an external data set of artemisinin analogues comprising the test set. As the experimental values ( $\text{pIC}_{50}$ ) for these inhibitors are already available, this set of molecules provides an excellent data set for testing the prediction power of the QSAR model for new ligands. Table 4.6 represents the predicted  $\text{pIC}_{50}$  values of the test set based on equation (2). The overall root mean square error (RMSE) between the experimental and predicted  $\text{pIC}_{50}$  value was 0.325 which revealed good predictability. The squared correlation coefficient between experimental and predicted  $\text{pIC}_{50}$  values for the test set is also significant ( $r^2 = 0.871$ ). The Figure 4.4 shows the quality of the fit. The estimated correlation coefficient

between experimental and predicted pIC<sub>50</sub> values with intercept ( $r^2$ ) and without intercept ( $r_0^2$ ) are 0.871 and 0.862 respectively. The value of  $[(r^2-r_0^2)/r^2] = (0.871- 0.862)/0.871 = 0.010$ , which is less than 0.1 (stipulated value) (Jaiswal *et. al.*, 2004). Also the values of  $k$  and  $k'$  were 1.028 and 0.847, which are well within the specified range of 0.85 and 1.15 (Jaiswal *et. al.*, 2004). Since the value of  $q^2_{test} = 0.876$  and  $rm^2 = 0.788$  were found to be in the acceptable range (Shapiro *et. al.*, 1998), thereby indicating the good external predictability of the QSAR model.

**Table 4.7.** Correlation matrix of the descriptors used in the QSAR model.

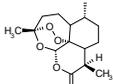
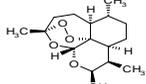
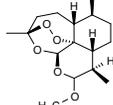
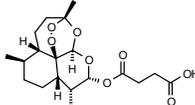
	V7CH	EMAX1	Log P	GEOM3	STRA6	STRA4	STRA2	L/B2	FVMN	HOMO	BOMX	MOLC9	V6C
V7CH	1.00												
EMAX1	-0.14	1.00											
Log P	0.05	0.17	1.00										
GEOM3	-0.06	0.15	0.20	1.00									
STRA6	0.06	-0.01	-0.07	0.28	1.00								
STRA4	0.29	0.09	-0.11	0.14	0.63	1.00							
STRA2	0.15	-0.01	0.04	0.22	0.25	0.12	1.00						
L/B2	-0.12	0.03	0.43	0.07	0.01	0.06	-0.16	1.00					
FVMN	0.06	-0.10	-0.09	-0.01	0.13	-0.01	-0.15	0.08	1.00				
HOMO	0.37	-0.02	0.21	-0.09	-0.26	-0.14	0.21	0.32	-0.08	1.00			
BOMX	0.01	-0.01	-0.03	0.07	-0.08	0.07	0.14	-0.06	-0.44	0.05	1.00		
MOLC9	-0.41	-0.04	-0.50	-0.13	-0.27	-0.35	-0.21	-0.18	0.12	-0.21	0.11	1.00	
V6C	0.22	-0.07	0.03	0.11	-0.09	0.03	-0.02	-0.13	-0.10	0.17	0.12	-0.05	1.00

To evaluate the accuracy of the QSAR model for antimalarial activities, we have taken a separate data set called validation set consisting of 4 analogues of artemisinin (Table 4.8). Their experimental activity and chemical structures were obtained from the literature (Darren *et. al.*, 2008). The experimental activity (IC<sub>50</sub> value) of these compounds obtained from in vitro study in parasitized whole blood (human) against drug resistant strains of *P. falciparum* (W-2 clone) (Desjardins *et. al.*, 1979; Milhous *et. al.*, 1985). The W-2 clone is chloroquinone-resistant. For all the compounds QSAR predictions produce exactly the same trend for antimalarial activity, even though the exact magnitudes of these values do not match very well to experimental values (Table 4.8). Coupled with the good predictive ability of the QSAR model developed in this study we believe that this model would perform well as rapid screening tools to uncover new and more potent anti-malarial drugs based on artemisinin derivatizations.



**Figure 4. 4.** Relationship between predicted and experimental activities as per equation (2).

**Table 4.8.** Observed and predicted activity against the drug-resistant malarial strain *P. falciparum* (W-2 clone) of validation set of artemisinin derivatives.

Compound name	W-2 Clone line inhibition (pIC <sub>50</sub> )			
	Structure	Observed	Predicted	Residual
artemisinin		1.004	0.228	0.776
Dihydroartemisinin		0.694	0.480	0.213
Artemether		1.638	0.606	1.03
Artesunate		0.259	1.119	0.861

#### 4.4. Conclusion

The QSAR analysis of a series of artemisinin derivatives enabled consistent models of structure- activity relationships to be obtained for several descriptors. The models that had the best predictive ability contained topological, thermodynamic, electronic, E-state indices and physicochemical descriptors. In this study, we used a more systematic way of variable selection in order of missing value test → zero test → simple correlation test → multicollinearity test → genetic algorithm to obtain QSAR models for 194 artemisinin derivatives. Using a combination of topological, electro-topological-state indices, electronic and thermodynamic descriptors of chemical structures, we have built several robust QSAR models with high values of  $q^2$  (for training sets) and predictive  $r^2$  (for test set). The high predictive ability of the models allows virtual screening of chemical databases or virtual libraries determined by either synthetic feasibility or commercial availability of starting materials to prioritize the synthesis of most promising candidates. Therefore, these models should facilitate the rational design of novel derivatives, guide the design of focused libraries based on the artemisinin skeleton and facilitate the search for related structures with similar biological activity from large databases.

## CHAPTER 5

### **The Binding Modes and Binding Affinities of Artemisinin Derivatives with *Plasmodium falciparum* Ca<sup>2+</sup>-ATPase (PfATP6)**

---

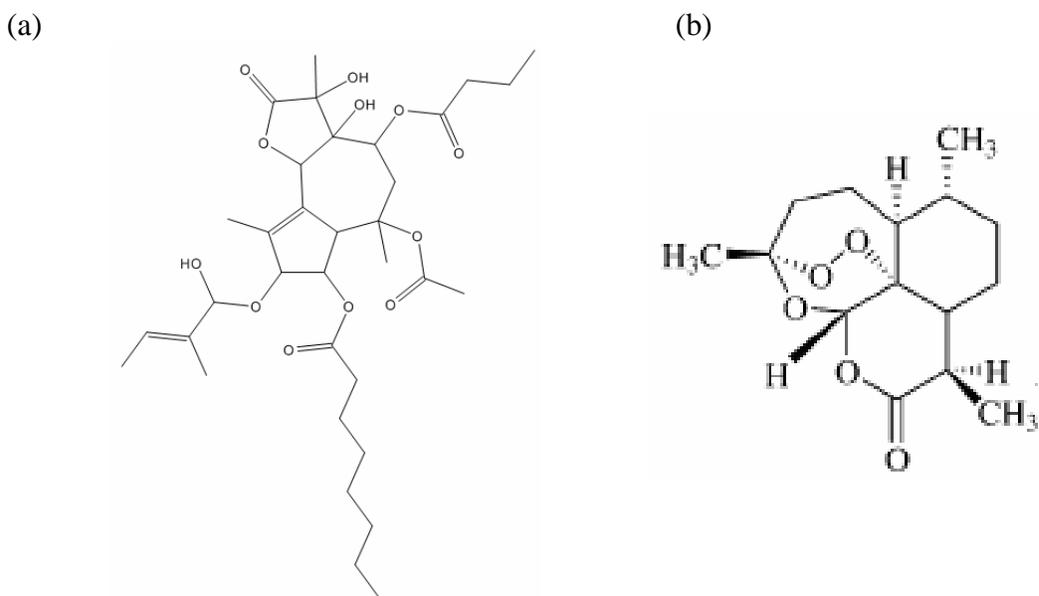
## Abstract

Noncompetitive inhibitors of sarco/endoplasmic reticulum  $\text{Ca}^{2+}$ -ATPase (SERCA) orthologue (PfATP6) of *P. falciparum* have important therapeutic value in the treatment of malaria. Artemisinin and its analogues are one such class of inhibitors that bind to a hydrophobic pocket located in the transmembrane region of PfATP6 near the biomembrane surface and interfere with calcium transport. The 3D structure of PfATP6 was modeled by homology modeling. A library of artemisinin analogues has been designed consisting of 154 analogues. Their molecular interactions and binding affinities with modeled PfATP6 protein have been studied using the docking, molecular mechanics based on generalized Born/surface area (MM-GBSA) solvation model and eMBrAcE. Docking and binding free energies scores show good relation with *in vitro* antimalarial activities. The main binding source of artemisinins to the PfATP6 is hydrophobic interaction and biologically important peroxide bonds were exposed to outside of the binding pocket. The study suggests binding of artemisinin to PfATP6 precedes activation of peroxide bond by  $\text{Fe}^{2+}$  species. Quantitative structure activity relationships were developed between the antimalarial activity ( $\log \text{RA}$ ) of these compounds and molecular descriptors like docking score and binding free energy. For both the cases the  $r^2$  was in the range of 0.538–0.688 indicating good data fit and  $r^2_{cv}$  was in the range of 0.525–0.679 indicating that the predictive capabilities of the models were acceptable. In addition, a scheme similar to Linear Response was used to develop a free energy of binding (FEB) relationship based electrostatic ( $\Delta G_{\text{ele}}$ ), van der Waal ( $\Delta G_{\text{vdw}}$ ) and surface accessible surface area (SASA), which can express the activity of these artemisinin derivatives. It can be seen that  $\Delta G_{\text{vdw}}$  has most significant correlation to the activity ( $\log \text{RA}$ ) and electrostatic energy ( $\Delta G_{\text{ele}}$ ) has less significant correlation to the activity. It indicates that the binding of these artemisinin derivatives to PfATP6 is almost hydrophobic.  $\Delta G_{\text{vdw}}$  may be a major drive force to their binding and contribution to their activity. Low levels of root mean square error for the majority of inhibitors establish the docking, Prime/MM-GBSA and eMBrAcE based prediction model as an efficient tool for generating more potent and specific inhibitors of PfATP6 by testing rationally designed lead compounds based on artemisinin derivatization.

## 5.1. Introduction

In the 1970s, Chinese scientists identified artemisinin (*qinghaosu*) from sweet wormwood (*Artemisia annua*), thereby giving us our most important class of antimalarial drugs. Use of artemisinin-containing therapies has increased exponentially (Klayman, 1985) but the mechanism of action of these sesquiterpene lactone endoperoxides is controversial (Arrow et. al., 2004). Some (Jefford et. al., 2001; Pandey et. al., 1999), but not all (Haynes et al., 2003; O'Neill et. al., 2000; Hawley et. al., 1998), studies suggest that artemisinin act by heme-dependent activation of an endoperoxide bridge occurring within the parasite's food vacuole. However, localization of artemisinins to parasite and not food vacuole membranes (Ellis et. al., 1985), and killing of tiny rings lacking haemozoin argue against the food vacuole being a major site for drug action (ter Kuile et. al., 1993).

An alternative hypothesis for the mode of action of artemisinin has been proposed, based on structural similarities between the sesquiterpene moieties of thapsigargin and in artemisinin (Figure 5.1).

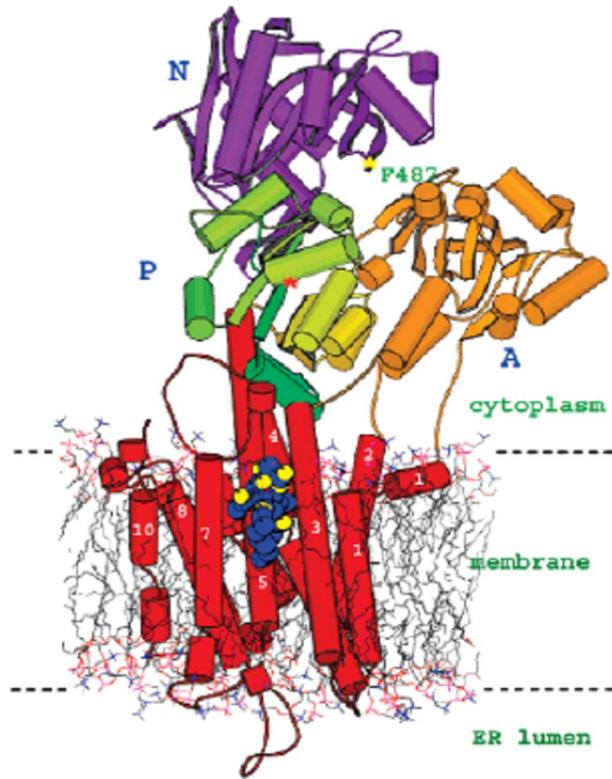


**Figure 5.1.** The 2D structure of (a) Thapsigargin (TG) and (b) Artemisinin showing the similarity between sesquiterpene moieties.

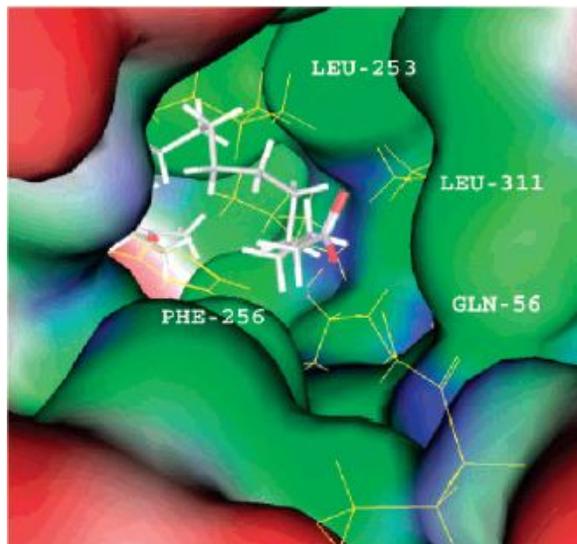
Thapsigargin (TG) (another plant product from *Thapsia garganica*) is an extremely potent inhibitor of  $\text{Ca}^{2+}$ -transporting ATPases (sarcoplasmic reticulum  $\text{Ca}^{2+}$ -transporting ATPases or SERCAs) from a wide variety of organisms. It was suggested that artemisinin may act in a

similar way, but more specifically to inhibit the SERCA of malarial (PfATP6) but not mammalian pumps (Eckstein-Ludwig et. al., 2003). PfATP6 is the only SERCA-type  $\text{Ca}^{2+}$ -ATPase sequence in the parasite's genome. Further the experimental studies revealed that artemisinin inhibit the sarco/endoplasmic reticulum  $\text{Ca}^{2+}$ -ATPase (SERCA) orthologue (PfATP6) of *P. falciparum* in *Xenopus* oocytes (Eckstein-Ludwig et. al., 2003). PfATP6 is thought to be the real molecular target of artemisinin in spite of some disagreements to be resolved (Ellis et. al., 1985).

The SERCA belongs to the family of P-type ATPases that are responsible for active transport of cations across biomembranes (Kuhlbrandt, 2004). The SERCA uses the energy released from hydrolysis of ATP to ADP for transporting calcium ions to the lumen of sarco- and endoplasmic reticulum (ER) against the electrochemical gradient. The publication of crystal structures of nucleotide free SERCA in calcium-bound form (Toyoshima et. al., 2000) (E1.2 $\text{Ca}^{2+}$ ) and in TG-bound form (Toyoshima et. al., 2002) (E2'.TG) as well as in complex with a nonhydrolyzable ATP analogue (E1.AMPPCP) (Toyoshima et. al., 2004; Sorensen et. al., 2004) and with ADP stabilized by aluminum fluoride (E1.AIFx.ADP) (Sorensen et. al., 2004) has elucidated the structures of key intermediates involved in the calcium transport cycle. The availability of such structural data has facilitated the understanding of conformational changes and dynamics involved at various steps of the transport cycle. The overall structure of SERCA consists of three cytoplasmic domains and 10 transmembrane helices (Figure 5.2). The two calcium binding sites in the E1 state of SERCA are about 5 Å apart and situated in the trans-membrane region, around 4 and 7 Å below the cytoplasmic surface of the membrane. The TG binding site is located in a hydrophobic cavity formed by transmembrane helices M3, M5, and M7. The polar end of the TG molecule is located near the membrane interface between residues Phe256 and Ile829 (Figure 5.3). Binding of TG to SERCA is mostly hydrophobic in nature with only one hydrogen bond formed between the Ile829 backbone and the carboxyl oxygen at the O-8 position of the TG molecule. The availability of structural information on SERCA facilitates understanding the structure-activity relationships (SAR) for SERCA inhibition and enables molecular modeling techniques to be applied for designing novel and more potent inhibitors.



**Figure 5.2.** Overall structure of SERCA. The thapsigargin molecule (shown in blue and yellow spheres) binds to SERCA in a cavity formed by residues on transmembrane helices 3, 5, and 7. The ATP binding site is located near residue F487 (yellow asterisk) in the nucleotide domain (N) in purple. The phosphorylating residue (Asp351) is situated near a location marked by an orange asterisk in the phosphorylation (P) domain, which is shown in green.



**Figure 5.3.** Close up view of TG binding site formed by mostly hydrophobic residues on M3 and M5 helices lining the channel.

The amino acid sequence of PfATP6 is known (Gardner et. al., 2002) but the three-dimensional structure is not available. In this study, therefore we have constructed the 3D structure of PfATP6 by homology modeling and taken for interaction study between artemisinin and PfATP6. Of utmost importance in a structure-based drug design is the reliable filtering of putative hits in terms of their predicted binding affinity (scoring problem) which is based on the *in silico*-generated near native protein-ligand configurations (docking problem). Most of scoring functions used in docking programs are designed to predict binding affinity by evaluating the interaction between a compound and a receptor. However, it should be noted that ligand receptor recognition process is determined not only by enthalpic effects but also by entropic effects. Moreover, the scoring functions have a simplified form for the energy function to facilitate high throughput evaluation of a large number of compounds in a single docking run. These functions may be problematic when used with contemporary docking programs, and can result in a decrease of virtual screening accuracy. To overcome this problem, more precise but time consuming computational methodologies are necessary. Here, we have used and evaluated several receptor-centric computational methodologies for computational modeling of artemisinin and its derivatives as potent inhibitor of PfATP6.

### **5.1.2. Overview of the methodologies tested in this study**

We applied several computational solutions from the Schrödinger software package (Schrödinger, LLC: Portland, OR). A brief overview of these methodologies is presented.

### **5.1.3. Glide Docking**

We used the Glide program (Friesner et. al., 2004) as our docking engine. The Glide docking algorithm performs a series of hierarchical searches for locations of possible ligand affinity within the binding site of a receptor. A rough positioning and scoring algorithm is applied during the initial search step, followed by torsional energy optimization on an OPLA-AA non-bonded potential energy grid for enduring candidate poses. The pose conformations of the very best candidates are further refined by using Monte Carlo sampling. Selection of the final docked pose is accomplished using a Glide score, which is a model energy function that combines empirical and force field based terms. The Glide score is a modified and extended version of the ChemScore function (Eldridge et. al., 1997).

#### 5.1.4. Multi-Ligand Bimolecular Association with Energetics (eMBrAcE)

The eMBrAcE (*MacroModel v9.1*) program calculates binding energies between ligands and receptors using molecular mechanics energy minimization for docked conformations. eMBrAcE applies multiple minimizations, during which each of the specified pre-positioned ligands is minimized with the receptor. For the energy-minimized structures, the calculation is performed first on the receptor ( $E_{protein}$ ), then on the ligand ( $E_{ligand}$ ), and finally on the complex ( $E_{complex}$ ). The energy difference is then calculated as:

$$\Delta E = E_{complex} - E_{ligand} - E_{protein}$$

#### 5.1.5. Prime MM-GBSA

This application is used to predict the free binding energy between a receptor and a ligand. MM-GBSA is a method that combines OPLS molecular mechanics energies ( $EMM$ ), surface generalized Born solvation model for polar solvation ( $G_{SGB}$ ), and a nonpolar solvation term ( $G_{NP}$ ). The  $G_{NP}$  term comprises the nonpolar solvent accessible surface area and van der Waals interactions. The total free energy of binding is calculated as:

$$\Delta G_{bind} = G_{complex} - (G_{protein} + G_{ligand})$$

$$G = E_{MM} + G_{SGB} + G_{NP}$$

### 5.2. Materials and Methods

#### 5.2.1. Sequence analysis

The protein sequence of PfATP6 of the organism *Plasmodium falciparum* was obtained from the PlasmoDB, the official database of the malaria parasite genome project (Toyoshima et. al., 2000; Toyoshima et. al., 2002). Gene PFA0310c located in *P. falciparum* chromosome 1 and annotated by Sanger encoded the only SERCA-type calcium transporting ATPase protein. This protein comprises of 1228 amino acids. The predicted amino acid sequence was downloaded from the web site (Toyoshima et. al., 2004). Sequence similarity search with BLAST in Protein Data Bank (PDB) database gives only one similar protein (43.5% identical), SERCA (PDB ID: 1IWO). This structure is determined at 1.3Å resolution and contains the highly specific inhibitor thapsigargin (TG). It has three functional domains, the  $\alpha$ -helix ion channel domain, where TG is located. The binding of TG to the ion channel domain is derived almost only through hydrophobic interaction with proposed hydrogen bonding of TG O8 and 1819 backbone amide hydrogen. We performed the pairwise alignment of PfATP6 with 1IWO as reference using

the homology module of PRIME (Schrodinger package). We initially build the structure of PfATP6 using 1IWO as template. The structure of the PfATP6  $\alpha$ -helix domain is very similar to the corresponding TG-binding site of SERCA. But the ATP-binding domain and calcium ion binding domain showed relatively low similarity to SERCA or other proteins. Therefore, we removed the mismatched sequence part (375-707) from the whole sequence and then constructed the three-dimensional structure of PfATP6. The sequence alignment after removing the part of mismatched sequence is shown in Figure 5.4.

### **5.2.2. Homology model construction**

The homology models of the proteins: PfATP6 built using Prime (Prime version 1.5, Macromodel version 9.1, Schrodinger, LLC, New York, NY, 2005) accessible through the Maestro interface (Schrodinger, Inc.). All water molecules were removed and the bound ligand (TG) was kept for the template. During the homology model building, Prime keeps the backbone rigid for the cases in which the backbone does not need to be reconstructed due to gaps in the alignment. The model was screened for unfavorable steric contacts and remodeled using a rotamer library database of PRIME. Explicit hydrogens were added to the protein and the protein model subjected to energy minimization using the Macromodel (Prime version 1.5) force-field OPLS 2005. Energy minimization and relaxation of the loop regions was performed using 300 iterations in a simple minimization method. Again the steepest descent was carried out until the energy showed stability in the sequential repetition. Model evaluation was performed in PROCHECK v3.4.4 (Laskowski et. al., 1993) producing plots that were analyzed for the overall and residue-by-residue geometry. Ramachandran Plot (Ramachandran et. al., 1963) provided by the program PROCHECK assured very good confidence for the predicted protein. There were only 0.3% residues in the disallowed region and 0.9% residues in generously allowed regions. Nevertheless, PROCHECK assured the reliability of the structure and the protein was subjected to VERIFY3D (Eisenberg et. al., 1997), available from NIH MBI Laboratory Servers.

1IWO	1	MEAAHSKSTEECLAYFGVSETTGLTPDQVKRHLEKYGHNELPAEEG	46
PfATP6	1	NEEVIKNAHTYDVEDVLKFLDVNKDNGLNKEELDRRLKYGLNELEVEKK	50
1IWO	47	KSLWELVIEQFEDLLVRILLAAACISFVLAWFEEGEETITA--FVEPFVI	94
PfATP6	51	KSIFELILNQFDDLLVKILLAAAFISFVLTLDDMKHKKIEICDFIEPLVI	100
1IWO	95	LLILIANAIVGVWQERNAENAIEALKEYEPEMGKVYRADRKSVDRIKARD	144
PfATP6	101	VLILILNAAVGVWQECNAEKSLEALKELOPTKARVLRDQKWEI--IDSKY	148
1IWO	145	IVPGDIVEVAVGDKVPADIRILSIKSTTLRVDSILTGESVSVIKHTEPV	194
PfATP6	149	LYVGDIIELSVGNKTPADARIIKIYSTSLKVEQSMLTGESCSVDKYAEKM	198
1IWO	195	PD--PRAVNQDKNNHLFSGTNIAAGKALGIVATTGVSTEIGKIRDQM--A	240
PfATP6	199	EDSYKNCEIQLKKNILFSSATAIVCGRCIAVVINIGMRTIEGHIQHAVIES	248
1IWO	241	ATEQDKTPLQOKLDEFGEQLSKVISLICVAVVLINIGHFNDPVHGGSWIR	290
PfATP6	249	NSEDQTPLQIKIDLFGQQLSKIIIFVICVTVWIIINFKHFSDFIH--GSFLY	297
1IWO	291	GAIYYFKIAVALAVAAIPEGLPAVITTCALGTRRMARKNAIVRSLPSVE	340
PfATP6	298	GCLYYFKISVALAVAAIPEGLPAVITTCALGTRRMVKKNAIVRKLQSV	347
1IWO	341	TLGCTSVICSDKTGTLTINQMSVCKMFIIDKVDGDFCSLNEFSITGSTYA	390
PfATP6	348	TLGCTTVICSDKTGTLTINQMT-----	369
1IWO	391	PEGEVLKNDKPIRSGQFDGLVELATICALCNDSSLDNFNETKGVYEKVG	440
PfATP6	370	-----	369
1IWO	441	TETALTLVEKMNVFNTEVRNLSKVERANACNSVIRQLMKKEFTLEFSRD	490
PfATP6	370	-----TTVFK-----	375
1IWO	491	RKSMNVYCSPAKSSRAAVGNKMFVKGAGEVIDRCNYVRVGTTRVPNTGP	540
PfATP6	376	-KEIILYC-----KGAPENIKNCKYVLTNRDIRPLNET	408
1IWO	541	VKEKILSVIKWGTGRDTRCLALATRDTPPKREEMVLDSSSRFMEYETD	590
PfATP6	409	LKNEIHNKIQ--NMGKRALRTLSFAYKLSK--DLNIKNTDDYYKLEQD	454
1IWO	591	LTFVGVVGHLDPPRKEVHGSIQLCRDAGIRVIMITGDNKGTAIACRRIG	640
PfATP6	455	LIVLGGGLIIDPPRKYVGRAIRLCHMAGIRVFMITGDNINTARATAIKEIN	504
1IWO	641	IFGENEEVADR-----AYTGREFDDLPLAEQREACRRA--CCF	676
PfATP6	505	ILNKNEGDDEKDNVTNNKNTQICCYNGREFEDFSLEKQKHILKNTPRIVF	554
1IWO	677	ARVEPSHKSKIVEYLQSYDEITANTGDGVNDAPALKKAEIGIANG--SGTA	725
PfATP6	555	CRTEPKHKKQIVKVLKDLGETVANTGDGVNDAPALKSADIGIANGINGTE	604
1IWO	726	VARTASEMVLADNDFSTIVAAVEEGRAIYNNHKQFIRYLISNVGEVVC	775
PfATP6	605	VAKEASDIVLADNDFNTIVEAIKEGRCIYNNHKAFIRYLISNIGEVASI	654
1IWO	776	FLTAALGLPEALIPVQLLWVNLVTDGLPATALGFNPPDLINDRPPRSK	825
PfATP6	655	FITALLGIPDSLAPVQLLWVNLVTDGLPATALGFNPPENDVMKCKPRHEN	704
1IWO	826	EPLISGWLFFRYMAIGGYVGAATVGAAMWFMYAEDGPG---VTYHQLTH	872
PfATP6	705	DNLINGLTLRLRYIIIGTVYGIATVSIYVYVWFLFYPSDMHTLINFYQLSH	754
1IWO	873	FMQC-----TEDHPHFEGLDCEIFEAP--EPNTHALSVLVT	906
PfATP6	755	YNQCKAOWNFRVNVKYDMSDEH-----CSYFSAGKIKASTLSLSVLV	797
1IWO	907	IEMCNALNSLENQSLMRNPPVWNIWLLGSICLSMSLHFLILYVDPLPMI	956
PfATP6	798	IEMFNALNALSEYNSLFEIPPWRNHYLVLATIGSLLLHVLILYIPPLARI	847
1IWO	957	FKLKALDLTQQLMVLKISLFPVIGLDEILKFIARNYLEG	994
PfATP6	848	FGVVPLSAYDWFVFLWSFPVVIILDEIKFYAKRKLKEEQRTKKIKID	895

Figure 5.4. Alignment of PfATP6 sequence with 1IWO as reference protein.

### 5.2.3. Ligand binding site prediction

Site directed mutagenesis studies in catalytic site of PfATP6 of *Plasmodium falciparum* have revealed that Leu 263 is the critical residue involved in binding of artemisinin with PfATP6 (Uhlemann et. al., 2005). In silico prediction of the binding site was done for the PfATP6 in *P. falciparum* using SiteMap (Schrodinger package). SiteMap treat entire proteins to locate binding sites whose size, functionality, and extent of solvent exposure meet user specifications. SiteScore, the scoring function used to assess a site's propensity for ligand binding, accurately ranks possible binding sites to eliminate those not likely to be pharmaceutically relevant. It identifies potential ligand binding sites by linking together "site points" that are suitably close to the protein surface and sufficiently well sheltered from the solvent. Given that similar terms dominate the site scoring function, this approach ensures that the search focuses on regions of the protein most likely to produce tight protein-ligand or protein-protein binding. Subsites are merged into larger sites when they are sufficiently close and could be bridged in solvent-exposed regions by ligand atoms. SiteMap evaluates sites using a series of properties. The binding site with highest site score was taken for docking of the artemisinin analogues. The algorithm proceeds as follows: the protein is projected onto a 3D grid with a step size of 1.0 Å; grid points are labeled as protein, surface, or solvent using certain rules. A grid point is marked as protein if there is at least one atom within 1.6 Å. After the solvent excluded surface is calculated the surface vertices' coordinates are stored. A sequence of grid points, which starts and ends with surface grid points and which has solvent grid points in between, is called a surface-solvent-surface event. If the number of surface-solvent-surface events of a solvent grid exceeds a minimal threshold of 6, then this grid is marked as pocket. Finally, all pocket grid points are clustered according to their spatial proximity. The clusters are ranked by the number of grid points in the cluster. The top three clusters are retained and their centers of mass are used to represent the predicted pocket sites. The binding pocket obtained by in silico studies on PfATP6 of *Plasmodium falciparum* was consistent with the site directed mutagenesis studies.

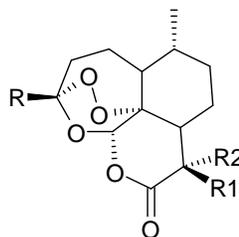
### 5.2.4. Preparation of the ligands

An initial dataset of 158 artemisinin analogues were collected from published data (Woolfrey et. al., 1998; Acton et. al., 1993; Lin et. al., 1989; Posner et. al., 1992; Avery et. al.,

1995; Avery et. al., 1996) in which several different ring systems were represented. All of the analogues were either peroxides or trioxanes, which should act via similar mechanisms of action and were categorized into 10 classes (Table 5.1). Each of these compounds had associated in vitro bioactivity values (IC<sub>50</sub> values reported in ng/ml) against the drug resistant malaria strain *P. falciparum* (W-2 clone). The log value of the relative activity (RA) of these compounds was used for analysis and was defined as:

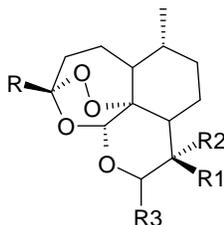
$$\text{Log(RA)} = \log[(\text{artemisinin IC}_{50}/\text{analogue IC}_{50})(\text{analogue MW}/\text{artemisinin MW})].$$

**Table 5.1.** Artemisinin analogues with anti-malarial activities against the drug resistant malarial strain *P. falciparum* (W-2 clone) used in this work.



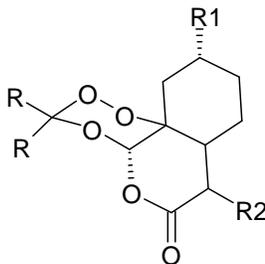
Sl No.	R	R1	R2	Log RA	pIC50 (ng/ml)
1	CH <sub>3</sub>	CH <sub>3</sub>	H	1.000	1.398
2	C <sub>4</sub> H <sub>8</sub> Ph	H	H	0.450	0.712
3	CH <sub>3</sub>	H	2-Z-Butenyl	-1.10	-0.760
4	CH <sub>3</sub>	H	H	0.790	1.188
5	CH <sub>3</sub>	H	2-E-Butenyl	-0.600	-0.260
6	CH <sub>3</sub>	Allyl	H	-0.100	0.260
7	CH <sub>3</sub>	C <sub>4</sub> H <sub>9</sub>	H	0.170	0.508
8	C <sub>4</sub> H <sub>8</sub> Ph	C <sub>4</sub> H <sub>9</sub>	H	-0.320	-0.117
9	C <sub>3</sub> H <sub>6</sub> (P-Cl-Ph)	C <sub>4</sub> H <sub>9</sub>	H	-0.280	-0.097
10	CH <sub>2</sub> CH <sub>2</sub> CO <sub>2</sub> Et	C <sub>4</sub> H <sub>9</sub>	H	1.360	1.595
11	CH <sub>3</sub>	C <sub>2</sub> H <sub>5</sub>	H	1.400	1.777
12	CH <sub>3</sub>	C <sub>6</sub> H <sub>13</sub>	H	0.860	1.162
13	CH <sub>3</sub>	i- C <sub>4</sub> H <sub>9</sub>	H	-0.550	-0.212
14	CH <sub>3</sub>	i-C <sub>6</sub> H <sub>13</sub>	H	-0.040	0.262
15	CH <sub>3</sub>	i-C <sub>3</sub> H <sub>7</sub>	H	-0.040	0.317
16	CH <sub>3</sub>	i-C <sub>5</sub> H <sub>11</sub>		0.070	0.389
17	CH <sub>2</sub> CH <sub>2</sub> CO <sub>2</sub> Et	H	H	0.370	0.669
18	C <sub>2</sub> H <sub>5</sub>	H	H	0.050	0.448
19	C <sub>3</sub> H <sub>7</sub>	H	H	0.830	1.207
20	CH <sub>3</sub>	C <sub>3</sub> H <sub>6</sub> (p-Cl-Ph)	H	1.370	1.595
21	CH <sub>3</sub>	CH <sub>2</sub> CH <sub>3</sub>	R <sub>1</sub> =R <sub>2</sub>	-0.360	-0.022
22	CH <sub>3</sub>	C <sub>5</sub> H <sub>11</sub>	H	1.020	1.339
23	CH <sub>3</sub>	C <sub>4</sub> H <sub>8</sub> Ph	H	0.630	0.876
24	CH <sub>3</sub>	C <sub>2</sub> H <sub>4</sub> Ph	H	0.120	0.398
25	CH <sub>3</sub>	C <sub>3</sub> H <sub>6</sub> Ph	H	0.780	1.042

**Table 5.1** (continued). 10-Substituted artemisinin derivatives with anti-malarial activities against the drug resistant malarial strain *P. falciparum* (W-2 clone) used in this work.



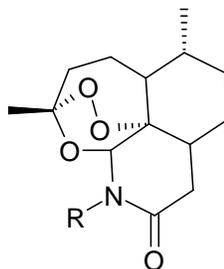
Sl. No.	R	R1	R2	R3	Log RA	pIC <sub>50</sub> (ng/ml)
26	CH <sub>3</sub>	CH <sub>3</sub>	H	H	0.750	1.170
27	CH <sub>3</sub>	CH <sub>3</sub>	H	OH	0.550	0.945
28	CH <sub>3</sub>	CH <sub>3</sub>	H	OEt	0.340	0.694
29	CH <sub>3</sub>	CH <sub>3</sub>	H	OH	0.960	1.295
30	CH <sub>3</sub>	CH <sub>3</sub>	H	OEt	-1.080	-0.740
31	CH <sub>3</sub>	H	Br	H	0.280	0.606
32	CH <sub>3</sub>	CH <sub>3</sub>	Br	NH-2-(1,3-thiazole)	0.660	0.874
33	CH <sub>3</sub>	CH <sub>3</sub>	Br	aniline	0.180	0.401
34	CH <sub>3</sub>	Br	CH <sub>3</sub>	NH-2-pyridine	-0.090	0.115
35	CH <sub>3</sub>	CH <sub>3</sub>	Br	NH-2-pyridine	-0.770	-0.564
36	CH <sub>3</sub>	CH <sub>3</sub>	H	OMe	0.280	0.654
37	CH <sub>3</sub>	CH <sub>3</sub>	H	α -OEt	0.320	0.674
38	CH <sub>3</sub>	C <sub>4</sub> H <sub>9</sub>	H	H	1.320	1.677
39	CH <sub>3</sub>	C <sub>2</sub> H <sub>5</sub>	H	H	0.670	1.068
40	CH <sub>3</sub>	C <sub>3</sub> H <sub>7</sub>	H	OEt	-0.040	0.277
41	CH <sub>3</sub>	C <sub>2</sub> H <sub>5</sub>	H	OEt	0.500	0.835
42	CH <sub>3</sub>	CH <sub>3</sub>	H	C <sub>3</sub> H <sub>6</sub> OH	0.780	1.115
43	CH <sub>3</sub>	CH <sub>3</sub>	H	OCH <sub>2</sub> CO <sub>2</sub> Et	0.520	0.800
44	CH <sub>3</sub>	CH <sub>3</sub>	H	OC <sub>2</sub> H <sub>4</sub> CO <sub>2</sub> Me	0.100	0.364
45	CH <sub>3</sub>	CH <sub>3</sub>	H	OC <sub>3</sub> H <sub>6</sub> CO <sub>2</sub> Me	-0.030	0.218
46	CH <sub>3</sub>	CH <sub>3</sub>	H	OCH <sub>2</sub> (4-PhCO <sub>2</sub> Me)	-0.070	0.143
47	CH <sub>3</sub>	CH <sub>3</sub>	H	(R)-OCH <sub>2</sub> CH(CH <sub>3</sub> )CO <sub>2</sub> Me	1.790	2.070
48	CH <sub>3</sub>	CH <sub>3</sub>	H	(S)-OCH <sub>2</sub> CH(CH <sub>3</sub> )CO <sub>2</sub> Me	2.250	2.530
49	CH <sub>3</sub>	CH <sub>3</sub>	H	(R)-OCH(CH <sub>3</sub> )CH <sub>2</sub> CO <sub>2</sub> Me	0.870	1.134
50	CH <sub>3</sub>	CH <sub>3</sub>	H	(S)-OCH(CH <sub>3</sub> )CH <sub>2</sub> CO <sub>2</sub> Me	1.700	1.964
51	CH <sub>2</sub> CH <sub>2</sub> CO <sub>2</sub> Et	H	H	H	0.700	1.017
52	C <sub>3</sub> H <sub>6</sub> (p-Cl-Ph)	H	H	H	-0.550	-0.295
53	C <sub>4</sub> H <sub>9</sub>	H	H	H	0.750	1.127
54	C <sub>2</sub> H <sub>5</sub>	H	H	H	-1.000	-0.580
55	i-C <sub>4</sub> H <sub>9</sub>	H	H	H	0.400	0.777
56	C <sub>3</sub> H <sub>7</sub>	H	H	H	0.840	1.238
57	C <sub>4</sub> H <sub>8</sub> Ph	H	H	H	0.580	0.858
58	CH <sub>3</sub>	-CH <sub>2</sub> O-		OOH	-0.570	-0.219
59	CH <sub>3</sub>	=CH <sub>2</sub>		OOH	-0.990	-0.616
60	-	CH <sub>3</sub>	OH	α-OH	-0.890	-0.519
61	CH <sub>3</sub>	C <sub>5</sub> H <sub>11</sub>	H	H	0.160	0.498
62	CH <sub>3</sub>	C <sub>3</sub> H <sub>6</sub> Ph	H	H	1.400	1.678
63	CH <sub>3</sub>	C <sub>3</sub> H <sub>7</sub>	H	H	0.740	1.117
64	-	CH <sub>3</sub>	OH	CH <sub>2</sub> CF <sub>3</sub>	0.330	0.615
65	-	OH	CH <sub>3</sub>	CH <sub>2</sub> CF <sub>3</sub>	-0.700	-0.415
66	-	CH <sub>3</sub>	OH	OEt	-0.440	-0.415
67	CH <sub>3</sub>	CH <sub>3</sub>	H	OOt-C <sub>4</sub> H <sub>9</sub>	0.920	1.217

**Table 5.1** (continued). Seco-artemisinin derivatives with anti-malarial activities against the drug resistant malarial strain *P. falciparum* (W-2 clone) used in this work.



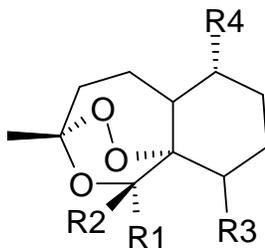
Sl. No.	R	R1	R2	Log RA	pIC <sub>50</sub> (ng/ml)
68	CH <sub>3</sub>	H	H	-2.370	-1.906
69	C <sub>2</sub> H <sub>5</sub>	H	H	-1.130	-0.713
70	-	-	-	-0.260	0.097

**Table 5.1** (continued). 11-Aza artemisinin derivatives with anti-malarial activities against the drug resistant malarial strain *P. falciparum* (W-2 clone) used in this work.



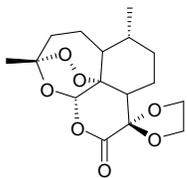
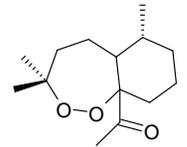
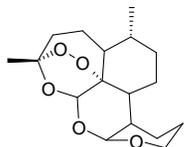
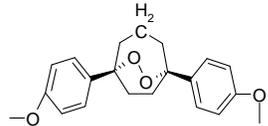
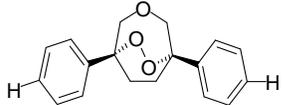
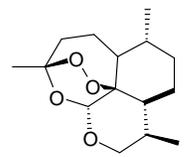
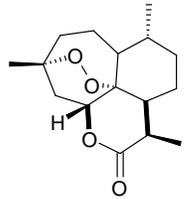
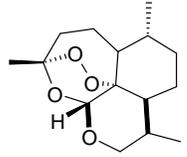
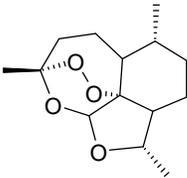
Sl. No.	R	Log RA	pIC <sub>50</sub> (ng/ml)
71	C <sub>3</sub> H <sub>6</sub> Ph	0.020	0.283
72	C <sub>2</sub> H <sub>4</sub> Ph	0.160	0.439
73	C <sub>3</sub> H <sub>11</sub>	-0.200	0.121
74	CH <sub>2</sub> (p-Cl-Ph)	-0.160	0.096
75	CH <sub>2</sub> Ph	0.340	0.636
76	CH <sub>2</sub> -(2-C <sub>3</sub> H <sub>4</sub> N)	1.460	1.487
77	2-Thiophene	0.170	0.458
78	Acetaldehyde	1.470	1.828

**Table 5.1** (continued). Artemisinin derivatives lacking the D-ring with anti-malarial activities against the drug resistant malarial strain *P. falciparum* (W-2 clone) used in this work.



Sl. No.	R1	R2	R3	R4	Log RA	pIC <sub>50</sub> (ng/ml)
79	-O <sub>2</sub> CCH <sub>2</sub> Ph	H	H	CH <sub>3</sub>	-0.510	-0.217
80	H	H	H	CH <sub>3</sub>	-0.320	0.202
81	H	OCH <sub>3</sub>	H	H	-0.310	0.180
82	C <sub>2</sub> H <sub>4</sub> OH	H	CH <sub>3</sub>		-1.800	-1.429
83	C <sub>2</sub> H <sub>4</sub> OH	CH <sub>3</sub>	H		0.230	0.601
84	C <sub>2</sub> H <sub>4</sub> OH	CH <sub>3</sub>	CH <sub>3</sub>		-1.800	-1.449
85	C <sub>2</sub> H <sub>4</sub> OCH <sub>2</sub> Ph	CH <sub>3</sub>	CH <sub>3</sub>		-1.800	-1.558
86	OCH <sub>3</sub>	H	C <sub>2</sub> H <sub>4</sub> O <sub>2</sub> CNEt <sub>2</sub>	H	0.650	0.929
87	OCH <sub>3</sub>	H	C <sub>2</sub> H <sub>4</sub> O <sub>2</sub> CNPh <sub>2</sub>		0.650	0.829
88	H	OCH <sub>3</sub>	C <sub>2</sub> H <sub>4</sub> OCH <sub>2</sub> Ph	H	0.750	1.039
89	H	OCH <sub>3</sub>	C <sub>2</sub> H <sub>4</sub> O-allyl	H	0.400	0.735
90	H	OCH <sub>3</sub>	C <sub>2</sub> H <sub>4</sub> O <sub>2</sub> Ph	H	-0.590	-0.319
91	H	OCH <sub>3</sub>	C <sub>2</sub> H <sub>4</sub> O <sub>2</sub> C(4-PhCO <sub>2</sub> C <sub>2</sub> H <sub>4</sub> NMe <sub>2</sub> )		-0.600	-0.446
92	H	OCH <sub>3</sub>	C <sub>2</sub> H <sub>4</sub> O <sub>2</sub> CCH <sub>2</sub> NCO <sub>2</sub> -(t-C <sub>4</sub> H <sub>9</sub> )	H	-0.040	0.174
93	OCH <sub>3</sub>	-	C <sub>2</sub> H <sub>4</sub> OCH <sub>2</sub> (4-F-Ph)		0.380	0.648
94	OCH <sub>3</sub>	-	C <sub>2</sub> H <sub>4</sub> OCH <sub>2</sub> (4-Py)		0.140	0.428
95	H	OCH <sub>3</sub>	C <sub>2</sub> H <sub>4</sub> OCH <sub>2</sub> (4-N-Me-pyridine)	H	-0.900	-0.647

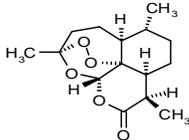
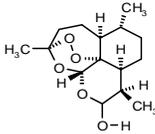
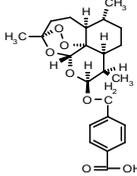
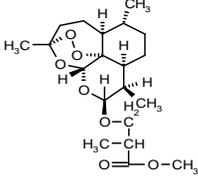
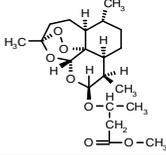
**Table 5.1** (continued). Miscellaneous artemisinin derivatives with anti-malarial activities against the drug resistant malarial strain *P. falciparum* (W-2 clone) used in this work.

Sl. No.	Ligand structure	Log RA	pIC <sub>50</sub> (ng/ml)
96		-2.090	-1.755
97		-1.270	-0.802
98		0.230	0.587
99		-0.670	-0.353
100		-2.260	-1.862
101		-0.240	0.180
102		-0.960	-0.559
103		-0.790	-0.370
104		-0.353	0.090

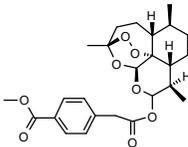
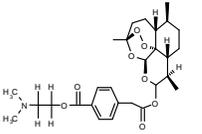
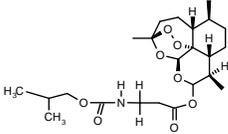
**Table 5.1** (continued). 9-substituted artemisinin derivatives with anti-malarial activities against the drug resistant malarial strain *P. falciparum* (W-2 clone) used in this work.

Sl. No.	Ligand structure	Log RA	pIC <sub>50</sub> (ng/ml)
105		-0.739	-0.365
106		-2.219	-1.821
107		-2.447	-2.106
108		-0.198	0.182
109		-0.717	-0.325
110		-1.487	-1.282
111		-0.460	-0.109
112		-0.409	-0.058
113		-0.361	0.013

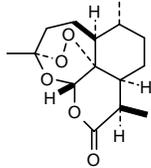
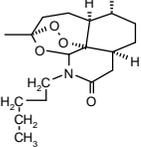
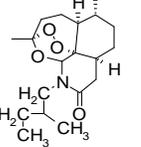
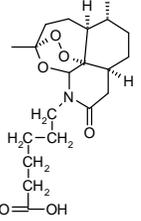
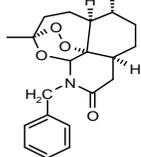
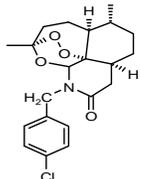
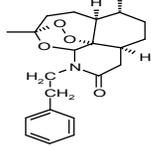
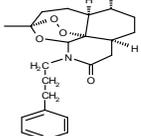
**Table 5.1** (continued). Dihydroartemisinin derivatives with anti-malarial activities against the drug resistant malarial strain *P. falciparum* (W-2 clone) used in this work.

Sl. No.	Ligand structure	Log RA	pIC <sub>50</sub> (ng/ml)
114		-0.269	0.129
115		0.310	0.705
116		0.176	0.404
117		1.524	1.788
118		0.599	0.863

**Table 5.1** (continued). Tricyclic 1,2,4-trioxane derivatives with anti-malarial activities against the drug resistant malarial strain *P. falciparum* (W-2 clone) used in this work.

Sl. No.	Ligand structure	Log RA	pIC <sub>50</sub> (ng/ml)
119		0.660	0.845
120		0.205	0.340
121		0.312	0.503

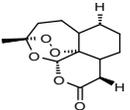
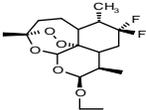
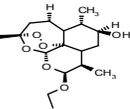
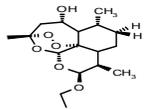
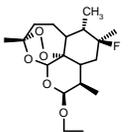
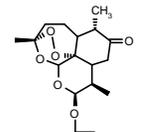
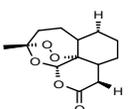
**Table 5.1** (continued). N-alkyl-11-aza-9-desmethylartemisinin derivatives with anti-malarial activities against the drug resistant malarial strain *P. falciparum* (W-2 clone) used in this work.

Sl. No.	Ligand structure	Log RA	pIC <sub>50</sub> (ng/ml)
122		0.000	0.398
123		0.041	0.362
124		0.173	0.494
125		-0.921	-0.652
126		0.276	0.572
127		0.045	0.301
128		0.294	0.573
129		0.312	0.574

**Table 5.1** (continued). 3C-substituted artemisinin derivatives with anti-malarial activities against the drug resistant malarial strain *P. falciparum* (W-2 clone) used in this work.

Sl. No.	Ligand structure	Log RA	pIC50 (ng/ml)	Sl. No.	Ligand structure	Log RA	pIC50 (ng/ml)
130		0.049	0.447	137		0.449	0.710
131		0.828	1.205	138		0.410	0.729
132		-0.745	-0.385	139		-0.481	-0.197
133		-0.347	0.010	140		-2.000	-1.769
134		0.365	0.665	141		-0.276	-0.093
135		-2.000	-1.706	142		-0.319	-0.116
136		0.104	0.343	143		1.359	1.594

**Table 5.1** (continued). Various derivatives of artemisinin and artemether with antimalarial activity against the drug resistant malarial strain *P. falciparum* (W-2 clone) used in the work.

Compound no.	Analogue structure	Log RA	pIC <sub>50</sub> (ng/ml)	Compound no.	Analogue structure	Log RA	pIC <sub>50</sub> (ng/ml)
144		0.437	0.083	148		1.549	0.497
145		2.188	0.672	149		0.054	-0.938
146		-0.120	0.192	150		0.160	0.495
147		0.016	-1.347				

Molecular models of the artemisinin and its analogues (Table 5.1) were built using the Builder feature in Maestro (Schrodinger package) and energy minimized in a vacuum using Impact. Each structure was assigned an appropriate bond order using ligprep script shipped by Schrödinger and optimized initially by means of the OPLS 2005 force field using default setting. Complete geometrical optimization of these structures was carried out with the HF/3-21G method (in this work) using the Jaguar (Schrodinger Inc.). In order to check the reliability of the geometry obtained, we compared the structural parameters of the artemisinin 1,2,4-trioxane ring with theoretical (Pinheiro et. al., 2001) and experimental (Leban et. al., 1988; Lisgarten et. al., 1998) values from the literature. All calculations reproduced most of the structural parameters of the artemisinin 1,2,4-trioxane ring seen in X-ray structures (Table 5.2). This applies especially to the bond length of the endoperoxide bridge which seems to be responsible for the antimalarial activity (Bernardinelli et. al., 1994; Posner et. al., 1995; Posner et. al., 1995; Haynes et. al., 1996; Rafiee et. al., 2005).

### 5.2.5 Docking of the ligands

All the ligands were docked to the PfATP6 receptor using Glide. After ensuring that protein and ligands are in correct form for docking, the receptor-grid files were generated

using grid-receptor generation program, using van der Waals scaling of the receptor at 0.4. The residues as mentioned in the ligplot (Figure 5.6) are included in the grid box. The default size was used for the bounding and enclosing boxes. The ligands were docked initially using the “standard precision” method and further refined using “xtra precision” Glide algorithm. For the ligand docking stage, van der Waals scaling of the ligand was set at 0.5. Of the 50,000 poses that were sampled, 4,000 were taken through minimization (conjugate gradients 1,000) and the 30 structures having the lowest energy conformations were further evaluated for the favorable Glide docking score. A single best conformation for each ligand was considered for further analysis.

### **5.2.6 Ligand & Structure-Based Descriptors (LSBD) protocol**

The eMBrAcE, Prime MM-GBSA and Liaison calculations were performed using the Ligand & Structure-Based Descriptors (LSBD) application of the Schrödinger software package. These calculations were applied the ligand-receptor complex structures obtained from Glide docking.

### **5.2.7 MM and Binding Free Energies**

After obtaining preferable binding structure from docking simulation, the complex was partially minimized by relaxing ligand and atoms of side chains that are within 7 Å away from the ligand while all other atoms were fixed. We compared the energies calculated based on pre-minimized structures and minimized structures, the later has better correlation with activity of ligands. So the minimized structures were used in all energy calculations. After all energies were calculated, factor analysis (FA) and multiple regression analysis (MRA) were used to derive a LRE-like equation, which produce a reasonable FEB which has good correlation with the activity of these compounds.

In order to explore the reliability of the proposed model we used the cross validation method. Prediction error sum of squares (PRESS) is a standard index to measure the accuracy of a modeling method based on the cross validation technique. The  $r^2_{cv}$  was calculated based on the PRESS and SSY (sum of squares of deviations of the experimental values from their mean) using following formula:

$$r_{cv}^2 = 1 - \frac{PRESS}{SSY} = 1 - \frac{\sum_{i=1}^n (y_{exp} - y_{pred})^2}{\sum_{i=1}^n (y_{exp} - \bar{y})^2}$$

where  $y_{exp}$ ,  $y_{pred}$  and  $y$  are the predicted, observed and mean values of the antimalarial activities of the artemisinin analogues. The cross validation analysis performed by using the leave one out (LOO) method in which one compound removed from the data set and its activity predicted using the model derived from the rest of the data points. The cross-validated correlation coefficient ( $q^2$ ) that resulted in optimum number of components and lowest standard error of prediction were considered for further analysis and calculated using following equations:

$$q^2 = 1 - \frac{\sum_y (y_{pred} - y_{observed})^2}{\sum_y (y_{observed} - y_{mean})^2}$$

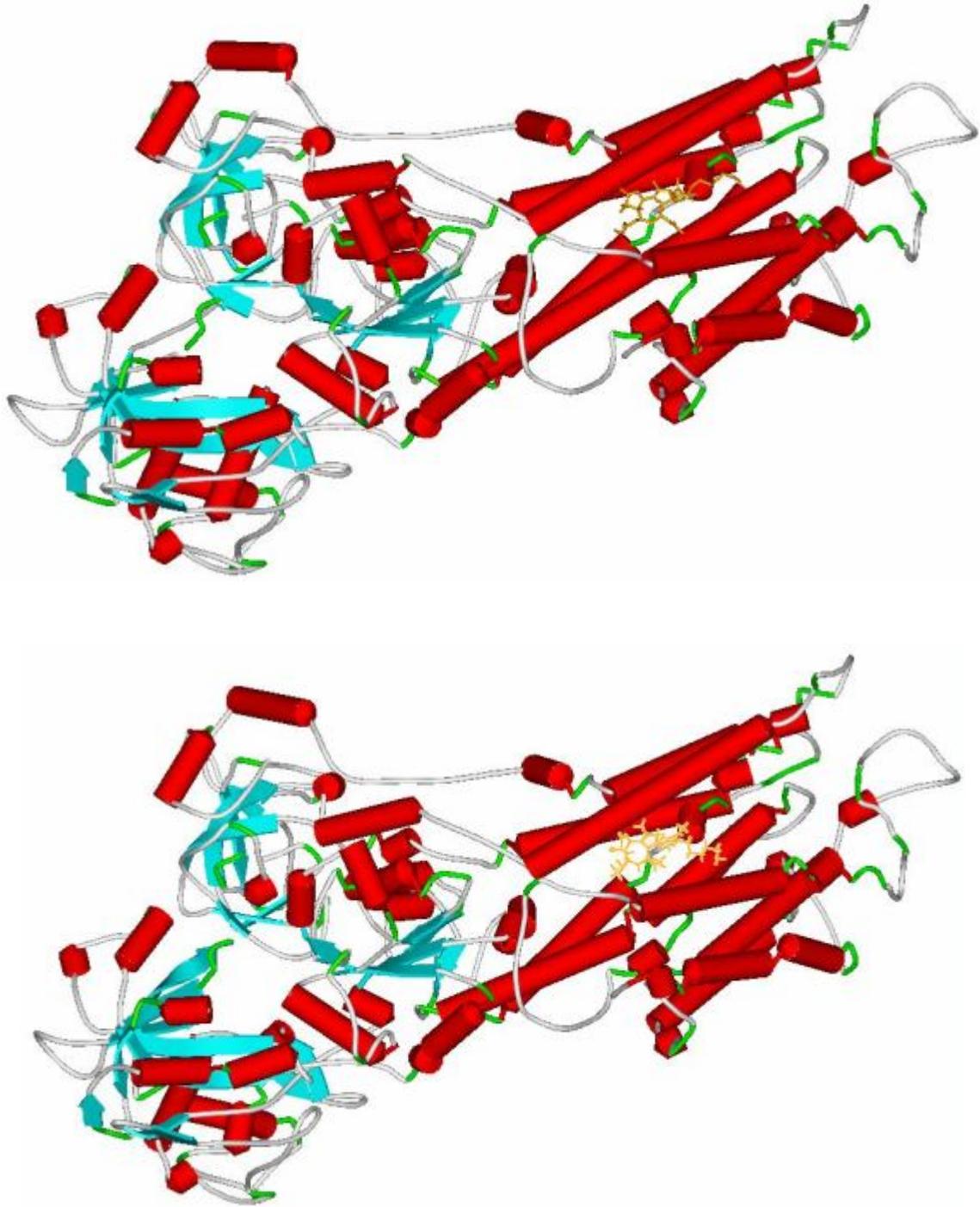
$$PRESS = \sum_y (y_{predicted} - y_{observed})^2$$

where  $y_{pred}$ ,  $y_{observed}$  and  $y_{mean}$  are the predicted, observed and mean values of the antimalarial activities of the analogues and PRESS is the sum of the predictive sum of squares.

### 5.3 Results and Discussion

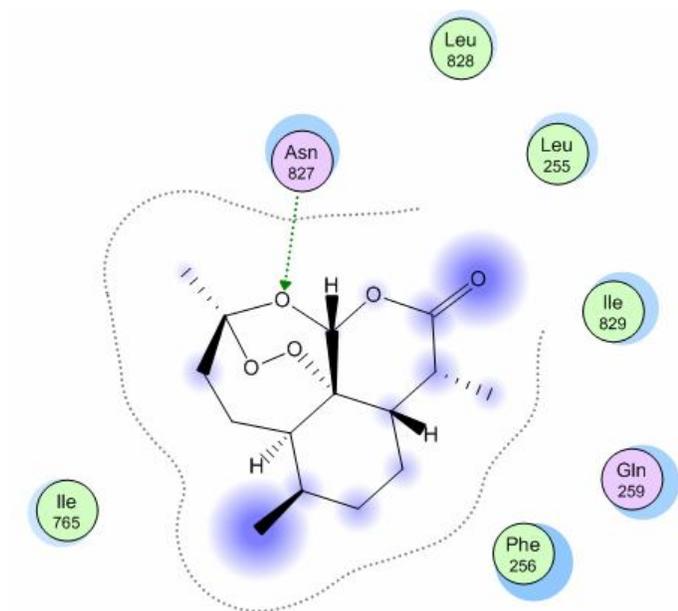
The atomic coordinates of PfATP6 for the organism *Plasmodium falciparum* was not available in Protein Data Bank, which necessitated developing a protein model. Homology modeling protocol was employed to predict the model of the protein. An identity of 43.5% with well-studied protein of SERCA (pdb id: 1iwo; resolution 1.3 Å) provided a great strength for modeling the protein. Three dimensional structure predictions by comparative modeling was done by PRIME 1.5 (Schrodinger package). Macromodel (Schrodinger package) was used in energy minimization embraces a range of force fields. Macromodel module is a molecular mechanics simulation environment offering energy minimization, dynamics and conformational search on molecular, aggregate or periodic systems. Due to some unfavorable steric contacts we had to edit the raw protein manually, with out disturbing the critical residues. Simple minimization followed by a detailed minimization was done until we could find no significant variation in energy. Atomic charges were assigned using OPLS-AA force field. The final model, which we took for further analysis, consisted of 895 amino acid residues. We used both PROCHECK and the VERIFY3D softwares to check the quality of the modeled protein. Ramachandran Plot obtained from the program PROCHECK, which checks the stereochemical quality of a protein structures, producing a number of postscript

plots, analyzing its overall and residue-by-residue geometry, assured the reliability of the modeled protein with 91.1% residues in most allowed region and 7.7% in additional allowed region. There were only 0.3% residues in disallowed region and 0.9% in generously allowed region. The assessment with VERIFY3D, which derives a ‘‘3D-1D’’ profile based on the local environment of each residue, described by the statistical preferences for: the area of the residue that is buried, the fraction of side-chain area that is covered by polar atoms (oxygen and nitrogen), and the local secondary structure, also substantiated the reliability of the three dimensional structure. The residues that deviated from the standard conformational angles of Ramachandran plot were the members of N terminal domain of the protein. This was an ignorable condition since the N-terminal end was not critical in our study. The distance of these residues to the active site residues also were found to be more than 10 Å, which suggested that those residues would interfere little with the binding of ligands in the active site region of PfATP6. Figure 5.4 compares the structures of SERCA and homology-modeled PfATP6. Active site was identified with reference to the studies done on SERCA (pdb id: 1IWO). We carried out in silico studies to confirm these active sites, using SiteMap algorithm. The output from the SiteMap program showed coherent active sites for the target protein as reported in site directed mutagenesis study (Uhlemann et. al., 2005). The structural comparison of template protein and PfATP6 model showed significant similarity in the binding site residues (Figure 5.5).

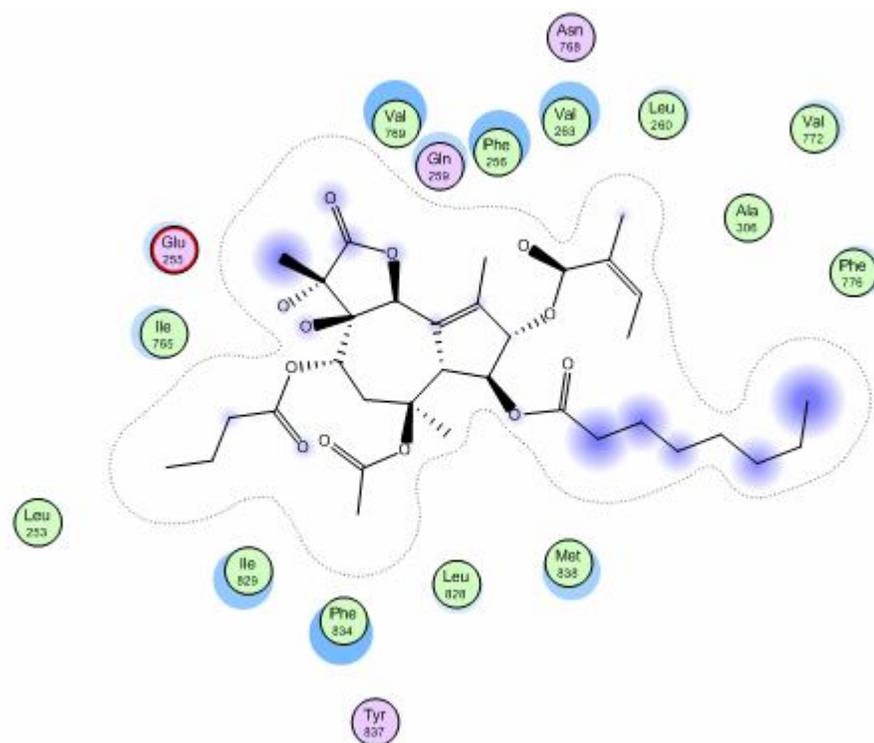


**Figure 5.5** Schematic display of SERCA (above) and PfATP6 (below) generated using DS Visualizer for windows. Helices and sheets are represented as red cylinders and cyan arrows, respectively. The ligand TG (in brown stick) is included in the structure.

(a)



(b)



**Figure 5.6:** Ligplot of (a) PfATP6- artemisinin binding site and (b) SERCA-TG binding site.

One of the key challenges in computer-aided drug discovery is to maximize the capabilities of the method in use for predicting and rank-ordering the binding affinities of compounds for a given target protein. The efficiency of a prediction method is predominantly determined by these capabilities. Various descriptors extracted from the structural information on ligand-receptor complex may provide an advantageous solution to creating a reliable binding-affinity-prediction model. Here, we combined the results obtained from a standard docking protocol with data from three different structure-based descriptors, and then investigated the utility of these descriptors on the virtual screening efficiency for artemisinin derivatives.

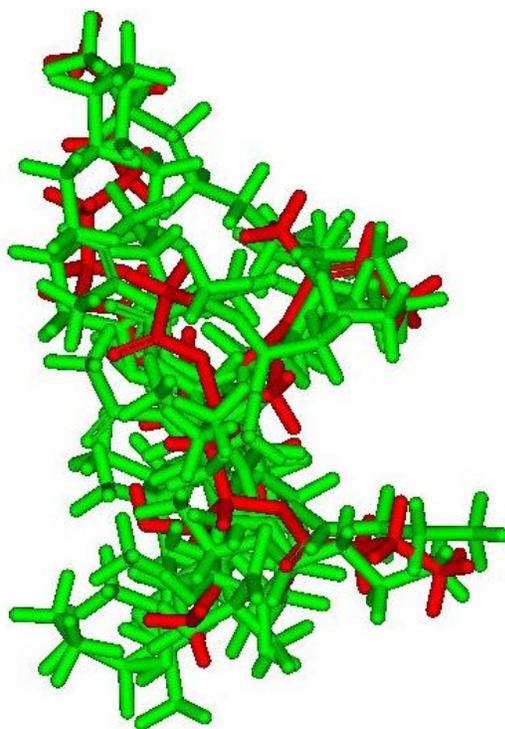
Docking simulation of artemisinin derivatives to the homology modeled PfATP6 was performed using the Glide program (Schrodinger package). The binding site of PfATP6 was constructed with thapsigargin as reference ligand. All the 154 artemisinin ligands with known antimalarial activities (W2 clone) and thapsigargin (TG) were docked into the defined binding site. The original crystal structure of TG was extracted from TG-SERCA complex (PDB ID: 1IWO) and was redocked into the binding site of PfATP6 in order to validate the Glide-XP docking protocol. The top 10 configurations after docking were taken into consideration to validate the result (Table 5.2).

**Table 5.2.** The RMSD and docking score from the docking simulation of 10 lowest configurations of crystal structure of TG with PfATP6.

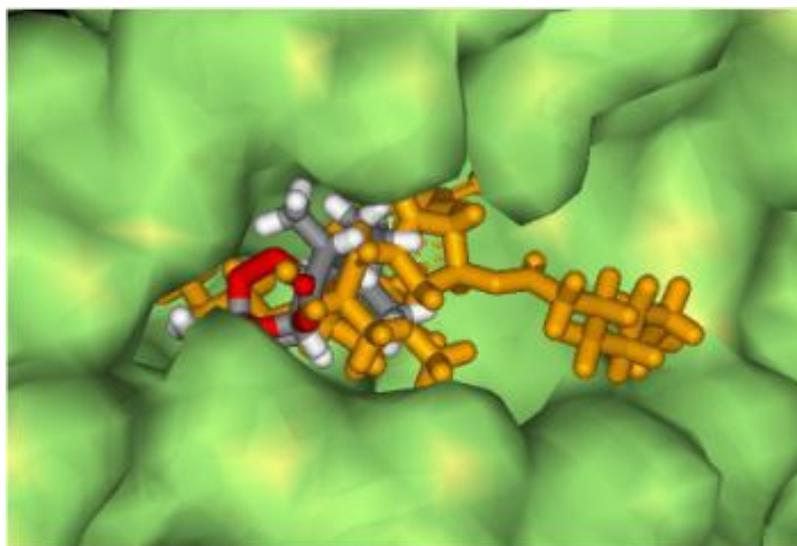
Configuration	Glide score	$\Delta G_{\text{score}}^a$	RMSD <sup>b</sup> (Å)	RMSD <sup>c</sup> (Å)
1	-9.33	0	0.37	0.04
2	-9.28		0.53	0.05
3	-8.04		0.57	0.02
4	-8.02		0.74	0.03
5	-8.01		0.81	0.02
6	-7.93		0.85	0.02
7	-7.53		1.02	0.02
8	-7.33		1.27	0.02
9	-6.99		1.36	0.01
10	-6.41		1.57	0.05

<sup>a</sup> $\Delta G_{\text{score}} = E_i - E_{\text{lowest}}$ , <sup>b</sup>RMSD, RMSD between docked and crystallographic thapsigargin structure, <sup>c</sup>RMSD, RMSD between docked poses corresponding to each configuration.

The RMSD was calculated for each configuration in comparison to the co-crystallized TG and the value was found to be in between 0.37–1.57 Å. Whereas the RMSD value calculated out of ten accepted poses for each configuration was found in between 0.02–0.05 Å. This revealed that the docked configurations have similar binding positions and orientations within the binding site and are similar to the crystal structure. The best docked structures, which is the configuration with the lowest Glide score is compared with the crystal structure as shown in Figure 5.7. These docking results illustrate that thapsigargin in PfATP6 maintained the same spatial coordinates as in SERCA. Docking of artemisinin derivatives to this binding site was performed using the standardized docking protocol. The binding mode of both TG and artemisinin within the binding site is represented in Figure 5.8. In this figure we can observe that both the molecule well fitted to the defined binding pocket. All the 154 artemisinin analogues were also found to be good binder with PfATP6. The binding modes of artemisinin and its derivatives showed hydrophobic interaction with PfATP6. This binding modes enable hypothesis that the artemisinin derivatives bind to PfATP6 with almost hydrophobic interactions and it should be the preorganized shape binding (Nogales et. al., 1998; Vander Velde et. al., 1993) between the rigid structure of artemisinin analogues and the binding pocket of PfATP6. As the Fe<sup>2+</sup>-dependent activation and antimalarial activity of artemisinin do not depend on the heme binding (Haynes et. al., 2004), we can propose that the production of the carbon centered free radical (Posner et. al., 1992) should not precede the binding to PfATP6. Therefore, artemisinin should be bound to PfATP6 before activation by Fe<sup>2+</sup> ion. For each ligand in the virtual library, the pose with the lowest Glide score was rescored using Prime/MM-GBSA and eMBrAcE approaches. These approaches predict the binding free energy for set of ligands to receptor.



**Figure 5.7** Superposition of all docked configurations of TG on crystal structure (*red-stick*).  
RMSD (heavy atom) = 0.37- 1.57Å.



**Figure 5.8** Binding mode of artemisinin and TG within the binding site of PfATP6. TG: brown stick model and Artemisinin: green stick model.

### 5.3.1 Building models for prediction of Log RA using Glide score and Prime/MM-GBSA energy

The mode of action of artemisinin structural derivatives is reported to be due to inhibition of PfATP6 protein. Thus in this study we have taken PfATP6 as the the molecular target and built prediction model for antimalarial activity by considering the Glide score and  $\Delta G_{\text{bind}}$  as descriptors. The docking score and the  $\Delta G_{\text{bind}}$  energy of the analogues are included in Table 5.3.

**Table 5.3.** Predicted antimalarial activities of (a) analogues using Glide score (XP) and Prime/MM-GBSA energy as a descriptor and experimental activity.

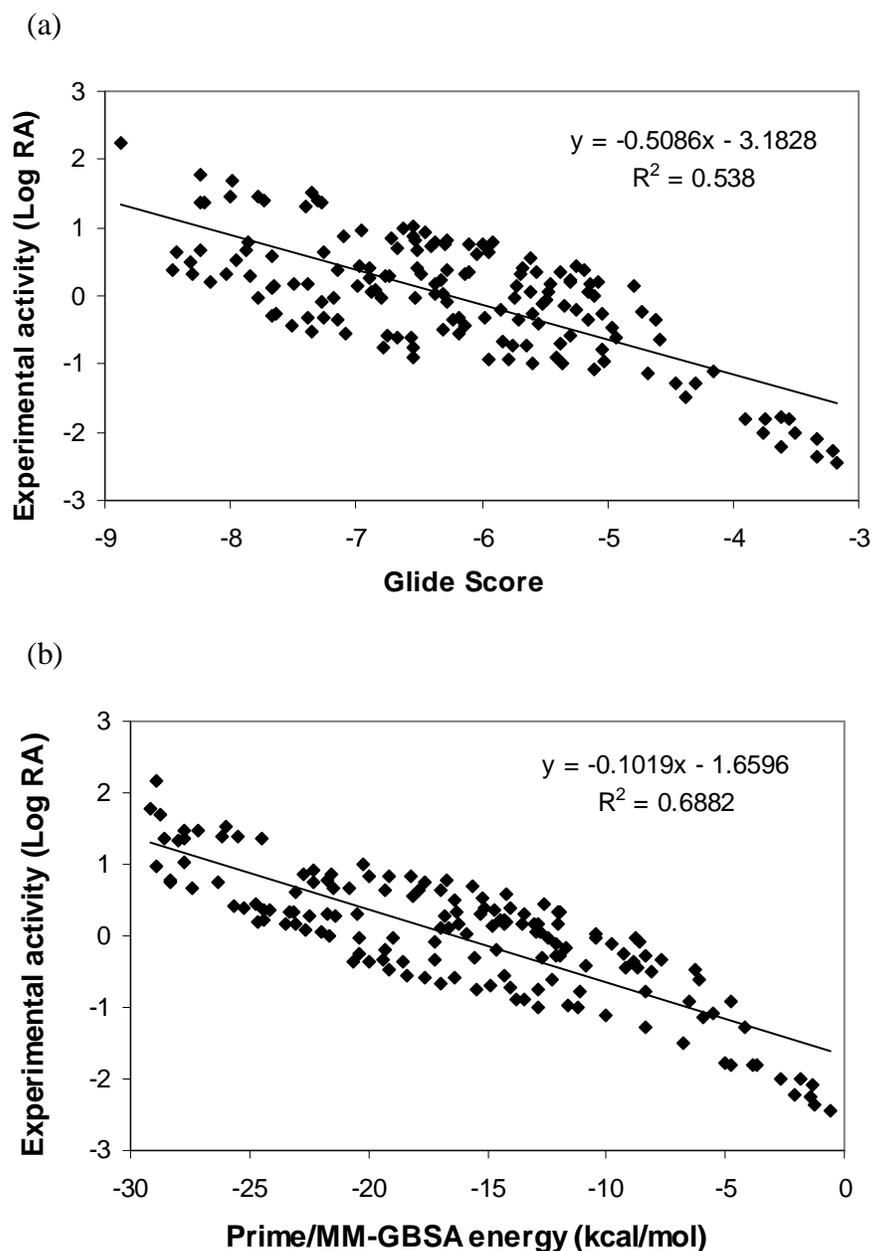
Ligand	Glide Score	$\Delta G_{\text{bind}}$ (kcal/mol)	Expt. Log RA	<sup>a</sup> Pred. Log RA (Gscore)	<sup>b</sup> Pred. Log RA ( $\Delta G_{\text{bind}}$ )
(a) Artemisinin derivatives					
1	-6.62	-20.21	1.00	0.19	0.40
2	-5.25	-12.66	0.45	-0.51	-0.37
3	-4.16	-10.00	-1.10	-1.06	-0.64
4	-6.37	-28.32	0.79	0.06	1.23
5	-6.56	-12.27	-0.60	0.16	-0.41
6	-6.28	-12.14	-0.10	0.02	-0.42
7	-7.50	-23.47	0.17	0.64	0.73
8	-7.25	-17.19	-0.32	0.51	0.09
9	-7.67	-12.10	-0.28	0.72	-0.43
10	-7.27	-24.49	1.36	0.52	0.84
11	-7.74	-26.16	1.40	0.76	1.01
12	-7.11	-22.75	0.86	0.44	0.66
13	-6.18	-18.36	-0.55	-0.03	0.21
14	-7.17	-12.49	-0.04	0.47	-0.39
15	-6.80	-18.99	-0.04	0.28	0.28
16	-6.88	-22.64	0.07	0.32	0.65
17	-6.28	-14.73	0.37	0.01	-0.16
18	-5.47	-12.79	0.05	-0.40	-0.36
19	-6.28	-19.12	0.83	0.01	0.29
20	-8.20	-28.54	1.37	0.99	1.25
21	-5.16	-20.63	-0.36	-0.56	0.44
22	-6.54	-27.74	1.02	0.15	1.17
23	-7.26	-17.00	0.63	0.51	0.07
24	-7.68	-16.62	0.12	0.73	0.04
25	-7.86	-21.75	0.78	0.82	0.56
(b) 10-Substituted artemisinin derivatives					
26	-6.10	-17.61	0.75	-0.08	0.14
27	-5.61	-18.11	0.55	-0.32	0.19
28	-5.57	-11.91	0.34	-0.35	-0.45
29	-6.96	-28.89	0.96	0.36	1.29

30	-5.10	-5.53	-1.08	-0.58	-1.10
31	-7.85	-22.45	0.28	0.82	0.63
32	-6.51	-21.46	0.66	0.13	0.53
33	-7.38	-12.05	0.18	0.58	-0.43
34	-7.28	-17.18	-0.09	0.52	0.09
35	-6.78	-8.39	-0.77	0.27	-0.80
36	-6.76	-21.43	0.28	0.26	0.53
37	-8.03	-23.15	0.32	0.91	0.70
38	-7.40	-27.96	1.32	0.59	1.19
39	-8.24	-20.83	0.67	1.01	0.46
40	-7.79	-20.43	-0.04	0.78	0.42
41	-8.32	-16.40	0.50	1.05	0.01
42	-5.92	-16.69	0.78	-0.17	0.04
43	-7.95	-15.18	0.52	0.87	-0.11
44	-6.84	-16.97	0.10	0.30	0.07
45	-6.53	-10.43	-0.03	0.14	-0.60
46	-5.49	-8.64	-0.07	-0.39	-0.78
47	-8.24	-29.13	1.79	1.01	1.31
48	-8.88	-28.91	2.18	1.34	1.29
49	-6.55	-21.56	0.87	0.15	0.54
50	-7.98	-28.76	1.70	0.88	1.27
51	-6.68	-15.59	0.70	0.22	-0.07
52	-7.09	-14.27	-0.55	0.43	-0.20
53	-5.99	-26.30	0.75	-0.13	1.02
54	-5.36	-11.18	-1.00	-0.45	-0.52
55	-5.67	-15.15	0.40	-0.29	-0.11
56	-6.72	-18.20	0.84	0.24	0.20
57	-7.67	-14.17	0.58	0.72	-0.21
58	-5.30	-17.61	-0.57	-0.48	0.14
59	-5.59	-12.91	-0.99	-0.33	-0.34
60	-5.40	-13.48	-0.89	-0.43	-0.29
61	-6.98	-13.50	0.16	0.38	-0.28
62	-7.31	-25.49	1.40	0.54	0.94
63	-6.41	-28.33	0.74	0.08	1.23
64	-6.13	-23.29	0.33	-0.06	0.72
65	-5.38	-14.84	-0.70	-0.44	-0.15
66	-6.14	-8.72	-0.44	-0.06	-0.77
67	-6.45	-22.30	0.92	0.10	0.61
(c) Seco-artemisinin derivatives					
68	-3.33	-1.26	-2.37	-1.49	-1.53
69	-4.68	-5.91	-1.13	-0.80	-1.06
70	-5.59	-20.41	-0.26	-0.33	0.42
(d) 11-Aza artemisinin derivatives					
71	-6.31	-15.84	0.02	0.03	-0.04
72	-7.66	-13.00	0.16	0.72	-0.33
73	-5.84	-14.60	-0.20	-0.21	-0.17
74	-5.35	-11.67	-0.16	-0.46	-0.47
75	-6.10	-16.28	0.34	-0.07	0.00
76	-7.78	-27.77	1.46	0.78	1.17

77	-5.14	-23.04	0.17	-0.56	0.69
78	-8.00	-27.17	1.47	0.89	1.11
(e) Artemisinin derivatives lacking the D-ring					
79	-7.36	-8.13	-0.51	0.56	-0.83
80	-6.19	-7.71	-0.32	-0.03	-0.87
81	-5.97	-12.73	-0.31	-0.14	-0.36
82	-3.56	-4.78	-1.80	-1.37	-1.17
83	-5.30	-14.26	0.23	-0.48	-0.21
84	-3.91	-3.66	-1.80	-1.19	-1.29
85	-3.74	-3.88	-1.80	-1.28	-1.26
86	-5.95	-19.29	0.65	-0.15	0.31
87	-8.42	-17.92	0.65	1.11	0.17
88	-6.30	-22.34	0.75	0.02	0.62
89	-6.52	-14.00	0.40	0.14	-0.23
90	-6.75	-16.35	-0.59	0.26	0.01
91	-6.67	-6.14	-0.60	0.22	-1.03
92	-5.74	-8.81	-0.04	-0.26	-0.76
93	-7.15	-25.27	0.38	0.46	0.92
94	-5.72	-14.82	0.14	-0.27	-0.15
95	-6.55	-13.79	-0.90	0.15	-0.25
(f) Miscellaneous artemisinin derivatives					
96	-3.33	-1.30	-2.09	-1.49	-1.53
97	-4.45	-8.39	-1.27	-0.91	-0.80
98	-6.33	-14.42	0.23	0.04	-0.19
99	-5.84	-16.94	-0.67	-0.21	0.07
100	-3.20	-1.43	-2.26	-1.55	-1.51
101	-4.73	-9.31	-0.24	-0.77	-0.71
102	-5.03	-11.61	-0.96	-0.62	-0.48
103	-5.04	-11.09	-0.79	-0.62	-0.53
104	-4.62	-20.00	-0.35	-0.83	0.38
(g) 9-sustituted artemisinin derivatives					
105	-4.38	-6.79	-1.49	-0.95	-0.97
106	-4.96	-6.29	-0.46	-0.66	-1.02
107	-5.55	-10.90	-0.41	-0.35	-0.55
108	-5.71	-18.53	-0.36	-0.28	0.23
109	-5.64	-12.88	-0.74	-0.31	-0.35
110	-3.62	-2.11	-2.22	-1.34	-1.44
111	-3.17	-0.56	-2.45	-1.57	-1.60
112	-5.25	-19.30	-0.20	-0.51	0.31
113	-5.76	-14.00	-0.72	-0.25	-0.23
(h) Dihydroartemisinin derivatives					
114	-5.04	-11.92	-0.27	-0.62	-0.44
115	-5.70	-21.73	0.31	-0.28	0.56
116	-6.37	-12.91	0.18	0.06	-0.34
117	-7.35	-26.00	1.52	0.56	0.99
118	-6.03	-23.09	0.60	-0.11	0.70

(i) Tricyclic 1,2,4-trioxane derivatives						
119	-7.88	-27.37	0.66	0.83	1.13	
120	-8.16	-24.67	0.21	0.97	0.86	
121	-8.30	-13.42	0.31	1.04	-0.29	
(j) N-alkyl-11-aza-9-desmethyartemisinin derivatives						
122	-5.10	-21.65	0.00	-0.58	0.55	
123	-6.37	-10.45	0.04	0.06	-0.59	
124	-5.45	-16.18	0.17	-0.41	-0.01	
125	-5.95	-6.54	-0.92	-0.15	-0.99	
126	-6.89	-16.82	0.28	0.33	0.06	
127	-5.61	-12.99	0.05	-0.32	-0.34	
128	-6.73	-20.51	0.29	0.24	0.43	
129	-6.48	-15.26	0.31	0.12	-0.10	
(k) 3C-substituted artemisinin derivatives						
130	-5.15	-22.00	0.05	-0.56	0.58	
131	-6.53	-20.00	0.83	0.14	0.38	
132	-6.54	-15.43	-0.74	0.15	-0.09	
133	-6.24	-19.37	-0.35	-0.01	0.32	
134	-5.19	-24.15	0.37	-0.54	0.80	
135	-3.50	-2.68	-2.00	-1.40	-1.39	
136	-8.47	-24.36	0.37	1.13	0.82	
137	-6.98	-24.72	0.45	0.37	0.86	
138	-7.51	-9.22	-0.43	0.64	-0.72	
139	-5.79	-4.78	-0.92	-0.23	-1.17	
140	-6.89	-25.68	0.41	0.33	0.96	
141	-6.31	-19.12	-0.48	0.03	0.29	
142	-3.76	-1.80	-2.00	-1.27	-1.48	
143	-7.64	-8.32	-0.28	0.71	-0.81	
130	-7.39	-15.56	-0.32	0.58	-0.07	
131	-8.25	-27.74	1.36	1.02	1.17	
(l) Various derivatives of artemisinin and artemether						
144	-7.15	-8.89	-0.36	0.46	-0.75	
145	-5.37	-12.07	0.34	-0.44	-0.43	
146	-3.61	-5.03	-1.79	-1.34	-1.15	
147	-5.07	-14.17	0.19	-0.60	-0.21	
148	-4.30	-4.21	-1.27	-0.99	-1.23	
149	-5.51	-9.81	-0.12	-0.37	-0.66	
150	-4.79	-14.26	0.16	-0.74	-0.21	

Both the docking scores and  $\Delta G_{\text{bind}}$  well explained the activities of artemisinin derivatives. Figure 5.9 (a & b) shows good correlations of Glide XP score and  $\Delta G_{\text{bind}}$  with relative antimalarial activities compared with artemisinin.



**Figure 5.9.** Models for predicting antimalarial activity (Log RA) of the artemisinin derivatives based on (a) Glide score and (b) Prime/MM-GBSA energy ( $\Delta G_{\text{bind}}$ ).

The equation (1) of the model and the corresponding statistics are shown below:

$$\text{Log RA} = - 3.18 (\pm 0.241) - 0.509 (\pm 0.038) \times \text{G-score}$$

(N = 154;  $r^2 = 0.538$ ; s = 0.601; F = 178.22;  $r^2_{cv} = 0.525$ ; PRESS = 56.811)

The root mean square error (RMSE) between the experimental RA and the predicted RA obtained by the regression model was 0.524, which is an indicator of the robustness of the fit and suggested that the calculated RA based on Glide score is reliable. The quality of the fit can also be judged by the value of the squared correlation coefficient ( $r^2$ ), which was 0.538 for the data set. Figure 5.8a graphically shows the quality of fit. The statistical significance of the prediction model is evaluated by the correlation coefficient  $r^2$ , standard error, F-test value, leave-one-out cross-validation coefficient  $r^2_{cv}$  and predictive error sum of squares PRESS. The regression model developed in this study is statistically ( $r^2_{cv} = 0.525$ ,  $r^2 = 0.538$ , F = 178.22) best fitted and consequently used for prediction of antimalarial activities (log RA) of the artemisinin analogues as reported in Table 5.4.

We have used Prime/MM-GBSA protocol for rescoring Glide XP poses of the artemisinin analogues. We did find a better correlation between  $\Delta G_{bind}$  energy and experimental RA ( $r^2 = 0.688$ ) (Fig. 5.8b). Rescoring using Prime/MM-GBSA leads to minor changes of the ligand conformations (due to energy minimization of the ligand in receptor's environment) and consequent stabilization of receptor and ligand complex. A linear regression model for prediction of predicted antimalarial activity (log RA) has been developed by considering analogues with known experimental activity. In this model we have taken  $\Delta G_{bind}$  energy as a descriptor. The Equation (2) of the model and the corresponding statistics are shown below:

$$\text{Log RA} = - 1.66 (\pm 0.098) - 0.102 (\pm 0.006) \times \Delta G_{bind} \quad (2)$$

(N = 154;  $r^2 = 0.688$ ; s = 0.495; F = 333.24;  $r^2_{cv} = 0.679$ ; PRESS = 37.997)

The statistical significance of the prediction model is evaluated by the correlation coefficient  $r^2$ , standard error, F-test value, leave-one-out cross-validation coefficient  $r^2_{cv}$  and predictive error sum of squares PRESS. The regression model developed based on  $\Delta G_{bind}$  energy is statistically ( $r^2_{cv} = 0.679$ ,  $r^2 = 0.688$ , F = 333.24) best fitted and consequently used for prediction of antimalarial activities (log RA) of the artemisinin analogues as reported in Table

5.4. The average root mean square error between predicted and experimental RA values was 0.445 by using leave-one-out cross validation technique which further revealed the reliability of the model for prediction of antimalarial activity. However, we may observe that model using  $\Delta G_{\text{bind}}$  descriptor is better for predicting antimalarial activity than model using Glide score as a descriptor.

### 5.3.2 Linear optimization of energy parameters vs Activity

One docking structure with better Glide score from each molecule docking result was picked up as final docked structure in PfATP6 for further calculations. As the Glide program treats a receptor rigidly during docking simulation, an energy minimization was performed to the docked complex. A vdW energy and electrostatic energy between ligand and receptor were calculated for each minimized complex. Also a desolvation energy and solvent accessible surface area (SASA) change were calculated using eMBrAcE (Schrodinger package). All these energies are listed in Table 5.5. By graphing these energies vs activity (log RA) of these ligands, all have bad correlation to experimental activity of the set of ligands. SASA has some degree correlation to activity for some subset ligands from all. A scheme similar to Linear Response was used to develop a free energy of binding (FEB) relationship based on these energies, which can express the activity of these artemisinin derivatives. A multiple regression was performed using Minitab statistical package. The properties of the final regression model are listed in Table 5.6. From the results of correlation factor analysis, it can be seen that  $\Delta G_{\text{vdW}}$  has most significant correlation to the activity (log RA) and electrostatic energy ( $\Delta G_{\text{ele}}$ ) has less significant correlation to the activity. It indicates that the binding of these artemisinin derivatives to PfATP6 is almost hydrophobic.  $\Delta G_{\text{vdW}}$  may be a major drive force to their binding and contribution to their activity.

The predicted activity (log RA) of these artemisinin derivatives are listed in Table 5.4. The correlation between predicted activity and actual activity is shown on Figure 5.10. The calculated activity has good correlation to the actual activity. The linear optimization of energy parameters represents the actual activity well. Theoretically, the binding affinity of drug molecules can be partitioned into several components: vdW, electrostatic, solvation and

entropy. The entropy is most difficult component to calculate normally. Different methods (56) have been suggested to estimate the entropy contribution. To relative rigid molecules, the entropy is relative small and normally is ignored or cancelled in relative free energy calculation. In the rational drug design, the calculation of relative binding free energy rather than absolute binding free energy is normally persuaded. Several papers have been reported, in which a reasonable correlation between calculated FEB and activity for a small set of ligands. Although these energy components are added directly together in most of these applications, it is still a challenge to apply these methods into large set of ligands. Normally, these different energy components (vdW, electrostatic, solvation) were calculated using more than one method. To same set of structure, different force field or different methods will produce different values of energy. This suggests that these energy components need to be scaled before an equation is obtained to get a better expression for these energy components. A set of weights can be used to scale these energies to get free energy expression by linearly combining these energies. Some scoring functions (73) used this strategy, which were optimized using a test set of molecules. In the work, a linear combination strategy was used to express FEB by four energy components calculated from different methods. An expression of free energy, whose weight coefficients were optimized by a multiple regression, was obtained and successfully predicted the activity of a large set of ligands. As stated early, the major interest in drug design is to express the variance of free energy over a set of active molecules. In the sense, the  $\Delta G_{vdW}$  is biggest contribution; SASA is next; electrostatic part and electrostatic part of solvation are smallest contribution to the free energy variance. The Equation (3) of the model and the corresponding statistics are shown below:

$$\text{Log RA} = -1.37 - 0.0035 \text{ SASA} - 0.0314 \Delta G_{vdW} + 0.0029 \Delta G_{ele} - 0.00398 \Delta G_{solv} \quad (3)$$

(N = 135;  $r^2 = 0.815$ ; s = 0.291; F = 141.99;  $r^2_{cv} = 0.802$ ; PRESS = 11.65)

The statistical significance of the prediction model is evaluated by the correlation coefficient  $r^2$ , standard error, F-test value, leave-one-out cross-validation coefficient  $r^2_{cv}$  and predictive error sum of squares PRESS. The regression model developed based on linear response scheme is statistically ( $r^2_{cv} = 0.802$ ,  $r^2 = 0.815$ , F = 141.99) best fitted and consequently used for prediction of antimalarial activities (log RA) of the artemisinin analogues as reported in Table 5.4. The average root mean square error between predicted and experimental RA values was 0.291 by using leave-one-out cross validation technique which further revealed the

reliability of the model for prediction of antimalarial activity. However, we may observe that model using linear response scheme is better for predicting antimalarial activity than model using Glide score and  $\Delta G_{\text{bind}}$  descriptor as descriptors.

**Table 5.4.** Predicted antimalarial activities of artemisinin derivatives based on linear response scheme of energy parameters and experimental activities for selected analogues.

Sl. No.	SASA	$\Delta G_{\text{vdW}}$	$\Delta G_{\text{ele}}$	$\Delta G_{\text{solv}}$	Log RA	Log RA <sub>pred</sub>
(a) Artemisinin derivatives						
1	459.0	-122.2	-116.4	106.2	1.00	1.02
2	624.5	-120.4	-97.4	104.9	0.45	0.01
3	591.7	-92.2	-138.6	118.4	-1.10	-0.69
5	512.5	-93.3	-66.3	65.4	-0.60	-0.16
6	436.8	-80.5	-87.7	103.2	-0.10	-0.12
7	633.3	-123.8	-31.7	87.8	0.17	0.09
8	690.3	-117.3	-29.0	61.4	-0.32	-0.44
9	741.4	-131.4	-123.1	144.6	-0.28	-0.35
12	609.6	-144.1	-203.2	199.6	0.86	0.76
13	521.3	-90.4	-152.6	127.5	-0.55	-0.34
14	477.4	-96.5	-36.5	66.4	-0.04	0.16
15	635.7	-119.6	-18.5	48.2	-0.04	-0.04
16	543.8	-105.0	-152.5	136.8	0.07	-0.03
17	475.0	-109.6	-76.8	90.1	0.37	0.56
18	479.9	-95.2	-94.7	61.0	0.05	0.08
19	508.2	-120.5	-48.3	83.0	0.83	0.71
20	654.6	-163.8	-77.2	107.1	1.37	1.17
21	521.8	-101.5	-80.9	95.7	-0.36	0.03
22	584.6	-142.0	-121.9	132.7	1.02	0.89
23	670.0	-141.4	-110.3	125.6	0.63	0.38
24	622.9	-119.5	-190.1	196.0	0.12	-0.06
25	647.4	-141.9	14.6	47.6	0.78	0.59
(b) 10-Substituted artemisinin derivatives						
26	497.3	-131.1	-133.6	120.4	0.75	1.06
27	475.3	-121.6	-93.8	85.2	0.55	0.92
28	663.6	-144.5	-670.9	620.4	0.34	0.20
29	574.3	-143.0	-221.9	198.8	0.96	0.93
30	532.7	-77.5	-68.6	59.1	-1.08	-0.76
31	637.0	-137.3	-41.0	63.3	0.28	0.48
32	532.3	-132.2	-47.9	56.8	0.66	0.93
33	574.9	-96.4	-121.2	103.3	0.18	-0.45
34	620.6	-126.7	-61.8	119.3	-0.09	0.24
35	654.1	-103.1	-68.0	76.2	-0.77	-0.68
36	711.9	-139.5	-134.4	156.8	0.28	0.06
37	694.7	-139.9	-183.3	205.5	0.32	0.15
39	827.3	-171.1	20.8	38.8	0.67	0.45
40	529.9	-111.2	-56.1	85.0	-0.04	0.30
41	658.8	-127.6	-168.1	165.6	0.50	-0.01

42	568.1	-131.4	-63.7	77.8	0.78	0.69
43	702.2	-141.0	-62.6	87.5	0.83	0.20
44	678.0	-126.2	-16.7	38.6	0.10	-0.08
45	688.1	-121.9	-42.0	57.6	-0.03	-0.29
46	669.7	-120.7	-34.1	62.1	-0.07	-0.21
49	504.0	-126.6	-51.9	69.0	0.87	0.92
51	575.1	-110.1	-42.9	54.4	0.70	0.01
52	673.1	-113.1	-38.7	97.9	-0.55	-0.47
53	536.8	-112.7	-109.5	107.7	0.75	0.27
54	581.8	-88.5	-22.5	45.5	-1.00	-0.68
55	501.5	-116.3	-33.5	35.7	0.40	0.63
56	512.4	-124.2	-107.3	98.8	0.84	0.77
57	659.9	-124.9	-56.1	118.8	0.58	-0.04
58	531.8	-96.8	-48.5	58.3	-0.57	-0.15
59	501.3	-77.6	-100.7	78.9	-0.99	-0.59
60	485.2	-72.4	-180.9	141.2	-0.89	-0.70
61	582.2	-105.5	-123.8	124.3	0.16	-0.22
62	627.7	-151.3	-17.9	70.3	1.40	0.98
64	535.3	-104.1	-76.4	56.7	0.33	0.04
65	548.8	-92.5	42.8	-12.0	-0.70	-0.33
66	548.9	-107.8	-78.7	88.1	-0.44	0.07
67	492.7	-117.8	-168.6	117.6	0.92	0.66
(c) Seco-artemisinin derivatives						
69	459.9	-56.6	-113.0	119.8	-1.13	-1.00
70	542.4	-118.8	-83.9	98.3	-0.26	0.44
(d) 11-Aza artemisinin derivatives						
71	651.8	-119.3	-77.7	119.6	0.02	-0.18
72	623.5	-119.2	-55.8	82.3	0.16	0.00
73	606.9	-119.2	-179.5	148.0	-0.20	0.03
74	628.8	-108.7	-14.2	47.2	-0.16	-0.33
75	581.9	-137.1	25.6	3.4	0.34	0.83
76	616.5	-157.6	-255.6	161.6	1.46	1.12
77	551.0	-118.1	-0.5	57.0	0.17	0.41
(e) Artemisinin derivatives lacking the D-ring						
79	611.6	-104.3	-360.5	411.6	-0.51	-0.57
80	418.9	-73.7	-87.9	77.2	-0.32	-0.22
81	419.0	-85.3	-73.7	94.6	-0.31	0.14
82	527.2	-52.5	-39.8	43.5	-1.80	-1.48
83	522.8	-116.9	-89.4	89.8	0.23	0.50
84	558.1	-63.0	-16.8	35.1	-1.80	-1.33
85	688.8	-136.4	-64.7	75.1	0.65	0.14
86	754.9	-144.2	-43.6	72.5	0.65	0.01
88	649.3	-129.9	-32.6	60.0	0.75	0.19
89	606.3	-115.5	-107.9	138.4	0.40	-0.05
90	659.8	-113.6	-59.8	69.5	-0.59	-0.39
91	874.4	-158.0	-480.8	531.2	-0.60	-0.51
92	695.1	-117.5	-159.8	155.8	-0.04	-0.53
93	657.1	-129.3	-74.0	89.3	0.38	0.10

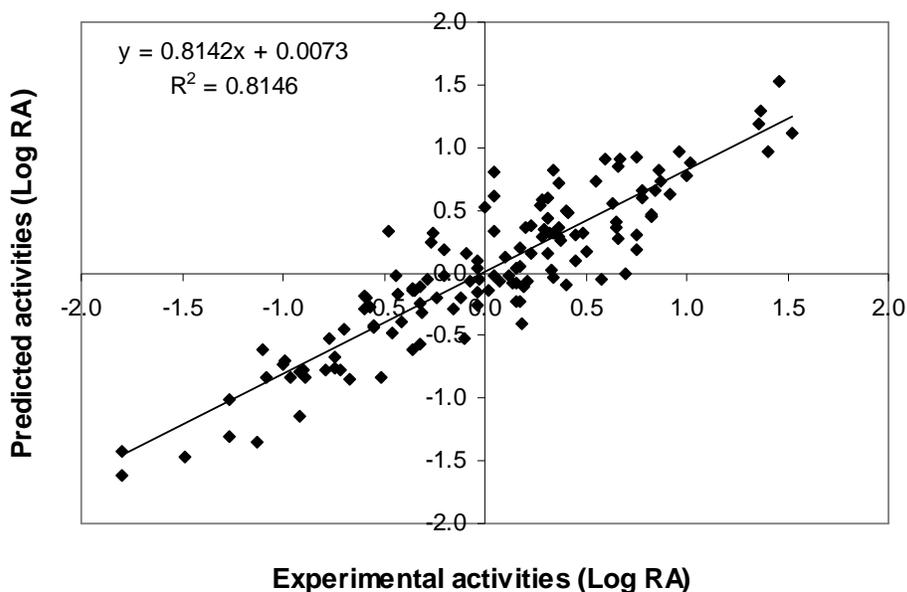
94	611.3	-112.3	-37.5	54.2	0.14	-0.13
95	701.5	-102.6	-63.1	90.8	-0.90	-0.97
(f) Miscellaneous artemisinin derivatives						
96	481.6	-59.6	-85.1	93.3	-1.27	-1.02
97	492.7	-106.4	-56.2	66.9	0.23	0.37
98	612.7	-91.5	-107.1	132.5	-0.67	-0.82
100	473.8	-92.0	-87.8	79.1	-0.24	0.02
101	486.5	-75.2	-61.5	77.4	-0.96	-0.56
102	472.5	-75.3	-59.8	75.1	-0.79	-0.47
103	444.9	-91.8	-54.5	67.1	-0.35	0.20
(g) 9-sustituted artemisinin derivatives						
105	508.1	-58.0	-155.3	151.3	-1.49	-1.26
106	484.3	-86.2	-142.6	135.4	-0.46	-0.25
107	490.3	-89.4	-141.0	131.1	-0.41	-0.19
108	470.1	-98.0	-144.3	153.8	-0.36	0.19
109	472.4	-73.5	-107.6	91.5	-0.74	-0.55
112	498.2	-101.8	-106.8	107.2	-0.20	0.17
113	464.5	-78.1	-159.8	174.5	-0.72	-0.40
(h) Dihydroartemisinin derivatives						
114	469.7	-106.4	-104.5	98.6	-0.27	0.48
115	472.4	-114.4	-72.1	48.6	0.31	0.72
116	465.8	-108.6	-22.2	43.4	0.49	0.61
117	651.1	-155.4	6.0	27.2	1.52	0.98
118	649.6	-147.6	-37.2	52.0	0.60	0.73
(i) Tricyclic 1,2,4-trioxane derivatives						
119	699.3	-141.2	-196.5	230.3	0.66	0.15
120	857.5	-162.5	-199.5	240.9	0.21	-0.12
121	765.0	-149.8	-194.1	199.7	0.31	0.04
(j) N-alkyl-11-aza-9-desmethylartemisinin derivatives						
122	473.4	-115.9	-63.4	70.2	0.00	0.77
123	578.7	-136.5	-98.0	142.1	0.04	0.76
124	591.5	-103.8	-44.3	46.9	0.17	-0.28
125	646.4	-109.2	-689.3	646.8	-0.92	-0.79
126	596.0	-136.6	-10.8	83.5	0.28	0.71
127	610.1	-141.0	-16.8	44.5	0.05	0.76
128	637.4	-131.4	-103.0	120.6	0.29	0.27
129	664.3	-128.2	-135.7	145.8	0.31	-0.01
(k) 3C-substituted artemisinin derivatives						
130	482.0	-112.9	-83.1	100.7	0.05	0.61
131	513.0	-121.3	-44.4	78.5	0.83	0.71
132	543.2	-87.5	-109.1	118.5	-0.74	-0.54
133	530.4	-102.0	-76.2	88.0	-0.35	0.00
134	610.6	-130.5	-174.8	185.1	0.37	0.35
136	642.6	-144.0	-23.4	65.6	0.37	0.66
137	622.4	-131.5	-250.5	252.9	0.45	0.28

138	812.2	-148.7	-325.0	396.0	-0.43	-0.35
139	786.8	-117.3	-324.9	413.9	-0.92	-1.17
142	593.5	-128.4	-151.0	136.0	0.41	0.40
143	619.9	-129.3	-79.7	107.3	-0.48	0.32
130	757.1	-129.4	-62.1	86.6	-0.32	-0.47
131	700.5	-163.9	-78.9	94.9	1.36	0.91
(l) Various derivatives of artemisinin and artemether						
144	435.4	-78.8	-141.0	151.5	-0.36	-0.20
145	555.6	-104.8	68.7	-45.2	0.34	0.02
146	567.1	-100.2	-78.1	58.3	-0.12	-0.27
147	531.8	-101.4	-65.7	57.8	0.16	-0.02
148	564.4	-104.7	-87.6	68.9	0.21	-0.12
149	554.0	-101.9	-54.6	42.3	0.19	-0.12
150	521.7	-73.3	-46.3	64.5	-1.27	-0.81

Predicted RA is calculated from optimized linear combination of  $\Delta G_{ele}$ ,  $\Delta G_{vdW}$ ,  $\Delta G_{solv}$ , and SASA from regression.

**Table 5.5.** Regression properties of energy parameters with experimental activities (Log RA).

	SASA	$\Delta G_{vdW}$	$\Delta G_{ele}$	$\Delta G_{solv}$
Correlation factor with log RA	0.018	0.587	0.009	0.001
Intercept (B)	-0.492	-3.24	0.087	0.099
SE of B	0.34	0.241	0.081	0.088



**Figure 5.10** Models for predicting antimalarial activity (Log RA) of the artemisinin derivatives based on free energy equation.

## 5.4 Conclusions

We have presented herein a FEB calculation on the binding affinity of 159 artemisinin derivatives with PfATP6. The binding structures of these ligands in PfATP6 were predicted by flexible docking simulations. The docking result demonstrated that the docking simulation could satisfactorily reproduce a binding structure from a crystal structure of a SERCA/TG complex. Superposition of the binding structure of whole set of ligands from docking simulations shows that these structurally similar ligand bind in a very similar pattern in PfATP6. They all bind at the same orientation, which have been found in crystal structures of SERCA/TG. They bind in a similar position inside the PfATP6 active site and try to fit the binding pocket well. The calculated FEB for these ligands reasonably predicted the activity of this set of ligands. The calculated activity has good correlation to experimental activity. The result shows that the linear combination of four energy terms: vdW, electrostatic, solvation (electrostatic part), and nonpolar energies optimized by regression has power to express the binding affinity of large set of ligands in receptor. The Dock-MM-GB/SA and eMBrAcE demonstrates a good ability on the binding structure prediction and binding energy determination of produce reasonable energies. The GB/SA method predicted a reasonable solvation energy terms to enable a satisfactory FEB expression was build. In the work, it is noticed that among these energy terms, the  $\Delta G_{vdW}$  has most significant correlation to the activity (log RA) and electrostatic energy ( $\Delta G_{ele}$ ) has less significant correlation to the activity. The binding modes of artemisinin and its derivatives showed hydrophobic interaction with PfATP6. This work suggests that in the relative FEB calculation, which is major interest in drug design, the contribution of different energy terms can be scaled by a set of weight factors to reach a good correlation. In practice, it is know that same energy term plays different role in different type of systems. This is one of reasons that a reasonable activity model can be obtained just based on some energy terms. The calculation of solvation effect upon a ligand binding in a protein is a challenge work. This work and many others have shown that solvation effect is any important driving force on ligand binding and a key factor in expression of activity of a set of ligands. In the work, GB and SASA methods were used to estimate the electrostatic and the nonpolar parts of solvation and produced satisfactory results in terms of good correlation with experimental activity.

# CONCLUSION

---

Different computational methods such as Dock, FEB, LIE-SGB and QSAR have been used in the work to study the binding structure and binding affinity of ligand/receptor of heme polymerization and PfATP6 and the inhibition mechanism of artemisinin. The effects of atomic charges, and ligand and heme structures on the docking results were investigated. Several charge schemes for both artemisinin and heme were employed for docking purpose. The docking results depended on the structures of both artemisinin and heme. Moreover, the atomic charges of heme have a significant effect on the docking configurations. The combined approaches of docking-molecular mechanics based on generalized Born/surface area (MM-GB/SA) solvation model showed that artemisinin and its structural derivatives approaches heme by pointing O1 and O2 at the endoperoxide linkage towards the iron center, a mechanism that is controlled by steric hindrance. Several sets of artemisinin analogues have been studied in the docking simulations. Results showed that these analogues bind in a very similar mode. The magnitude of the binding affinity can be a key factor that decides the activeness of an individual inhibitor. An energetic evaluation of the binding affinity will provide a way to estimate the activity of inhibitors. In any binding energy calculation, the correct binding structure of each ligand has to be determined first prior to binding energy estimation. Very similar binding structures were obtained for a set of analogues. This makes a credible prediction model of the antimalarial activity ( $pIC_{50}$ ) calculation possible. The calculated Glide score and binding free energy value of a set of structural analogues demonstrate excellent linear correlation to the experimental antimalarial activity thus these models could be useful to predict the range of activity for new artemisinin analogues. We also found that refinement of poses and consequent rescoring using PRIME/MM-GBSA leads to better predictivity of  $pIC_{50}$ .

We have demonstrated that the SGB-LIE method can be applied to estimate the free energy of binding with a high level of accuracy for a range of compounds with varying inhibition potencies. Despite the limitation imposed by the insufficient sampling inherent in the MD and HMC protocols, the methods have reproduced experimental data with reasonably small error for the majority of artemisinin analogues. The close estimation of inhibition potencies of a wide range of compounds has established the LIE methodology as an efficient tool for screening novel compounds with very different structures. Compared to the empirical

methods, such as scoring function approaches, the LIE method is more accurate due to the semiempirical approach adopted in which experimental data are used to build the binding affinity model.

The QSAR analysis of a series of artemisinin derivatives enabled consistent models of structure- activity relationships to be obtained for several descriptors. In this study, we used a more systematic way of variable selection in order of missing value test → zero test → simple correlation test → multicollinearity test → genetic algorithm to obtain the meaningful descriptors leading to QSAR model development. The high predictive ability of the model developed here in this study allows virtual screening of chemical databases or virtual libraries determined by either synthetic feasibility or commercial availability of starting materials to prioritize the synthesis of most promising candidates. Therefore, these models should facilitate the rational design of novel derivatives, guide the design of focused libraries based on the artemisinin skeleton and facilitate the search for related structures with similar biological activity from large databases.

The atomic coordinates of PfATP6 for the organism *Plasmodium falciparum* was not available in Protein Data Bank, hence homology modeling protocol was employed to predict the model of the protein. A library of artemisinin analogues has been designed and their molecular interactions and binding affinities with modeled PfATP6 protein have been studied using the docking, molecular mechanics based on generalized Born/surface area (MM-GBSA) solvation model and eMBrAcE. In the docking simulations, the flexible docking reproduced well the binding structure of crystal structures. These experiments verified the docking protocol adapted in the work. Also the docking simulations of structurally similar inhibitors showed that the docking simulation could dock inhibitors into a receptor with similar binding positions and orientations within the binding site. Results show that these analogues bind in a very similar mode. This suggests that they interact with the enzyme in a very similar way. Docking and binding free energies scores show good relation with in vitro antimalarial activities. In addition, a scheme similar to Linear Response was used to develop a free energy of binding (FEB) relationship based electrostatic ( $\Delta G_{ele}$ ), van der Waal ( $\Delta G_{vdw}$ ) and surface accessible surface area (SASA), which can express the activity of these artemisinin

derivatives. It was seen that  $\Delta G_{\text{vdw}}$  has most significant correlation to the activity (log RA) and electrostatic energy ( $\Delta G_{\text{ele}}$ ) has less significant correlation to the activity. The binding modes of artemisinin and its derivatives showed hydrophobic interaction with PfATP6. This binding mode enable hypothesis that the artemisinin derivatives bind to PfATP6 with almost hydrophobic inateractions and it should be the preorganized shape binding between the rigid structure of artemisinin analogues and the binding pocket of PfATP6. As the  $\text{Fe}^{2+}$ -dependent activation and antimalarial activity of artemisinin do not depend on the heme binding we can propose that the production of the carbon centered free radical should not precede the binding to PfATP6. Therefore, artemisinin should be bound to PfATP6 before activation by  $\text{Fe}^{2+}$  ion. The calculated FEB of a set of artemisinin analogues demonstrates excellent linear correlation to the experimental activity. Low levels of root mean square error for the majority of inhibitors establish the docking, Prime/MM-GBSA and eMBrAcE based prediction model as an efficient tool for generating more potent and specific inhibitors of PfATP6 by testing rationally designed lead compounds based on aremisinin derivatization.

The satisfactory results obtained for virtual screening of artemisinin analogues and prediction of antimalarial activity based on screening methodology of Docking-MM-GBSA, SGB-LIE and QSAR will help a lot to design new generation inhibitors. We hope that the knowledge and insight on the screening models learnt from the work will help a lot on the battle against the malaria and benefit human health and life.

# Bibliography

---

- Acton N, Karle JM, Miller RE, (1993). Synthesis and antimalarial activity of some 9-substituted artemisinin derivatives. *J. Med. Chem.* 36, 2552-2557.
- Ahmad A, Misra LN (1994) Terpenoids from *Artemisia annua* and constituents of its essential oil. *Phytochemistry.* 37 (1): 183-186.
- Aqvist J, Marelius J, (2001) The linear interaction energy method for predicting ligand binding free energies. *Combin. Chem. High Throughput Screening.* 4: 613-626.
- Aqvist J, Medina C, Samuelsson JE, (1994) New Method for Predicting Binding-Affinity in Computer-Aided Drug Design. *Protein Eng.* 7: 385-391.
- Arav-Boger R, Shapiro TA, Molecular (2005) mechanisms of resistance in antimalarial chemotherapy: The unmet challenge. *Annu. Rev. Pharmacol. Toxicol.*, 45: 565-585.
- Arrow KJ, Panosian CB, Geltband H (2004) (eds.) Saving Lives, Buying Time: Economics of Malaria Drugs in an Age of Resistance. *National Academic Press, Washington, DC.*
- Asawamahesakda W, Benakis A., Meshnick SR, (1994) The interaction of artemisinin with red cell membranes. *J Lab Clin Med* 123: 757-62.
- Augustijns P, D'Hulst A, Van Daele J, Kinget R, (1996) Transport of artemisinin and sodium artesunate in Caco-2 intestinal epithelial cells. *J Pharm Sci* 85:577-9.
- Avery MA, Alvin-Gaston M, Rodrigues CR, Barreiro EJ, Cohen FE, Sabnis YA, Woolfrey JR, (2002). Structure Activity Relationships of the Antimalarial Agent Artemisinin. The Development of Predictive InVitro potency Models Using CoMFA and HQSAR Methodologies. *J Med.Chem.* 45:292-303.
- Avery MA, Bonk JD, Chong WKM, Mehrotra S, Miller R. et al. (1995) Structure-Activity Relationships of the Antimalarial Agent Artemisinin. 2. Effect of Heteroatom Substitution at O-11: Synthesis and Bioassay of N-Alkyl-11-aza-9-desmethylartemisinins. *J. of Med. Chem.*, 38: 5038-5044.
- Avery MA, Gao F, Chong WK, Mehrotra S, Milhous WK. (1993). Structure-activity relationships of the antimalarial agent artemisinin. 1. Synthesis and comparative molecular field analysis of C-9 analogs of artemisinin and 10-deoxyartemisinin. *J Med Chem.*; 36: 4264-4275.
- Avery MA, Mehrotra S, Bonk JD, Vroman JA, Goins DK. et al. (1996) Structure-Activity Relationships of the Antimalarial Agent Artemisinin. 4. Effect of Substitution at C-3. *J. of Med. Chem.* 39: 2900-2906.
- Bagchi GD (2003). Essential oil constituents of *Artemisia annua* during different growth periods at monsoon conditions of subtropical north Indian plains. *J. Essent. Oil Res.* 15(4): 248 – 250.

- Bannister LH, Hopkins JM, Fowler RE, Krishna S, Mitchell GH (2000). A brief ultrastructural guide to asexual blood stages of *Plasmodium falciparum*. *Parasitol Today* 16: 427-433.
- Bernardinelli G, Jefford CW, Maric D, Thomson C, Weber J, (1994). Computational studies of the structures and properties of potential anti-malarial compounds based on the 1, 2, 4-trioxane ring structure: I. Artemisinin-like molecules, *Int. J Quant Chem: Quant Biol Symp.*; 21: 117–131.
- Bhakuni RS et al. (2001). Secondary metabolites of *Artemisia annua* and their biological activity. *Curr. Sci.*, 80 (1): 35-48.
- Bhattacharjee AK, Hartell MG, Nichols DA, Hicks RP, Stanton B, van Hamont JE, Milhous WK, (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. *Eur. J. Med. Chem.* 39: 59–67.
- Bowman S, Lawson D, Basham D, Brow D, Chillingworth T, Churcher CM, Craig A, Davies RM, Devlin K, Feltwell T, Gentles S, Gwilliam R, Hamlin N, Harris D, Holroyd S, Hornsby T, Horrocks P, Jagels K, Jassal B, Kyes SJ, McLean S, Moule S, Mungall, K, Murphy L, Olive RK, Quail MA, Rajadream MA, Rutter S, Skelton J, Squares R, Squares S, Sulston JE, Whitehead S, Woodward JR, Newbold C, Barrell BG, (1999). The complete nucleotide sequence of chromosome 3 of *Plasmodium falciparum*. *Nature*. 400: 532–538.
- Brenwald NP, Gill MJ, Wise R, (1998). Prevalence of a Putative Efflux Mechanism among Fluoroquinolone-Resistant Clinical Isolates of *Streptococcus pneumoniae*, *Antimicrob. Agents Chemother.* 42: 2032-2035.
- Brossi A, Venugopalan B, Dominguez Gerpe L, Yeh HJ, Flippen-Anderson JL, Buchs P, Luo XD, Milhous W, Peters W (1988). Arteether, a new antimalarial drug: synthesis and antimalarial properties. *J Med Chem* 31:645-50.
- Carlson HA, Jorgensen WL, (1995) An Extended Linear-Response Method for Determining Free-Energies of Hydration. *J. Phys. Chem.*, 99: 10667-10673.
- Chawira AN, Warhurst DC (1987). The effect of artemisinin combined with standard antimalarials against chloroquine-sensitive and chloroquine-resistant strains of *Plasmodium falciparum* in vitro. *J Trop Med Hyg.* 90: 1-8.
- Chen B, Harrison RF, Papadatos G, Willett P, Wood DJ, Lewell XQ, Greenidge P, Stiefl N, (2007). Evaluation of machine-learning methods for ligand-based virtual screening. *Journal of computer-aided molecular design* 21: 53-62.

- Cheng F, Shen J, Luo X, Zhu W, Gu J, Ji R, Jiang H, Chen K, (2002) Molecular docking and 3-D-QSAR studies on the possible antimalarial mechanism of artemisinin analogues. *Bioorg. Med. Chem.* 10: 2883-2891.
- Cherkasov A, Ban F, Li Y, Fallahi M, Hammond GL, (2006) Progressive docking: a hybrid QSAR/docking approach for accelerating in silico high throughput screening. *Journal of medicinal chemistry* 49: 7466-7478.
- Cleves AE, Jain AN, (2006). Robust ligand-based modeling of the biological targets of known drugs. *Journal of medicinal chemistry* 49:2921-2938.
- Cox-Singh J, Davis TM, Lee KS, Shamsul SS, Matusop A, Ratnam S, Rahman HA, Conway DJ, Singh B (2008). Plasmodium knowlesi malaria in humans is widely distributed and potentially life threatening. *Clin Infect Dis* 46: 165-171.
- Cramer RD III, Patterson DE, Bunce JD, (1988). Comparative Molecular Field analysis (CoMFA). I. Effect of shape on Binding of Steroids to carrier Proteins. *J. Am. Chem.Soc.* 110: 5959-5967.
- Cramer RD III, Patterson DE, Bunce JD, Bunce JD, Frank IE, (1998) Crossvalidation, Bootstrapping and Partial Least Squares compared with Multiple Regression in Conventional QSAR Studies. *Quant. Struct-Act. Relat. Pharmacol. Chem. Soc.* 7: 18-25.
- Cumming JN, Ploypradith P, Posner GH, (1997) Antimalarial activity of artemisinin (qinghaosu) and related trioxanes: mechanism(s) of action. *Adv. Pharmacol.* 37: 253-297.
- Darren JC, William NC, Francis CKC, Richard JP, Yuxiang D, Jonathan LV, Susan AC, (2008). Relationship between Antimalarial Activity and Heme Alkylation for Spiro- and Dispiro-1,2,4-Trioxolane Antimalarials. *Antimicrob. Agents and Chemo.* 52: 1291-1296.
- Davis TME, Karunajeewa HA, Ilett KF (2005) Artemisinin-based combination therapies for uncomplicated malaria. *Med. J. Austr.* 182: 181-185.
- Deng DA et. al. (1981) Studies on the structure of artemisic acid. *Kexue Tonbao.* 19: 1209-1211.
- Desjardins RE, Canfield CJ, Haynes DE, Chulay JD, (1979). Quantitative assessment of antimalarial activity in vitro by a semiautomated microdilution technique. *Antimicrob. Agents Chemother.* 16: 710-718.
- Deswal S, Roy N, (2006). Quantitative structure activity relationship studies of aryl heterocycle-based thrombin inhibitors, *J. Med. Chem.* 41(11): 1339-1346.

- Duane S, Kennedy AD, Pendleton BJ, Roweth D, (1987). *Hybrid Monte Carlo. Phys. Lett. B*: 195-216.
- Eckstein-Ludwig U, Webb RJ, Van Goethem ID, East JM, Lee AG, Kimura M, O'Neill P M, Bray PG, Ward SA, Krishna S, (2003) Artemisinin target the SERCA of *Plasmodium falciparum*. *Nature* 424: 957-961.
- Eckstein-Ludwig U, Webb RJ, van Goethem IDA, East JM, Lee AG, Kimura M, O'Neill PM, Bray PG, Ward SA, Krishna S (2003). *Nature*. 424: 957.
- Eisenberg D, Luthy R, Bowie JU (1997). VERIFY3D: assessment of protein models with three-dimensional profiles. *Methods Enzymol*, 277, 396-404.
- Eldridge MD, Murray CW, Auton TR, Paolini GV, Mee RP (1997). Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *Journal of computer-aided molecular design* 11: 425-445.
- Eliopoulos GM, Moellering RC Jr., (1996). Antimicrobial combinations: *In V. Lorian (ed.), Antibiotics in laboratory medicine* (4<sup>th</sup> edn). 330–396.
- Ellis DS et al. (1985). The chemotherapy of rodent malaria, XXXIX. Ultrastructural changes following treatment with artemisinin of *Plasmodium berghei* infection in mice, with observations of the localization of [3H]-dihydroartemisinin in *P. falciparum* in vitro. *Ann. Trop. Med. Parasitol.* 79: 367–374.
- Fersht AR, (1984) Basis of biological specificity *Trends Biochem. Sci.*, 9: 145.
- Fersht AR, Shi JP, Knill-Jones J, Lowe DM, Wilkinson AJ, Blow DM, Brick P, Carter P, Waye MM, Winter G, (1985). Hydrogen bonding and biological specificity analyzed by protein engineering. *Nature*. 314(6008): 235–238.
- FirstDiscovery2.7*, 2.7 ed.; Schrodinger Inc.: Portland, 2004.
- Foote SJ, and Cowman AF (1994). The mode of action and the mechanism of resistance to antimalarial drugs. *Acta Trop* 56: 157-71.
- Francis SE, Sullivan Jr. DJ, and Goldberg DE, (1997). Hemoglobin metabolism in the malaria parasite *Plasmodium falciparum*. *Annu Rev Microbiol* 51: 97-123.
- Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK, Shaw DE, Francis P, Shenkin PS (2004). Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *Journal of medicinal chemistry* 47: 1739-1749.

- Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, Carlton JM, Pain A, Nelson KE, Bowman S, Paulsen IT, James K, Eisen JA, Rutherford K, Salzberg SL, Craig A, Kyes S, Chan MS, Nene V, Shallom SJ, Suh B, Peterson J, Angiuoli S, Pertea M, Allen J, Selengut J, Haft D, Mather MW, Vaidya AB, Martin DM, Fairlamb AH, Fraunholz MJ, Roos DS, Ralph SA, McFadden GI, Cummings LM, Subramanian GM, Mungall C, Venter JC, Carucci DJ, Hoffman SL, Newbold C, Davis RW, Fraser CM, Barrell B (2002). Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature*. 419: 498-511
- Golbraikh A, Tropsha A, (2002). Beware of  $q^2$ . *J. Mol. Graph. Model.* 20(4): 269-276
- Goldberg DE, Slater AF, Cerami A, Henderson GB, (1990) Hemoglobin degradation in the malaria parasite *Plasmodium falciparum*: an ordered process in a unique organelle. *Proc Natl Acad Sci U S A*. Apr;87(8):2931–2935.
- Gregorio de C, Kier LB, Hall LH, (1998). QSAR modeling with the electrotopological state indices: Corticosteroids, *J. Comput. Aid. Mol. Des.* 12:557-561.
- Gupta S, Singh M, Madan AK, (1999). Superpendentic index: a novel topological descriptor for prediction of biological activity, *J. Chem. Inf. Comput. Sci.* 39: 272-277.
- Hansch C, Kurup A, Garg R, Gao H, (2001). Chem-Bioinformatics and QSAR: A Review of QSAR Lacking Positive Hydrophobic Terms. *Chem. Rev.* 101: 619-672.
- Hawkins PC, Skillman AG, Nicholls A (2007). Comparison of shape-matching and docking as virtual screening tools. *Journal of medicinal chemistry* 50: 74-82.
- Hawley SR et al. (1998). Relationship between antimalarial drug activity, accumulation, and inhibition of heme polymerization in *Plasmodium falciparum* in vitro. *Antimicrob. Agents Chemother.* 42: 682–686.
- Hay PJ, Wadt WR, (1985) Ab initio effective core potentials for molecular calculations. Potentials for K to Au including the outermost core orbitals. *J. Chem. Phys.*, 82: 299-310.
- Haynes RK et. al., Artemisinin antimalarials do not inhibit hemozoin formation. *Antimicrob. Agents Chemother.* 47: 1175 (2003).
- Haynes RK, Krishna S, (2004) Artemisinins: activities and actions. *Microbes Infect.* 6: 1339-1346.
- Haynes RK, Vonwille SC, (1997). From Qinghao, Marvelous Herb of Antiquity, to the Antimalarial Trioxane Qinghaosu – and Some Remarkable new Chemistry. *Acc. Chem. Res.* 30: 73-79.

- Haynes RK, Vonwiller SC, (1996). The behaviour of qinghaosu (artemisinin) in the presence of hemin iron (II) and (III). *Tetrahedron Lett.* 37: 253–256.
- Haynes RK, Vonwiller SC, (1997) From Qinghao, marvelous herb of antiquity, to the antimalarial trioxane qinghaosu- and some remarkable new chemistry. *Acc Chem Res.*, 30: 73-79.
- Hien TT, White NJ (1993). Qinghaosu. *Lancet* 341: 603-8.
- Homewood CA, Warhurst DC, Peters W, Baggaley VC (1972). Lysosomes, pH and the Anti-malarial Action of Chloroquine *Nature* 235: 50-52.
- Hong YL, Yang YZ, Meshnick SR, (1994). The interaction of artemisinin with malarial hemozoin. *Mol. Biochem. Parasitol.* 63: 121–128.
- [http://plasmodb.org/plasmodb/servlet/sv?page=gene&source\\_id=PFA0310c&detail=proteinQ.http://www.plasmodb.org/](http://plasmodb.org/plasmodb/servlet/sv?page=gene&source_id=PFA0310c&detail=proteinQ.http://www.plasmodb.org/).
- International Artemisinin Study Group. (2004) Artesunate combination for treatment of malaria: Meta-analysis. *Lancet.* 363: 9–17.
- Jaguar, version 4.1: *Schrodinger.* (2000). Inc.: Portland, OR.
- Jain N et. al. (2002). Essential oil composition of *Artemisia annua* L. ‘Asha’ from the plains of northern India *J. Essen. Oil Res.* 14 (4): 305 – 307.
- Jaiswal M, Khadikar PV, Scozzafava A, Supuran CT, (2004). Carbonic anhydrase inhibitors: the first QSAR study on inhibition of tumor-associated isoenzyme IX with aromatic and heterocyclic sulfonamides. *Bioorg. Med. Chem. Lett.*; 14: 3283-3290.
- Janse CJ, Waters AP, Kos J, Lugt CB (1994). Comparison of in vivo and in vitro antimalarial activity of artemisinin, dihydroartemisinin and sodium artesunate in the *Plasmodium berghei*-rodent model. *Int J Parasitol* 24: 589-94.
- Jefford CW (2001). Why artemisinin and certain synthetic peroxides are potent antimalarials. Implications for the mode of action. *Curr. Med. Chem.* 8: 1803–1826.
- Jefford CW, Vicente MGH, Jacquier Y, Favarger F, Mareda J, Millasson-Schmidt P, Brunner G, Burger U, (1996) The deoxygenation and isomerization of artemisinin and artemether and their relevance to antimalarial action. *Helv. Chim. Acta* 79: 1475–1487.
- Jung M (1990). Artemisinic acid: a versatile Chiral synthon and bioprecursor to natural products. *Planta Medica.* 56: 56.
- Jung M (1997). Synthesis and cytotoxicity of novel artemisinin analogs. *Bioorg. Med.Chem.Lett* 7 (8):1091-1094.

- Jung M, (1994) Current developments in the chemistry of artemisinin and related compounds. *Curr Med Chem* 1:35-49
- Kamchonwongpaisan S, and Meshnick SR, (1996) The mode of action of the antimalarial artemisinin and its derivatives. *Gen Pharmacol* . 27: 587-92.
- Kamchonwongpaisan S, Samoff E, Meshnick SR, (1997) Identification of hemoglobin degradation products in *Plasmodium falciparum*. *Mol. Biol. Parasitol.*, 86: 179-186.
- Kannan R, Kumar K, Sahal D, Kukreti S, Chauhan VS, (2005). Reaction of artemisinin with haemoglobin: implications for antimalarial activity. *Biochem. J.* 385: 409-418.
- Katritzky AR, Mu Lan, Lobanov VS, Karelson M, (1996). Correlation of Boiling Points with Molecular Structure. I A Training Set of 298 Diverse Organics and a Test Set of 9 Simple Inorganics, *J. Phys. Chem.* 100: 10400-10407.
- Kier L B, Hall L H, (1976). Molecular Connectivity in Chemistry and Drug Research, *Academic Press. New York.*
- Kier LB, Hall LH, (1997). The E-State as an Extended Free Valence. *J. Chem. Inf. Comput. Sci.* 37: 548-552.
- Kimura M, Yamaguchi Y, Takada S, Tanabe K, (1993) Cloning of a Ca(2+)-ATPase gene of *Plasmodium falciparum* and comparison with vertebrate Ca(2+)-ATPases. *J Cell Sci* 104 ( Pt 4): 1129-1136.
- Kissinger JC, Brunk BP, Crabtree J, Fraunholz MJ, Gajria B, Milgram AJ, Pearson DS, Schug J, Bahl A, Diskin SJ, Ginsburg H, Grant GR. Gupta D, Labo P, LiL, Mailman MD, McWeeney SK, Whetzel P (2002). *Nature.* 419: 490.
- Kitchen, DB, Decornez H, Furr JR, Bajorath J, (2004). Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat. Rev. Drug Discov.* 3: 935-949.
- Klayman DL (1985) Qinghaosu (artemisinin): an antimalarial drug from China. *Science* 228: 1049-1055.
- Klebe G, (2006). Virtual ligand screening: strategies, perspectives and limitations. *Drug Discovery. Today* 11: 580-594.
- Kondrachine AV, Trigg PI (1997). Global overview of malaria. *Indian J Med Res* 106: 39-52.
- Kuhlbrandt W (2004). Biology, structure and mechanism of P-type ATPases. *Nat. Rev. Mol. Cell Biol.*, 5: 282-295.

- Kumar S, Srivastava S, (2005) Establishment of artemisinin combination therapy as first line treatment for combating malaria: *Artemisia annua* cultivation in India needed for providing sustainable supply chain of artemisinin. *Curr. Sci.*, , 89: 1097–1102.
- Lai H, Singh NP (1995). Selective cancer cell cytotoxicity from exposure to dihydroartemisinin and holotransferrin. *Cancer Letters*, 91: 41-46.
- Laskowski RA, MacArthur MW, Moss DS, Thornton JM, (1993). *PROCHECK*: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* 26:283-291.
- Le WJ et al. (1982). Studies on the efficacy of artemether in experimental schistosomiasis. *Acta Pharmaceu.Sin.* 17: 187-193.
- Le WJ et. al. (1980) Artemisin derivatives against *Schistosoma japonicum* in animals. *Chin.*
- Leban I, Golic L, Japelj M, (1988) Crystal and molecular structure of qinghaosu: a redetermination. *Acta. Pharm. Jugosl.* 38: 71–77.
- Lee IS, Hufford CD (1990). Metabolism of antimalarial sesquiterpene lactones. *Pharmacol Ther* 48: 345-55.
- Lee MR, (2002) Plants against malaria Part I: Cinchona or the Peruvian bark. *J. R. Coll. Physicians Edinburgh*, 32: 300–305.
- Leiros HKS, Brandsdal BO, Andersen OA, Os V, Leiros I, Helland R, Otlewski J, Willassen N P, Smalas AO, (2004). Trypsin specificity as elucidated by LIE calculations, X-ray structures, and association constant measurements. *Protein Sci.* 13: 1056-1070.
- Li GQ, Guo XB, Fu LC, Jian HX, Wang XH (1994). Clinical trials of artemisinin and its derivatives in the treatment of malaria in China. *Trans R Soc Trop Med Hyg* 88 Suppl 1:S5-6.
- Lin AJ, Lee M, Klayman DL (1989). Antimalarial Activity of New Water-Soluble Dihydroartemisinin Derivatives, *J. Med. Chem.* 32: 1249-1252.
- Lin AJ, Li LQ, Andersen SL, Klayman DL, (1992) Antimalarial activity of New Dihydroartemisinin Derivatives. 5. Sugar Analogues. *J. Med. Chem.* 35: 1639-1642.
- Lin AJ, Margaret L, Klayman DL, (1989). Antimalarial Activity of New Water-Soluble Dihydroartemisinin Derivatives. 2. Stereospecificity of the Ether Side Chain. *J. Med. Chem.* 32: 1249-1252.

- Lipinski A, Lombardo F, Dominy B, Feeney P, (2001). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Del. Reviews.* 46(1-3): 3-26.
- Lisgarten JN, Potter BS, Bantuzeko C, and Palmer RA, (1998). Structure, Absolute Configuration, and Conformation of the Antimalarial Compound, Artemisinin, *J. Chem. Cryst.* 28: 539–543.
- Liu JM, (1979). Structure and reaction of arteannuin. *Acta Chim. Sin.* 37:129.
- Liu SS, Cao CZ, Li ZL, (1998). Approach to estimation and prediction for normal boiling point (NBP) of alkanes based on a novel molecular distance edge (MDE) vector  $\lambda$ , *J. Chem. Inf. Comput. Sci.* 38: 387-394.
- Livingstone DJ, (2000). The Characterization of Chemical Structures Using Molecular Properties. A Survey, *J. Chem. Inf. Comput. Sci.* 40: 195-209.
- Luo XD, Shen CC (1987). The chemistry, pharmacology, and clinical applications of qinghaosu (artemisinin) and its derivatives. *Med Res Rev* 7: 29-52.
- Lyne PD, Lamb ML, Saeh JC, (2006). Accurate prediction of the relative potencies of members of a series of kinase inhibitors using molecular docking and MM-GBSA scoring. *J. Med. Chem.*, 49:4805-4808.
- Maeno Y, Toyoshima T, Fujioka H, Ito Y, Meshnick SR, Benakis A, Milhous WK, Aikawa M, (1993) Morphologic effects of artemisinin in *Plasmodium falciparum*. *Am J Trop Med Hyg* 49:485-91.
- Meneses-Marcel Y, Marrero- Ponce Y, Machado-Tugores A, Monterro-Torres DM, Pereira JA, Escario JJ, Nogel-Ruiz C, Ochoa VJ, Aran AR, Martinez-Fernandez RN, Garcia Sanchez, (2005). A linear discrimination analysis based virtual screening of trichomonacidal lead-like compounds: Outcomes of in silico studies supported by experimental results, *Bioorg. Med. Chem. Lett.* 15: 3838-3843.
- Meshnick SR, (2002). Artemisinin: mechanisms of action, resistance and toxicity. *Int. J. Parasitol.* 32: 1655-1660.
- Meshnick SR, Thomas A, Ranz A, Xu CM, Pan HZ, (1991) Artemisinin (qinghaosu): the role of intracellular hemozoin in its mechanism of antimalarial action. *Mol Biochem Parasitol.* 49(2):181–189.
- Meshnick SR, Tsang TW, Lin FB, Pan HZ, Chang CN, Kuypers F, Chiu D, Lubin B, (1989). Activated oxygen mediates the antimalarial activity of qinghaosu. *Prog Clin Biol Res.* 313:95–104.

- Meshnick SR, Yang YZ, Lima V, Kuypers F, Kamchonwongpaisan S, Yuthavong Y, (1993) Iron-dependent free radical generation from the antimalarial agent artemisinin (qinghaosu). *Antimicrob Agents Chemother* 37: 1108-14.
- Meshnick, SR, Thomas A, Ranz A, Xu CM, Pan HZ (1991) Artemisinin (qinghaosu): the role of intracellular hemin in its mechanism of antimalarial action. *Mol. Biochem. Parasitol.* 49(2): 181–189.
- Meylan WM, Howard PH. (1995). Atom/fragment contribution method for estimating octanolwater partition coefficients. *J. Pharm. Sci.* 84(1): 83-92.
- Milhous WK, Kyle DE, Sherman IW (1998). ed. Malaria Parasite Biology, Pathogenesis, and Protection. Washington, D.C. *ASM Press*, pp 303-316.
- Milhous WK, Weatherley NF, Bowdre JH, Desjardins RF, (1985). In vitro activities of and mechanisms of resistance to antifolate antimalarial drugd. *Antimicrob. Agents Chemother.* 27:525-530.
- Misra LN et. al. (1993). Bisnor-cadinanes from *Artemisia annua* and definitive C-13 NMR assignments of beta- arteether. *Phytochemisrty.* 33 (6): 1461- 1464.
- Mockenhaupt FP, (1995) Mefloquine resistance in Plasmodium falciparum. *Parasitol. Today* 11: 248.
- Moore DV, Lanier JE (1961). Observations on two Plasmodium falciparum infections with an abnormal response to chloroquine *Am. J. Trop. Med. Hyg.* 10: 5-9.
- Moriguchi I, Hirono S, Liu Q, Nakagome I, Matsushita Y, (1992). Simple method of calculating Octanol/Water Partition Coefficient. *Chem. Pharm. Bull.* 40(1):127- 130.
- O'Neill P et al. (2000). Biomimetic Fe(II)-mediated degradation of arteflene (Ro-42–1611). The first EPR spin-trapping evidence for the previously postulated secondary carbon-centered cyclohexyl radical. *J. Org. Chem.* 65: 1578–1582.
- Olliaro P, (2001). Mode of action and mechanisms of resistance for antimalarial drugs. *Pharmacol. Ther.*, 89: 207-219.
- Olliaro PL, Trigg PI (1995). Status of antimalarial drugs under development. *Bull World Health Organ* 73: 565-71.
- Oloff S, Mailman RB, Trospha A, (2005). Application of Validated QSAR Models of D<sub>1</sub> Dopaminergic Antagonists for Database Mining, *J. Med. Chem.* 48: 7322-7332.
- Oprea TI, Matter H, (2004). Integrating virtual screening in lead discovery. *Curr. Opin. Chem. Biol.* 8: 349-358.

- Ostrovsky D, Udier-Blagovic M, Jorgensen WL, (2003) Analyses of activity for factor Xa inhibitors based on Monte Carlo simulations. *J. Med. Chem.* 46: 5691-5699.
- Padmanaban G, Rangarajan, PN (2001). Emerging targets for antimalarial drugs *Expert Opin. Ther. Targets* 5: 423-441.
- Pandey AV, Tekwani BL, Singh RL, (1999). Chauhan VS, Artemisinin, an endoperoxide antimalarial, disrupts the hemoglobin catabolism and heme detoxification systems in malarial parasite. *J. Biol. Chem.* 274: 19383-19388.
- Peters W, Li ZL, Robinson B L, Warhurst D C, (1986) The chemotherapy of rodent malaria, XL. The action of artemisinin and related sesquiterpenes *Ann. Trop. Med. Parasitol.* 80: 483-489.
- Pinheiro JC, Ferreira MMC, Romero OAS, (2001) "Antimalarial activity of dihydroartemisinin derivatives against *P. falciparum* resistant to mefloquine: a quantum chemical and multivariate study". *Theochem-J. Mol. Struct.*, 572: 35-44.
- Posner G, Oh CH, (1992). Regiospecifically oxygen-18 labeled 1,2,4-trioxane: a simple chemical model system to probe the mechanism(s) for the antimalarial activity of artemisinin (Qinghaosu). *J. Am. Chem. Soc.* 114: 8328.
- Posner GH, Cumming JN, Ploypradith P, Oh CH, (1995) Evidence for Fe(IV)=O in the Molecular Mechanism of Action of the Trioxane Antimalarial Artemisinin. *J. Am. Chem. Soc.* 117: 5885-5886.
- Posner GH, Oh CH, Gerena L, Milhous WK, (1992). Extraordinarily Potent Antimalarial Compounds: New, Structurally Simple, Easily Synthesized, Tricyclic 1,2,4-Trioxanes. *J. Med. Chem.* 35: 2459-2467.
- Posner GH, Oh CH, Wang D, Gerena L, Milhous WK, Meshnick SR, Asawamahasadka W. (1994) Mechanism-based design, synthesis, and in vitro antimalarial testing of new 4-methylated trioxanes structurally related to artemisinin: the importance of a carbon-centered radical for antimalarial activity. *J Med Chem.* 37(9):1256-1258.
- Posner GH, O'Neill PM, (2004) Knowledge of the proposed chemical mechanism of action and cytochrome p450 metabolism of antimalarial trioxanes like artemisinin allows rational design of new antimalarial peroxides. *Acc Chem Res* 37: 397-404.
- Posner GH, Park SB, Gonzalez L, Wang D, Cumming JN, Klinedinst D, Shapiro TA, Bachi MD, (1996) "Evidence for the Importance of High-Valent Fe=O and of a Diketone in the Molecular Mechanism of Action of Antimalarial Trioxane Analog of Artemisinin", *J. Am. Chem.*, 118: 3537-3538.
- Posner GH, Wang D, Cumming JN, Ploypradith P, Oh CH, French AN, Bodley AL, Shapiro TA, (1995). Further evidence supporting the importance of and the restrictions

- on a carbon-centered radical for high antimalarial activity of 1, 2, 4-trioxanes like artemisinin, *J. Med. Chem.* 38: 2273–2275.
- Price RN, Nosten F, Luxemburger C, ter Kuile FO, Paiphun L, Chongsuphajaisiddhi T, White N J (1996). Effects of artemisinin derivatives on malaria transmissibility. *Lancet* 347: 1654-8.
- Rafiee MA, Hadipour NL, Naderi-manesh H, (2005). The Role of Charge Distribution on the Antimalarial Activity of Artemisinin Analogs. *J. Chem. Inf. Model.* 45: 366-370.
- Ramachandran GN, Ramakrishnan C, Sasisekharan V (1963). Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* 7: 95-99.
- Rasooli L et. al. (2003) Microbial sensitivity to and chemical properties of the essential oil of *Artemisia annua* L. *J. Essen. Oil Res.* 15 (1): 59 – 62.
- Ridley RG (2002). Medical need, scientific opportunity and the drive for antimalarial drugs. *Nature* 415: 686-93.
- Rizzo RC, Tirado-Rives J, Jorgensen WL, (2001). Estimation of Binding Affinities for HEPT and Nevirapine Analogs with Reverse Transcriptase via Monte Carlo Simulations. *J. Med. Chem.* 44: 145-154.
- Robert A, (1998) Meunier B. Is alkylation the main mechanism of action of the antimalarial drug atemisinin? *Chem,Soc,Rev.* 27: 273-274.
- Ronn AM, Msangeni HA, Mhina J, Wernsdorfer W H, and Bygbjerg I C (1996). High level of resistance of Plasmodium falciparum to sulfadoxine-pyrimethamine in children in Tanzania. *Trans R Soc Trop Med Hyg.* 90: 179-81.
- Roy PP, Roy K,(2008). On some aspects of variable selection for partial least squares regression models. *QSAR Comb. Sci.* 27: 302-313.
- Sachs J, Malaney P (2002). The economic and social burden of malaria. *Nature.* 415: 680-685.
- Santana L, Uriarte H, Gonzalez-Diaz H, Zagotto R, Soto-Otero E, Mendez-Alvarez J, (2006). A QSAR Model for in Silico Screening of MAO-A Inhibitors. Prediction, Synthesis, and Biological Assay of Novel Coumarins. *J. Med. Chem.* 49: 1149-1156.
- Shapiro S, Guggenheim B. (1998). Inhibition of Oral Bacteria by Phenolic Compounds. Part 1. QSAR Analysis using Molecular Connectivity. *Quant. Struct-Act. Relat.* 17: 327-337.
- Sharma VP, (1996) Reemergence of malaria in india *Indian J. Med. Res.* 103: 26–45.
- Sharma VP, (1998) Fighting malaria in India *Curr. Sci.* 75: 1127–1140.

- Sharma VP, (1999) Current scenario of malaria in India *Parasitologia*. 41: 349–353.
- Sharma, VP, (2003) Malaria and poverty in India. *Curr. Sci*. 84: 513–515.
- Sherman IW (1998) (Ed.) In Malaria. *Am Soc Microbiol Press, Washington, DC*. pp. 1-556.
- Shi LM, Fan Y, Myers TG, Paul JN, (1998). Mining the NCI Anticancer Drug Discovery Databases: Genetic Function Approximation for the QSAR Study of Anticancer Ellipticine Analogs. *J. Chem. Inf. Comput. Sci*. 38: 189-199.
- Shukla KL, Gund TM, Meshnick SR, (1995). Molecular modeling studies of the artemisinin (qinghaosu)-hemin interaction: Docking between the antimalarial agent and its putative receptor. *Journal of Molecular Graphics*. 13: 215-222.
- Singh NP, Lai H, (2001) Selective toxicity of dihydroartemisinin and holotransferrin toward human breast cancer cells. *Life Sci*. 70 (1): 49-56.
- Slater AF, (1993) Chloroquine: mechanism of drug action and resistance in *Plasmodium falciparum*. *Pharmacol. Ther.* 57: 203-235.
- Sorensen TLM, Moller JV, Nissen P (2004) Phosphoryl transfer and calcium ion occlusion in the calcium pump. *Science*. 304: 1672-1675.
- Sousa SF, Fernandes PA, Ramos MJ, (2006). Protein-ligand docking: current status and future challenges. *Proteins* 65: 15-26.
- Sowunmi A, Oduola A M, Ogundahunsi O A, and Salako LA (1998). Comparative efficacy of chloroquine plus chlorpheniramine and pyrimethamine/sulfadoxine in acute uncomplicated falciparum malaria in Nigerian children. *Trans R Soc Trop Med Hyg* 92: 77-81.
- Stahura FL, Bajorath J, (2004). Virtual screening methods that complement HTS. *Comb. Chem. High Throughput Screen*. 7: 259-269.
- Street IP, Armstrong CR, Withers SG, (1986) Hydrogen bonding and specificity. Fluorodeoxy sugars as probes of hydrogen bonding in the glycogen phosphorylase-glucose complex. *Biochemistry*., 25: 6021-6027.
- Surolia N, RamachandraRao S R, Surolia A (2002). Paradigm shifts in malaria parasite biochemistry and anti-malarial chemotherapy. *BioEssays*, 24: 192-196.
- Sy LK et. al. (1998) A novel endoperoxide and related sesquiterpines from *Artemisia annua* which are possibly derived from allylic hydroperoxides. *Tetrahedron*. 54(17): 4345-4356.

- Sy, LK et. al. (2001) Structure elucidation of artannuin-O, a novel cadinane diol from *Artemisia annua*, and the synthesis of arteannuin- K, arteannuin- L, arteannuin- M, arteannuin- O. *Tetrahedron*, 57 (40): 8481-8493.
- ter Kuile F, White NJ, Holloway PH, Pasvol G, Krishna S (1993). Plasmodium falciparum: In vitro studies of the pharmacodynamic properties of drugs used for the treatment of severe malaria. *Exp. Parasitol.* 76: 85–95.
- Tewari A, Bhakuni RS (2003). Terpenoid and lipid constituents from *Artemisia annua*. *Indian J. Chem. Sect B*, 42 (7):1782-1785.
- Tominaga Y, Jorgensen WL, (2004). General model for estimation of the inhibition of protein kinases using Monte Carlo simulations. *J. Med. Chem.* 47: 2534-2549.
- Tommuphean S, Kokpol S, Parasuk V, Wolschann P, Winger RH, Liedl KR, Rode BM, (1998). Comparative molecular field analysis of artemisinin derivatives: ab initio versus semiempirical optimized structures. *J. Comput.-Aided Mol. Des.* 12: 397-409.
- Tonmunphean S, Parasuk V, Kokpol S (2000) QSAR Study of Antimalarial Activity and Artemisinin – Heme Binding Properties Obtained from Docking Calculation. *Quant. Struct–Act. Rel.* 19: 475.
- Toyoshima C, Mizutani T (2004). Crystal structure of the calcium pump with a bound ATP analogue. *Nature*. 430, 529-535.
- Toyoshima C, Nakasako M, Nomura H, Ogawa H(2000). Crystal structure of the calcium pump of sarcoplasmic reticulum at 2.6 angstrom resolution. *Nature*. 405: 647-655.
- Toyoshima C, Nomura H (2002) Structural changes in the calcium pump accompanying the dissociation of calcium. *Nature* 418: 605-611.
- Traditional Chinese Medicine (1979). Qinghaosu, Antimalarial, Coordination, R and Group. Antimalarial studies on qinghaosu. *Chinese Medical Journal* 92:6.
- Traditional Chinese Medicine (1992). Rediscovering wormwood: qinghaosu for malaria. *Lancet* 339:649-51.
- Tu YY et. al. (1981) Chemical constituents in *Artemisia annua* L. and derivatives of artemisinin. *Zhong Yao Tong Bao*. 6 (2):31.
- Tu YY et. al. (1981) Studies on the constituents of *Artemisia annua* L. *Acta Pharmaceutica Sinica*, 16 (5): 366-370.
- Tu YY et. al. (1982). Studies on the constituents of *Artemisia annua*. Part II. *Planta Medica*. 44: 143- 145.

- Tuckerman M, Berne BJ, Martyna GJ, (1992) Reversible multiple time scale molecular dynamics. *J. Chem. Phys.*, 97: 1990.
- van Agtmael MA, Eggelte TA, and van Boxtel CJ (1999). Artemisinin drugs in the treatment of malaria: from medicinal herb to registered medication. *Trends Pharmacol Sci.* 20: 199-205.
- van Lipzig MMH, Vermeulen NPE, Wamelink M, Geerke D, Jongejan A, ter Laak T M, Meerman JHN, (2003) Molecular dynamics simulations with the estrogen receptor: Prediction of ligand binding affinity and orientation by the linear interaction energy method. *Drug Metab. Rev.* 35: 107- 107.
- Varadi A, Tusnady GE, Sarkadi B, (2003) Membrane topology of the human ABC transporter proteins. In: ABC Proteins From Bacteria to Man. I. B. Holland, S. P. C. Cole, K. Kuchler & C. F. Higgins (eds). London/San Diego: *Academic Press*, pp. 37-46.
- Vathsala, P. G. *et al.*, Widespread occurrence of the *plasmodium falciparum* chloroquine resistance transporter (pfcr) gene haplotype svmt in *p. falciparum* malaria in India *Am. J. Trop. Med. Hyg.*, 2004, 70, 256–259.
- Wallaart E (2000) “Investigations on biosynthesis of the novel antimalarial drug artemisinin in the plant *Artemisinin annua*” Ph. D. thesis Rijkuniversiteit Groningen, 16.
- Ward SA, Bray PG, Mungthin M, Hawley SR (1995). Current views on the mechanisms of resistance to quinoline-containing drugs in *Plasmodium falciparum*. *Ann Trop Med Parasitol* 89: 121-4.
- Warren GL, Andrews CW, Capelli AM, Clarke B, LaLonde J, Lambert MH, Lindvall M, Nevins N, Semus SF, Senger S, Tedesco G, Wall ID, Woolven JM, Peishoff CE, Head MS, (2006). A critical assessment of docking programs and scoring functions. *Journal of medicinal chemistry* 49: 5912-5931.
- Wernsdorfer WH, and Payne D (1991). The dynamics of drug resistance in *Plasmodium falciparum*. *Pharmacol Ther* 50: 95-121.
- White N (1999a). Antimalarial drug resistance and combination chemotherapy. *Philos Trans R Soc Lond B Biol Sci.* 354: 739-49.
- White NJ (1998). Drug resistance in malaria. *Br Med Bull* 54: 703-15.
- White NJ (2008). *Plasmodium knowlesi*: the fifth human malaria parasite. *Clin Infect Dis* 46: 172-173.

- White NJ, (1999) Antimalarial drug resistance and combination chemotherapy. *Philos. Trans. R. Soc. London, Ser. B*, 354: 739–749.
- White NJ, (2004) Antimalarial drug resistance. *Clin Invest.* 113: 1084-1092.
- WHO (2008a). Global malaria control and elimination: report of a technical review. In.: World Health Organisation, pp. 1-47.
- WHO (2008b). Treatment policies. In: World Health Organisation. pp. 1–52.
- WHO, (2008c). World Malaria Report 2008. In.: World Health Organisation, pp. 1-190.
- WHO, First Meeting of Regional Technical Advisory Group on Malaria, Manesar, Haryana, 15–17 December 2004, World Health Organization, Regional Office for South-East Asia, New Delhi, SEA-MAL-239, pp. 1–38.
- WHO, Report, 2001, [http://rbm.who.int/partnership/wg/wg\\_management/docs/RBMConsensusStatement\\_act.pdf](http://rbm.who.int/partnership/wg/wg_management/docs/RBMConsensusStatement_act.pdf)
- WHO, Report, Regional Office for South-East Asia, New Delhi, SEA-Mal- 237, October 2004, pp. 1–25.
- WHO. Roll Back Malaria. *A Global Partnership*. Geneva: World Health Organization. (1998).
- Wirth DF (2002). Biological revelations. *Nature*. 419: 495-496.
- Woolfrey JR, Avery MA, Doweiko AM, (1998). Comparison of 3D quantitative structure-activity relationship methods: Analysis of the in vitro antimalarial activity of 154 artemisinin analogues by hypothetical active-site lattice and comparative molecular field analysis. *Journal of Computer-Aided Molecular Design*. 12(2): 165-181.
- World Health Organization. *The World Malaria Report 2005*.
- Wu LJ et al. (1995) The experimental studies of early treatment of artesunate against schistosomiasis. *Chin. J. Schistosom. Control*, 7 (3): 129-133.
- Wu WM, Wu Y, Wu YL, Yao ZJ, Zhou CM, Li Y, Shan F, (1998). Unified Mechanistic Framework for the Fe(II)-Induced Cleavage of Qinghaosu and Derivatives/Analogues. The First Spin-Trapping Evidence for the Previously Postulated Secondary C-4 Radical. *J. Am. Chem. Soc.* 120: 3316-3325.
- Wu Y, (2002). How might qinghaosu (artemisinin) and related compounds kill the intraerythrocytic malaria parasite? A chemist's view. *Acc Chem Res* 35: 255-9.

- Wu YL, Li Y (1995). Study on the chemistry of qinghaosu (artemisinin). *Med Chem Res* 5: 18.
- Wu ZH, Wang YY (1984). Structure and synthesis of arteannuin and related compounds XI. Identification of epoxy arteannunic acid. *Acta Chim. Sinica.* 42: 596.
- Yang YZ, Aswamahasakda W, Meshnick SR, (1993). Alkylation of human albumin by the antimalarial artemisinin. *Biochem. Pharmacol.* 46: 336-339
- Yoon S, Smellie A, Hartsough D, Filikov A, (2005a). Computational identification of proteins for selectivity assays. *Proteins* 59: 434-443.
- Yoon S, Smellie A, Hartsough D, Filikov A, (2005b). Surrogate docking: structure-based virtual screening at high throughput speed. *Journal of computeraided molecular design* 19: 483-497.
- Zaman SS, Sharma RP (1991). Some aspects of chemistry and biological activity of artemisinin and related antimalarials. *Heterocycles.* 32 (8): 1593-1637.
- Zhang F, Gosser DK Jr, Meshnick SR, (1992). Hemin-catalyzed decomposition of artemisinin (qinghaosu). *Biochem Pharmacol.* 43: 1805-1809.
- Zheng GQ et. al. (1994). Cytotoxic terpenoids and from artemisia. *Planta Medica.* 60: 54-57.
- Zhou R H, Friesner R A, Ghosh A, Rizzo RC, Jorgensen WL, Levy RM, (2001) New linear interaction method for binding affinity calculations using a continuum solvent model. *J. Phys. Chem. B,* 105: 10388-10397.
- Zhou R, Berne BJ, (1995) A new molecular dynamics method combining the reference system propagator algorithm with a fast multipole method for simulating proteins. *J. Chem. Phys.* 103: 9444.
- Zhou R, Berne BJ, (1997). Smart Walking: a new method for Boltzmann sampling of protein conformations. *J. Chem. Phys.* 107: 9185.
- Zhu DY et. al. (1984) Structure of artemisilactone. *Acta Chim.Sinica.* 42 (9): 937-939.
- Zhu DY. et. al. (1982). Study on antibacterial constituents of Qinghaosu (*Artemisa annua* L.) *Zhongcaoyao.* 13 (2): 54.

## List of Publications

1. **M. Srivastava**; H. Singh; P. K. Naik, Application of the linear interaction energy method for rational design of artemisinin analogues as haeme polymerisation inhibitors journal of SAR and QSAR in Environmental Research, Volume 20,(2009) , pp: 327 – 355.
2. **M. Srivastava**; H. Singh; P. K. Naik, Quantitative structure-activity relationship (QSAR) of the artemisinin: the development of predictive in vitro antimalarial activity models Journal of Chemometrics (2009) (In Press).
3. **M. Srivastava**; H. Singh; P. K. Naik, Molecular modeling evaluation of the antimalarial activity of artemisinin analogues: molecular docking and rescoring using prime/MM-GBSA approach journal of current research in Biological Sciences (2009) (In Press).
4. P. K. Naik; **M. Srivastava**; A. Dubey and P. Bajaj, The binding modes and binding affinities of artemisinin derivatives with *Plasmodium falciparum* Ca<sup>2+</sup>-ATPase (PfATP6). Communicated to Journal of Molecular Modeling (2009).